P2.22

RECONSTRUCTION OF GRIDDED MODEL DATA RECEIVED VIA NOAAPORT

Lloyd A. Treinish*, Anthony P. Praino and Craig Tashman
IBM Thomas J. Watson Research Center, Yorktown Heights, NY

1. INTRODUCTION

Our on-going work focuses on systems for and applications of operational mesoscale numerical weather prediction. In particular, our goal is to provide weather forecasts at a level of precision and fast enough to address specific weather-sensitive operations. Hence, we are addressing problems of high-performance computing, visualization, and automation while designing, evaluating and optimizing an integrated system that includes receiving and processing data, modelling, and post-processing analysis and dissemination (Treinish and Praino, 2004).

In addition to considering both business and meteorological value of such mesoscale models in a number of application areas, we are also addressing whether a practical and usable system can be implemented at reasonable cost? To begin to answer these questions, a prototype system, dubbed "*Deep Thunder*", has been implemented for several metropolitan areas (initially, New York City, followed more recently by Chicago, Kansas City and Baltimore/Washington) with nested forecasts to 1 or 2 km resolution.

1.1 Data and NOAAport

One aspect of the integration and automation is the creation of appropriate initial and boundary conditions for the nested domain for each geographic area for which we produce an operational, model-based forecast. In that regard, we are using the results from the Eta synoptic-scale model operated by the National Centers for Environmental Prediction (NCEP), which covers all of North America and surrounding oceans at 12 km resolution with 60 vertical levels. Originally, these data were made available via the National Weather Service NOAAport satellite-based data transmission system after sampling to 40 km resolution on the AWIPS 212 grid every three hours and interpolated to 27 isobaric levels for the continental United States in a Lambert-Conformal projection. A subset of the surface fields are available at 20 km resolution on the AWIPS 215 grid, which we also utilize.

The NOAAport system provides a number of different data sources as disseminated by the National Weather Service (NWS). These include *in situ* and remotely sensed observations used currently for forecast verification as well as the aforementioned Eta data for model boundary and initial conditions. We currently operate a four-channel facility manufactured by Planetary Data, Incorporated (PDI., 2004). It was installed at the IBM Thomas J. Watson Research Center in Yorktown Heights, NY in 2000, and upgraded in 2003 and again in 2004. This NOAAport receiver system, based upon Red Hat Enterprise Linux (Workstation Version 3.0), has a very flexible design, enabling the type of customization and integration necessary to satisfy our project goals. The various messages transmitted via NOAAport are converted into conventional files in Unix filesystems in their native format, accessible via NFS mounting on both Linux and AIX systems via a private gigabit ethernet.

2. APPROACH AND RATIONALE

We have decomposed our processing of the Eta data into two parts. The first is essentially a parsing of the data received via NOAAport into usable formats to be used by the second part -- analysis and visualization. For the aforementioned Eta-212 and Eta-215 grids, the data are received in the compressed GriB (GriB-1) format. The data are uncompressed (deGriBbed) via an automated process, which is run as a periodic Unix cron job for each of the four Eta runs per day (0Z, 6Z, 12Z and 18Z). It provides a set of flat binary files (one per each three-hour time step) as input to several other processes, and a set of summary statistics. These processes include isentropic analysis, which leads to appropriate initial and boundary conditions for mesoscale model execution, forecast verification and comparison, and interactive and production, web-based three-dimensional visualization. The model lateral boundaries are nudged every three hours, using the processed Eta-212 grids. The isentropic analysis also incorporates other static and dynamic data (Treinish and Praino, 2004).

These consumers of the Eta data are relatively sensitive to degradation of the content of the GriB files that may occur due to problems at transmission or reception. For example, information about the former is not readily available a priori, although issues do arise on an irregular basis, which lead to retransmissions of the data at a later time. Problems at reception are typically due to aperiodic adjustments in the spacecraft signal strength, severe local weather or nearby flyovers of US Air Force AWACS planes. Since these factors are outside of our control, we determined that a simple format translation and related transformations (i.e., units, projections) is inadequate. Instead, we designed and implemented an approach that treats the input arrays of volumetric and surface fields from the GriB files as potentially incomplete samplings of a model atmosphere, and the output as a reconstructed representation of the intended Eta results. Depending on the extent of the degradation, different algorithms are introduced to compensate for data dropouts, intentionally incomplete population of specific arrays and data outside of a reasonable dynamic range.

An additional justification for this approach has emerged over the last year. NWS has plans in place to change the how data will be transmitted via NOAAport, some of which will be incompatible with current data such as the output from Eta (Klet, 2004). In particular, the NOAAport Direct Video Broadcast (DVB-S) chan-

*Corresponding author address: Lloyd A. Treinish, IBM T. J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598, `lloydt@us.ibm.com`, `http://www.research.ibm.com/people/l/lloydt`

nel will be used to transmit Eta data in a new format on the AWIPS 218 grid. This data stream will eventually replace the GriB-1-based 212 and 215 grids. Independent of the fact that the older, simplistic processing code used originally in our forecast operations does not handle the data degradation, it would be rendered useless after the NOAAport changes are put into operation. Hence, there is a clear need to migrate away from processing of the older data format.

At high-level, the data and procedural flow of the processes that utilize the NOAAport data streams within the *Deep Thunder* system are outlined in Figure 1. Most of them run serially on, although compiler-optimized for IBM Power processors (Treinish and Praino, 2004).

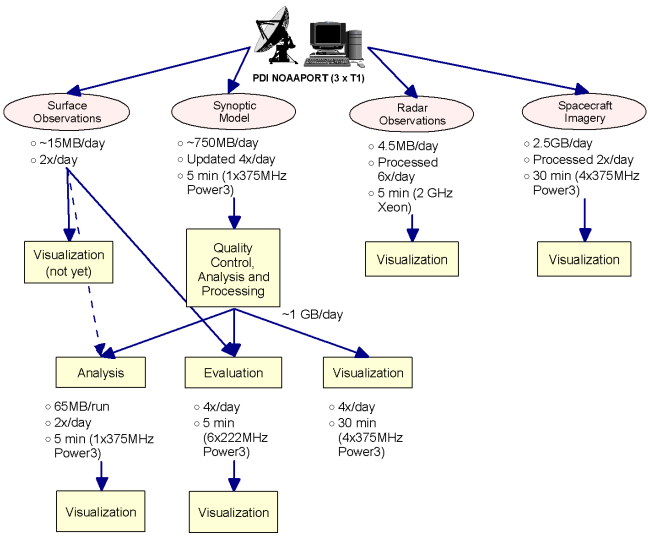

Figure 1. Pre-Processing Procedural and Data Flow.

3. IMPLEMENTATION

The original GriB-1 processing code was replaced with more flexible and portable software written in Java to address the aforementioned issues. Given the changes being made for the transmission of the Eta data, this same philosophy was used to develop a second implementation to apply to the new data stream.

The software for each implementation is structured into four basic layers, which are shown in Figures 2 and 3, respectively. Each of the layers, enumerated below utilizes data generated from the next layer.

1. *GribCode*: File management (logging, polling for input, determine need to generate output, etc.)
2. *Layer1*: Create actual output (".ETA" file) and compute statistics
3. *Layer2*: Check for errors, apply gross quality control via range checking, do reconstruction
4. *Decode*: Decompress, parse and decode GriB messages

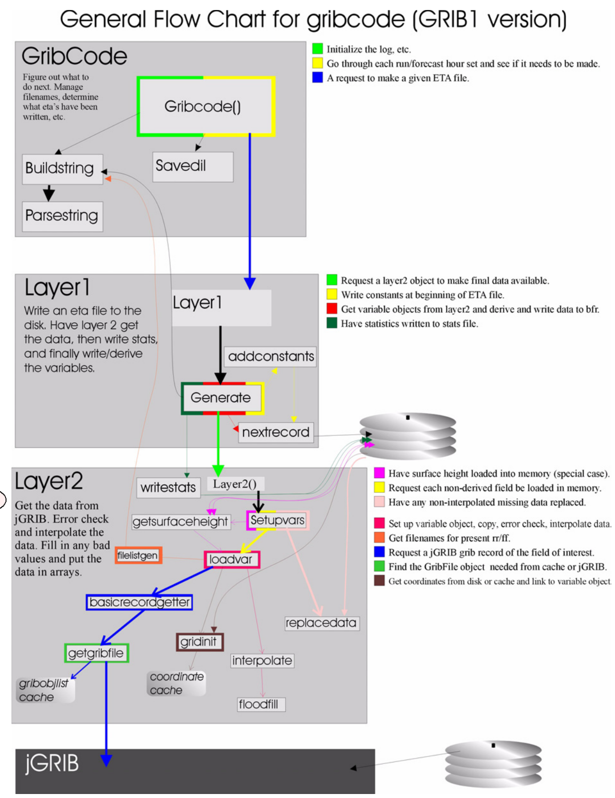

Figure 2. GriB-1 Processing Procedural and Data Flow.

3.1 GriB-1 (NWSTG)

The initial implementation only works with the Eta data in the 212 and 215 grids compressed as GriB-1. The details are outlined in Figure 2. The PDI NOAAport system receives each of the individual GriB-1 messages composing a time step of either of the two grids and concatenates them into a single file, which will contain surface only or surface and upper air data (interpolated to 27 isobaric levels).

Each of these files must be converted into a usable form for the aforementioned processing. Starting from the bottom, the first step is to uncompressed the arrays for subsequent analysis, in response to a request for data from *GribCode->Layer1->Layer2*. An open source Java class library *(jgrib)* has been employed to provide this capability, including some modifications to integrate it into this system (Stark, 2003).

At the next step (*Layer2*), each variable of interest (e.g., winds, temperature, etc.) for each atmospheric layer (surface or upper air) is a two-dimensional array. Each element of the each array is checked for validity (either inappropriate values or missing entirely). The collection of "bad" elements are then characterized to determine an appropriate remedial step under the assumption that the arrays are an incomplete sampling of the model atmosphere. For example,

- If a small number of elements (less than ten percent) of any areas are "bad", then interpolate from available good elements that are nearby via a flood-fill algorithm. If there are insufficient number of points in a region (e.g., near a boundary), then use linear interpolation.

• If too much of a layer is considered “bad”, then interpolate from surrounding layers via vertical log-linear interpolation (logarithm of pressure). A bi-directional vertical search is done to find the nearest good set of elements. If the vertical deviation is beyond two layers, then extrapolation is done, although that clearly is not preferred. In both cases, a lapse-rate correction is applied as appropriate (e.g., to surface data). This approach also reconstructs layers that are intentionally missing from the input data stream (e.g., upper air humidity).

• If there are too few good elements to reconstruct missing arrays, then a static representation is used, and adjusted using the above mechanisms.

• For volumetric data, stack the (reconstructed) isobaric layers into a three-dimensional array.

The request for a “valid” array comes from Layer1, which will then collect all of them, compute the minimum and maximum of each layer as well as for the volume for upper air variables, and the number of bad and missing elements. The collection of arrays are written to a single flat binary file with the statistics written to a flat ASCII file.

3.2 GriB-2 (DVB-S)

The implementation for GriB-1 was extended to accommodate full-resolution Eta results at 12 km. These data are being transmitted via the NOAAport Direct Video Broadcast channel using the new GriB-2 format (WMO, 2004) on the AWIPS 218 grid. Currently, these data are resampled from the original 12-km-resolution grid to another at 12 km, but on a different map projection covering a smaller area, focused on the continental United States, interpolated from 60 vertical levels to 30 isobaric levels, and then sampled every three hours in time

Given the much higher resolution of the 218 grid, more efficient compression is required. The GriB-1 approach is many decades old and does not yield particularly high compression rates. As part of a new standard, GriB-2, an image-based lossless compression scheme utilizing JPEG2000 was selected (JPEG, 2004).

This data stream will eventually replace the GriB-1-based 212 and 215 grids. Unfortunately, the combination of a new structure, metadata, physical format and compression was implemented in a way that is completely incompatible with GriB-1 data or software that utilizes data stored in GriB-1 data directly (NCEP, 2004a).

The approach with the GriB-1 data was used with the GriB-2 data, which is summarized in Figure 3. Unfortunately, there is no class library analogous to *jgrib* or other decompression and parsing tool for Grib-2 data. In addition, the PDI NOAAport software only provides raw GriB-2 messages (one for each time step for each layer for each variable). Therefore, an analogous concatenator and decoder needed to be implemented.

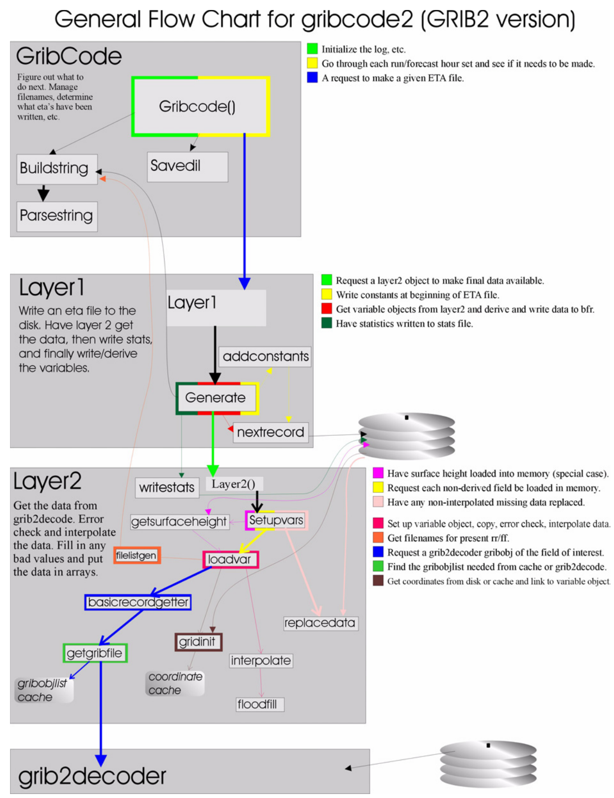

Figure 3. GriB-2 Processing Procedural and Data Flow.

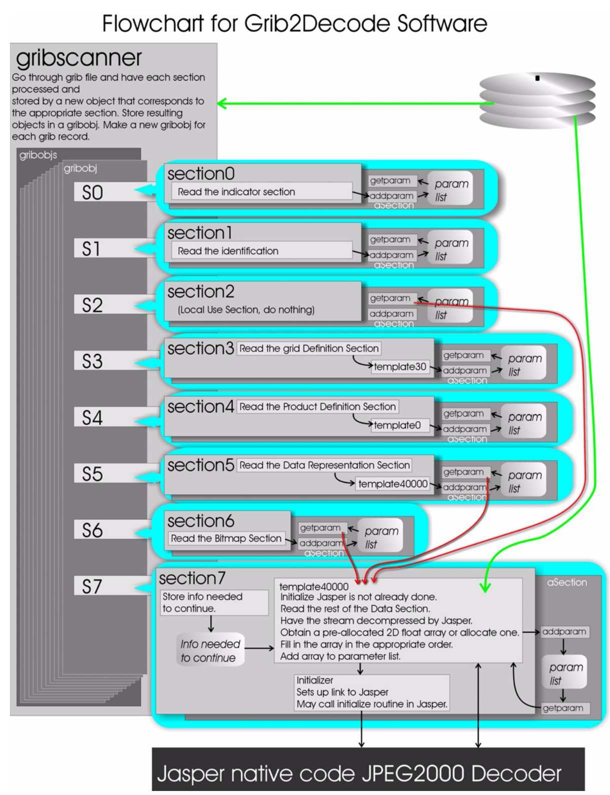

Figure 4. GriB-2 Decoding Process.

The new decoding process is summarized in Figure 4. It takes as input a Grib-2 file, which has all of the data of interest for a single time step, which has been concatenated from the individual files that contain a single GriB-2 message. This software utilizes a layer of C code provided by NCEP, which employs an open source toolkit (*Jasper*) to decompress images stored as JPEG2000 (NCEP, 2004b and Adams, 2004). The result is a collection of GRIB-2 message in memory, which then must be parsed in software following the metadata and structure outlined by WMO, 2004 and NCEP, 2004a. This yields the raw two-dimensional arrays of data, which can then be analyzed, repaired and collected in a manner similar to the GriB-1 data, discussed earlier.

4. EXAMPLE PRODUCTS

To primary use of the reconstructed Eta data is for subsequent analysis to yield background fields and lateral boundaries for mesoscale numerical weather predicitions. These capabilities and examples are illustrated elsewhere (e.g., Treinish and Praino, 2004, and Treinish and Praino, 2005). In addition, there are some other products generated by the *Deep Thunder* system that utilize these Eta data.

One example is shown in Figure 5 as a type of comprehensive, three-dimensional visualization. It shows a scene in vertical pressure coordinates in a Lambert Conformal projection covering most of North America. It is using the native AWIPS 218 coordinate system at full resolution, which implies over 7.5 million data points for each upper air variable, a few of which are shown. A number of specific visualization strategies are employed (*Class IV* from Treinish, 2001).

Near the top there is a translucent orange surface, which is an isosurface of horizontal wind speed at 30 meters/second. Its shape and movement in animation is indicative of the jet stream. Closer to the bottom are white translucent surfaces representing "clouds" as an isosurface of relative humidity at 95%. At the bottom is a colored surface, whose height corresponds to the forecasted surface pressure. That surface coloring uses contour bands of surface temperature following the legend at the lower left. The surface is overlaid with a map of political boundaries and coastlines in black.

These visualization are presented in an animated fashion on the operational web site used to disseminate products generated by *Deep Thunder* as well as a complementary interactive application used for diagnostic purposes.

Figure 6 is an example of a verification visualization that is produced as part of automated post-processing, which is partially derived from Eta data (Treinish and Praino, 2004). After each *Deep Thunder* model run, the results are bilinearly interpolated to the locations of the National Weather Service metar stations, whose data are available through the NOAAport receiver. An analogous process is applied to each of the reconstructed Eta

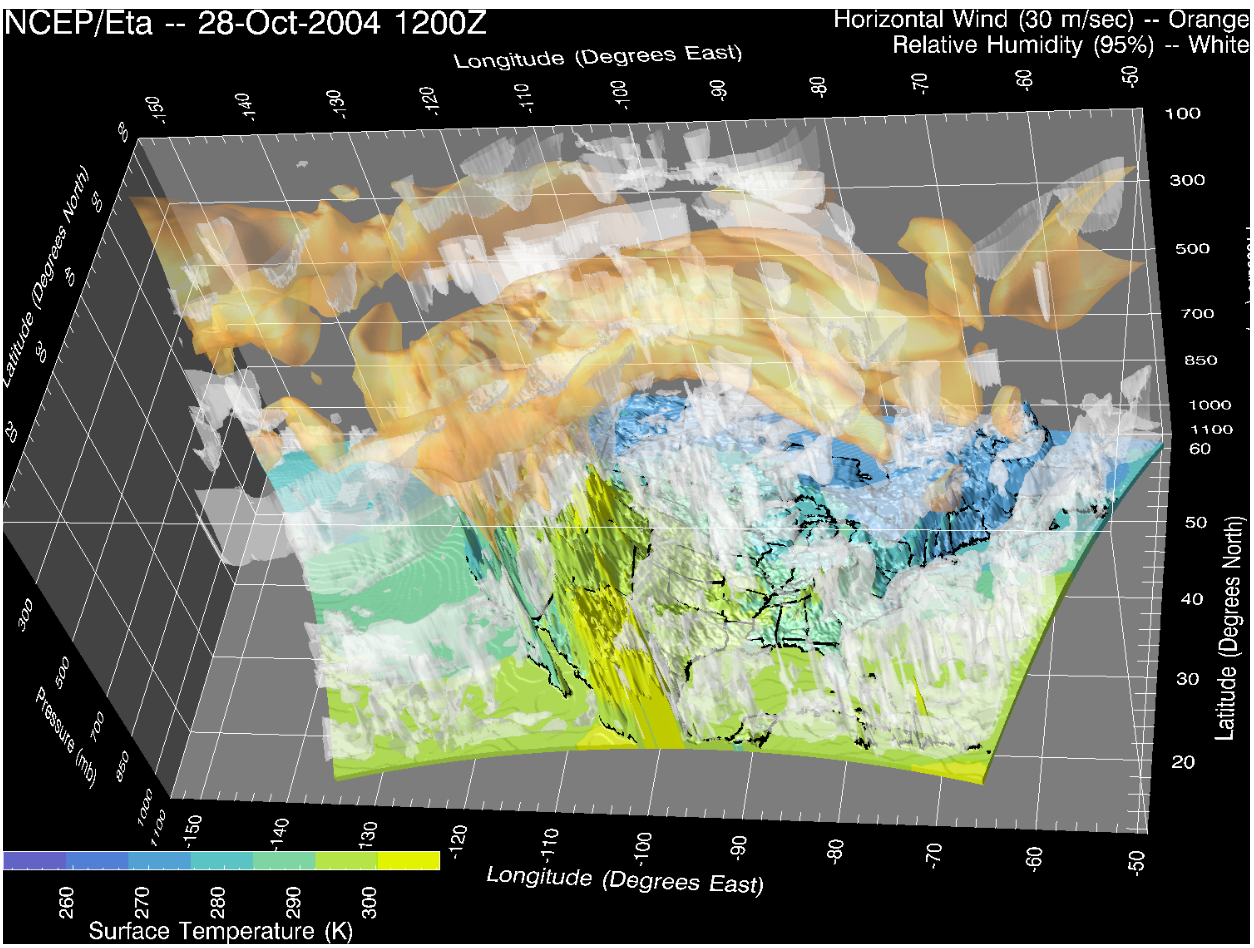

Figure 5. Example Three-Dimensional Visualization of Eta-12 in Native 218 Grid.

Grib-1 grid. After the observations corresponding to each model run become available, a verification process is initiated in which these spatially interpolated results are statistically analyzed and compared to parsed and quality-checked surface observations. This yields a set of evaluation tables as well as visualizations for each model run as well as the aggregation of all model runs during the previous week. The latter are presented via web pages in a manner following that of the model visualizations. The details of this approach and examples are discussed in Praino et al, 2003.

Figure 6 shows summary statistics for one week's worth of *Deep Thunder* (4 km and 1 km nests for the New York City metropolitan area) and Eta model surface results in comparison to 55 metar observations. Following the legend to the upper right, four curves are shown for temperature and dew point results plotted as a function of forecast time (x-axis), bias (y-axis) and root mean square error (z-axis). To enable perspective viewing in this coordinate system, interaction is available for this visualization on the operational *Deep Thunder* web site, which enables forecast error and biases to be examined simultaneously.

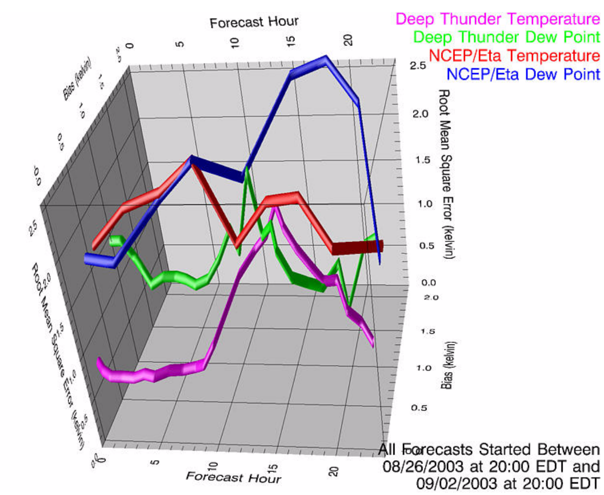

Figure 6. Example Verification Visualization.

5. CONCLUSIONS AND NEXT STEPS

Even though the overall approach and its integration in the *Deep Thunder* system and implementation is still evolving, the results to date have been very encouraging. The original and new GriB-1 processing were run simultaneously for several months with comparable level of computational performance before the former was retired. Many cases of sufficiently corrupt data occurred that the old system could not handle, which would have prevented operational forecasts from being produced. Hence, the new implementation provides more reliable and robust results, especially when the quality of the received GriB files has been compromised.

Both the GriB-1 and GriB-2 systems are run in parallel at the present time because the GriB-2 DVB-S broadcasts are not expected to be operational until January 2005. Some testing of the GriB-2-based results has been done with both the isentropic analysis and subsequent modelling with encouraging results. Little additional computational cost has been seen despite the higher resolution and more detailed content. The testing will continue, including simultaneous processing until the GriB-2 products are deemed operational at which time, the GriB-1 system will also be retired.

6. ACKNOWLEDGEMENTS

This work is supported by the Mathematical Sciences Department and the Exploratory Server Systems Department at the IBM Thomas J. Watson Research Center.

7. REFERENCES

Adams, M. D. *Jasper*. (**http://www.ece.uvic.ca/~mdadams/jasper**), 2004.

JPEG 2000 Committee. *JPEG 2000*. (**http://www.jpeg.org/jpeg2000**), 2004.

Klet, L. *NOAAPORT Broadcast Changes*. **Proceedings of the 20th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology**, January 2004, Seattle, WA.

National Centers for Environmental Prediction. *NCEP GRIB2 Documentation*. (**http://www.nco.ncep.noaa.gov/pmb/docs/grib2/download/index.shtml**), 2004.

National Centers for Environmental Prediction. *NCEP codes for decoding and encoding data in GRIB2 format*. (**http://www.nco.ncep.noaa.gov/pmb/codes/GRIB2**), 2004.

Planetary Data, Inc. (**http://www.planetarydata.com**), 2004.

Praino, A. P., L. A. Treinish, Z. D. Christidis and A. Samuelsen. *Case Studies of an Operational Mesoscale Numerical Weather Prediction System in the Northeast United States.* **Proceedings of the 19th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology**. February 2003, Long Beach, CA.

Stark, B. *JGRIB*. (**http://jgrib.sourceforge.net**), 2003.

Treinish, L. A. *How Can We Build More Effective Weather Visualizations?* **Proceedings of the Eighth ECMWF Workshop on Meteorological Operational Systems**, November 2001, Reading, England, pp. 90-99.

Treinish, L. A. and A. P. Praino. *Applications and Implementation of a Mesoscale Numerical Weather Prediction and Visualization System*. **Proceedings of the 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction**, January 2004, Seattle, WA.

Treinish, L. A. and A. P. Praino. *The Potential Role For Cloud-Scale Numerical Weather Prediction for Terminal Area Planning and Scheduling*. **Proceedings of the 21st International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology**. January 2005, San Diego, CA.

World Meteorological Organization. *Guide to the WMO Table Driven Code Form Used for the Representation and Exchange of Regularly Spaced Data In Binary Form: FM 92 GRIB Edition 2*. (**www.wmo.ch/web/www/DPS/Migration/Guide-GRIB2.doc**), 2004.