3.9

IMPLEMENTATION OF MESOSCALE NUMERICAL WEATHER PREDICTION FOR WEATHER-SENSITIVE BUSINESS OPERATIONS

Lloyd A. Treinish*, Anthony P. Praino and Zaphiris D. Christidis
IBM Thomas J. Watson Research Center, Yorktown Heights, NY

1. INTRODUCTION

For many applications, expected local weather conditions during the next day or two are critical factors in planning operations and making effective decisions. Typically, what optimization that is applied to these processes to enable proactive efforts utilize either historical weather data as a predictor of trends or the results of synoptic-scale weather models. Alternatively, mesoscale numerical weather models operating at higher resolution in space and time with more detailed physics may offer greater precision and accuracy within a limited geographic region for problems with short-term weather sensitivity (e.g., Mass et al, 2002; Gall and Shapiro, 2000). Such forecasts can be used for competitive advantage or to improve operational efficiency and safety. To evaluate this hypothesis, a prototype system, dubbed "*Deep Thunder*", has been implemented for the New York City area.

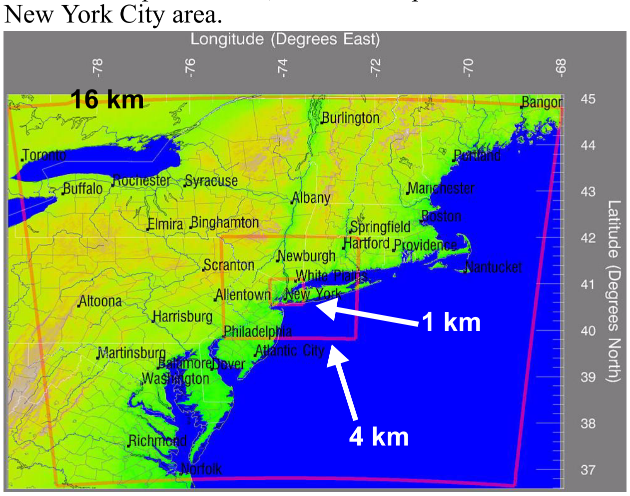

Figure 1. Model Nesting Configuration.

2. FORECAST MODEL DESCRIPTION

The model used for this effort is non-hydrostatic with a terrain-following coordinate system and includes interactive, nested grids. It is a highly modified version of the Regional Atmospheric Modeling System or RAMS (Pielke et al, 1992), which is derived from earlier work supporting the 1996 Centennial Olympic Games in Atlanta (Snook et al, 1998). It includes explicit parameterization of cloud microphysics (e.g., liquid and ice) to enable prediction of precipitation. Operationally, a 3-way nested configuration is utilized via stereographic projection. Each nest is a 62 x 62 grid at 16, 4 and 1 km resolution, respectively (i.e., 976 x 976 km^2, 244 x 244 km^2 and 61 x 61 km^2), focused on New York City, which is illustrated in Figure 1. The specific location was chosen to include the major airports operating in the New York City metropolitan area within the 1 km nest. In addition, it was desirable to have good coverage for a number of weather-sensitive applications in that geographic region as well as for the locations of the authors' homes and office. The three nests employ 48, 12 and 3 second time steps, respectively. Each nest employs the same vertical grid using 31 stretched levels with the lowest level at 48 m above the ground, a minimum vertical grid spacing of 100 m, a stretch factor of 1.12 and and a maximum grid spacing of 1000 m. At the present time, two 24-hour forecasts are produced daily, typically initiated at 0Z and 12Z.

Currently, the data for both boundary and initial conditions for each model execution are derived 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 and 60 vertical levels. These data are made available via the National Weather Service NOAAport data transmission system after sampling to 40 km resolution on the AWIPS 212 grid and interpolated to 27 isobaric levels for the continental United States in a Lambert-Conformal projection. In addition, the model lateral boundaries are nudged every three hours, using the Eta-212 grids, which are available via NOAAport. Static surface coverage data sets provided by the United States Geological Survey at 30-second resolution are used to characterize topography and vegetation coverage. Similar but lower-resolution data are used to define land use and coverage (at 10-minute resolution) and sea surface temperature (one-degree resolution). The latter is updated to use data corresponding to the particular month in which the forecast is made. The static and dynamic data are processed via an isentropic analysis package to generate three-dimensional data on the model nested grids for direct utilization by the modelling code.

3. ARCHITECTURE AND IMPLEMENTATION

This effort began with building a capability sufficient for operational use in 2001 after the appropriate hardware and basic software infrastructure was implemented. In particular, the goal was to provide weather forecasts at a level of precision and fast enough to serve as a testbed to address specific business problems. Hence, the focus has been on 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.

Part of the rationale for this focus is practicality. Given the time-critical nature of weather-sensitive business decisions, if the weather prediction can not be completed fast enough, then it has no value. Such predictive simulations need to be completed at least an order of magnitude faster than real-time. But rapid computation

*Corresponding author address: Lloyd A. Treinish, IBM T. J. Watson Research Center, P. O. Box 218, Yorktown Heights, NY 10598, **lloydt@us.ibm.com**, **http://www.research.ibm.com/people/l/lloydt**

is insufficient if the results can not be easily and quickly utilized. Thus, a variety of fixed and highly interactive flexible visualizations focused on the applications have also been implemented.

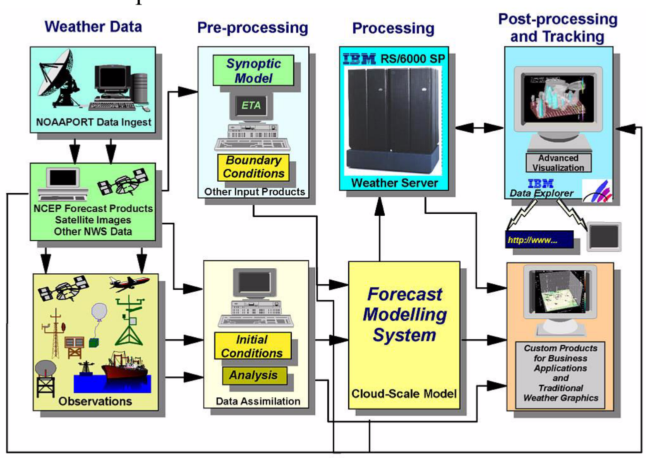

Figure 2. ***Deep Thunder*** **Architecture.**

With such goals, the system has also evolved from its initial implementation. Hence, the discussion herein outlines the current approach, whose components are shown schematically in Figure 2, and are described below from left to right.

3.1 Data

The NOAAport system provides a number of different data sources as disseminated by the National Weather Service. These include *in situ* and remotely sense observations used currently for forecast verification as well as the aforementioned Eta data for model boundary and initial conditions. For the *Deep Thunder* system, a two-channel facility manufactured by Planetary Data, Incorporated, is utilized, which was installed at the IBM Thomas J. Watson Research Center in Yorktown Heights, NY in February 2000.

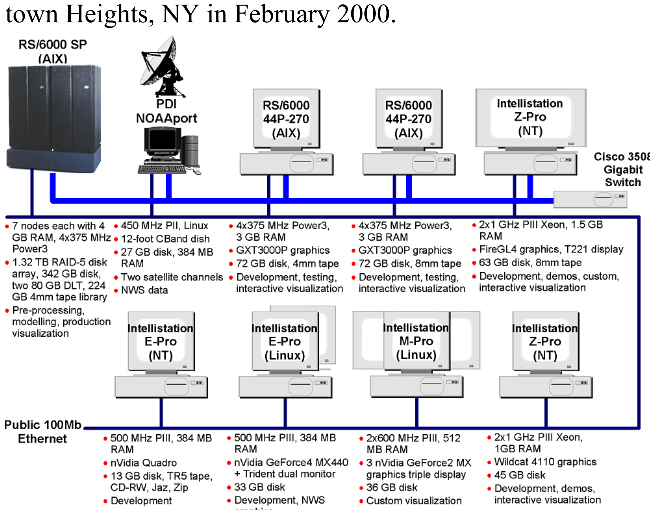

Figure 3. ***Deep Thunder*** **Hardware Environment.**

The NOAAport and other hardware that supports this project is shown in Figure 3. This NOAAport receiver system, based upon Linux, has a very flexible design, enabling the type of customization and integration necessary to satisfy the project goals. The various files transmitted via NOAAport are converted into conventional files in Unix filesystems in their native format, accessible via NFS mounting on other hardware systems via a private gigabit ethernet.

3.2 Pre-Processing

The pre-processing consists of 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. In all cases, gross quality control is applied via range checking. For the aforementioned Eta-212 grids, the data are received in the compressed GriB 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. One is the isentropic analysis discussed earlier. Another is to support forecast verification outlined below. A third is a set of summary, three-dimensional animations available on the operational web site used to disseminate products generated by *Deep Thunder* as well as a complementary interactive application used for diagnostic purposes. Similar visualizations can also be generated from the output of the isentropic analysis. Given expected changes in the near future for NOAAport-received NCEP model data, the GriB-processing code is being replaced with more flexible and portable software written in Java. The data and procedural flow of these processes is outlined in Figure 4. Most of them run serially on, although compiler-optimized for an IBM Power3 processor. Other aspects related to forecast verification and product visualization are discussed in subsequent sections.

3.3 Processing

To enable timely execution of the forecast models, which is required for operations, the simulation is parallelized on a high-performance computing system. For this effort, an IBM RS/6000 Scalable Power Parallel (SP) is employed. This is IBM's previous generation of supercomputer systems, which is in common use at many operational centers for numerical weather prediction. It is a distributed memory MIMD computer, consisting of two to 512 RS/6000 processor nodes, that communicate via a high-speed, multi-stage interconnect (the SP Switch). Each node has an SMP configuration of two to 16 Power3 processors. In the current implementation, there are seven nodes of four 375 MHz Power3 processors, as shown in Figure 3. The modelling software is parallelized using the Scalable Modelling System/Nearest-Neighbor Tool described by Edwards et al, 1997 for single model domains. It has been extended to support multiple nests. The modelling domain for all nests is spatially decomposed for each processor to be utilized, which is mapped to an MPI task. Within each node, there are four MPI tasks, which communicate via shared memory. The SP switch fabric enables communications between nodes. None of these tasks do I/O. Instead an additional processor is utilized to collect results from the MPI tasks and perform disk output asynchronously. This enables an efficient utilization of the SP platform for the modelling code. For current operations, six nodes of four processors each are used for computing and a single cpu of a similar node is used for I/O. A typical model run with the aforementioned configuration requires about two hours to complete.

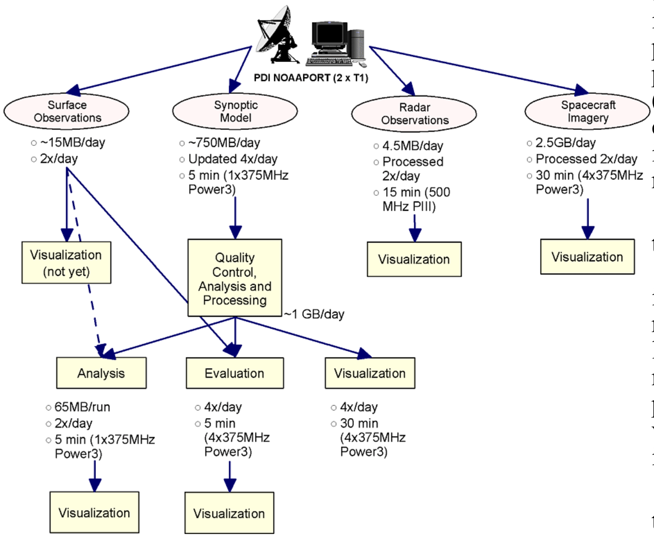

Figure 4. Pre-Processing Procedural and Data Flow.

3.5 Post-Processing

Post-processing essentially operates on the raw model output to provide useful products. There are several aspects of post-processing, the most important of which is visualization, as suggested earlier. Since large volumes of data are produced, which are used for a number of applications, the use of traditional graphical representations of data for forecasters can be burdensome. Alternative methods are developed from a perspective of understanding how the weather forecasts are to be used in order to create task-specific designs. In many cases, a "natural" coordinate system is used to provide a context for three-dimensional analysis, viewing and interaction.

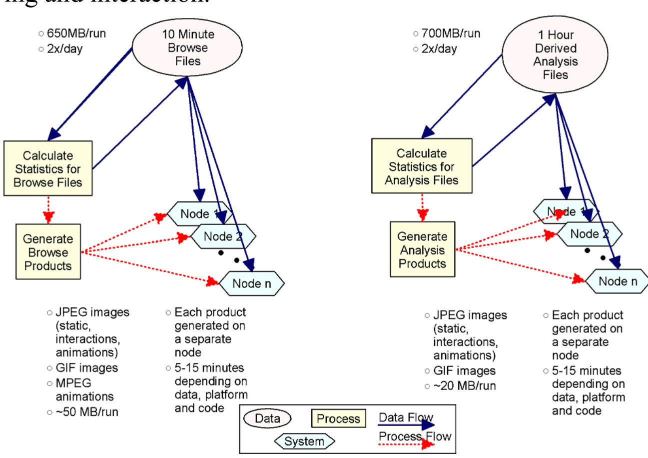

Figure 5. Visualization Post-Processing Procedural and Data Flow.

These visualizations provide representations of the state of the atmosphere, registered with relevant terrain and political boundary maps. This approach for *Deep Thunder* and details of its implementation are discussed in Treinish, 2001.

To enable timely availability of the visualizations, the parallel computing system used for the model execution is also utilized for post-processing. This approach is outlined schematically in Figure 5. Two types of output data are generated by the model. The first, is a comprehensive set of variables at hourly resolution (analysis) files for each nest, which are further processed to generate derived products and interpolated to isobaric levels from the model terrain-following coordinates.

The second output is a subset of variables relevant to the applications of the model output produced every 10 minutes of forecast time. The finer temporal spacing is required to better match the model time step in all nests as well as to capture salient features being simulated at the higher resolutions. A subset is chosen to minimize the impact of I/O on the processing throughput. These browse files are also generated to enable visualization of model results during execution for quality control and simulation tracking.

Two classes of visualizations are provided as part of the *Deep Thunder* system. The first is a suite of highly interactive applications utilizing the workstation hardware shown in Figure 3, including ultra-high-resolution and multi-panel displays (Treinish, 2001).

The second is a set of web-based visualizations, which are generated automatically after each model execution via a set of hierarchical scripts (Treinish, 2002). That work is also illustrated schematically in Figure 5. In addition, this processing utilizes additional SMP workstations clustered via a private Gigabit or public 100MB Ethernet as shown in Figure 3. The work to create individual products (i.e., JPEG or MPEG files) is split up among the available nodes to run simultaneously. This simple parallelism, including intranode parallelism, enables the independent generation of various products for placement on a web server to be completed in five to fifteen minutes.

An approach similar to that used for visualization is employed for forecast verification. After each model run, the results of all three nests combined in a multiresolution structure (Treinish, 2000) 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 Eta-212 grid as part of the automated preprocessing. 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 later 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.

3.6 Integration

All of the components are operated by a master script, implemented in Perl. Model executions are set up via a simple spreadsheet identifying basic run characteristics such as start time, length, location, resolution, etc. A Unix crontab is used to initiate the script. In addition to bookkeeping and quality control and logging, it polls input data availability whose arrival via the NOAAport is variable, does all the necessary pre-processing steps, initiates the parallel modelling job and then launches the parallel visualization post-processing.

4. EXAMPLE RESULTS

To illustrate some of the range of capabilities that have been implemented, a few visualization products that *Deep Thunder* can generate automatically are shown herein. Additional ones can be seen at **`http://www.research.ibm.com/weather/NY/NY.html`**.

Figure 6 represents a class of meteogram that is oriented toward interpretation by the non-meterologist. It consists of four panels showing surface data and supplemented with two panels to illustrate upper air winds. In all cases, the variables are shown as a function of time interpolated to a specific location (White Plains Airport within the 1 km nest). The upper and middle plots on the left each show two variables while the rest each show one. The top left plot presents temperature (blue) and pressure (red). The middle left panel shows humidity (blue) and total precipitation (red). Since the precipitation is accumulated through the model run, the slope of the curve will be indicative of the predicted rate of precipitation. Therefore, when the slope is zero, it is not raining (or snowing). In addition, the model calculations require some time to "spin-up" the microphysics to enable precipitation. Therefore, there will typically be no precipitation in the first hour or two of model results. The top right plot illustrates forecasted winds -- speed (blue) and direction (red). The wind direction is shown via the arrows that are attached to the wind speed plot. The arrows indicate the predicted (compass) direction to which the wind is going. The middle right plot is a colored contour map of forecasted total (water and ice) cloud water density as a function of elevation and time. This "cross-sectional" slice can provide information related to storms, fog, visibility, etc. predicted at this

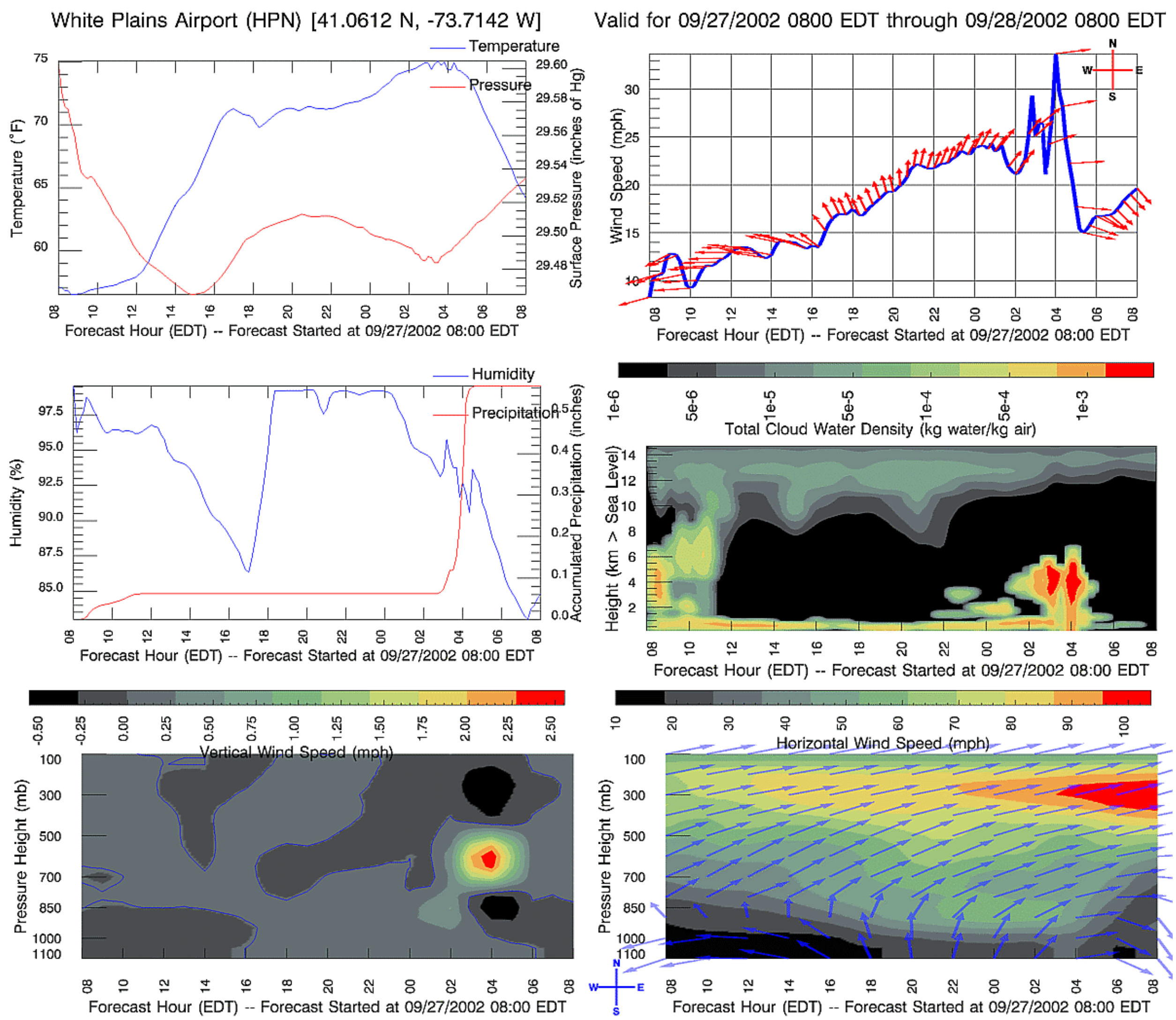

Figure 6. Example One- and Two-Dimensional Visualizations.

location. Portions of the plot in black imply time or elevations where there are little or no clouds. Areas in yellow, orange and red imply when and where the relatively densest clouds are forecasted, following the color legend on the top of the panel. The bottom two panels show upper air winds using some of the same techniques. The lower left shows contours of vertical winds as a function of time and pressure following the legend above it. In addition, the zero velocity contour is shown in blue. At the lower right, is a contour map of horizontal wind speed also as a function of time and pressure. It is overlaid with arrows (blue) to illustrate the predicted compass wind direction. These example plots illustrate the forecast of the remnants of an extra-tropical event followed by the passage of a cold front, including convection induced by the latter.

Figure 7 shows predicted accumulated snow as simple two-dimensional map in south-central New York State (within the 16 km nest). A set of colored contour bands following the legend to the upper right are overlaid with the location of cities and county boundaries. In addition, the path of a major highway, the New York State Thruway, is also indicated and colored by predicted snowfall. This type of map is designed for easy interpretation for road transportation applications. This visualization illustrates a prediction of the unusual late spring snowfall observed in May 2002 in New York State.

Figure 8 is an example of a qualitative, yet comprehensive, three-dimensional visualization. It shows a terrain map, colored by a forecast of total precipitation, where darker shades of blue indicate heavier accumulations. The map is marked with the location of major cities or airports as well as river, coastline and county boundaries within the 4 km nest. In addition, there are colored arrows indicating predicted winds, with the lighter color being faster winds and the arrow direction corresponding to the direction to which the wind is flowing. Above the terrain is a forecast of clouds, represented by a three-dimensional translucent white surface of total cloud water density (water and ice) at a threshold of 10^{-4} kg water/kg air. This approximation of a cloud boundary shows the typical "anvil" shaped structure of a cluster of thunderstorm cells. Within the cloud surface are translucent cyan surfaces of forecast reflectivities at a threshold of 50 dbZ, that correspond to rain shafts for individual convection cells. This image is part of an animation sequence illustrating a forecast of frontal-induced convective storms.

5. DISCUSSION

Even though the overall system and implementation is still evolving, the type of products that *Deep Thunder* can generate has provided a valuable platform to inves-

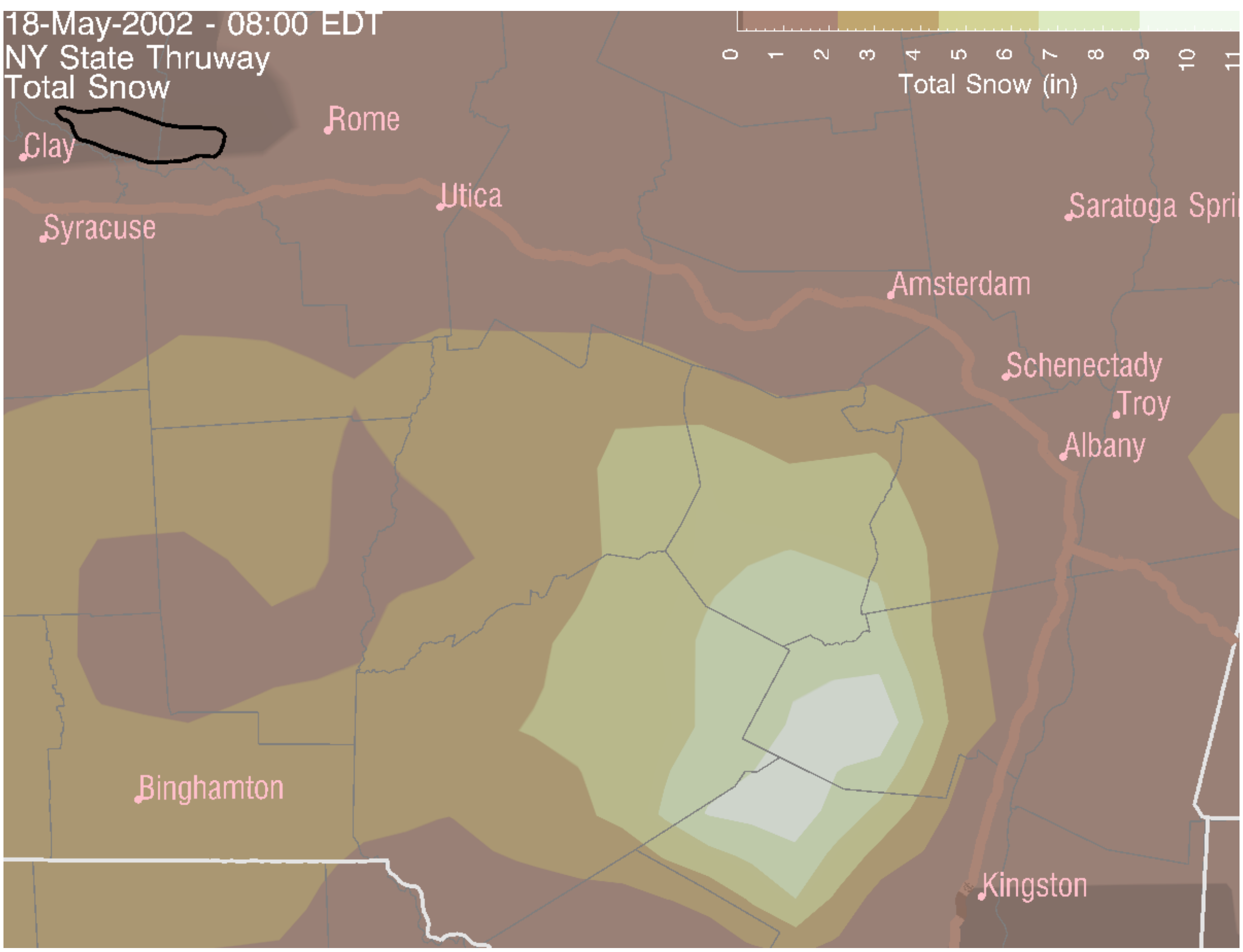

Figure 7. Example Two-Dimensional Visualization.

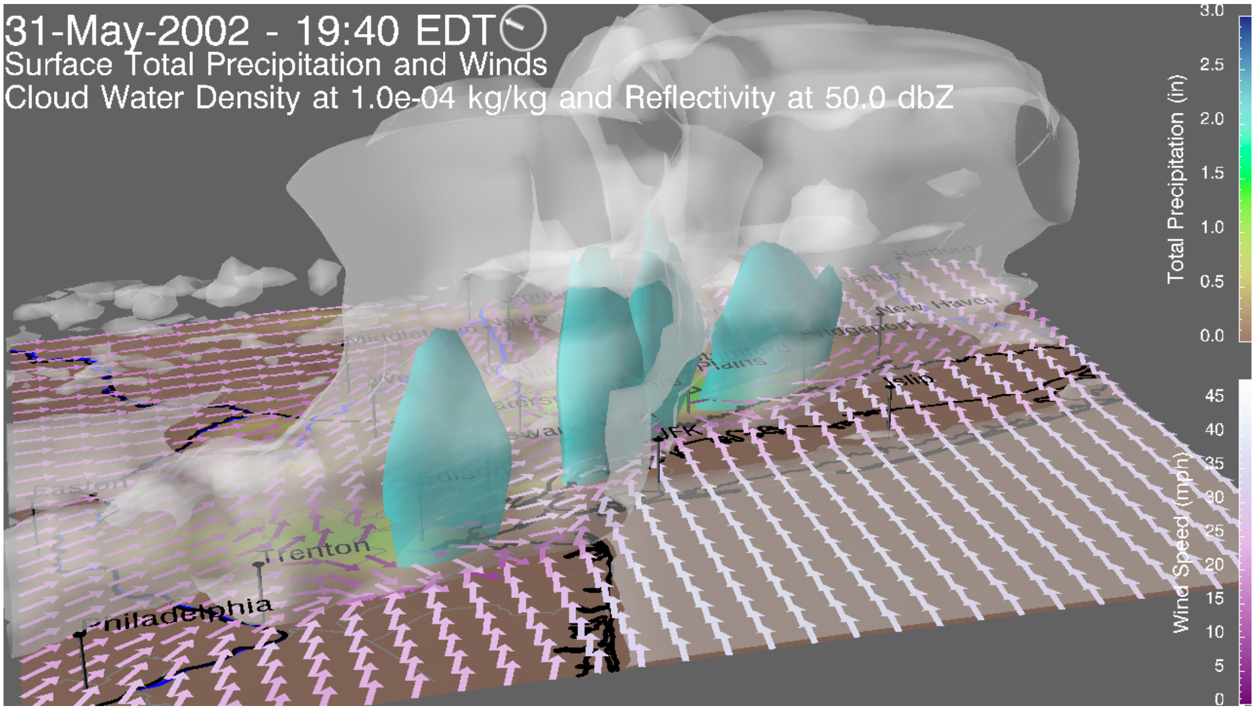

Figure 8. Example Three-Dimensional Visualization.

tigate a number of practical business applications. To aid in that evaluation, the products have been made available to a number of local agencies to assist in their operational decision making with various weather-sensitive problems in surface transportation, emergency response and electricity distribution. Aspects of the visualization and web-based access to these products are discussed in Treinish, 2003.

The feedback from these users coupled with more rigorous verification (Praino et al, 2003) has raised a number of comments and issues. In general, there has been a very favorable view of the ability of the overall system to provide useful and timely forecasts of severe weather including convective events and high winds with greater precision. The user-driven design of visualization products has enabled effective utilization of the model output. However, improved throughput is required to enable more timely access to the forecast products, which need to cover broader areas at higher resolution. These suggest the direction for continued work and improvement of the utility of the system.

6. CONCLUSIONS AND FUTURE WORK

This is an on-going effort. The results to date illustrate a practical and useful implementation with automatically generated user-application-oriented forecast visualizations on the world-wide-web. But they also point to several next steps besides refining the quality of the model results, improving the degree of automation, and developing new methods of visualization and dissemination.

To enable more timely availability of forecast products, the number of model runs each day will expand, including coverage of other geographic areas at high resolution. Overall throughput is limited by the capacity of the current hardware. Although plans for a modest increase will be applied, the model configuration will need to be adjusted to somewhat lower resolution with broader geographic coverage focused on the specific areas of concern expressed by the current set of users at local government agencies.

To aid in the improvement of overall forecast quality, the ability to leverage the expected availability of full-resolution 12 km Eta results on the AWIPS 218 grid as well as some assimilation of observations will be implemented to enhance both initial and boundary conditions. Although all of these changes will result in up to an order of magnitude increase in data production and processing, the current hardware environment does have the capacity to support it.

As these customized capabilities are made available to assist in weather-sensitive business operations, efforts will also be addressed to determine and apply appropriate metrics for measuring business value. These will serve to provide an evaluation of *Deep Thunder* that is complementary to the traditional meteorological verification.

7. ACKNOWLEDGEMENTS

The authors wish to acknowledge the contributions to this work by two students. Annette Samuelsen, a Ph.D. candidate in oceanography at Florida State University, implemented much of the software that is used for the verification portion of the *Deep Thunder* system during her summer 2002 internship at the IBM Thomas J. Watson Research Center.

Craig Tashman, an undergraduate physics student at Pace University, has an on-going internship at IBM Research. He began by working on techniques of graphics compression for the dissemination of model results, which are not discussed herein. More recently, he has expanded his work to include the implementation

of the Java-based processing of NOAAport-received Eta GriB files.

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

8. REFERENCES

Edwards, J., J. S. Snook and Z. Christidis. *Forecasting for the 1996 Summer Olympic Games with the SMS-RAMS Parallel Model.* **Proceedings of the Thirteenth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology**, February 1997, Long Beach, CA, pp. 19-21.

Gall, R. and M. Shapiro. *The Influence of Carl-Gustaf Rossby on Mesoscale Weather Prediction and an Outlook for the Future*. **Bulletin of the American Meteorological Society**, **81**, no. 7, pp. 1507-1523, July 2000.

Mass, C. F., D. Owens, K. Westrick, B. A. Colle. *Does Increasing Horizontal Resolution Produce More Skillful Forecasts*. **Bulletin of the American Meteorological Society**, **83**, no. 3, pp. 407-430, March 2002.

Pielke, R. A., W. R. Cotton, R. L. Walko, C. J. Tremback, W. A. Lyons, L. D. Grasso, M. E. Nicholls, M.-D. Moran, D. A. Wesley, T. J. Lee and J. H. Copeland. *A Comprehensive Meteorological Modeling System - RAMS*. **Meteorology and Atmospheric Physics**, **49**, 1992, pp. 69-91.

Praino, A. P., L. A. Treinish and Z. D. Christidis. *Evaluation of an Operational Mesoscale Numerical Weather Prediction System in the Northeast U.S.* To be published in **Proceedings of the Nineteenth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology**, February 2003, Long Beach, CA.

Snook, J. S., P. A. Stamus, J. Edwards, Z. Christidis, J. A. McGinley. *Local-Domain Mesoscale Analysis and Forecast Model Support for the 1996 Centennial Olympic Games*. **Weather and Forecasting**, **13**, no. 1, pp. 138–150, January 1998.

Treinish, L. *Multi-Resolution Visualization Techniques for Nested Weather Models*. **Proceedings of the IEEE Visualization 2000 Conference**, October 2000, Salt Lake City, UT, pp. 513-516, 602.

Treinish, L. *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. *Interactive, Web-Based Three-Dimensional Visualizations of Operational Mesoscale Weather Models*. **Proceedings of the Eighteenth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology**, January 2002, Orlando, FL, pp. J159-161.

Treinish, L. *Web-Based Dissemination and Visualization of Mesoscale Weather Models for Business Operations*. To be published in **Proceedings of the Nineteenth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology**, February 2003, Long Beach, CA.