Next-Generation Global and Mesoscale Atmospheric Models

Francis X. Giraldo
Department of Applied Mathematics
Naval Postgraduate School
833 Dyer Road
Monterey, CA 93943-5216
phone: (831) 656-2293 fax: (831) 656-2355 e-mail: fxgirald@nps.edu

LONG-TERM GOALS

The long-term goal of this research is to construct a unified global and mesoscale nonhydrostatic numerical weather prediction (NWP) models for the U.S. Navy using new numerical methods specifically designed for modern computer architectures; this unified model is called the Non-hydrostatic Unified Model of the Atmosphere or NUMA. To take full advantage of distributed-memory computers, the global domains of these new models are partitioned into local sub-domains, or elements, that can then be solved independently on multiple processors. The numerical methods used on these sub-domains are local, high-order accurate, fully conservative, highly efficient, and geometrically flexible. Using these ideas we are developing global and mesoscale nonhydrostatic atmospheric models that will improve on the operational models currently used by all U.S. agencies including the U.S. Navy.

OBJECTIVES

The objective of this project is to construct new high-order local methods for the Navy's next-generation global and mesoscale nonhydrostatic NWP models. The high-order accuracy of these methods will improve the accuracy of the dynamics for the current global spherical harmonics model (NOGAPS) and the finite difference local-area model (COAMPS). It is conjectured that improving the accuracy of the dynamics (including tracers) will increase the overall accuracy of the forecast; however, such improvements will also have to be made to both the physics and the data assimilation system but these two topics are currently beyond the scope of this project. The objective of this project is to improve the accuracy of the dynamics while increasing the geometric flexibility in order to use any grid as well as to increase the efficiency of these models on large processor-count distributed-memory computers. Higher efficiency means that the new models will require less computing time that then allows for increasing the number of ensemble members and/or increasing the resolution of the NWP models. The methods that we propose to use for these models are state-of-the-art and are not being used by either current or newly emerging NWP models.

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **2012**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Next-Generation Global and Mesoscale Atmospheric Models**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Department of Applied Mathematics Naval Postgraduate School 833 Dyer Road Monterey, CA 93943-5216**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**11**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

APPROACH

To meet our objectives we explore:

1. Unified high-order continuous Galerkin (CG) and discontinuous Galerkin (DG) spatial discretization methods;

2. Unified high-order implicit-explicit (IMEX) time-integrators with adaptive time-stepping for vastly improved efficiency;

3. unified global and mesoscale models based on the nonhydrostatic equations;

4. fully unstructured and adaptive grids; and

5. scalable parallel implementation of NUMA.

The power of CG and DG methods is that they are high-order accurate yet are completely local in nature – meaning that the equations are solved independently within each individual element and, on parallel computers, may reside on separate processors. Furthermore, high-order methods have minimal dispersion error – this is an important property for capturing fine-scale atmospheric phenomena (e.g., tropical cyclones, Kelvin and Rossby waves). The theoretical development of CG and DG methods are now well established and these methods are currently the two most successful methods found in the literature for fluid flow problems.

Semi-implicit (SI) and fully-implicit (FI) time-integrators offer vast improvements in efficiency due to the longer time steps that they permit; semi-implicit methods can be classified under the heading of implicit-explicit (IMEX) methods that has garnered much attention in the computational mathematics literature. Furthermore, in order to reap the full benefits of the high-order spatial discretization methods requires increasing the order of accuracy of the time-integration methods as well; this is a topic that too often has been ignored by most scientific computing communities, including the NWP community, but has been at the forefront of research efforts in this project.

Before committing resources towards the development of new NWP models, it is important to identify the form of the governing equations that is most capable of conserving all quantities deemed important. We have been performing studies on this topic for the past few years – that is, to identify the form of the governing equations capable of representing conservation of either mass, energy, or both. In addition, we have analyzed various forms of the governing equations with respect to robustness, flexibility, and efficiency in the context of implicit-explicit (IMEX) time-integration methods. Two of the papers developed within this project are slowly becoming the standard papers for identifying the proper equation sets and test cases for nonhydrostatic atmospheric modeling (Giraldo-Restelli JCP 2008 and Giraldo et al. SISC 2010).

One final area that needs to be explored is the concept of adaptive grids. In the past few years, adaptive grids have gained considerable momentum in the atmospheric modeling community – I am currently participating in a four month program on this topic at the Newton Institute in Cambridge University, England.

WORK COMPLETED

In this section, we describe the work completed this fiscal year. The work can be categorized into four sections: unified global/mesoscale governing equations, unified IMEX time-integration methods, unified spatial discretization methods, and parallel implementation.

Unified Global/Mesoscale Equations. In FY 2010, we developed a fully parallel 3D nonhydrostatic model for limited-area (mesoscale) applications using the CG method. This model has been verified using standard mesoscale tests such as inertia-gravity waves, rising thermal bubbles, and isolated three-dimensional mountains. However, in order to construct a truly unified model requires that the equations also be able to handle global dynamics. To do this, let us introduce the following governing equations (that we refer to as Set 2NC in the nomenclature of Ref. [1])

$$\frac{\partial \rho}{\partial t} + \nabla \bullet (\rho \mathbf{u}) = 0$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \bullet \nabla \mathbf{u} + \frac{1}{\rho}\nabla P + g\hat{\mathbf{r}} + f(\hat{\mathbf{r}} \times \mathbf{u}) = 0$$

$$\frac{\partial \theta}{\partial t} + \mathbf{u} \bullet \nabla \theta = 0$$

$$P = P_A \left(\frac{\rho R \theta}{P_A} \right)^{\gamma}$$

where the main (new) strategy introduced here is that of using a radial vector $\hat{\mathbf{r}}$ that denotes the direction in which gravity acts. For example, for mesoscale flow (flow in a box) $\hat{\mathbf{r}}$ becomes $\mathbf{k}$, the unit vector along the z-coordinate direction. On the sphere, $\hat{\mathbf{r}}$ becomes the radial unit vector. This then requires that the hydrostatic balance be defined as follows:

$$\nabla P_0 = -\rho_0 g\hat{\mathbf{r}}.$$

The second critical step in constructing a unified global/mesoscale model requires deriving a general form for the reference fields that satisfy the required background conditions. For example, the equation above imposes the constraint on hydrostatic balance. That is, the prognostic variables are split into the following reference fields and perturbations from this reference field as follows:

$$\rho(x,y,z,t) = \rho_0(x,y,z) + \rho'(x,y,z,t)$$

$$\theta(x,y,z,t) = \theta_0(x,y,z) + \theta'(x,y,z,t)$$

$$P(x,y,z,t) = P_0(x,y,z) + P'(x,y,z,t)$$

where the reference field satisfies the hydrostatic balance. Now, if we wish to impose additional constraints on the reference fields (e.g., geostrophic balance) then we must retain some terms that are usually omitted in the equations. We arrive at the following set of general equations:

$$\frac{\partial \rho'}{\partial t} + \nabla \bullet (\rho \mathbf{u}) = 0$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \bullet \nabla \mathbf{u} + \frac{1}{\rho} \left(H \nabla P_0 + \nabla P' + \rho' g \hat{\mathbf{r}} \right) = -f \left(\hat{\mathbf{r}} \times \mathbf{u} \right)$$

$$\frac{\partial \theta'}{\partial t} + \mathbf{u} \bullet \nabla \theta = 0$$

where H is a projector that nullifies the component of the gradient operator in the direction of the radial component $\hat{\mathbf{r}}$. This then leaves the remainder of the gradient operator components untouched in order to satisfy geostrophic balance. To our knowledge, this form of the governing equations have not been derived before nor have the equations been solved in this fashion. We developed this idea at the end of 2011 and were able to show preliminary results for a balanced initial state and baroclinic instability problem using a reference field that is both hydrostatically and geostrophically balanced. In addition, in 2011 we showed preliminary results for global applications such as the acoustic wave propagation and inertia-gravity wave problems. This year, we analyzed the results of this approach quantitatively and the results are illustrated in the Results section.

Unified IMEX Time-Integrators. Directly connected with our choice of equation sets is the resulting Implicit-Explicit (i.e., semi-implicit) operators. In studies performed in 2010-2011, we compared IMEX time-integrators both in their Schur (i.e., pseudo-Helmholtz) and No Schur (i.e., full system) forms and compared them to the types of explicit time-integrators currently being used in explicit models. Our results show that if the Schur form is used, then the IMEX methods are always faster than explicit models. Furthermore, we have analyzed the types of problems that are amenable to an IMEX approach in all directions (which we call 3D-IMEX) versus IMEX approaches only in the vertical (which is known as HEVI, Horizontal Explicit – Vertically Implicit, or 1D-IMEX). This study is contained in a manuscript that has been submitted to the SIAM Journal of Scientific Computing (see Ref. [2]).

One of the main results of that paper is that we have been able to extract the radial direction of our model in a general way in order to build the 1D-IMEX methods. This general 1D-IMEX is the clear winner for most global modeling problems that we have considered due to the disparate horizontal versus vertical scales typically used in global modeling (e.g., the horizontal resolution is typically around 50km whereas the vertical resolution is 1km). An additional advantage of 1D-IMEX is that it scales as perfectly as a fully explicit method if the domain decomposition is done across the horizontal direction (which is explicit) while the points along the vertical column are on processor and so the IMEX method incurs no MPI communication per solve.

Unified Spatial Discretization Methods. We have been arguing in the course of this project that the best next-generation models will be those based on element-based Galerkin (EBG) methods such as the spectral element (SE/CG) and discontinuous Galerkin (DG) methods. However, we have only partly showed the benefits of this approach such as: high parallel efficiency and high-order accuracy. Last year, we began studying adaptivity on triangular grids (see Ref. [3]). This year, we began a study of adaptive non-conforming quadrilateral grids. One of the attractions of NUMA is that it uses either CG or DG methods. In other words, the CG and DG spatial discretizations have been written in a unified way which allows the model to use either method. Currently, only the DG-NUMA is able to use non-

4

conforming adaptive mesh refinement (AMR); we are currently working on extending AMR to CG-NUMA.

Parallel Implementation. In 2010, we were able to extend the 2D serial CG models that we developed for the compressible Navier-Stokes equations to 3D parallel. This year we extended the DG models to 3D and MPI so that we now have two 3D MPI models. We have begun to compare and contrast the models in terms of accuracy and efficiency and this study is documented in Ref. [4] and is still underway.

RESULTS

The main results for this year can be divided into the following sections: 1) Study of the IMEX Time-Integrators; 2) a rigorous Analysis of the Parallel Scalability of both the CG and DG models; and 3) positivity preserving algorithms for CG-NUMA and dispersion analysis of the CG method.

Unified IMEX Time-Integrators. In Ref. [1] we were able to derive all Implicit-Explicit methods to be written in the exact same way using the usual semi-implicit formulation used by all operational centers. The advantage of this is that any IMEX method (multistep, multistage, etc.) can be incorporated easily into an existing model using our approach. NUMA currently contains the following time-integrators: explicit Runge-Kutta, Adams-Bashforth, Backward Difference Formulas, and Leapfrog and also implicit-explicit Runge-Kutta, Adams-Moulton, Backward Difference Formulas, and Leapfrog. All these methods are defined in a similar manner that the standard Leapfrog-Trapezoidal rule is defined. This allows NUMA to add new time-integrators quite easily. In Ref. [1] we compared all of these time-integrators and have found the IMEX Runge-Kutta methods to be the clear winners in terms of accuracy versus cost.

a) *Accuracy versus Courant Number*   b) *Accuracy versus Wallclock Time*
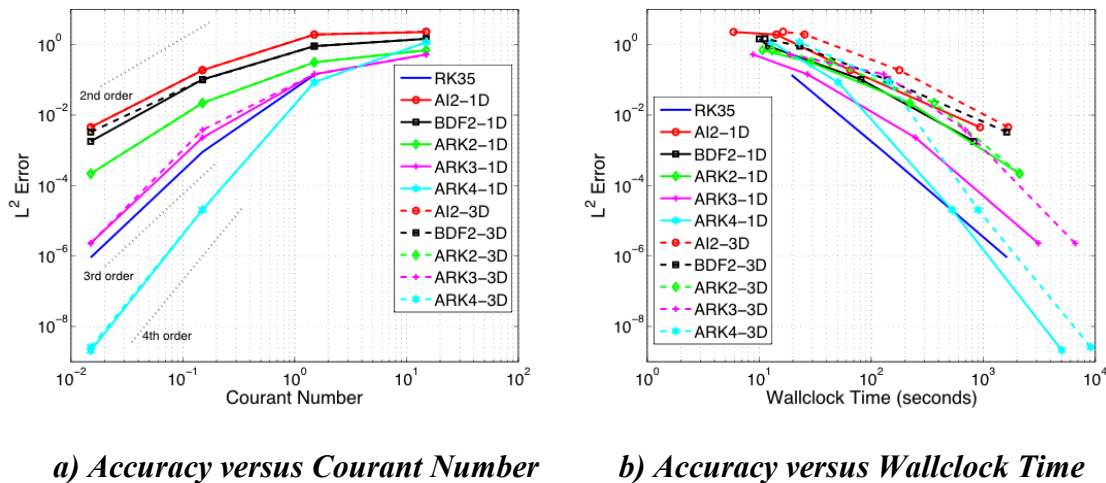
Figure 1: The a) accuracy and b) efficiency of NUMA using various time-integrators for a global atmospheric problem (propagation of an acoustic wave).

Figure 1 shows that the Runge-Kutta methods (ARK) yield high-order rates of convergence (left panel) while delivering these quality solutions very efficiently (right panel). For orders of accuracy below

10^-1 the 3rd and 4th order RK methods are the clear winners. Another interesting topic that is rarely mentioned is the effect of the time-integrator on the conservation metrics of a model.

*a) Mass Conservation*   *b) Energy Conservation*
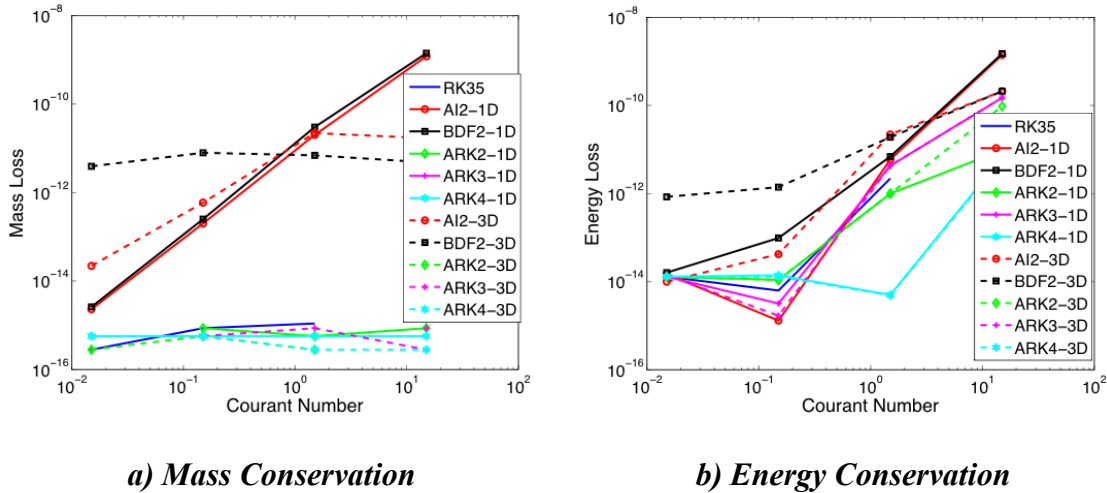
Figure 2: Conservation of a) mass and b) energy for NUMA using various time-integrators for a global atmospheric problem (propagation of an acoustic wave).

For example, the EBG methods used in NUMA should allow the model to conserve mass but not energy (energy is only conserved up to time-truncation error). Figure 2 shows that all of the Runge-Kutta methods conserve mass (left panel) while the BDF and Adams methods do not. Admittedly we are talking about errors on the order of machine zero but is still worth noting. The right panel shows the energy loss of the model where the Runge-Kutta methods yield far smaller errors in energy.

Since the equation sets used in NUMA can only conserve energy to time-truncation error then it makes sense to use high-order time-integrators especially for very long time-integrations. In addition, we have not discussed preconditioners and iterative solvers but these are discussed in detail in Ref. [2] where we derive a new class of preconditioners specifically designed for our IMEX methods.

Parallel Scalability. Building on previous scalability studies using smaller clusters located at NPS, we have deployed NUMA on several TeraGrid machines, include Kraken at NICS and Ranger at TACC. Optimized explicit versions of both CG-NUMA and DG-NUMA were deployed on Kraken, and a scalability study was run using processor counts ranging from 32 to 32,000 cores. Figure 3 displays the scalability of CG and DG models. A 32 x 32 x32 rising thermal bubble problem with 4th and 8th order polynomials resulting in 200,000 and 1.7 million grid points, respectively.
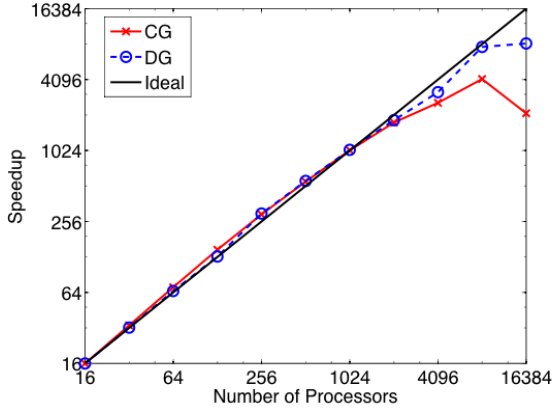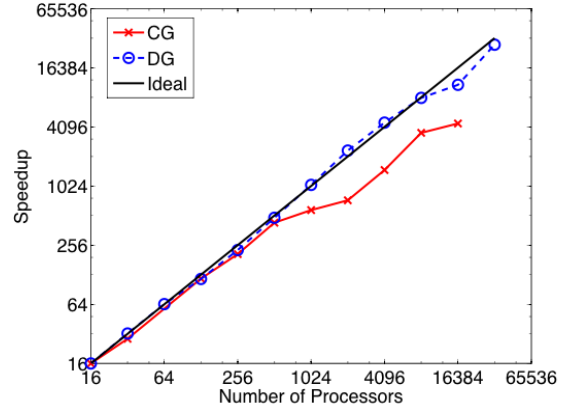
a) 4ᵗʰ Order Polynomials   *b) 8ᵗʰ Order Polynomials*

Figure 3: Scalability of CG-NUMA and DG-NUMA on TACC-Ranger using a 32 x 32 x32 rising thermal bubble problem (8ᵗʰ order) and 1.7 million grid points. Results are a) using 4ᵗʰ order polynomials and b) using 8ᵗʰ order polynomials.

The results in Fig. 3 show that at 4ᵗʰ order polynomials (left panel) both models shows strong scaling up to 8,000 cores; at this point, there are only 4 elements per processor. On the right panel we show the results using 8ᵗʰ order polynomials which shows that both models scale strongly up to 16,000 processors with the DG model continuing to scale strongly up to 32,000 cores; at this number of processors, there is only one element per processor. Note that we expect a similar behavior for the CG method but ran out of HPC time to complete this run. The results shown in these figures show a very important conclusion, that is, that in order to achieve strong scaling one must be able to do more work on the processor than is communicated. This can only be achieved for high-order methods and we have rigorously analyzed this in the Journal of Computational Physics paper in Ref. [3]. The analysis that we show is essentially a geometric argument whereby one can compute the computational volume and compare this amount of on-processor work to the surface area of the communication stencil; we show that this ratio has to be of a certain size in order to see the benefits of strong scaling. It should be noted that strong scaling is rarely achieved – most models only show weak scaling where the cost of running a model remains flat as both the problem size and number of processors are increased at the same rate. Strong scaling means that for a fixed problem size one will see perfect (linear) scaling as the number of processors are increased. The results in Fig. 3 show strong scaling for NUMA.

Positivity Preserving and Dispersion of CG. In Ref. [4] we included moist processes to NUMA and showed that NUMA, although a high-order model, is more efficient than a 2ⁿᵈ order finite difference model (such as COAMPS) when one considers not only the wallclock time but also the quality of the solution obtained. However, in order to handle moist processes, we required applying severe cut-off filter whereby any negatives are completely clipped. This will not conserve moisture in the atmosphere and so other strategies were explored.

7
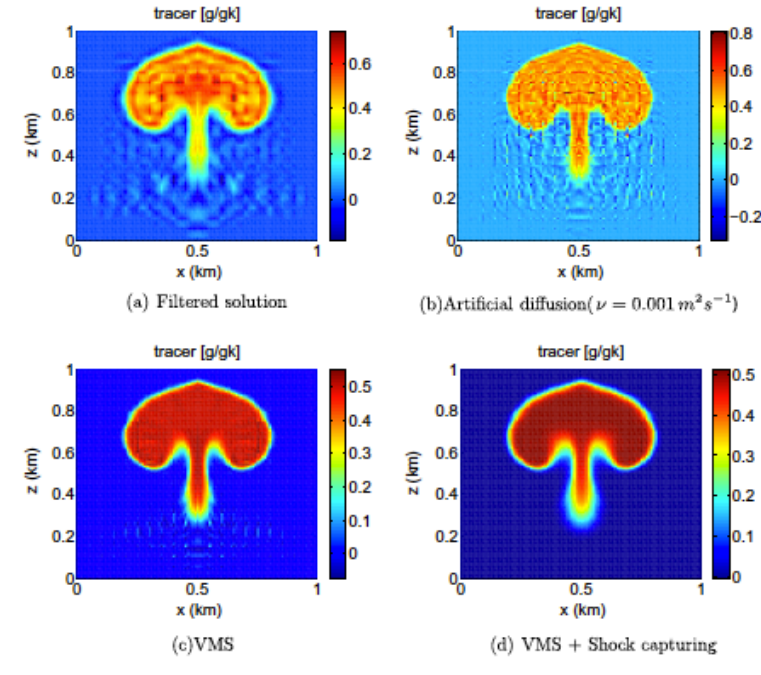
Figure 4: Transport of a discontinuous tracer using a) filtered solutions, b) artificial viscosity, c) variational multi-scale method, and d) variational multiscale method with shock capturing.

To this end, in Ref. [5] we explored various diffusivity strategies including classical CG filters, artificial viscosity, and a relatively new approach known as the Variational Multi-scale (VMS) method. In Fig. 4 we show the results after a long time for a discontinuous tracer embedded in the fully compressible equations for a rising thermal bubble problem. The results show that neither filtering nor artificial viscosity will help maintain positivity of a fully discontinuous tracer. On the other hand, VMS goes a long way in doing this and with the addition of shock capturing it can indeed keep the solution positive. We are currently exploring this approach for a moist physics problem in the next few months.

In order to understand why CG and DG methods capture wave phenomena so well, we decided to perform some dispersion analyses of these methods. This year we began with CG methods. We begin with studying the dispersion relations for CG for the 1D wave equation. In Fig. 5 we plot the dispersion relations for an 8th order polynomial (red dots) comparing with the analytic dispersion relation (blue line). We can see that at high-frequencies (far right of the plots) there is a gap in the dispersion spectrum which means that large dispersion errors are committed. On the right panel we perform the same analysis but adding a little dissipation (any of the methods from Fig. 4) and show that the dispersion behavior of CG looks like what one would expect from an accurate high-order method.

a) No Dissipation   **b) With Dissipation**
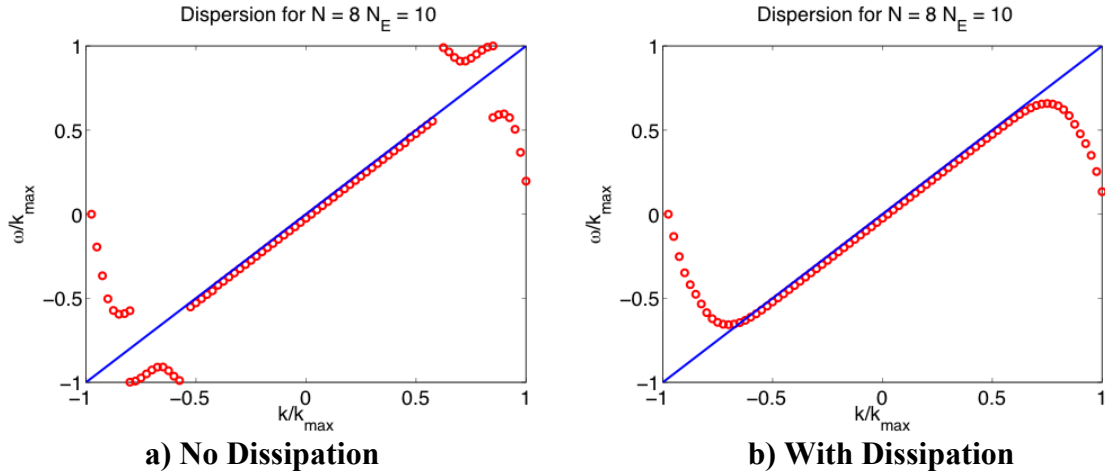
Figure 5: Dispersion analysis of the CG method: a) with no dissipation and b) with dissipation. The results shown are for an 8th order polynomial CG method.

Although the left and right panels in Fig. 5 look different, the overall result is the same, that is, that CG is extremely accurate (with no dispersion errors) for waves from the low-frequency (long waves) to the mid-frequency waves. It is only the waves in the far right (high-frequency waves) that look a bit different between the two plots. However, these waves are completely unresolved in any model and therefore insignificant. To see how the unfiltered and filtered solutions behave, we now show the time-series results for the 1D wave equation.

a) No Dissipation   **b) With Dissipation**
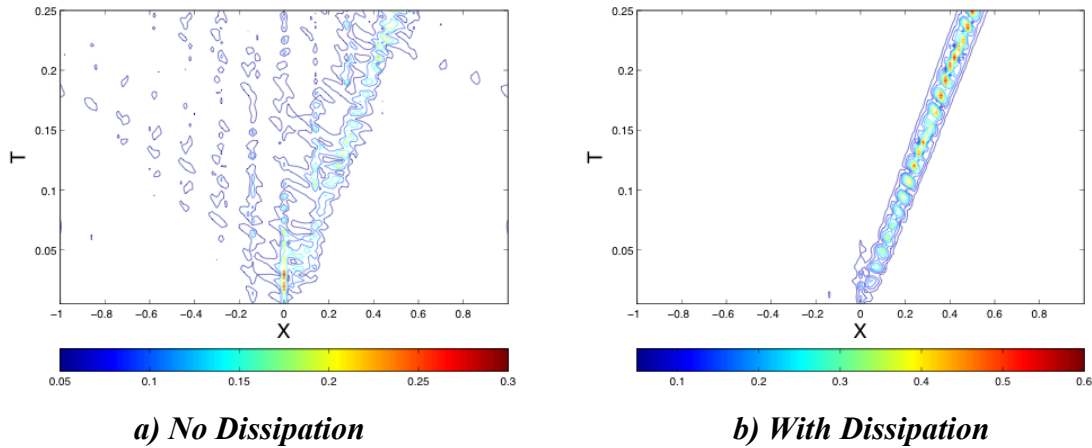
Figure 6: Dispersion analysis of the CG method: a) with no dissipation and b) with dissipation. The results shown are for an 8th order polynomial CG method for the 1D wave equation.

Figure 6 shows the time-behavior of a right-moving wave. The left panel shows the results for an 8th order polynomial with no dissipation whereas the right panel has the same result using one of the dissipation mechanisms shown in Fig. 4. Clearly, CG with dissipation yields an ideal result. One may argue that methods should be perfect with no dissipation but in reality no models are ever used without any dissipation mechanisms. In fact, they are essential to the proper cascading of energy from the large to the small scales.

IMPACT/APPLICATIONS

NOGAPS and COAMPS are run operationally by FNMOC and is the heart of the Navy's operational support to nearly all DOD users worldwide. This work targets the next-generation of these systems for massively parallel computer architectures. NUMA has been designed specifically for these types of computer architectures while offering more flexibility, robustness, and accuracy than the current operational systems. Additionally, the new models are expected to conserve many important quantities such as mass and use state-of-the-art time-integration methods that will greatly improve the capabilities of the Navy's forecast systems.

TRANSITIONS

Improved algorithms for model processes will be transitioned to 6.4 as they are ready, and will ultimately be transitioned to FNMOC.

RELATED PROJECTS

Some of the technology developed for this project could be used to improve NOGAPS in other NRL projects. The work on the mesoscale models will help improve COAMPS. An example is the time-integration methods that we are exploring for the new models may well be incorporated into the current operational version of COAMPS. In a separate Department of Energy (DoE) project, the Mathematics and Computational Science group at Argonne National Laboratory is working on interfacing NUMA with their highly scalable software PETSc (Portable Extensive Toolkit for Scientific Computing). This union will vastly increase the capability of NUMA through the access to the large suite of time-integrators, preconditioners, and solvers contained in PETSc. Furthermore, having NUMA being scrutinized by software and parallel computing scientists will allow us to improve NUMA.

PUBLICATIONS

Journals
1. Giraldo, F.X., Kelly, J.F., and Constantinescu, E.M., 2012: Implicit-Explicit Formulations for a 3D Nonhydrostatic Unified Model of the Atmosphere (NUMA). *SIAM Sci. Comp.* (in review).

2. Carr, L.E., Borges, C.F., and Giraldo, F.X., 2012: An element-based spectrally-optimized approximate inverse preconditioner for the Euler equations. *SIAM J. Sci. Comp.* (in press).

3. Kelly, J.F., Giraldo, F.X., 2012: Continuous and Discontinuous Galerkin Methods for a Scalable 3D Nonhydrostatic Atmospheric Model: limited-area mode. *J. Comp. Phys.*, Vol. 231, 7988–8008.

4. Gabersek, S., Giraldo, F.X., and Doyle, J.D., 2012: Simple Microphysics Experiments with a Spectral Element Model. *Monthly Weather Review* (in press).

5. S. Marras, J.F. Kelly, F.X. Giraldo, and M. Vazquez, 2012: Variational Multiscale Stabilization and p-adaptivity of High-Order Spectral Elements for the Convection-Diffusion Equation, J. Comp. Phys. Vol. 231, *7187-7213.*

6. Mueller, A., Behrens, J., Giraldo, F.X., and Wirth, V., 2012: Adaptive Discontinuous Galerkin Nonhydrostatic Model for Studying Convection. *Journal of Computational Physics (*in review*).*

Theses
1. Mugg, P. 2011: *Multi-rate Time-Integrators for Hyperbolic Systems of Equations*. Master's Thesis, Naval Postgraduate School (June 2012).

Conference Abstracts
1. Kelly, J.F., Giraldo, F.X., and G. Jost: Scalability of semi-implicit time-integrators for nonhydrostatic Galerkin-based atmospheric models on large-scale clusters. *Cluster Computing 2011* (September 2011).

Invited Talks
1. Giraldo, F.X., 2012: High-Order Methods for the Construction of Next-Generation Unified Atmospheric Models, Oregon State University, General Colloquium (February 2012).

2. Giraldo, F.X., 2012: Construction of the Schur Form for the Discontinuous Galerkin Method, Oregon State University, Department of Mathematics (February 2012).

3. Giraldo, F.X., 2012: Development of the Nonhydrostatic Unified Model of the Atmosphere: challenges in Adaptive Mesh Refinement. Plenary talk at the Newton Institute, Cambridge University (August 2012).

4. Giraldo, F.X., 2012: Discontinuous Galerkin Methods for Geophysical Fluid Dynamics. Department of Mathematics, University of Warwick UK (October 2012).