6B.3 A NOVEL APPROACH TO DETECT REGIONS OF PHENOMENA IN NAM MODEL OUPTUT

Xiang Li*, Rahul Ramachandran, Sara Graves, Bradley Zavodsky
University of Alabama in Huntsville, Huntsville, AL

Keith Brewster
University of Oklahoma, Norman, OK

Steven Lazarus
Florida Institute of Technology, Melbourne, FL

1. Introduction

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 (Droegemeier et al., 2004) 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 (Droegemeier et al., 2005). One of the key components for dynamic and adaptive forecasts is the detection of current regions experiencing severe weather and prediction of regions of future threats. Current regions of significant weather events can be identified from real time observational data. For instance, individual thunderstorms can be detected from radar reflectivity measurements using a simple thresholding technique. Normally, regional severe weather events emerge as clusters of local storms, and it is of more interest for forecast model domain selection to detect the regions of storm clusters rather than individual storms. Using a spatial data clustering technique, Li et al. (2008) developed an algorithm for automated detection of storm clusters from WSR88-II radar measurements. On the other hand, the prediction of future weather events relies on numerical weather prediction output. 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 variables. 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 of the University of Alabama Huntsville, is used to detect regions of interest from each model variable and its corresponding difference field. The PEA can identify regions of interest characterized by abnormal intensity and local variance based only on image data statistics. Using the regions identified from each individual model output variable, a composite image is generated, with higher values for those regions identified in multiple individual variables. The PEA is applied once more to the resulting composite image in order to identify regions of greatest interest for weather events. The rest of this paper is organized as follows: Section 2 describes the model data used in this study; Section 3 describes the phenomena detection approach; Section 4 discusses experiment and result analyses; finally, Section 5 concludes the investigation.

2. Data

The purpose of this study is to assist in ODM activities, focusing on regions with emerging severe weather. Here, North American Mesoscale (NAM)

* *Corresponding author address*: Xiang Li, Information Technology and Systems Center, Univ. of Alabama in Huntsville, Huntsville, AL 35899; email: xli@itsc.uah.edu

model forecasts interpolated to a grid with 40-km resolution (NAM 212 grid) were analyzed using the PEA to determine areas of interesting weather. The NAM is produced by the National Centers for Environmental Prediction (NCEP) and is initialized every 6 hours. Each initialization spawns forecasts for three hour intervals out to 84 hours from the forecast initialization time. Theoretically, NAM analyses could be used in the same way for nowcasting purposes. At each output time, a total of 617 model variable 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 one winter month, consisting of 29 forecasts ranging from 17 January 2007 to 22 February 2007.

For this study, individual parameter fields for a given model run are directly obtained from an in-house web service except the total precipitation field and convective precipitation field, a total of 615 fields.

3. Approach

The PEA is a statistical feature extraction algorithm that automatically identifies phenomena of interest (Ramachandran et al, 2006). A “phenomenon” in a geospatial data set can be identified as a region significantly different from the rest of the scene. Statistically, therefore, a region of geophysical phenomenon can be characterized as having a value higher or lower than the average background intensity value and/or having higher intensity variations or gradients as compared to the remaining data points. The PEA uses hypothetical T- and F-tests to iteratively decompose an image (in this case data from a model pressure level or other variables) and identify regions of interest. For more details refer to Ramachandran et al (2006). Previously, the PEA was used in the LEAD project to identify heavy precipitation from the NAM precipitation output as a LEAD capability demonstration (Clark et al. 2006).

For each 3-hour forecast initialized from the daily 00Z model run, the PEA is applied to each of the 615 NAM fields to obtain regions of interest for that particular field. The PEA output for a field is a binary mask image, indicating whether a pixel belongs to a region of interest or not. A composite image is then generated by averaging all 615 mask images for a given forecast. Large intensity values in the composite image correspond to regions which are identified in multiple individual NAM fields. As a result, regions of large intensity in the composite image suggest regions of interest contributed from many NAM fields. The PEA is then applied to the composite image to extract final regions of interest. A statistical score S is used to evaluate the relative importance of the individual NAM fields in detecting the overall regions of interest. Assuming P is the set of pixels identified in regions of interest, P_i indicates pixels from model variable field i and P_{comp} indicates pixels from composite image. Then the score S_i is defined as $P_{i,i}$ / P_{comp} where $P_{i,i}$ is the intersection of P_{comp} and P_i This statistical score is calculated for each parameter field over each of the 29 model runs. The mean scores for individual parameters are ranked with the top 10 NAM fields listed in Table 1. This table shows that the pressure vertical velocity fields in the mid-troposphere are the highest contributors to overall regions of interest.

The difference field characterizes rapid changes in weather conditions exhibited in the model for each model output variable and is defined as the absolute difference between the 3-hour forecasts from the 00Z initialized model run and the 15-hour forecasts from the 12Z initialized model runs from the previous day. Large differences may indicate where weather is changing rapidly or is most sensitive to variations in model initial conditions. Therefore, it is assumed that difference fields contain valuable information on regions of significant weather events of interest for further refined forecasts. Similar to the approach described above for NAM fields, the PEA is applied to the difference fields to further detect regions of interest. Table 2 shows the top 10 NAM difference fields that contribute most to detected regions. From Table 2, it is apparent that the most significant contributors are parameters such as precipitation fields and vorticity fields, which are known to be associated with significant weather events.

Based on the analyses described above, the algorithm for automated detection of regions of weather interest is as follows:

1. Apply PEA to each NAM field to obtain mask field for regions of interest.
2. Apply PEA to each NAM difference field to obtain mask field for regions of interest.
3. Calculate a composite field as the weighted average of mask fields from each NAM and NAM difference field. The weights are the relative scores with top 10 shown in Tables 1 and 2.
4. Apply PEA to composite field to get mask field for overall regions of interest.
5. Apply standard region growing technique to extract these regions.
6. Apply higher threshold to regions with size larger than 300 grid points in order to obtain mesoscale phenomena. The threshold is determined based on data statistics of the particular region.

4. Experiments and Results

Figure 1 shows the results of this method using a 3-hour 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 shown in Figure 1d. From the mean sea level pressure field shown in Figure 1a, 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. In order to better understand the phenomena of interest, the detected regions are overlaid with mean sea level pressure contour lines and wind vectors at 1000 mb (Figure 1b), 500 mb (Figure 1c) and 850 mb (Figure 1d), respectively. Identified regions are shown as dots in the figures. Among other detected regions, the algorithm successfully identified regions corresponding to the three low pressure systems, as shown in these figures.

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. Shading in Figure 2a indicates the detected regions of interest from Figure 1, with the NEXRAD national mosaic reflectivity image from the same time shown in Figure 2b for comparison. 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. Note that the identified region over the central U.S., where there is good radar coverage, correlated well with areas of high radar reflectivity. However, 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 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 do not correlate as well with radar echoes; this could be due to greater variation of numerical weather forecasts over the eastern Pacific where upstream data are sparse.

The algorithm did not perform as well in another instance. Figure 3 shows the regions of interest identified in a 3-hour NAM forecast initialized at 00Z on 17 February 2007 and the corresponding radar reflectivity image. Compared to the radar reflectivity image in Figure 3b, the algorithm successfully identified an interesting region over Washington State. It also approximately identified the regions of interest over Colorado, Wyoming and Nebraska. However, it failed to identify the core weather systems over Illinois. More effort is required to examine the NAM model output in order to figure out what caused this failure; it could be due to the relatively predictable nature of the large weather system and its location in the middle of country where abundant observations can result in small run-to-run variations in numerical models.

5. Conclusion

This paper presents the preliminary results from a novel method of model-based weather event detection. This method is based on the PEA, a statistical data mining algorithm. Based on preliminary examination of 29 days of 3-hour 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 where higher resolution model runs are needed. This method can be fully automated with no human intervention required. More effort is needed to examine if and to what extent regions are falsely identified. Further study is also required 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.

References

Clark, R. D., D. Fitzgerald, T. Baltzer, R. Ramachandran, E. Joseph and S. Chiao, 2006: EARLY LEAD: A WRF ENSEMBLE DEMONSTRATING A DATA MINING CAPABILITY, *22nd International Conference on Interactive Information Processing Systems (IIPS)*, 86th Americal Meteorological Society Annual Meeting, Atlanta, Georgia

Droegemeier, K. K. and Co-Authors, 2004: Linked environments for atmospheric discovery (LEAD): A cyberinfrastructure for mesoscale meteorology research and education, *20th Conf. on Interactive Info. Processing Systems for Meteorology, Oceanography, and Hydrology*, Seattle, WA , Amer. Meteor. Soc.

Droegemeier, K. K., D. Gannon, D. Reed, B. Plale, J. Alameda, T. Baltzer, K. Brewster, R. Clark, B. Domenico, S. Graves, E. Joseph, V. Morris, D. Murray, R. Ramachandran, M. Ramamurthy, L. Ramakrishnan, J. Rushing, D. Weber, R. Wilhelmson, A. Wilson, M. Xue, and S. Yalda, 2005: Service-Oriented Environments in Research and Education for Dynamically Interacting with Mesoscale Weather. *IEEE Computing in Science & Engineering*, 7, 24-32.

Li, X, R. Ramachandran, S. Movva, S. Graves, B. Plale, and N. Vijayakumar, 2008: Storm Clustering for Data-driven Weather Forecasting, *24th International Conference on Interactive Information Processing Systems (IIPS)*, 88th AMS Annual Meeting, New Orleans, Louisiana

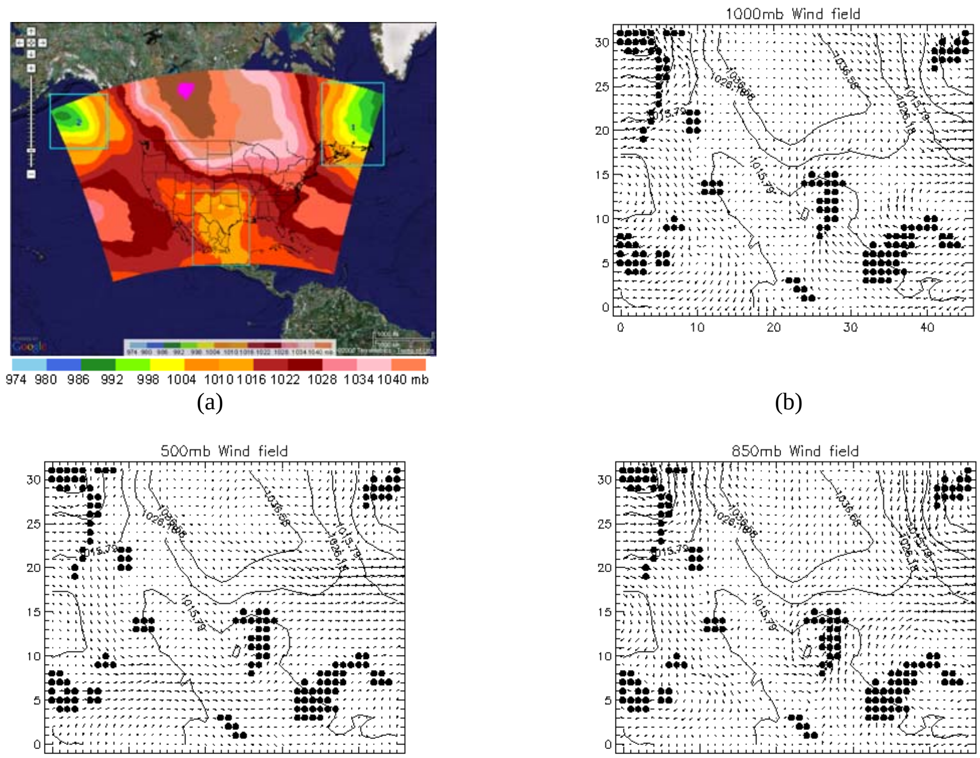
Ramachandran, R., X. Li, S. Graves, R. D. Clark, and D. Fitzgerald, 2006: PEA: Phenomena Extraction Algorithm, *22nd International Conference on Interactive Information Processing Systems (IIPS)*, 86th AMS Annual Meeting, Atlanta, Georgia.

Table 1. Top 10 list of relative score of NAM fields that contributes most to regions of interest

NAM Field	Score
650 hpa Pressure Vertical Velocity	0.5374
600 hpa Pressure Vertical Velocity	0.5366
625 hpa Pressure Vertical Velocity	0.5361
575 hpa Pressure Vertical Velocity	0.5340
675 hpa Pressure Vertical Velocity	0.5320
700 hpa Pressure Vertical Velocity	0.5288
725 hpa Pressure Vertical Velocity	0.5246
550 hpa Pressure Vertical Velocity	0.5243
180 hpa -150 hpa Pressure Vertical Velocity	0.5206
500 hpa Pressure Vertical Velocity	0.5176

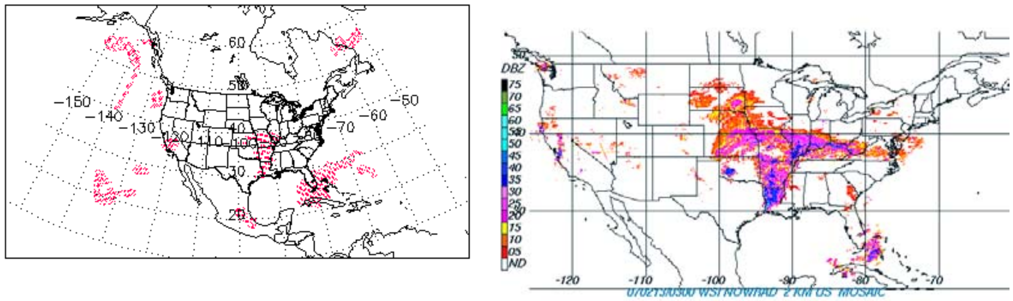
Table 2. Top 10 list of relative score of NAM difference fields that contributes most to regions of interest

NAM difference field	Score
Total column condensate	0.6678
Total column snow	0.6336
Precipitation rate	0.6017
Total precipitation	0.5919
725 hpa Pressure vertical velocity	0.5909
700 hpa Pressure vertical velocity	0.5856
675 hpa Pressure vertical velocity	0.5829
775 hpa Pressure vertical velocity	0.5813
700 hpa Absoluate vorticity	0.5805
650 hpa Pressure vertical velocity	0.5761



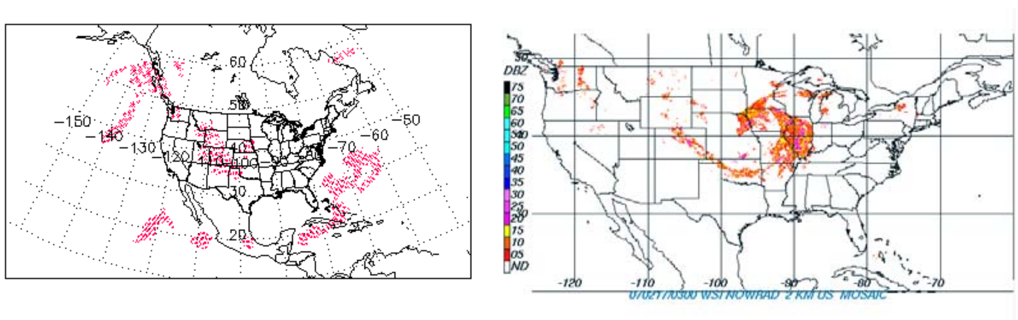
(c) (d)

Figure 1 Results for NAM forecast at 02/13/2007, 03Z with model run at 02/13/2007, 00Z. (a) Mean sea level pressure field. Identified regions of interest shown as dots with contour overlay of mean sea level pressure and wind field at (b) 1000mb, (c) 500mb and (d) 850mb, respectively.



(a) (b)

Figure 2 (a) identified regions of interest in cylindrical map projection at 03Z, 02/13/2007, (b) NEXRAD national mosaic radar reflectivity image at 03Z, 02/13/2007.



(a) (b)

Figure 3 (a) identified regions of interest in cylindrical map projection at 03Z, 02/17/2007, (b) NEXRAD national mosaic radar reflectivity image at 03Z, 02/17/2007.