6.5 THE NOAA OPERATIONAL MODEL ARCHIVE AND DISTRIBUTION SYSTEM (NOMADS): A STATUS REPORT

Glenn K. Rutledge *[1], Jordan Alpert[2], Ron Stouffer[3], Steve Hankin[4], Brian Doty[5], Ethan Davis[6], Bryan Lawrence[7], Mike Fiorino[8], Dean Williams[8], Lola Olsen[9], and Lawrence Buja[10]

[1]National Climatic Data Center
Scientific Services Division
Asheville, NC 28801

[2]National Centers for Environmental Prediction
Environmental Modeling Center
Camp Springs, MD 20736

[3]NOAA Office of Atmospheric Research
Geophysical Fluid Dynamics Laboratory
Princeton, NJ 08542

[4]NOAA Office of Atmospheric Research
Pacific Marine Environmental Laboratory
Seattle, WA 98115

[5]Institute of Global Environment and Society, Inc.
Center for Ocean-Land-Atmosphere Studies
Calverton, MD 20705

[6]University Corporation for Atmospheric Research
Unidata Program Center
Boulder, CO 80301

[7]British Atmospheric Data Centre
Rutherford Appleton Laboratory
Chilton, Didcot, OX11 0QX, U.K

[8]Lawrence Livermore National Laboratory
Program for Climate Model Diagnosis and Intercomparison (PDMCI), Livermore, CA 94550

[9]National Aeronautics and Space Administration
Goddard Spaceflight Center
Greenbelt, MD 20771

[10]University Corporation for Atmospheric Research
National Center for Atmospheric Research
Boulder CO. 80305

1. BACKGROUND

To address a growing need for real-time and retrospective General Circulation Model (GCM) and Numerical Weather Prediction (NWP) model output and model input, the National Climatic Data Center (NCDC) along with the Geophysical Fluid Dynamics Laboratory (GFDL) and the National Centers for Environmental Prediction (NCEP) initiated the collaborative NOAA Operational Model Archive and Distribution System (NOMADS) (Rutledge, et al., 2002). This collaboration of systems, allows access to weather and climate model data sets. A new paradigm for sharing data among climate and weather modelers is evolving. It takes advantage of the Internet and relatively inexpensive computer hardware.

Initiated in 2000, the NOMADS pilot project supports, leverages, and advances established and emerging technologies to access and integrate model and other data stored in geographically distributed repositories in heterogeneous formats. Today, this model for data access is generally accepted with both providers and users adopting and promoting the distributed data access philosophy. In this framework, scientists place their data into a computer on the Internet. Software allows outside users to see not only their local data but also data on other computers running this same software.

* *Corresponding author address*: Glenn Rutledge, National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801; e-mail: Glenn.Rutledge@noaa.gov.

The goals of NOMADS are:

- Provide access to NWP (weather) and GCM (climate, including ocean related) models.
- Provide the observational and model data assimilation products for Regional model initialization and forecast verification for use in both weather and climate applications.
- Develop linkages between the research and operational modeling communities and foster collaborations between the climate and weather modeling communities.
- Promote product development and collaborations within the geo-science communities (ocean, weather, and climate) to improve operational weather and climate forecasts by allowing more users to interact with the model data.
- Foster inter-disciplinary research to study multiple earth systems using collections of distributed data under sustainable system architectures.
- Ensure permanent stewardship of select agreed upon model data sets.

NOMADS promotes the sharing and inter-comparing of model results and the comparing of model results with observations. It is a major collaborative effort, spanning multiple government agencies and academic institutions. The data available under the NOMADS framework include NCEP data assimilation model input for National and Regional modeling activities; gridded NWP output, and Atmospheric-Ocean General Circulation Models (AOGCM) and simulations from GFDL, the National Center for Atmospheric Research (NCAR), and other leading

institutions from around the world. The effort has gained many international partners, as the need for a convergence of emerging yet similar distributed data efforts become clear to organizations. The NCDC NOMADS is one of many distributed data access projects gaining popularity across the world.

2. WHY NOW?

A major transition in weather and climate prediction is now occurring, one in which real-time and retrospective Numerical Weather Prediction (NWP) and global climate AOGCM research is spreading from a handful of national centers to dozens of groups, both public and private across the world. This growth of global and regional scale NWP and GCM model development is in part now possible due to (1) the availability of low-cost multiprocessor workstations; (2) the availability of regional scale models that run on these workstations (e.g., WRF, MM5); and (3) the availability of climate simulations, analysis and forecast grids from NCEP, GFDL, NCAR, the Center for Ocean-Land-Atmosphere Studies (COLA), and other institutions.

NOMADS addresses model data access needs as outlined in the U.S. Weather Research Program (USWRP) Implementation Plan for Research in Quantitative Precipitation Forecasting and Data Assimilation to "redeem practical value of research findings and facilitate their transfer into operations." And more recently, the Strategic Plan for the U.S. Climate Change Science Program (CCSP). The NOMADS framework was also developed to facilitate climate model and observational data inter-comparison issues as discussed in documents such as the Intergovernmental Panel on Climate Change (IPCC 1990, 1995, 2001) and the U.S. National Assessment (2000). NOMADS is being developed as a unified climate and weather archive so that users can make decisions about their specific needs on time scales from days (weather), to months (El Nino), to decades (global warming).

One of the primary objectives of the NOMADS project is to promote collaboration between institutions, and a feedback mechanism between the new regional modeling community and the operational data centers such as NCEP. Another objective is to provide vertical and lateral boundary conditions from the NCEP data assimilation analysis cycling systems for use in regional models for initial conditions, and for local and climate related efforts such as downscaling.

To ensure that agency and institutional requirements are being met, the NOMADS collaborators have established science and technical expert teams. These newly established teams comprised of NOMADS members will ensure system and data inter-operability; and develop data archive requirement recommendations to the NESDIS Data Archive Board (DAB).

3. SYSTEM DEVELOPMENT

Early in the development of the NOMADS system at NCDC, one of the known challenges facing the NOMADS Team was the design of a data management structure for high volume binary data. A consistent archive with quality control was desired, and a backup for data ingest was required to ensure for continuity of the data archive. Along with distributed access to the on-line data, access to the entire archive would also be required.

For interoperable data access across the collaborators the team finally focused on an existing but now mature binary transport protocol called OPeNDAP (formally called DODS, (Davis, et al., 1999).

Once the transport and format independent metadata model was agreed upon by the NOMADS collaborators, an internal team within NCDC composed of representatives from four Branches cooperated to implement this fully automated ingest and access system. Under the NOMADS umbrella the receipt, error checking/quality control, logging, monitoring, access and archive of these new data within NCDC's archive is now operational. These Branches include the Active Archive Branch, the Data Access Branch, the Information Technology Branch and the Climate Analysis Branch which leads the project at NCDC. The cooperation across so many branches within NCDC for the ingest and servicing of these new data, and in a new and open distributed environment, is noteworthy.

For further information on the collaboration, and access to various models see the model resources and NOMADS sites at NCDC at: www.ncdc.noaa.gov/oa/model/model-resources.html.

4. CURRENT STATUS

Each day the NOMADS at NCDC ingests over 50K NWP grids in real-time from three sources: the NWS NOAAPort broadcast (Rutledge, et al., 2000), the Unidata Internet Data Distribution (IDD) network, and NCEP. The three sources of data, two being redundant provide the highest level of quality in the archive as real-time QC can be performed on the duplicative data feeds, and provide for a once per day, "Last Look" to ensure for a serially complete model archive.

Upon receipt at the NCDC NOMADS servers, the model data are scanned and indexed for quality control and ease of access from the archive. Several data management tasks are also performed before the data are archived, to expedite access such as alignment of the forecast projections within a single model and directory. This more closely resembles how most users access the data. It also better aggregates the data for bulk time-series access used predominately for climate analysis and monitoring. Active data monitoring tools were developed specifically for the NOMADS system and archive, and provide a robust automated ingest and archive system. For example as part of the quality control routine for NOMADS, the monitoring software discovered WMO header errors contained within several models on the NOAAPort broadcast. This fact was reported to the NWS and an AWIPS-wide modification had to be implemented nationally.

During 2003 the NOMADS project exceeded its stated objectives for the year. One of the most significant of those goals was the establishment of a permanent U.S. archive of NCEP NWP model output. Another objective

met was of course access to that archive in both traditional web, and distributed client-server technologies.

Three data access portals have been built into the NOMADS systems at NCDC to accommodate most of the known users of these data. Traditional data access methods are available such as http (Web Browse and plotting), ftp (raw grib), and wget etc. Users can use traditional on-line web services to browse, plot, print, plot long term time series on any selected parameter; Hovmoller diagrams on-line using both near real-time data or archived data.

Non-traditional access is provided using the GrADS-DODS Server (Doty, et al., 2001), developed at COLA. GDS allows users to subset data in space and time, and perform limited remote calculations on data using the host NCDC server. The number of users continues to grow while knowledge on how to use these non-traditional interfaces also grows.

5. DATA AVAILABILITY

5.1 *Numerical Weather Prediction Output*

Data that are currently accessible both on-line and in NCDC archive include GFS, Eta, and RUC-II models. The models on the NOAAPort broadcast are limited in space and temporal resolutions, and in physical parameter. However the NOAAPort is the NWS operational feed. The ingest, cataloging (metadata generation for fast, open, and distributed access), and archiving of these the NOAAPort gridded products acted as the stepping stone and as a proof-of-concept for the higher resolution models generated at NCEP- both now and in the near future. Model data are available from May 2002 to the present, with data extending to 1999 upon request.

Beginning in November 2003 NCDC NOMADS began the ingest and servicing of the higher resolution models as provided by the NCEP NOMADS system called R/T NOMADS (Alpert, et al., 2002). This service is provided for a seamless access to the high resolution real-time data sets at NCDC, and retrospective data sets at NCDC. The completeness of the high resolution archive is being determined, but it is expected that approximately 10-20 TB of data will be archived per year, on a five year rotation. The mirrored NCDC site was also installed for a backup to the NCEP R/T NOMADS.

5.2 *Model Input: Data Assimilation from NCEP*

Publicly available on the NOMADS servers at NCEP and NCDC is the Global Data Assimilation System (GDAS) dataset. GDAS consists of the minimum set necessary to re-generate NCEP analysis and forecast products (model re-start and initialization files). GDAS includes the Global Spectral Forecast Model (GSM), and the Spectral Statistical Interpolation (SSI) Cycling Analysis System (CAS) with triangular truncation (T) 170 and 28 sigma levels. To start the CAS, model spectral coefficients are provided on Gaussian grid in a sigma vertical coordinate system. These data represent the model's "ground truth", and the best estimate - in terms of analyzed fields - for scientific study. Data that are restricted may not be available.

"Post" is a FORTRAN program available from NCEP that will convert spectral coefficients to gaussian grid, sigma to pressure, and gaussian to latitude and longitude. Work continues to couple Post to the NOMADS user interface for source and executable downloads. The GDAS dataset under NOMADS will include the GSM and the SSI-CAS binary files and contains ~2.5Gb per day (4 cycles/day: 00Z, 06Z, 12Z, and 18Z). The binary files are raw data, which are acted on by NOMADS servers to produce useful grids.

The archived analysis data sets serve as model verification as well as the source for model reruns and retrospectives. Including the observations allows for cycling analysis systems to re-analyze the observations. Never before has this model input data and information been available to the public. A partial list of NOMADS planned available observations (with associated data format) include:

- Analysis Bias Corrected Information (ASCII)
- Ship / Buoy Observations (BUFR)
- Guess prep / guess output (BUFR)
- Observational Toss List (ASCII)
- Bogus Observations (BUFR)
- ACARS and Aircraft (BUFR)
- Wind Observations (BUFR)
- Analysis Ready Obs. (prepBUFR)
- Surface Analysis Restart Files (BUFR)
- Surface and Upper-Air observations (BUFR)
- Fixed Snow Field (GRIB)
- Previous 6 hour forecast (BUFR)
- "Post" Guess Output (spectral binary)
- Profiler (BUFR) / SST's (GRIB)
- MSU 14 and HIRS 14/15/16 (IEEE)
- SSMI Satellite obs (BUFR)
- NOAA (satellite) 15/16 AMSU - A/B
- TOVS 1B Radiances (IEEE)
- TOVS Satellite Obs (BUFR)
- GOES Satellite Obs (BUFR)
- O3 Sat Obs (binary) and ERS Sat obs
- SBUV: Satellite Wind Observations
- Radar VAD Winds (BUFR)

The format of these data sets is generally dictated by the necessity to run models efficiently on modern computers. NOMADS converts the formats and structure to the users requested form but also allows the raw data to be directly accessed.

5.3 *North American Regional Reanalysis*

The new North American Regional Reanalysis (NARR) became available at the NCDC NOMADS servers in December 2003. The NARR is a 32km, 45 layer, EDAS derived system, from 1979 thru 2003, updated monthly. Features include a new land surface model, direct satellite radiance retrievals, and for the first time in any reanalysis, precipitation. It is expected that this data set, in early 2004, will be widely used across the weather, climate and oceanographic communities.

5.4 *GFDL AOGCM*

AOGCM models available under the NOMADS framework include the GFDL R-30 climate model. R-30 is a coupled AOGCM. Its four major components are an atmospheric spectral GCM, and ocean GCM, and relatively simple models of sea ice and land surface processes. The name "R30" is derived from the resolution of the atmospheric spectral model (rhomboidal truncation at wave number 30). The R30 model is identified as GFDL_R30c in the Third Assessment Report (TAR) of the Intergovernmental Panel on Climate Change (IPCC). See Chapter 9, Table 9.1 of "Climate Change 2001: The Scientific Basis" IPCC Working Group I (2001).

The model output that is stored on the GFDL NOMADS server are taken from six experiments conducted at GFDL using the GFDL_R30c model. Data files produced by one long-running control integration (one with no changes in external forcings, e.g., constant CO_2) and five climate change scenario experiments have been made accessible to interested researchers. More information regarding the individual experiments is found in the Control & Transient Forcing Experiments section of the GFDL NOMADS Web site. The GFDL R-30 numerical model was developed and the experiments conducted by members of the Climate Dynamics and Prediction Group at the GFDL in Princeton, New Jersey. For further information on current available GCM data see the NOMADS site at GFDL at http://nomads.gfdl.noaa.gov.

6. RELATED PROJECTS AND OUTREACH

6.1 *The Weather Research and Forecast Model*

NOMADS is a technology for collaboration and model development as exemplified by the Weather Research and Forecast (WRF) model. The contribution of the scientific community to the WRF development effort requires the transfer of data sets of model retrospective results for testing and comparison. NOMADS enables both retrospective and real time access to the suite of digital products from reanalysis and operational results to give a range of users, from commercial to university scientist, model comparison access and supply operational grade initialization for forecast model study. For more information regarding the emerging WRF model see www.wrf-model.org.

6.2 *CLASS and NOMADS*

The NOMADS is actively partnering with existing and development activities especially with the Comprehensive Large Array Stewardship System (CLASS); also with the National Oceanographic Partnership Program's (NOPP) National Virtual Ocean Data System (NVODS); the Department of Energy's Earth System Grid (ESG); and the Thematic Real-time Environmental Data Distributed Services (THREDDS) project being developed through the National Science Foundation and Unidata.

Given the momentum, and success of NOMADS, NCDC has just initiated a process to analyze potential systems and data overlap between NOMADS and the CLASS systems. At the very least, a synergy between the two projects will assist both projects to leverage existing efforts, and future requirements.

6.3 *CEOS-Grid*

The Committee on Earth Observation Satellites established a pilot network of data servers to prototype Grid technologies. NOMADS was selected as one of the test bed applications for CEOS-Grid. The NOMADS CEOS-Grid effort successfully installed the inter-operable Globus Grid toolkit on a Linux based server now installed within NCDC's firewall. Globus is the grid software package from the Department of Energy's ESG project to communicate over the Internet or Internet-2 using secure and sub-setting client server distributed processes. NCDC's CEOS- Server communicated with the CEOS-Grid Team central facility and demonstrated transmission testing.

During FY03, the CEOS-Grid Team revised it's FY03-FY04 goals for application partnerships to better define align itself with known agency data and application requirements. A proof-of-concept final demonstration by FY05 is now required (vice FY04). The NCDC NOMADS will now develop prototype application services with the CEOS-Grid partners University of Alabama (Huntsville); the NPOESS; and the CEOP (an international hydrologic experiment) project to provide NOMADS model data over CEOS-Grid.

6.4 *NASA's Global Change Master Directory*

NASA's Global Change Master Directory (GCMD) is a NOMADS collaborator and provides descriptions of Earth science data sets and services relevant to global change research. The GCMD database includes descriptions of data sets covering agriculture, the atmosphere, biosphere, hydrosphere and oceans, snow and ice, geology and geophysics, paleoclimatology, and human dimensions of global change. The DODS portal at the GCMD can be reached at the following url: gcmd.gsfc.nasa.gov/Data/portals/dods/index.html.

The GCMD has made progress on three pivotal fronts, which positively affect discovery and access to data and model output. They have (1) provided for access directly to data through the directory and have provided access through an Open Applications Programmer Interface (API); (2) developed an operational distributed database with partners from the United Nations Environment Program (UNEP) and the Joint Committee on Antarctic Data Management (JCADM); (3) developed the capability of creating customized virtual subsets ("portals") from their extensive holdings. With attention to the needs of the modeling community, the directory now offers a portal for model output data sets. See: http://gcmd.nasa.gov/Data/portal_index.html and more specifically http://gcmd.nasa.gov/Data/portals/models/.

GCMD representatives also actively participate in the Earth Science Portal (ESP) effort to bring the modeling community together in improving access to modeling data.

In the coming year, the GCMD plans to (1) release an improved design of the web site; (2) offer a new spatial

and temporal resolution search; (3) upgrade the geographic search; and (4) finalize the ISO19115 version of the metadata standard.

6.5 *The Earth System Portal*

The ESP is a collaboration designed to build the infrastructure needed to create web portals to provide access to observed and simulated data within the climate and weather communities. The infrastructure created within ESP will provide a flexible framework that will allow interoperability between the front-end and back-end software components. The ESP collaboration was formed as an extension of the original NOMADS collaboration to address a common portal design.

6.6 *Outreach and Briefing Activities*

During 2003 NOMADS was invited and briefed the following projects and organizations:

- NOAA Climate and Global Change WG
- NOAA Climate Monitoring WG
- NOAA Science Advisory Board
- NVODS/IOOS/GCOS
- NWS Office Level Directors and NWS CIO
- Earth Science Portal (ESP)
- NOAA Tech and Web forums
- Cooperative Agreement for Climate and Weather Impacts on Society and the Environment (CWISE).

7. THE FUTURE

Even with the emergence of new web based services, the comparison of GCM results with the observational climate record is still difficult for several reasons. One limitation is the global distributions of a number of basic climate quantities, such as precipitation or clouds, are not well known. Similarly, observational limitations exist with model re-analysis data. Both the NCEP/NCAR (Kistler, Collins, Kalnay, et al., 2001) and the ECMWF (Gibson, et al., 1997) re-analysis eliminate the problems of changing model analysis systems but observational data also contain time-dependant biases by changing observational networks, station moves, and the assimilation of various remotely sensed data (Rutledge, et al., 2001) using differing sensor instruments, or calibrations. These changes in input data are blended with the natural variability making estimates of true variability uncertain. The need for data homogeneity is critical to study questions related to the ability to evaluate simulation of past climate. One approach to correct for time-dependant biases and data sparse regions is the development and use of high quality "reference" data sets (Karl, et al., 2000).

Now that the underlying system design is in place other data sets can be easily be added to the NOMADS lists of data services. At NCDC during 2004 this includes Ocean and Ensemble forecasting models, and observational data sets such as the Smith-Reynolds Extended Reconstructed SST's; the CARDS upper air reference data set, and the Global Historical Climatology Network (GHCN) data set.

During 2004 NCDC NOMADS will install and configure the new Live Access Server (LAS) (Hankin, et al., 2001). The LAS group is a key player in the development and promotion of the distributed data access philosophy. LAS at NCDC will also allow for enhanced data manipulation and browse capabilities.

8. CONCLUSION

Both researchers and policy-makers now expect national data assets to be easily accessible and interoperable with each other, regardless of their physical location. As a result, an effective interagency distributed data service requires the coordination of data infrastructure and management extending beyond traditional organizational boundaries.

With NOMADS and its collaborators, users at any skill level will be able to obtain weather and climate information. This will allow the users to make better, informed decisions about how nature will impact their future, either in their life, or business decisions.

9. ACKNOWLEDGMENTS

The authors gratefully acknowledge the combined efforts of all the NOMADS Team members for the growing success of NOMADS. Thanks also to Don Hooper, Jennifer Adams, and Howard Diamond. Special thanks to the internal team at NCDC: Ted Smith, Alan Hall, Ken Schmidt, David Bowman, John Fauerbach and Danny Brinegar. Pilot funding has been provided by ESDIM 02-444E, and the NESDIS CIO's office; and finally direct support from the Director's of NCDC, NCEP, and GFDL.

10. BIBLIOGRAPHY

Alpert, J.C., 2002, Rutledge, G.K., Williams, D., Stouffer, R., Buja, L., Doty, B, Hankin, S., Domenico, B., Kafatos, M., 2002, "The Plan to Access Real-Time NWP Operational Model Data Sets using NOMADS", Proceedings of the 18th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, American Meteorological Society, Orlando FL. 73 pp.

Davis, E. R., J. Gallagher, J., 1999, "Using DODS to Access and Deliver Remote Data", Proceedings of the 15th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, American Meteorological Society, Dallas, TX. 571 pp.

Doty, B.E., Wielgosz, J., Gallagher, J., Holloway, D., 2001, "GrADS and DODS", Proceedings of the 17th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, American Meteorological Society Albuquerque, NM. 385 pp.

Gibson, J. K., P. Kallberg, S. Uppala, A. Noumura, A.

Hernandez, and E. Serrano, 1997, "ERA Description. ECMWF Re-Analysis Project Report", Series 1, ECMWF, Reading, UK, 77 pp.

Hankin, S., J. Callahan, and J. Sirott, 2001, "The Live Access Server and DODS: Web visualization and data fusion for distributed holdings", 17th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, American Meteorological Society, Albuquerque, NM. 380 pp.

Karl, T., D. Easterling, P. Groisman, et al., March 2000, "Observed Variability and Trends in Extreme Climate Events: A Brief Review", Bulletin of the American Meteorological Society, 81, 417 pp.

Kistler, R., W. Collins, E. Kalnay, R. Reynolds, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, J. Janowiak, M. Kanamitsu, K. Mo, C. Ropelewski, R. Jenne, D. Joseph and M. Fiorino, 2001: The NCEP/NCAR 50-year Reanalysis: Monthly-means CD-ROM and Documentation. Bulletin of the American Meteorological Society, 92, 247 pp.

Rutledge, G.K., A. Stanley, E. Page, L. Spayd, and J. Brundage, 2000, "Implementation of the NOAAPort Data Archive and Retrieval System (NDARS) at the National Climatic Data Center", Proceedings 16th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Long Beach, American Meteorological Society, 492 pp.

Rutledge, G.K., T. Karl, D. Easterling, L. Buja, R. Stouffer, 2001 "Evaluating Transient Global and Regional Model Simulations: Bridging the Model/Observations Information Gap", (Invited), American Geophysical Union, Spring Meeting Boston MA S35 pp.

Rutledge, G.K., D. Williams, R. Stouffer, J. Alpert, L. Buja, B. Doty, S. Hankin, B. Domenico, M. Kafatos, 2002, "The NOAA Operational Model Archive and Distribution System (NOMADS)", Proceedings 13th Symposium on Global Change and Climate Variations, American Meteorological Society, Orlando FL. J76 pp.