

Darien L. Davis, U. Herb Grote, Carl S. Bullock
NOAA/OAR/Forecast Systems Laboratory,
Boulder, Colorado

Tom B. Kent, G. Joanne Edwards, James E. Ramer
Cooperative Institute for Research in the Atmosphere,
Colorado State University, Fort Collins, Colorado

Edward J. Mandel, Jason P. Tuell, Timothy Hopkins
NOAA/NWS/Office of Science and Technology,
Silver Spring, Maryland

1. INTRODUCTION

The explosion of forecast model data produced by the National Centers for Environmental Prediction (NCEP) as well as associated efforts (such as the WRF model development) creates a unique problem for the path to operations in the Weather Forecast Office (WFO). Software developed in the mid-1990's has reached it's capacity towards support of a proliferation of gridded data sets in the forecast office. A need for new hardware and software technology infusion exists.

Current metrics show that the NOAAPORT Satellite Broadcast Network broadcasts over one million products (expanding to over 15GB of data) daily. Of these data, approximately 150,000 grib-encoded grids are transmitted and processed each day at a forecast office. A case study at the Boulder, CO weather forecast office, monitoring every key click a forecaster makes on the D2D workstation, shows that only 575 grids are accessed more than ten times in a month. (Roberts, et al). This begs the question: why do forecast office need to ingest all of grids if only a small number of are actually accessed? The quantity of model data will only increase rapidly, creating challenges for the current data storage and retrieval schema. The file storage sizes are starting to exceed maximum sizes for operating systems. A new way to store and access data is necessary to keep up with growing demands.

2. BACKGROUND

The Forecast Systems Laboratory and the National Weather Service are supporting an exploratory development effort trying to infuse new technologies into the NWS forecast office. One of three major components of this development is defining a distributed data infrastructure that can be used in the operational forecasting environment. Current investigations point us toward the use of the NOAA Operations Model Archive and Distribution System (NOMADS) concept of web-based gridded data sets. NOMADS uses the OPen source project for Network Data Access Protocol (OPeNDAP) as well as the GrADS Data Server (GDS) developed by the Center for

Ocean-Land- Atmosphere Studies (COLA) to post data and enable visualization packages of gridded data sets. Servers exist at NCEP and the National Climatic Data Center (NCDC) providing retrospective as well as real-time data.

The interest in using these types of technologies are many-fold. A groundswell of software development exists supporting a web-based, data "pull" technology. The hydromet community, aided by the Unidata development community within the University Corporation for Atmospheric Research, supports an open-source development environment as well as source code repositories and a network for exchanging scientific data. It is no longer efficient to try to create in-house, application specific data management schemas. The cost and inflexibility of implementing and maintaining these efforts is prohibitive.

The National Climatic Data Center has spearheaded an effort to employ OPeNDAP for archiving data. NOMADS services an archive of model data with OPeNDAP and GDS. Similar NOMADS servers exist throughout NOAA and NASA, including a system at Forecast Systems Laboratory (FSL) serving RUC data and the Meteorological Assimilation Data Ingest System (MADIS). To encourage a streamlined, enterprise type of data management for NOAA, it behooves the National Weather Service to investigate the AWIPS program to closely examine the use of the "pull" technology supported at NCDC and NCEP. Although an enterprise architecture is not a charter for AWIPS, the leveraging power provided by more experienced NOAA users of web-based data access technologies is invaluable in these times of tight resources.

The overwhelming reason to use the OPeNDAP and GDS, however, is the support for netCDF (network Common Data Format). AWIPS has heavily relied on this UNIDATA developed software for storing and retrieving data. In fact, the netCDF interface provides a high level application interface for storage and retrieval of data. The possibility of accessing data, using the same netCDF calls, via a web address, was intriguing.

3. DEVELOPMENT

The FSL development is proposed as a two year

*Corresponding author address: D. Davis, Forecast Systems Laboratory, NOAA/R/FS4 325 Broadway, Boulder, CO 80305-3328

project starting spring, 2004. The first year has been spent investigating use of netCDF files of model data accessed via a web address. The second year examines the use of GDS and decoding grib-encoded grids. This is a very powerful alternative to current AWIPS data storage schema, and could be an extremely valuable alternative for the ever-growing amount of model data.

FSL has approached this effort with three self-defined constraints. First, the proposed software architecture must support a hybrid system. The current AWIPS data schema relies on locally-resident data for access and use at the forecaster workstation. The system is based solely on Network File System (NFS) for sharing data among workstations. Data is ingested, decoded, and stored in a hierarchical file system, using directory names as indexes to data type, scale, and other data-defining characteristics. Filenames are used to reflect the time coverage of data. A filename such as the following:

```
$FXA_DATA/radar/kftg/V/elev0_5/res0_25/level256/20041103_1639
```

is from a the kftg WSR-88D radar, velocity product, .5 degree elevation, 1/4 km radial resolution, 8bit color, radar scan starting at 16:39 on November 3, 2004. Although this has been an effective data storage/description schema, it has outgrown its usefulness. The data is constrained to one top node (defined by environment variable FXA_DATA), which makes distribution of data across several file systems impossible. NWS has experienced an obse infusion of data. The system designed in 1995 for NOAAPORT and radar data is now bursting at the seams with a dire need for infrastructure change. Any changes need to accommodate current data access as well as introduce new schemes. Therefore, the need for a hybrid system.

Second, the exploratory development is solely concentrating on accessing model data with the "pull" technology. In particular, the development is only being applied to GRIB encoded messages, not the new GRIB2 encoded data. The reason for these related controls are based on the amount of software available from the community. The GRIB2 has simply not been available long enough on NOAAPORT for non-NWS developers to have created the necessary support software. This would then require in-house developed efforts. Also, most software investigated was heavily skewed toward model data access; therefore, the initial development would take advantage of these efforts.

Third, FSL would apply this technology to locally created model data. Grids can be created by local models, Office of Hydrology applications, and output from the Graphical Forecast Editor (GFE). Currently, there is no elegant method to inherit these grids for display by the D2D application of the forecaster workstation. The D2D application has myriad tools available for perusing and fusing data. The need exists to merge these data sets.

Currently, FSL has staged a system using OPeNDAP and THREDDS, a UNIDATA developed cataloging application on an AWIPS data server. The storage software required a few changes, one of which was the use of symbolic links on every model file with a ".nc" extension due to constraints of the OPeNDAP software.

To access the data on a forecaster workstation, the AWIPS software uses a series of entries in tables to locate data. The current entry in a table has the following line in a configuration table:

```
|1|Grid/SBN/netCDF/CONUS211/NGM |ngm211  
|grid211 ...
```

This entry is the NGM 211 grid. To access these data over the web, the following modification is made:

```
|1|www |ngm211 |grid211
```

A new configuration file has been added to redirect data access to a web address. The associated entry in this file is:

```
NGM |dx1-alps |cgi-bin/nph-dods  
|gridNetcdf/CONUS211/NGM
```

When the "www" field is encountered, the software builds an address and opens the data file remotely. The netcdf calls to access data are then applied.

Two problems have arisen using with our current access of data. The AWIPS software uses the UNIX readdir function to return a listing of files in a directory. This listing has the following format:

```
20041004_0000  
20041004_1200
```

The software can then open these filenames to retrieve data for display. Since the data is remote and the site is unknown, software is required to retrieve the same information. The THREDDS catalog on the data server can provide this service. The catalog creates and XML document with the following entry excerpted:

```
<dataset name="gridNetcdf/CONUS211/Eta/">  
<dataset name="2004-10-04 12:00:00 GMT"  
urlPath="gridNetcdf/CONUS211/Eta/20041004_1200.n  
c"/>  
<dataset name="2004-10-04 00:00:00 GMT"  
urlPath="gridNetcdf/CONUS211/Eta/20041004_0000.n  
c"/>  
</dataset>
```

The access software for AWIPS has to be engineered to extract the listings from the XML document from the AWIPS C++ code.

The second problem is the need for a parameter inventory to expedite data access. Each parameter,, forecast time, and level has a binary byte (in the netCDF file) signifying existence of the data. Since a

model arrives over a period of time, there is often a sparse matrix of data available for a run. Using the OPeNDAP access of netCDF files, this is transparent and the information is available. However, the current netCDF file structure will soon be unusable. The size of a model run for 12km data requires clipping in order to store. More granularity in temporal and spatial areas will soon create huge system problems.

GDS accesses data from grib-encoded grids that do not require netCDF on the data server. A storage schema can be used to access these grids and decode on-the-fly from a web server. The data can still be accessed using the netcdf interface calls embedded in the AWIPS software. The challenge using the GDS is again replicating the inventories needed (as described in the previous paragraph).

The second year of development will also focus on populating information for updating menus and display. Current forecaster systems "auto-update" display of data -- no user intervention is necessary to populate new data displayed on a screen. No software exists for real-time inventory services using XML documents.

In order to implement this exploratory development

on more than an intra-office use, the bandwidth of the AWIPS wide-area network needs to be upgraded. Accessing data remotely and making sure that performance is optimized for a forecaster is a priority. These are areas that need to be addressed outside of this development effort.

4. SUMMARY

In summary, the efforts underway at the Forecast Systems Laboratory are promising. Introducing and positioning the future AWIPS systems for a data"pull" technology are underway, and a hybrid system is feasible for deployment in the next two years. Continued efforts in this arena can create a much more seamless introduction of model data from NCEP.

5. REFERENCES

Roberts, W.F. and L. K. Cheatwood, 2005: Examples of GFESuite and D2D Use in Operations During 2004. Preprints, 21st Conf. On Interactive Information and Processing Systems, San Diego, CA, Amer. Meteor. Soc., Boston, Mass. (in this volume).