MEAD (A Modeling Environment for Atmospheric Discovery)

[1,11]Robert Wilhelmson, [1]Jay Alameda, Kelvin Droegemeier, [1]Michael Folk, [3]Rob Fowler, [4]Dennis Gannon, [5]Sara Graves, [6]Dale Haidvogel, [7]Parry Husbands, [8]Charles Lee Isbell, Jr., [2]Dan Weber, [9]Paul Woodward, [10]Bryant W. York, [9]Sarah Anderson, [11]Brian Jewett, [12]Christopher Moore, [13]David Nolan, [9]David Porter, [1]Dave Semeraro, and [5]Steve Tanner.

[1]National Center for Supercomputing Applications
University of Illinois
Champaign-Urbana, Illinois

[2]Center for Analysis and Prediction of Storms
University of Oklahoma
Norman, Oklahoma

[3]Rice University
Houston, Texas

[4]University of Indiana
Bloomington, Indiana

[5]University of Alabama in Huntsville
Huntsville, Alabama

[6]Rutgers University
New Brunswick, New Jersey

[7]Lawrence Berkeley National Laboratory
Berkeley, California

[8]Georgia Institute of Technology
Atlanta, Georgie

[9]University of Minnesota
Minneapolis, Minnesota

[10]Portland State University
Portland, Oregon

[11]University of Illinois
Champaign-Urbana, Illinois

[12]Pacific Marine Environmental Laboratory
Seattle, Washington

[13]University of Miami
Miami, Florida

The goal of the MEAD Expedition is the development and adaptation of Grid and TeraGrid-enabled cyberinfrastructure for enabling ensemble or very large domain model simulations coupled with data handling, analysis, data mining, and visualization services. This includes a dynamic workflow and data management environment applicable in a variety of fluid flow modeling environments. The specific applications chosen for MEAD are mesoscale storm and hurricane research and education. The MEAD Expedition is a cyberinfrastructure proving ground that has been funded for two years by the National Computational Science Alliance, an NSF PACI program. The MEAD project is

documented at http://www.ncsa.uiuc.edu/AboutUs/FocusAreas/MEADExpedition.html . An extended abstract with information on what has already been accomplished during MEAD's first year will be available at this web site in January along with plans for the final year of MEAD.

More specifically, Portal Grid and Web infrastructure in MEAD will enable launching of hundreds of individual Weather Research and Forecasting (WRF), Regional Ocean Modeling System (ROMS), or coupled WRF/ROMS simulations in either ensemble or parameter mode. Discovery and use metadata coupled to the resulting terabytes of data will then be made available to enable further exploration. Thus, a user of the MEAD workflow will be able to configure and integrate model simulations, manage resulting model and derived data (10's to 100's of terabytes), and analyze, mine, and visualize large model data suites in a distributed research (not predictive) context. Finally, very large domain research fault-tolerant simulations will be enabled through grid services and decomposition techniques that can be utilized efficiently on the new TeraGrid architecture.

MEAD efforts have been organized into seven areas and associated working groups. In addition to those funded through the MEAD expedition, there are participants from four other Alliance expeditions, EOT (PACI Education, Outreach, and Training Program), and other unfunded efforts at NCAR, ANL, and NOAA. An overview of activities in each area will be presented. They are

Collaborative Analysis and Visualization (CAV)

Existing visualization tools are being adapted for use in the MEAD environment and include stereo GeoWall display, simultaneous display of multiple simulations on a tiled display wall, workflow-driven visualizations, and collaborative analysis and visualization using personal Access Grid technology in collaboration with SWOF. Software includes NCAR graphics, the Java-based Interactive Data Viewer (IDV) built on VisAD, the parallel ParaView visualization system, the Hierarchical Volume Renderer, TeraVision for streaming animation, collaborative Access Grid 2, and a shared image markup tool for visualization and annotation of image data.

Coupling and I/O (CIO)

The Weather Research and Forecasting Model (WRF) and the Regional Ocean Modeling System (ROMS) are being coupled for execution in a distributed grid environment with the aid of functions within the Model Coupling Toolkit (MCT). HDF5 is being added as an I/O option in addition to netCDF in both models. Reading of either type of file will be facilitiated using ESML (Earth System Markup Language). ADAS (ARPS Data Assimilation System) has been adapted for interpolation to WRF grids and includes assimilation of Doppler radar data.

Data Mining and Machine Learning (DMML)

Data mining and machine learning tools are being adapted to help researchers analyze data from 10's to 100's of model simulations (ensembles and parameter studies) distributed across the grid. This includes current and new functionality provided

through datamining services from ADaM that includes a variety of algorithms and some visualization capabilities. Feature-detection algorithms are being adapted and new approaches are being explored such as the use of the grid-enabled MATLAB/Parallel Problems Solver environment and application of a hypergraph clustering algorithm.

Education (ED)

Understanding of model uncertainty in weather, including the prediction of hurricane strength, behavior, and human impact, will be aided through curriculum development and provision of simplified MEAD related model, analysis and visualization tools. The materials focus on inquiry-based learning with hands-on experimentation and bring students into the world of the scientific researcher and weather forecaster. This work is being done in collaboration with PACI Education, Outreach, and Training investigators.

Metadata and Data Management (MDM)

Stored data from model simulations carried out in the MEAD environment will be accessible through user-specified and automatically-generated metadata provided through catalogue and Grid Data Services. The user will eventually be able to publish derived data, visualizations, and textual reports/annotations using this system. This work is being done in collaboration with the PACI Data Quest Expedition and in consort with the Earth System Grid effort (NCAR, ANL, and LLNL), the Metadata Catalogue Services (MCS) effort (ISI), and other efforts.

High Performance Grid Computing (HPGC)

Improvement of model performance on parallel clusters for WRF and ROMS is being carried out in collaboration with the Performance Expedition using LoopTool for loop fusion, HPCview for compile time performance improvement and Prophesy for run-time performance improvement. The MEAD workflow will include allocation of grid resources using a form of GrADS. A fault-tolerant grid computing paradigm is being implemented using SHMOD for carrying out very large fluid simulations.

Portal Interface and Workflow (PIW)

The MEAD-PI interface is being developed for specifying model parameters for suites of simulations that will be executed on the Grid and subsequently analyzed/mined and visualized. Workflow management, including support for coordination of job submission, scheduling of model simulations, analysis, and visualization, and collation of model simulations, is being carried out with software developed in collaboration with the Alliance Portal Expedition in a framework referred to as the MEAD Application Grid Services Framework (MEAD AGSF).

The MEAD project will officially be over at the end of FY2004. Key deliverables include the coupled WRF/ROMS model; a portal and workflow environment for carrying out multiple simulations, data mining, machine learning, and visualization activities; a framework for building fault tolerant numerical codes, and educational modules exploring the concept of uncertainty. The work on portals and workflow in MEAD will

be continued in a recently funded large ITR (Information Technology and Research) project entitled LEAD (Linked Environments for Atmospheric Discovery). This five year project will focus on carrying out research that will enable the construction of a real time, on-demand, fault tolerant forecasting environment for the university community (see LEAD extended abstract immediately preceding this one in the IIPS schedule).. Whereas MEAD focuses primarily on modeling in a research environment (not real-time), LEAD will focus on real- time environments and associated observational data and will include WRF (not WRF/ROMS) as the primary modeling engine. .