MM5V3-Based 4DVAR: Current Status and Future Plans

Frank H. Ruggiero*, Donald Norquist
Air Force Research Laboratory, AFRL/VSBL, 29 Randolph Road, Hanscom AFB, MA, 01731

Thomas Nehrkorn, George D. Modica, Mark Cerniglia
Atmospheric and Environmental Research, Inc., Lexington, MA 02421

John G. Michalakes
National Center for Atmospheric Research, NCAR/MMM Boulder, CO 80301

Xiaolei Zou
Florida State University, Tallahassee, FL 32306

1. Introduction

While research studies have shown the utility of four-dimensional variational analysis (4DVAR) in assimilating data (e.g., Zou et al. 2001), its high computational cost has all but prohibited its use at operational weather forecasting centers. Previous experiments have been limited to very small domains or relatively coarse spatial resolutions (e.g., Zou and Kuo, 1996). The European Center for Medium Range Weather Forecasts is currently the only operational center in the world running 4DVAR (Rabier et al. 2000), albeit on a global scale with coarse spatial resolution and simple model physics for the data assimilation model.

To address the issue of computational cost constraining 4DVAR applications, the DoD High Performance Modernization Office (HPCMO) has funded, under its Common High Performance Computing Software Support Initiative (CHSSI) program, a team of collaborators to develop a scalable version of a 4DVAR system based on the National Center for Atmospheric Research/Penn State University Fifth Generation Mesoscale Model (MM5; Grell et al. 1995). MM5 was chosen as the model to base this system on because it is the current operational forecast model used by the Air Force Weather Agency. The objective is to create a version of the code that would be optimized for distributed memory, multiple processor machines. The goal of the project is achieve sufficient wall-clock time speed up so as to allow 4DVAR's use at operational forecasting centers. The emphasis for observation operators will be remotely sensed satellite radiances so that the 4DVAR system can be used in otherwise data-denied regions.

*Corresponding author address: Frank H. Ruggiero AFRL/VSBL, 29 Randolph Rd, Hanscom AFB, MA 01731-3010; (781) 377-2940; FAX: (781) 377-3661 email: frank.ruggiero@hanscom.af.mil

2. Previous work

A 4DVAR system based on MM5 already exists (Zou et al. 1997). However it runs only in serial mode and is based on a version of MM5 (v1) that is several releases behind the current version (3.5). The option to parallelize the existing MM5v1 based 4DVAR code was carefully considered. However substantial code changes have been implemented into MM5 between v1 and the current version. The most notable is that a scalable version of the MM5 forward model can now be compiled and run (Michalakes, 2000). Given this and the fact that it was perceived that a 4DVAR system based on a more current release of MM5 would find greater user acceptance, it was decided to build a new version of the MM5 4DVAR system compatible with the current version of MM5.

There are five major components of a 4DVAR system. Two are the gradient solver and the observation operators. The other three components are related to the NWP model on which the system is based, the forward forecast or non-linear model, the tangent-linear version of the forecast model, and the adjoint of the tangent-linear model. Since the 4DVAR system being developed is to be based on MM5v3, the forward forecast model is already able to run efficiently on distributed memory, multiple processor computers (Michalakes 2000). The parallelization was implemented in such a way as to maintain the same source architecture for the serial and parallel versions of the code. The basic principle employed in the parallelization is horizontal domain decomposition where the model's horizontal grid is broken up into sub-domains for processing on individual processors. Ruggiero et al. (2001) described the successful development and testing of the MM5v3-based tangent-linear model. Its coding implementation followed the same design as the forecast model. Development of the tangent-linear model was greatly aided by the use of the Tangent-linear and Adjoint Model Compiler (TAMC) (Giering and Kaminski 1998). This code takes a forward forecast model as input and generates either the tangent-linear version of the model or its adjoint.

3. Current status

The MM5v3-based adjoint model has been successfully developed and tested. The adjoint was developed with the use of TAMC as described by Nehrkorn et al (2002). The adjoint was tested for accuracy first at the subroutine level and then as an entire system. The check was carried out using the definition of the adjoint following the procedure of Zou et al (1997),

$$\langle y, Lx \rangle = \langle L^{*}y, x \rangle,$$

where x is the tangent-linear model input, y=L(x) is the tangent-linear model output, and L and L^* are the tangent-linear model and adjoint respectively.

The adjoint contains the same set of physics parameterizations that are in the MM5v1-based 4DVAR system. This includes the Kuo (1974) convective parameterization, explicit grid-scale precipitation, and a bulk planetary boundary layer parameterization.

The MM5v3 adjoint was initially developed as a serial mode version. After this was completed, a parallelized build capability was added to the code, again making use of the same source architecture used in MM5. The same source architecture used in MM5 and the tangent-linear model is maintained in the adjoint.

When a 4DVAR system is executed the forward forecast model is run forward in time and typically its output from each time step is written to disk. The adjoint of the model is then executed and it essentially runs backwards in time. However since it is based on the linearized version of the forward model, it needs to read the time step information from the forward model as it retraces the forward models steps. This situation does not arise when running a forward model to generate a forecast and thus has never been an issue when parallelizing MM5. Initially for the tangent-linear model and adjoint, the thought was to write and read the forward model data needed for each sub-domain to and from the scratch disk for the processor associated with it. In reality this was not practical given the amount of disk space and I/O needed and the other system needs of the scratch disk. Therefore all the trajectory information was written to a single file using the parallel file system such as the General Parallel File System on the IBM SP P3. This was immediately recognized as bottleneck in the code's execution. In order to mitigate the amount of I/O calls for the time step data was buffered so that a read or write to disk was executed for every five time steps. While this cut down on the number of I/O calls and the I/O was done asynchronously and only for data needed for a particular processor, the result was that the parallel file server was still being bombarded by simultaneous I/O requests and slowing down the program execution. An additional mitigation measure was then added in which the I/O for the sub-domains was staggered so that at any particular time step only one in four of the sub-domains need to write to the disk file. While these steps have helped alleviate the I/O bottleneck substantially, it is still impacting the scaling efficiency. Currently, the code updates the trajectory information at every time step. However the code has an option that allows updating of trajectory information at a given number of time steps. In the next couple of months tests will be run to see what is the minimum update frequency needed to maintain the quality of the analyses. We expect to be able to significantly reduce the number of I/Os for the trajectory information and thus substantially improve the scaling results.

Two observation operators have been added to the system. Both of these have leveraged off previous efforts at NCAR. One of the operators is for rawinsondes. Since rawinsondes provide direct measurements of model variables, the only function of the observation operator is to interpolate the model data to the observation locations. The second observation operator is for Geostationary Orbiting Environmental Satellite (GOES) 8 sounder data. The GOES-8 sounder measures top-of-atmosphere radiances, therefore in addition to interpolating the model data to the measurement locations, the observation operator must also execute radiative transfer codes to convert the model variables into radiances.

4. Performance examples

Along with the unit and system testing described above, checking of the accuracy of the 4DVAR system was accomplished by comparisons of analyses generated by the MM5v1 and MM5v3 based 4DVAR systems. One of the comparisons was done using a case from the Severe Environmental Storms and Mesoscale Experiment (SESAME). The case in question was a severe weather outbreak in the Red River Valley region of Texas on 10-11 April 1979 (Alberty et al. 1980). Figure 1. contains the comparisons of the analysis low-level temperature and wind fields at 1800 UTC 10 April 1979. The analyses were constructed using a horizontal grid domain of 19x17 points with a grid spacing of 120 km. One can see that the analyses are quite similar. There are some very small differences in the position the isotherms and the wind barbs. This is to be expected since in the course of the migration from the v1 to v3 MM5 model some fixes and restructuring have been made to the basic model code. However the closeness of the results indicates that the new version of MM5 4DVAR is qualitatively performing similar to the previous version.

While the driver for running the 4DVAR system in MPP mode is still being tested, timing results for the tangent-linear model and the adjoint components running separately have been carried out. Using a dataset from Hurricane Felix (Zou et al. 2001) that uses a domain of 84x75 horizontal grid points, wall-clock time speed-up of approximately 20 times faster were noted when using 24 processors versus one processor. These results were done updating the base state information at every time step using the I/O mitigation measures mentioned above. Comparison of the MPP runs of tangent-linear model and adjoint with differing number of processors ranging from 1 to 24 resulted in bit for bit agreement in the output.

5. Future plans

With the completion of the adjoint a full 4DVAR system based on MM5v3 with observation operators for rawinsondes and GOES-8 sounder radiances is now complete. The other major component of the 4DVAR system is the gradient solver. The purpose of the gradient solver is to find the optimal initial conditions that produce a best fit to the observations that are distributed in time and space. The gradient solver employed in the MM5v3-based 4DVAR system is the same as that used in the MM5v1 version and is based on the method of Liu and Nocedal (1989). As part of a separate 3DVAR effort at the National Center for Atmospheric Research, the Liu and Nocedal gradient solver has been parallelized and has shown to achieve significant speed up in wall-clock time. We attempt will leverage that work to the current 4DVAR system.

Several physics upgrades are planned for the system. One upgrade is to give the user the option to run the multi-level PBL scheme of Hong and Pan

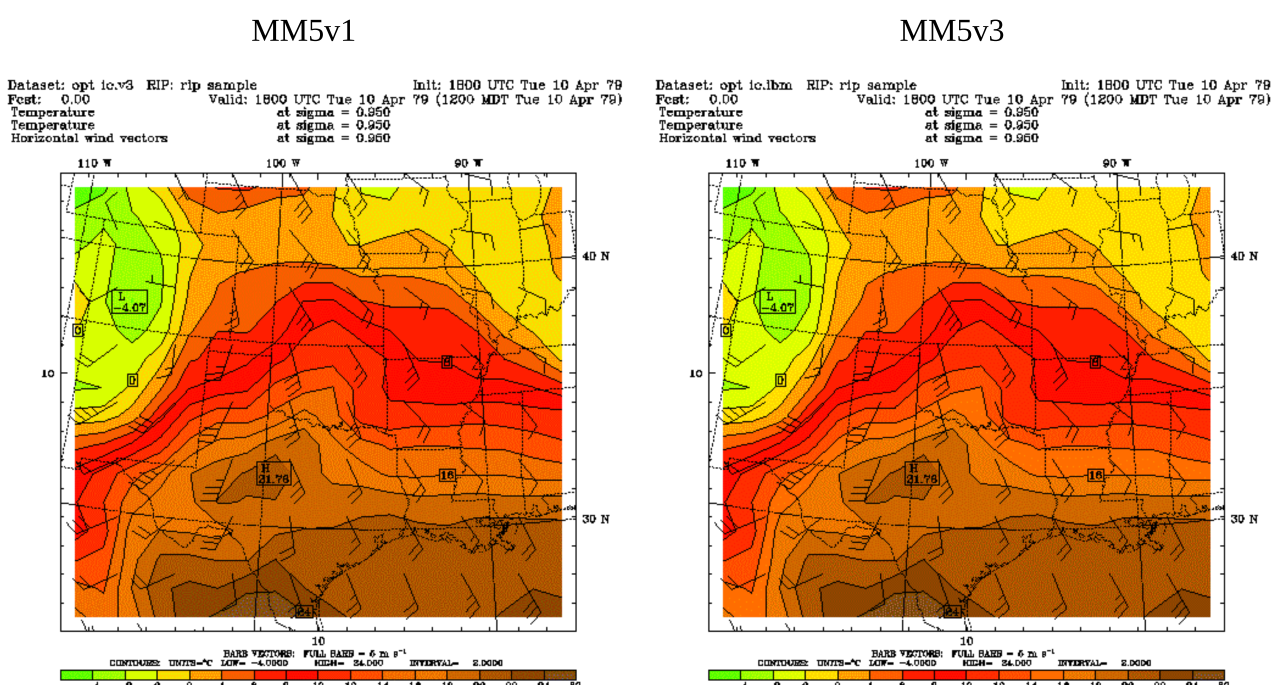

FIG. 1. Comparison of low-level temperature and winds for 1800 UTC 10 April 1979 from MM5v1 4DVAR and MM5v3 4DVAR analyses

(1996). This would allow the user to use a greater number of vertical levels than is practical with the single-level bulk PBL scheme. A second upgrade is the installation of the tropical cyclone vortex bogussing scheme of Zou and Xao (2000). Tests of the bogussing scheme in the MM5v1 4DVAR system revealed that it produced superior forecasts of tropical cyclone positions for Hurricane Felix (Zou et al. 2001).

Another major modification to the 4DVAR system will be the development of a driver for incremental 4DVAR. The use of incremental 4DVAR has been explained by Courtier et al (1994). The premise of incremental 4DVAR is that one would use a simplified forward model such as the tangent-linear version of the forward model. The use of the simplified forward model should shorten the execution time of the 4DVAR system. Another advantage of employing incremental 4DVAR is that it allows the use of the analysis generated by 4DVAR for another forecast model. Thus the 4DVAR system developed here could be compatible with COAMPS or WRF.

The bulk of this project is to be completed by early 2003 and it is planned to make the code available for community use at that time. Since the beginning of this year, we have worked with members of the data assimilation group at NCAR. This has resulted in the leveraging of prior NCAR work to our effort and extensive testing of the code we are developing. The purpose of the testing is to insure that the code is robust enough and documented sufficiently to meet the standards for community use.

Acknowledgements: This work was supported by the Department of Defense (DoD) High Performance Computing (HPC) Modernization Office's Common High Performance Computing Software Support Initiative (CHSSI). Additional support was provided from the DoD Major Shared Resource Center at Wright-Patterson AFB for time on the IBM SP P3.

REFERNCES

Alberty, R. L., D. W. Burgess, and T. T. Fujita, 1980: Severe Weather Events of 10 April 1979. *Bull. Am. Meteorol. Soc.*, **61**, 1033-1034.

Courtier, P., J. N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using and incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367-1387.

Giering, R., and T. Kaminski, 1998: Recipies for adjoint code construction. *Trans. Math. Software,* **24**, 437-474.

Grell, G. A., J. Dudhia, and D. R. Stauffer, 1995: A description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note NCAR/TN-398+STR, 122pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO, 80307.]

Hong, S. –Y., and H. –L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev*., **124**, 2322-2339.

Kuo, H. L., 1974: Further studies of the influence of cumulus convection on larger-scale flow. *J. Atmos. Sci.*, **31,**1232-1240.

Liu, D. C., and J. Nocedal, 1989: On the limited memory BFGS method for large scale optimization. *Mathematical Programming,* **45**, 503-528.

Michalakes, J., 2000: The same-source parallel MM5. *Sci. Programming,* **8**, 5-12.

Nehrkorn, T., G. D. Modica, M. Cerniglia, F. H. Ruggiero, J. G. Michalakes, X. Zou, 2002: 4DVAR development using an automatic code generator: Application to MM5v3. Preprints, *12th PSU/NCAR Mesoscale Modeling System Users' Workshop, Boulder, CO, June 24-25, 2002.*

Rabier, F., H. Järvinen, E. Klinker, J. –F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation, I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126(**A), 1143-1170.

Ruggiero, F. H., G. D. Modica, T. Nehrkorn, M. Cerniglia, J. Michalakes, and X. Zou, 2001: Development of an MM5-based four-dimensional variational analysis system for distributed memory multiprocessor computers. Preprints, *HPCMP 2001 User's Group Conference, Biloxi, MS, June 18-22, 2001.*

Zou, X., and Y. –H. Kuo, 1996: Rainfall assimilation throughout an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Mon. Wea. Rev.*, **124**, 2859-2882.

–––, F. Vandenberghe, M. Pondeca, and Y.-H. Kuo, 1997: Introduction to adjoint techniques and the MM5 adjoint modeling system. NCAR Tech. Note NCAR/TN-435-STR, 110pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO, 80307.]

–––, Q. Xiao, A. E. Lipton, and G. D. Modica, 2001: A numerical study of the effects of GOES sounder cloud-cleared brightness temperatures on the prediction of Hurricane Felix. *J. Appl. Meteor.*, **40**, 34-55.