An automated approach to detect phenomena from NAM model outputs

Rahul Ramachandran
University of Alabama in Huntsville
rramachandran@itsc.uah.edu

Advances in real-time observation, information technology and modeling techniques are fueling a paradigm shift in short-term weather forecasts from static numerical model forecasts to dynamic and adaptive model forecasts, where, for example, regional models can be run at high resolution for areas with rapidly evolving severe weather. A timely and accurate severe weather forecast requires timely detection or prediction of weather events, easy access to real time observational data, advanced weather models and large computing resources in an integrated framework. The Linked Environments for Atmospheric Discovery (LEAD) project is a large-scale, interdisciplinary NSF-funded research project that aims at developing such a cyber-infrastructure to enable identifying, accessing, decoding, assimilating, analyzing, mining, and visualizing a broad array of meteorological data and model output necessary for dynamic and adaptive weather forecasts. One of the key components for dynamic and adaptive forecasts is the detection of current regions and prediction of future regions of severe weather phenomena. Current regions of significant weather events can be identified from real time observational data. On the other hand, the prediction of future weather events relies on model forecasts. A direct method is to apply a thresholding technique to one or more model output fields to detect regions of severe weather. These regions may represent low pressure systems, strong vorticity, strong vertical updrafts, large convective available potential energy (CAPE), or a combination of any of these phenomena. However, identifying appropriate numerical model output fields and corresponding optimal thresholds can pose a big challenge.

In this paper, we describe a novel data mining approach to automatically identify potential regions of severe weather for on-demand modeling (ODM). This approach makes some assumptions for phenomena detection. First, we assume that significant values in any model output field suggest regions of interest in that field. Second, we assume that a difference field – defined as the absolute difference between model output fields from two consecutive runs valid at the same time – contains important weather phenomena information. That is, the difference fields indicate where weather is changing rapidly or is most sensitive to variations in model initial conditions. For this study, the Phenomena Extraction Algorithm (PEA), developed at the Information Technology and Systems Center (ITSC) at the University of Alabama Huntsville, is used to detect regions of interest from each output and its corresponding difference field. The PEA can automatically identify regions of interest characterized by abnormal intensity and local variance based only on image data statistics. Using the regions identified from each individual field of model outputs, a composite image is generated, with higher values for those regions identified in multiple individual fields. The PEA is applied once more to the resulting composite image in order to identify regions of greatest interest for weather events.

The data used in this study were 40-km North American Mesoscale (NAM) model (NAM-212) forecasts. The NAM is operated by the National Centers for Environmental Prediction (NCEP) and is initialized every 6 hours. Each initialization spawns forecasts that are output every 3 hours out to 84 hours. At each output time, a total of 617 parameter fields are reported. These outputs include 2D fields (e.g. mean sea level pressure, CAPE, and precipitation) and 3D fields (e.g. geopotential height, temperature, moisture, wind, and vorticity). The data used in this study are from 29 forecasts ranging from 17 January 2007 to 22 February 2007.

A case study was conducted using a 3-h NAM model forecast initialized at 00Z on 13 February 2007. In this case, there was a significant active weather system in the central United States. This system can be identified from the 850 mb wind pattern. From the mean sea level pressure field, two strong low pressure systems existed: one off the west coast of Canada and the other off the east coast of Canada. There was also a weaker area of low pressure in the central United States, which produced weather events. Among other detected regions, the algorithm successfully identified regions corresponding to the three low pressure systems.

To further validate the algorithm, the identified regions of interest were compared with the NEXRAD radar reflectivity images obtained from the National Climate Data Center. Because the NEXRAD radar network only covers most of the contiguous U.S., the regions of interest corresponding to the low pressure systems over both Canadian coasts, the Atlantic and Pacific Oceans, and Mexico are not covered by radar. However, the identified region over the central U.S., where there is good radar coverage, correlated well with areas of high radar reflectivity. The identified region did not extend as far north and east as the radar image. The identification of only the core area could be caused by the design of the algorithm, which forcibly breaks down the large regions of interest into mesoscale systems. The algorithm also successfully identified weather regions over south Florida and central California. The algorithm did successfully identify the location and size of weather regions over southern Florida and central California. The regions identified by the algorithm over the coastline of Washington and Oregon appear to be off the mark as compared with radar map.

Based on preliminary examination of 3-hr NAM forecasts, the proposed method is very promising in identifying regions of interest, which may be used to predict regions of future weather events. The outcome from this method may be helpful in assisting meteorologists define areas of weather focus to guide higher resolution model runs. This method is fully automatic, and no human intervention is required. More efforts will be made to examine if and how much the regions are falsely identified. Efforts will also be made to analyze individual regions of interest and to understand what parameter fields contribute to the identification of these regions. This will aid in filtering out falsely identified regions if such regions exist.