NASA Contractor Report NASA/CR-2006-214199

Evaluation of Mesoscale Model Phenomenological Verification Techniques

Prepared by:
Winifred Lambert
The Applied Meteorology Unit

Prepared for:
Kennedy Space Center
Under Contract NAS10-01052

NASA
National Aeronautics and Space Administration

Office of Management
Scientific and Technical Information Program

2006

Attributes and Acknowledgments

NASA/KSC POC:
Dr. Francis J. Merceret
KT-C-H

Executive Summary

Forecasters at the Spaceflight Meteorology Group, 45th Weather Squadron, and National Weather Service in Melbourne, FL use mesoscale numerical weather prediction model output in creating their operational forecasts. These models aid in forecasting weather phenomena that could compromise the safety of launch, landing, and daily ground operations and must produce reasonable weather forecasts in order for their output to be useful in operations. Considering the importance of model forecasts to operations, their accuracy in forecasting critical weather phenomena must be verified to determine their usefulness. The currently-used traditional verification techniques involve an objective point-by-point comparison of model output and observations valid at the same time and location. The resulting statistics can unfairly penalize high-resolution models that make realistic forecasts of a certain phenomena, but are offset from the observations in small time and/or space increments. Manual subjective verification can provide a more valid representation of model performance, but is time-consuming and prone to personal biases. An objective technique that verifies specific meteorological phenomena, much in the way a human would in a subjective evaluation, would likely produce a more realistic assessment of model performance.

Such techniques are being developed in the research community. The Applied Meteorology Unit (AMU) was tasked to conduct a literature search to identify phenomenological verification techniques being developed, determine if any are ready to use operationally, and outline the steps needed to implement any operationally-ready techniques into the Advanced Weather Information Processing System (AWIPS).

The AMU conducted a search of all literature on the topic of phenomenological-based mesoscale model verification techniques and found 10 different techniques in various stages of development. Six of the techniques were developed to verify precipitation forecasts, one to verify sea breeze forecasts, and three were capable of verifying several phenomena. The AMU also determined the feasibility of transitioning each technique into operations and rated the operational capability of each technique on a subjective 1–10 scale:

- 1 indicates that the technique is only in the initial stages of development,
- 2–5 indicates that the technique is still undergoing modifications and is not ready for operations,
- 6–8 indicates a higher probability of integrating the technique into AWIPS with code modifications, and
- 9–10 indicates that the technique was created for AWIPS and is ready for implementation.

Eight of the techniques were assigned a rating of 5 or below. The other two received ratings of 6 and 7, and none of the techniques a rating of 9–10.

At the current time, there are no phenomenological model verification techniques ready for operational use. However, several of the techniques described in this report may become viable techniques in the future and should be monitored for updates in the literature. The desire to use a phenomenological verification technique is widespread in the modeling community, and it is likely that other techniques besides those described herein are being developed, but the work has not yet been published. Therefore, the AMU recommends that the literature continue to be monitored for updates to the techniques described in this report and for new techniques being developed whose results have not yet been published.

Table of Contents

List of Tables

1. Introduction

Forecasters at the Spaceflight Meteorology Group (SMG), 45th Weather Squadron (45 WS), and National Weather Service in Melbourne, FL (NWS MLB) use numerical weather prediction (NWP) model output on a daily basis in creating their operational forecasts. Models such as the Rapid Update Cycle (RUC), North American Mesoscale (NAM, formerly Eta) model, Global Forecast System (GFS), and the Advanced Regional Prediction System (ARPS) aid in forecasting weather phenomena that could compromise the safety of launch, landing, and daily ground operations. Such phenomena include low- and upper-level winds, cloud cover, timing and strength of the sea breeze, convection, and precipitation. Although no model can produce a flawless forecast of these phenomena, it must produce a reasonable depiction of the future state of the weather in order for its output to be useful for operational forecasting. Considering the importance of model forecasts to operations, their accuracy in forecasting critical weather phenomena must be verified properly to determine their actual usefulness.

The quality of a model forecast can be assessed through several verification methods. Given the large amount of model and observational data required to produce meaningful verification results, automated objective techniques are needed. However, it is well known in the modeling community that traditional objective techniques often fall short of providing an accurate depiction of model performance in forecasting mesoscale and convective-scale phenomena. Traditional techniques involve a point-by-point comparison of model output and observations valid at the same time and location. The resulting statistics can unfairly penalize high-resolution models that make realistic forecasts of certain phenomena that are offset from the observations in small time and/or space increments. Manual subjective verification can provide a more valid representation of model performance; however, subjective techniques are costly, time-consuming, and prone to personal biases. An objective technique that verifies specific meteorological phenomena, much in the way a human would in a subjective evaluation, is often needed to produce a more realistic assessment of model performance for a given application. Such techniques are being developed in the research community.

The Applied Meteorology Unit (AMU) was tasked to conduct a literature search to identify the phenomenological verification techniques being developed, assess if any are ready to use operationally, and determine the steps needed to implement any operationally-ready techniques into the Advanced Weather Information Processing System (AWIPS). In all, 10 candidate verification techniques were found through a literature search. For each technique, the AMU identified the meteorological phenomenon or phenomena that each technique was developed to verify, determined the method used to verify the phenomena, and assessed the operational readiness for incorporation into AWIPS. The report is organized as follows. Section 2 provides a survey of all the literature containing information about phenomenological-based (aka event- or object-based) verification techniques. The feasibility of implementing each technique into operations in AWIPS is given in Section 3, and a summary is provided in Section 4.

2. Relevant Literature

The AMU conducted a search for literature on the topic of phenomenological-based verification techniques. A total of 21 journal articles, preprints, presentations, and web sites were identified that described an event-based verification technique or the need to develop one. Of those, 13 described an actual technique that had been or was about to be developed. A few of the articles described the same technique used in different studies. In all, 10 unique techniques were identified in the literature. Tables were created that contain detailed summaries about the articles, which are in Appendices A–C.

The descriptions of the techniques were organized by the type of phenomenon being verified. There were three phenomenon categories:

- Precipitation,
- Sea Breeze, and
- Multiple Phenomena.

Considering the goal of the task, the AMU also determined the stage of development for each technique to determine its level of readiness for operations and implementation in AWIPS. This determination was strictly subjective and based on several factors:

- Whether an actual routine was developed or if it was just proposed as a possible routine,
- If a developed routine was an initial version that needed further testing,
- How many cases were used to test the routine and if it had been used in multiple studies as a valid verification method, and
- Whether it was developed for real-time operations or AWIPS.

Within each phenomenon category in this section, the techniques are listed in order of increasing operational-readiness.

2.1. Precipitation

Six of the ten techniques were developed to verify model forecasts of precipitation only. A brief summary of each technique is given below, with more detailed summaries provided in Appendix A.

2.1.1. Automated Rainfall System Classification

The work in Baldwin et al. (2005, Table A.1) focused on identifying characteristics of precipitation areas to be used in developing an automated rainfall system classification. Such a classification was expected to be helpful for future use in an object-based verification tool. The classification was to discern between stratiform and convective precipitation, with convective precipitation subdivided into linear and cellular types.

The authors conducted a statistical analysis of observed hourly rainfall rates and rain-area shapes using a training data set of 48 cases to determine features important to each classification. These statistical attributes were used to determine the classification of rainfall areas in a testing data set of 100 cases. The procedure was able to distinguish between stratiform and convective systems in 89% of the cases, and between stratiform, linear convective, and cellular convective in 85% of the cases. The conclusions state, however, that further work needs to be done to refine the procedure by using more cases in the training data set. This may reveal other features important in classifying rainfall systems.

This is a rainfall system classification procedure and not a verification method. The authors state that it is a precursor to a verification method. The procedure needs more refinement and is not ready for operational classification of rainfall systems.

2.1.2. Storm-Scale Statistical Measures

The goal of Zepeda-Arce et al. (2000, Table A.2) was to create a verification technique that would help model developers determine deficiencies in microphysical parameterization schemes by analyzing the quality of the model precipitation forecasts. The technique consists of four algorithms that provide different statistical measures of model performance in predictions of storm-scale precipitation. The first algorithm calculates a threat score (TS) as

$$TS = A_c/(A_o + A_f - A_c),$$

where A_c is the size of the correctly forecast area of rainfall bounded by a defined threshold amount, A_o is the observed area, and A_f is the forecast area. This shows the ability of the model to predict the area size of precipitation that has amounts exceeding a given threshold. The second is a depth-area-duration curve, in which rainfall depth (accumulated amount) versus the size of the area over which that depth is exceeded is plotted for a fixed time duration. This measure shows the areal variance in precipitation amount for the forecasts and observations of precipitation individually. The third algorithm is based on results from previous studies that showed a constant increase in rainfall variance with spatial scale on a logarithmic scale. This method was used to determine if the observed and forecast precipitation exhibited similar characteristics. The fourth algorithm expanded on the third by determining the variance in rainfall intensity over different space and time scales.

These routines were developed for post-analysis of model forecasts to help model developers determine the strengths and weaknesses of the model microphysical parameterizations. The mathematics are complicated and it is not clear how the output could be useful in operations. It would require considerable effort to transform the routines and create output useful to operations.

2.1.3. Intensity/Spatial Scale Verification

An intensity-scale verification technique for precipitation forecasts was described in Casati et al. (2004, Table A.3). The goal of the technique was to allow the user to assess the skill of the forecast in terms of precipitation rate and spatial scale errors. The observed and forecast precipitation data were pre-processed in several steps before the verification took place. A small amount of uniformly distributed noise was added to non-zero precipitation values in the analysis and forecast fields. This helped compensate for the effects caused by digitizing the archived data values. The new precipitation rate values were normalized in a logarithmic (base 2) transformation. This reduced the skewness of the rainfall rate distribution due to the large amount of small values, producing more normally distributed values. Finally, the forecasts were re-calibrated by substituting each value in the forecast image with the value in the analysis image having the same cumulative probability, which is the probability that the precipitation rate will be a certain value or less.

The forecast and analysis were converted to binary images based on rainfall rate threshold: a "1" for values greater than and a "0" for values less than the threshold. A binary error image was created by subtracting the analysis binary image from the forecast binary image, and errors on different spatial scales were determined through a wavelet decomposition analysis. The error values from the wavelet analysis were used in calculations of mean square error (MSE) and skill score (SS), which revealed model performance in both spatial and intensity scales.

The authors believe that phenomena other than precipitation can be verified with this technique. It is a new technique and needs more testing to determine its viability and usefulness in verifying phenomena other than precipitation.

2.1.4. Acuity-Fidelity

Marshall et al. (2004) describes a method in which the two metrics of acuity and fidelity were used to verify model precipitation forecasts (Table A.4). Acuity was calculated by finding the best matching forecast for every observation, and fidelity by finding the best matching observation for the forecast. The best match was not necessarily the object that was closest in time or space. Acuity and fidelity were calculated separately by minimizing a cost function with four components: spatial difference, temporal difference, intensity difference, and missed events. The authors conducted sensitivity studies to determine the best value for these parameters, and then used these values in verifying the performance of the models.

This is another proof-of-concept approach that is still being tested. The cost function component multipliers are configurable, but sensitivity tests would have to be conducted to determine their appropriate magnitudes based on the model, verification data, and phenomenon of interest.

2.1.5. Convolved Object Matching

Bullock et al. (2004) and Chapman et al. (2004) both describe a convolved-object-matching process for use in evaluating quantitative precipitation forecasts. The basic concept was described in Bullock et al. (2004, Table A.5) and examples using real data were provided in Chapman et al. (2004, Table A.6). In this method, both the forecast and observed fields of precipitation were resolved into objects to be compared. The objects were created by applying a filter to the raw data, called a convolving function. The convolving function attenuated the large gradients in the data while maintaining the small gradients. The convolved data were then filtered by a threshold value found in the precipitation field. A 'mask' field was created by setting all values above the threshold to 1 and all values below to 0, resulting in a field of object shapes. A new grid was created in which the original data values were restored where the mask field was 1, and set to 0 where the mask field was 0. Shape attributes (e.g. centroid and axes) of each of the objects were calculated, and shapes (e.g. band aids and ellipses) were fit to the objects. The closer the forecast object attributes matched the observed object attributes, the better the forecast was said to be. Forecast quality could be summarized by statistics of object attribute differences.

Two different tests were conducted in Chapman et al. (2004). In the first, the authors determined a threshold number of grid points between two or more objects below which the objects were merged into one and above which the objects were deemed separate. This required a human subjective analysis, but the results indicated that an automated matching technique might be useful in some cases. In the second test, data from two other cases were used to illustrate the utility of looking at smaller scale objects to help match larger scale objects. It was possible to consider two areas related if their small and large scale features were similar.

This technique was able to detect distinct objects and match forecast objects to observed objects. However, more sophisticated merging and object matching routines were needed for complex cases as described in Bullock et al. (2004). This technique needs further refining before it can be used in operations with confidence.

2.1.6. Contiguous Rain Area

The technique described in Ebert and McBride (2000, Table A.7) was derived from that of Hoffman et al. (1995; Section 2.3.3, Table C.3) in which the entire forecast object is translated as one entity. A Contiguous Rain Area (CRA) was defined as the union of the observed and forecast precipitation areas that exceeded a user-specified rainfall amount. In keeping with the ideas of Hoffman et al. (1995), the technique was tested with forecasts from a larger (regional) scale model. The forecast precipitation area was shifted incrementally by grid points over the observed area until the total MSE between the forecast and observations was minimized. The errors due to displacement, rain amount, and rain pattern were then calculated. The authors also conducted tests to determine how many grid points within a rain system were needed to obtain realistic verification results when an observed or modeled precipitation area was cut off by an observational (e.g. coastline) or model boundary. For their parameter settings, they found at least 20 grid points were needed within these rain areas.

The CRA technique was used by Grams et al. (2004) to determine how well mesoscale models predicted precipitation systems based on their morphology (Table A.8). They used the CRA method but modified it to account for the higher spatial resolution of the models and the shorter time period over which precipitation accumulated. The authors further defined different morphology types in linear and non-linear groups, and were able to create CRAs based on these criteria.

This technique appears to be ready to use in a post-analysis mode, although the user must determine appropriate parameter settings through testing. The technique's parameters can be changed to accommodate any time and space resolution as evidenced in Grams et al. (2004, Table A.8). However, it may take considerable effort to incorporate this technique into AWIPS.

2.2. Sea Breeze

The Contour Error Map (CEM) technique was described in Case et al. (2004, Table B.1), and is the one technique in this study that verifies sea breeze forecasts exclusively. In this method, data from observations and the model were interpolated to the same grid, and then a binary threshold was applied to distinguish between onshore (easterly) and offshore (westerly) flow. A time estimation filter was applied to the sine of the wind direction to determine the timing of transition from offshore to onshore winds at each of the grid points. The transition times at each grid point were displayed in a 2-D plot. A negative time gradient in the east–west direction indicated a west-to-east progression of a boundary. This was assumed to be the result of a non-sea breeze phenomena such as a river breeze or convective outflow boundary and was eliminated from the analysis in a process called image erosion. The verification statistics included determining the

- Fractional area of the grid with only observed sea breeze transition times,
- Fractional area of grid with only forecast sea breeze transition times, and
- Average and standard deviation of the sea breeze transition time errors at grid points that experienced both observed and forecast sea breeze transition.

This technique was developed as a post-process, proof-of-concept technique. The technique is tuned to detect sea breezes in the Kennedy Space Center (KSC)/Cape Canaveral Air Force Station (CCAFS) area using high temporal and spatial resolution data. It was not tested on other phenomena. The authors believe this technique could be more fully automated and transitioned into AWIPS for real-time operations with moderate effort.

2.3. Multiple Phenomena

There were three techniques designed to verify multiple phenomena, including rain and wind. Summaries containing more details of these studies are provided in Appendix C.

2.3.1. Mesoscale Verification Tool

The Mesoscale Verification Tool (MVT) along with an associated Mesoscale Data Manipulator (MDP) were in the beginning stages of development at the University of Washington as described by Sandgathe and Heiss (2004, Table C.1). The authors stated that a verification tool should be automated and flexible; adaptable to issues concerning forecast parameters, timing, and intensity; capable of evaluating distortion, timing, and amplitude errors; able to address large numbers of cases and multiple models rapidly; be statistically sound; and present results that are easy to interpret.

The MVT separated the forecast error into amplitude and timing components using the procedure defined by Hoffman et al. (1995, Section 2.3.3, Table C.3), which uses a search technique to find a forecast object that is similar to an observed object on a grid. The authors found that the full grid point-by-grid point search was too slow to be operationally feasible, so they employed accelerated search techniques based on image matching algorithms used by the motion picture industry. The main focus of the paper was a web-based graphical user interface (GUI) developed to display graphical representations of the model output, and how to use the MDP in selecting dates, initialization times, model domain, verifying field and forecast hour, and other items.

An email conversation with the author indicated that the verification techniques are still being developed and tested and the system is not ready for use. It may be quite some time before the technique is developed and working.

2.3.2. Composite Verification

Nachamkin (2004, hereafter N1) and Nachamkin et. al. (2004, hereafter N2) both discuss the same verification technique tested on two different phenomena: mistral wind (N1) and heavy precipitation (N2) events (both in Table C.2). The mistral is a strong northerly wind event that occurs over the northern Mediterranean Sea and the heavy precipitation events were those of 25 mm or more over the contiguous United States. Each event occurred over a defined spatial area.

Once identified, the forecast and observed events were translated to separate relative grids, one for the forecast and one for the observations, with the center of each event-object positioned at the center point of the grids. The relative grids had spacing equal to that of the model data. The observed objects associated with the forecast objects were superimposed on the forecast relative grid in the same relative location as on the regular grid, and vice versa. In the forecast relative grid, the observations were conditional on the existence of a forecast event, and in the observation relative grid, the forecasts were conditional on the existence of an observed event. The two grids in this method allow two general questions to be answered: What was observed if an event was forecast and what was forecast when an event was observed? Many events were composited together to determine the general statistical properties of the observed and forecast events separately and in relation to each other. Statistics were calculated and displayed to demonstrate the magnitude and location differences between the forecasts and observations in each conditional grid.

This method requires an archive of events to get overall results of how well a model predicts the attributes of a specific phenomenon, and was not developed for real-time verification. It would be more useful for a climatological verification. The mechanics of the method, which involves moving both forecast and observations to a relative grid, are more cumbersome than complicated.

2.3.3. Object Distortion

This technique was introduced in Hoffman et al. (1995, Table C.3) and was developed for verification of phenomena on the synoptic scale, not the mesoscale. The distortion of an object was defined as the combination of a spatial displacement error and an amplitude error. Any error not accounted for by the distortion was deemed residual error. The displacement was defined to be a smooth transformation of a field without modification to the magnitude of the data. The transformation could include translation, stretching, and rotation of the object through movement of the individual grid points within an object. Limits were imposed on how far from the original grid point the displaced data could be moved. The values of the data were then multiplied by an amplification factor to fit the observed field as closely as possible. The displacement and amplification took place until the root mean square (RMS) error was minimized.

The authors discussed another method in which the entire forecast field was displaced and amplified as one object to match as closely as possible the location of the observed object. No distortion to the shape of the object took place. Then, all the data values were multiplied by one amplitude factor that minimized the error. As in the previous method, the displacement and amplitude factors were chosen such that the total RMS error in the analysis area was minimized.

This technique was developed as a prototype to verify forecasts of synoptic-scale phenomena. It still needs testing and development to determine appropriate parameter settings, but it performed well on the cases presented in the paper. Several articles referencing this technique have been published, most notably Ebert and McBride (2000, Table A.7). It appears to be a seminal paper on the topic of model object-based verification techniques.

3. Operational Readiness

The AMU determined the feasibility of transitioning each technique identified in this report into operations by whether it could

- Be automated for real time,
- Provide graphical displays of the verification information, and
- Be integrated into AWIPS.

3.1. Technique Ratings

The AMU rated the operational capability of each technique on a 1–10 scale. A "1" indicates that the technique is only in the initial stages of development and needs much more testing and modification. A "2"–"5" indicates that the technique is still undergoing modifications and is not ready for transition into operations, but future literature on the technique should be monitored. A "6"–"8" indicates a higher probability of integrating the technique in AWIPS with moderate to significant modifications to the code. A "9"–"10" would indicate that the technique was created for AWIPS and that it is ready or almost ready for implementation. The ratings for each technique are shown in Table 1.

Table 1. A list of all the techniques discussed in this report, their operational readiness ratings on a scale from 1–10, and references to their description locations in the report. These techniques should be monitored for further testing and development.

Technique	*Rating*	*Reference Location in Report*
Precipitation		
Automated Rainfall System Classification	1	Section 2.1.1, Table A.1
Storm-Scale Statistical Measures	2	Section 2.1.2, Table A.2
Intensity/Spatial Scale Verification	3	Section 2.1.3, Table A.3
Acuity-Fidelity	3	Section 2.1.4, Table A.4
Convolved Object Matching	4	Section 2.1.5, Tables A.5 and A.6
Contiguous Rain Area	6	Section 2.1.6, Tables A.7 and A.8
Sea Breeze		
Contour Error Map	7	Section 2.2, Table B.1
Multiple Phenomena		
Mesoscale Verification Tool	2	Section 2.3.1, Table C.1
Composite Verification	3	Section 2.3.2, Table C.2
Object Distortion	5	Section 2.3.3, Table C.3

3.2. Integration into AWIPS

Although none of the techniques were ready for transition into real-time operations in AWIPS, it would still be helpful for future reference to outline the steps needed for AWIPS implementation. The procedure discussed in the following two paragraphs provides only a general idea of the steps to be taken in AWIPS-implementation of a technique. The ability to make the techniques available in AWIPS is dependent on AMU and AMU customer expertise in modifying AWIPS, and the state of AWIPS when a technique is finally ready for operations. If and when a phenomenological verification technique has been determined to be ready for operational use, the idea of making the technique available in AWIPS must be revisited and the exact steps to do so determined at that time.

The first step would be to ensure that the code for the technique is written in a programming language compatible with AWIPS. At this time, those languages include C, Python, Perl, and C++, FORTRAN, Java, Tcl/Tk, and Motif. If the code is not written in any of these languages, it would have to be translated into one of them. Beyond that, the technique must be able to process the mesoscale model forecast data available in AWIPS at the time and space resolution of that data. Observational data used to verify model performance must be analyzed to a grid prior to being used in the technique, preferably to a grid size similar to that of the model data. This can be done using a tool such as the ARPS Data Analysis System (ADAS), already in use at SMG and NWS MLB. Another possible source for gridded observational data is the Real-Time Mesoscale Analysis (RTMA) to be developed in the near future by NOAA's Environmental Modeling Center. This will be an hourly analysis of surface observations on a 5 X 5 km grid over the Continental U.S. (CONUS). There are certain parameter values that must be tuned in several of the techniques described in this report. Some of the studies conducted tests with model and observational data to determine the optimal values for these parameters, which can depend on the phenomenon being verified, time of year or day, the model space and time resolution, and other issues. Depending on the technique, this type of testing may have to be done to determine the parameter settings prior to automated implementation of the technique.

Once the code language, model and observational data, and parameter setting issues have been ironed out, there should be a way for users to run the technique automatically with minimal user input. The localization capabilities in AWIPS can be used to create menu items in which the user can define the model and observational data sets to use in the verification. The user could also choose a verification technique to use if two or more are made available. AWIPS can also accommodate graphical and textual output of the verification results.

4. Summary and Recommendations

This report provided summaries of articles that describe phenomenological model verification techniques, and discussed the feasibility of using any of the techniques operationally. Forecasters at SMG, 45 WS, and NWS MLB all use model output for guidance in their daily forecast operations. Considering the importance of model data to these forecasts, the accuracy of models in forecasting critical weather phenomena must be verified to determine their actual usefulness. The most common verification techniques involve a point-by-point comparison of model output and observations valid at the same time and location. These techniques are believed to unfairly penalize high-resolution models that make realistic forecasts of phenomena, but may be offset from the observations in small time and/or space increments. The consensus opinion from the mesoscale modeling community is that a verification technique created specifically to identify phenomena, or objects, in the model and observed data is likely to provide a more accurate portrayal of model performance.

All of the articles summarized in this report, save one, were published on or after the year 2000, indicating the relative newness of this technology. However, mentions on the need for such technology have appeared in papers dating back into the 1970's. This also speaks to the inherent difficulty in creating automated, objective algorithms that can make decisions similar to that of humans. Phenomenological model verification is an inherently complex seven-dimensional problem (Mr. William Roeder, 45 WS, Personal Communication):

1) Occurrence (yes/no);
2) x-,
3) y-, and
4) z-location;
5) Timing;
6) Areal coverage; and
7) Intensity.

There can be some ambiguity for each dimension, and each dimension can have several metrics to fully describe the errors. In the articles found for this report, all of these dimensions are accounted for in different ways, but no technique accounted for all seven dimensions at once.

4.1. Summary of Techniques

The AMU identified 10 different event-based verification techniques through a literature search. Of the 10 techniques,

- Six were created to verify precipitation events, although 2 of the 6 stated that their technique could be used for other phenomena,
- One was created to verify sea breeze events, and
- Three were created for any phenomena that could be defined over a specific geographical area (e.g. pressure fields, localized wind events, etc.).

All of the techniques were still undergoing some level of development and testing at the time the work was published. Eight of the techniques received a subjective rating of 5 or below in Table 1, indicating that these techniques are still undergoing modifications and are not ready for transition into operations, but future literature describing them should be monitored. The other two received ratings of 6 and 7, signifying that they have a higher probability of being integrated into AWIPS with some level of modifications to the code. None of the techniques were developed specifically for real-time use in AWIPS.

The AMU consulted with Ms. Jennifer Mahoney of the NOAA Global Systems Division (formerly Forecast Systems Laboratory), a well-known expert in the field of model verification. Ms. Mahoney stated that, while there is much research being conducted in this area, no one phenomenological verification technique has proven robust or reliable enough to verify operational or archived model data with confidence. Many issues still remain, such as how to identify a specific phenomenon or event objectively, what parameters should be used, and what threshold values are appropriate. Ms. Mahoney estimated that such a reliable technique may be available in 5–10 years given the current rate of advancements in the research. She also stated that the CRA technique described in Ebert and McBride (2000) is gaining favor among several groups (Section 2.1.6, Table A.7).

4.2. Recommendations

At the current time, there are no phenomenological model verification techniques ready for operational use either for AWIPS or any other platform and, therefore, none can be transitioned for operational use in the short-term. Several of the techniques described here may become robust and reliable techniques in the future and should be monitored for updates in the literature. The desire to develop a phenomenological verification technique is widespread in the modeling community, and it is likely that other techniques besides those described herein are being developed but the work has not yet been published. Based on the findings in this report, the following actions are recommended:

- Monitor the progress of all techniques that received a rating of 2 or higher in Table 1,
- Monitor conference preprints and journal articles for new techniques that show promise,
- Closely monitor studies that use the CRA technique since it is considered one of the better techniques by the model verification community, and
- Determine the amount of work needed to transition the CEM technique into AWIPS. Even though it only verifies sea breeze transitions, the sea breeze is a critical weather generator in the KSC/CCAFS area.

Appendix A

Table A.1. Detailed summary of Baldwin et al. (2005), discussed in Section 2.1.1.	
Reference	Baldwin, M. E., J. S. Kain, and S. Lakshmivarahan, 2005: Development of an automated classification procedure for rainfall systems. *Mon. Wea. Rev.,* **133**, 844-862.
Weather Element	Rain Systems
Model	None, developed for future use in model verification and other applications.
Data	National Centers for Environmental Prediction (NCEP) Stage IV rainfall analysis (national 1-hour rainfall estimates using radar and rain gauge data on a 4 X 4 km grid)
Time Period	2002
Name of Technique	Automated Rainfall System Classification
Description	The goal of this work was to develop an automated rainfall system classification for future use in a possible model phenomenological verification tool and to develop climatologies. The classification was to discern between convective and stratiform precipitation, and subdivided convective precipitation into linear and cellular types. A training data set was created by manually choosing 48 precipitation "objects" from the data set. The selection was based on typical rainfall systems that occur in the US throughout the year. The systems were divided into convective and non-convective events based on rainfall rates, and then the convective cases were subdivided into linear and cellular cases using subjective techniques. Two groups of attributes were created for the classification system. The first was based on rainfall intensity. The distributions of rainfall amount were determined for each object and fit to a theoretical gamma distribution. The gamma scale and shape parameters of the distribution for each object were used as possible attributes. The second group was based on spatial continuity. A correlogram for each object was constructed, which showed the correlation between all possible pixels separated by a distance lag value. The area and eccentricity of various rainfall contour values were used as possible attributes. Tests using a hierarchical cluster analysis showed that the combination of the gamma scale parameter and the correlogram eccentricity was the best discriminator between precipitation classes. A cluster analysis using these two parameters was created and tested on the 48 objects in the training set, then on 100 objects in the testing data set.
Results	On the training data set, the cluster analysis was able to discriminate between convective and non-convective precipitation in 100% of the cases. It was able to classify between linear convection, cellular convection, and stratiform precipitation in 90.5% of the cases. Values were lower for the testing data set: 89% for the convective/non-convective distinction, and 85% for linear, cellular, and stratiform.
Operational Capability	The conclusions state that the procedure needs more refinement through use of more data in the training data set. More data may also reveal additional attributes for distinguishing between precipitation classes. It is not ready for operational classification of rainfall systems as more research is needed to refine the process.

Table A.2. Detailed summary of Zepeda-Arce et al. (2000), discussed in Section 2.1.2.	
Reference	Zepeda-Arce, J, E. Foufoula-Georgiou, and K. Droegemeier, 2000: Space-time rainfall organization and its roll in validating quantitative precipitation forecasts. *J. Geophys. Res.*, **105 No. D8**, 10 129-10 146.
Weather Element	Storm-scale Precipitation
Model	ARPS 6 km inner/18 km outer grid
Data	Hourly and 15-min accumulated precipitation forecasts from 6 km ARPS, hourly and 18-min rainfall accumulations estimated from local WSR-88D radars on a 4 km grid from a multiple squall-line storm system.
Time Period	May 7-8, 1995
Name of Technique	Storm-Scale Statistical Measures
Description	This technique is made up of four algorithms designed to help model developers determine deficiencies in microphysical parameterizations by analyzing the quality of model precipitation forecasts. The first algorithm is a threat score (TS) that measures the model's ability to predict the size of the precipitation area that has amounts exceeding a given threshold. The formula is $TS=A_c/(A_o+A_f-A_c)$ where A_c is the correctly forecast area size, A_o is the observed area size, and A_f is the forecast area size. It can be computed over any desired grid scale. The second algorithm is a depth-area-duration curve, in which accumulated rainfall amount versus the size of the area over which that depth is exceeded is plotted for a fixed time duration. This shows the areal variance in amount for the forecast and observed precipitation individually, allowing a comparison of the internal precipitation structure between the forecast and the observations. The third algorithm uses results from previous studies on rainfall variability as a function of scale. These studies showed an increase in rainfall variance with spatial scale that remains logarithmically constant. Given that this scale invariance was often found to exist in observed precipitation, this method was used to determine if model forecast precipitation exhibited the same characteristics. The fourth algorithm expands on the idea of consistent spatial scale variance. It determines the variance in rainfall intensity over different space and time scales. As with the third method, consistencies in the graphs of the variance were found in past studies. Therefore, the method was used to determine if the forecast and observed variances were similar.
Results	Except for the TS, all routines were developed specifically for verification of precipitation forecasts. The results show that observations and forecasts are in greater agreement at larger spatial and temporal scales. While this result is probably intuitive, the routines show the values of the spatial and time scales at which they do come into close agreement. The last two measures are represented by lines with slopes. Model output compares well with the observations when the slopes and values of their lines are close in value.
Operational Capability	These routines were not developed for real-time operations, but for post-analysis of model forecasts to help model developers determine the weaknesses and strengths of the microphysical parameterizations. The mathematics are rather complicated and it is not clear how the output could be useful in an operational capacity. Three of the routines could only be used for precipitation. It would require considerable effort to transform the routines and create output useful to operations.

Table A.3. Detailed summary of Casati et al. (2004), discussed in Section 2.1.3.	
Reference	Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Met. App.,* **11**, 141-154.
Weather Element	Precipitation
Model	Nimrod (very short-range mesoscale NWP system at UK Met Office)
Data	Nimrod analyses of rainfall estimated from UK radar images, satellite, and surface data, and Nimrod 3-hour forecasts.
Time Period	6 precipitation events in 1999
Name of Technique	Intensity Scale Verification
Description	This technique assesses the skill of the model precipitation forecasts as a function of spatial scale and intensity. The six cases were chosen to include a variety of precipitation features on different spatial scales, and to highlight the typical Nimrod forecast errors. The data needed processing before the verification technique was applied in order to obtain more reliable data (according to the authors). A small amount of uniformly distributed noise in the range -1/64 to +1/64 mm/hour was added to non-zero precipitation values in the analysis and forecast fields. This helped compensated for the effects caused by discretizing the archived data in multiples of 1/32 mm/hour. Then the precipitation rate values were normalized in a logarithmic (base 2) transformation. This normalization reduced the skewness of the rainfall rate distribution, due to the large amount of small values, and produced more normally distributed values. Finally, the forecasts were re-calibrated by substituting each value in the forecast image with the value in the analysis image having the same cumulative probability, which is the probability that the precipitation rate will be a certain value or less, i.e. P[precipitation rate $\leq$ x], where x is in the range of the observed or forecast precipitation rates. The forecast and analysis were converted to binary images based on rainfall rate threshold: a "1" for values greater than and a "0" for values less than the threshold. A binary error image was created by subtracting the analysis binary image from the forecast binary image. Errors on different spatial scales were determined through a wavelet decomposition analysis. The error values from the wavelet analysis were used in calculations of MSE and SS, which revealed the performance of the model in both spatial and intensity scales.
Results	The contour graphs of the spatial and intensity MSE and SS are intuitive and show the spatial scales at which intensity errors are greatest, and the intensity scales at which spatial errors are greatest. The process of replacing forecast values with observed values having the same cumulative probability seems to "fudge" the forecast to be more like the observations. This does not appear to create a fair assessment of model performance. However, a graph of the re-calibration function values used to make the transformation shows the forecast bias at different rainfall rates. In this case the model shows systematic behavior in forecasting too many low precipitation rate events and not enough high precipitation rate events. The authors contend that parameterization of these recalibration functions could be used to help calibrate future precipitation forecasts.
Operational Capability	The math appears to be simple enough to be able to run in real time. The authors believe that phenomena other than precipitation can be verified with this technique. Although the technique does not analyze the spatial displacement error mathematically, the binary error image could be used to determine that aspect.

Table A.4. Detailed summary of Marshall et al. (2004), discussed in Section 2.1.4.	
Reference	Marshall, S. F., Pl. J. Sousounis, and T. A Hutchinson, 2004: Verifying mesoscale model precipitation forecasts using an acuity-fidelity approach. *Preprint J13.3, Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 8 pp.
Weather Element	Precipitation
Model	Eta, RUC, two different grid configurations of WRF
Data	3-hourly Eta/RUC and 12-min WRF precipitation forecasts, NCEP Stage IV hourly precipitation, 36.5 – 44 degrees N, 103 – 92 degrees W (Midwest US)
Time Period	April – May 2003
Name of Technique	Acuity-Fidelity
Description	The two metrics of acuity and fidelity were used to determine model performance. Acuity was calculated by finding the best matching forecast for every observation, and fidelity by finding the best matching observation for the forecast. The best match was not necessarily the observation or forecast that was closest in time or space. Acuity and fidelity were calculated separately by minimizing a cost function with four components: spatial difference, temporal difference, intensity difference, and missed events. The components were all converted to common units of distance through constant multipliers in order to calculate a cost function value. The authors assigned initial values to the constants, and then conducted sensitivity studies to determine the best value for these parameters by holding three of them constant while varying one.
Results	Graphic representations of the acuity and fidelity cost functions showed the locations and extent of the model errors. Graphs of the individual components of the cost function showed where most of the error was in the model (location, timing, intensity, and/or missed events). Mean acuity and fidelity values can be calculated for the entire grid to determine the overall model performance. The goal of the authors was to develop a technique that measured the skill of model precipitation forecasts much in the way a subjective analysis would in considering the distance, timing, and intensity errors. They believe this technique can be applied to phenomena other than precipitation.
Operational Capability	This appears to be another proof-of-concept approach that is not ready to transition into operations. The cost function component multipliers are configurable and the values used in the study could be used as a first guess in verifying precipitation forecasts, but sensitivity tests would have to be conducted to determine their appropriate magnitudes based on the chosen model, verification data, and phenomenon of interest.

Table A.5. Detailed summary of Bullock et al. (2004), discussed in Section 2.1.5.	
Reference	Bullock, R., B. G. Brown, C. A Davis, M. Chapman, K. W. Manning, and R. Morss, 2004: An object-oriented approach to the verification of quantitative precipitation forecasts: Part I – Methodology. *Preprint J12.4, Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 6 pp.
Weather Element	Precipitation
Model	Weather Research and Forecasting (WRF)
Data	CONUS precipitation forecasts
Time Period	Summer 2001
Name of Technique	Convolved Object Matching (AMU–selected name)
Description	The goal of this research was to develop and test an object-oriented method of evaluating quantitative precipitation forecasts. Both the forecast and observed fields of precipitation were resolved into objects, or regions-of-interest, which were then compared. The objects were created by first applying a filter to the raw data, called a convolving function, and then a threshold was applied to the convolved field to reveal the objects. The convolving function attenuated the large gradients in the data while retaining the small gradients. The authors contend that a threshold applied to the raw precipitation data field, in which there are large and varying spatial gradients, would not produce representative objects and give a convincing example in their Figures 1 and 2. The convolved data were filtered by a chosen threshold value equal to some value in the original precipitation field. A 'mask' field was created by setting all values above the threshold to 1 and all values below to 0. This resulted in a field of object shapes. A new grid was created from the original data in which the original data values were restored where the mask field was 1, and set to 0 where the mask field was 0. Shape attributes (e.g. centroid and axes) of each of the objects were calculated, and shapes (e.g. band aids and ellipses) were fit to the objects. The closer the forecast object attributes matched the observed object attributes, the better the forecast was said to be. Forecast quality could be summarized by statistics of object attribute differences.
Results	This paper did not show examples of model verification with observed data. Examples are shown in Part II and discussed in Table A.6.
Operational Capability	This article described the basic concept of this technique, but did not show any test results. It appears to be under development and still in the proof-of-concept stage.

Table A.6. Detailed summary of Chapman et al. (2004), discussed in Section 2.1.5.	
Reference	Chapman, M., R. Bullock, B. G. Brown, C. A Davis, K. W. Manning, R. Morss, and A. Takacs, 2004: An object oriented approach to the verification of quantitative precipitation forecasts: Part II – Examples. *Preprint J12.5, Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 9 pp.
Weather Element	Precipitation
Model	22 km WRF, 4 km WRF
Data	CONUS 22 km WRF precip, Stage IV analysis data smoothed to match 22 km model, BAMEX 4 km WRF and unsmoothed Stage IV analysis
Time Period	Summer 2001, 20 May – 6 July 2003 (Bow echo And MCV EXperiment - BAMEX)
Name of Technique	Convolved Object Matching (AMU–selected name)
Description	This was a two-part test of the technique described in Table A.5 using model and observational data. In the first part, verification results from three cases in the 22 km WRF and smoothed Stage IV analysis data archive were shown. In the second part, two cases from BAMEX were shown to illustrate the utility of looking at smaller scale objects to help match larger scale objects. In part one, the authors conducted tests to determine a threshold number of grid points between two or more objects below which the objects were merged into one and above which the objects were deemed separate. This required a human subjective analysis, but the results indicated to the authors that an automated matching technique would be useful in some cases. In one case they could not split a large forecast object over the CONUS as they were able to do with the observed objects. Another case showed that adjusting the grid point number threshold for convolution and/or object merging/separation would cause one precipitation area to be properly analyzed while causing another area to be split or merged improperly. In part two, the objects from the BAMEX data were created using thresholds that would analyze them on a larger scale. Then, the thresholds were changed to resolve smaller scale objects within the large scale objects. By adding the smaller scale features it was possible to consider two areas related if their small and large scale features were similar. They did not attempt to match objects and compare their attributes, but showed two useful displays comparing the model objects to the observed objects. One showed objects of forecast precipitation overlaid with the area of the objects for which there was corresponding observed precipitation. The other was just the opposite showing observed objects overlaid with the area of the object that was forecast.
Results	This study confirmed the ability of the technique to separate objects and match forecast objects to observed objects. More sophisticated merging and object matching routines are needed in the complex cases described in the first part of the article. Higher resolution data can be used to resolve smaller scale objects in larger synoptic scale objects, but comparisons must be done on the larger scale before analyzing the smaller scale. Work will continue to develop more sophisticated object identification and matching techniques
Operational Capability	This shows that the technique is beyond the proof-of-concept stage, but still needs further refining before it can be used in operations with confidence.

Table A.7. Detailed summary of Ebert and McBride (2000), discussed in Section 2.1.6.	
Reference	Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrology*, **239**, 179-202. More information on this technique can also be found at: http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/CRA/CRA_verification.html
Weather Element	Precipitation
Model	Australian Bureau of Meteorology Limited Area Prediction System (LAPS), 0.75° X 0.75° lat/lon resolution
Data	24-hour accumulated precipitation from 23 UTC LAPS, Australian rain gage network daily rainfall analysis
Time Period	July 1995 – June 1999
Name of Technique	Contiguous Rain Area (CRA) technique
Description	This technique was based on that of Hoffman et al. (1995; Section 2.3.3, Table C.3). A CRA was defined as the union of the observed and forecast precipitation areas that exceeded a user-specified rainfall amount. The authors isolated individual precipitation systems over a smaller area rather than several systems over a larger area. In keeping with the ideas of Hoffman et al. (1995), the technique was tested with forecasts from a larger (regional) scale model. In this technique, the forecast area was shifted incrementally by grid points over the observed area until the total MSE in the verification domain was minimized. The verification domain was defined as the union of the original forecast area, the observed area, and the new shifted forecast area. The errors due to displacement, rain volume, and rain pattern were then calculated. Problems arose when observed rain occurred across a boundary such as a coastline, between observation-rich and observation-deprived areas, or at model grid boundaries. The authors conducted tests to determine how many grid points were needed in a truncated observed or modeled precipitation area in order to obtain realistic verification results. For their parameter settings, they found at least 20 grid points were needed.
Results	The user determines the size of the search area, or the maximum number of grid points beyond an observed (forecast) CRA to search for a forecast (observed) CRA. In this case, the authors used 5°, or ~700 km. Event verifications required CRAs that contained at least 10 grid points within a 5 mm/day isohyet. Model systematic errors were only determined from CRAs where the observed area contained at least 20 grid points. Different criteria to verify smaller systems would likely yield different results. Out of over 1811 CRAs in the 4-year period, the algorithm was able to match 695. This was determined by the authors as enough to determine systematic errors in the model with confidence.
Operational Capability	This appears to be ready to use in a post-analysis mode to determine overall model performance, although the user must determine appropriate parameter settings through tests. CRA matching could also be done in real time over a smaller grid (see Table A.8). It may also have utility in verifying other phenomena that cover a specified area, e.g. a low pressure center. Consideration must be given to systems that cross data boundaries, as stated in the Description.

Table A.8. Detailed summary of Grams et al. (2004), discussed in Section 2.1.6.	
Reference	Grams, J. S., W. A. Gallus, L. S. Wharton, S. Koch, E. E. Ebert, and A. Loughe, 2004: Use of a modified Ebert-McBride technique to verify IHOP QPF as a function of convective system morphology. *Preprint J13.4, Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 9 pp.
Weather Element	Convective precipitation
Model	12 km Eta, 12 km Penn State/National Center for Atmospheric Research Mesoscale Model version 5 (MM5)
Data	From International H_2O Project (IHOP): 6-hour accumulated precipitation in first 6 hours of each model run, Stage IV 6-hour accumulated precip, 2 km 30-min NEXRAD Information Dissemination Service (NIDS) radar images. Area over the Great Plains.
Time Period	9 May – 26 June 2002
Name of Technique	CRA (Table A.7)
Description	The goal of this study was to determine how well the mesoscale models predicted precipitation systems based on their morphology. They used the CRA method described in the previous table but modified it to account for the higher spatial resolution of the models and the shorter time period over which precipitation accumulated. The modifications included reducing the percentage of grid points allowed to shift outside the domain, and reducing the critical precipitation mass threshold by a factor of 4 to reflect shorter 6-hour accumulation periods rather than 24 hours. The authors further defined different morphology types into linear and non-linear groupings, and were able to create CRAs based on these criteria.
Results	The CRA technique was used with success after the modifications to the technique. The authors tested the modifications thoroughly to ensure they would yield the desired results for this work.
Operational Capability	Again, this technique was used in a post-analysis study, not on real-time model output and observations. This study shows that the technique's parameters can be changed to accommodate any time and space resolution, and that it could be possible to use the technique on phenomena other than convective rain. It still has the same border issues as described in Table A.7.

Appendix B

Table B.1. Detailed summary of Case et al. (2004), discussed in Section 2.2.	
Reference	Case, J. L., J. Manobianco, J. E. Lane, C. D. Immer, and F. J. Merceret, 2004: An objective technique for verifying sea breezes in high-resolution numerical weather prediction models. *Wea. Forecasting*, **19**, 690-705.
Weather Element	Sea breeze, all other boundaries not considered
Model	Regional Atmospheric Modeling System (RAMS)
Data	KSC/CCAFS wind tower network and RAMS grid forecasts
Time Period	July and August 2000
Name of Technique	Contour Error Map (CEM)
Description	The technique interpolated observed data to the model grid, and then used a binary threshold to distinguish between onshore (easterly) and offshore (westerly) flow. A time estimation filter was applied to the sine of the wind direction to determine the timing of the transition from offshore to onshore at each grid point. The transition times could be displayed in a 2-D contoured plot. The east-west transition time gradient was computed in order to determine the progression of the inland-propagating sea breeze boundary (defined as a positive gradient). A negative time gradient in the east to west direction was assumed to be the result of a river breeze or convective outflows and was eliminated from the analysis in a process called image erosion. The verification statistics included a fractional area of the grid with only observed sea breeze transition times (forecast miss), fractional area of grid with only forecast sea breeze transition times (false alarm), and the average and standard deviation of the sea breeze transition time errors at grid points that experienced both observed and forecast sea breeze transition.
Results	The CEM identified the observed or forecast sea breeze occurrence or non-occurrence correctly 93% of the time. The reasons for failures 7% of the time included false identification of the observed/forecast sea breeze due to precipitation outflows, the observed/forecast sea breeze ending prematurely because of precipitation outflows, or the observed/forecast sea breeze barely propagating inland due to strong synoptic-scale westerly flow.
Operational Capability	The CEM technique was designed as post-process, proof-of-concept software. The filter is not designed to run in real-time, but the authors indicate that it could be modified to do so. The technique is tuned to detect sea breeze occurrence in the specific geographical area of KSC/CCAFS using high temporal and spatial resolution data. A subjective evaluation is still needed with this technique due to possible contamination of results from observed or forecast precipitation outflow or other boundaries. Furthermore, the algorithm may not perform well in data-sparse regions.

Appendix C

Table C.1. Detailed summary of Sandgathe and Heiss (2004), discussed in Section 2.3.1.	
Reference	Sandgathe, S. A. and L. Heiss, 2004: MVT – An automated mesoscale verification tool. *Preprint J13.1, Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 4 pp.
Weather Element	No element was tested
Model	University of Washington Short Range Mesoscale Ensemble Forecast (SREF)
Data	Model analyses, SREF output
Time Period	No results were discussed or shown in the paper
Name of Technique	Mesoscale Verification Tool (MVT) and Mesoscale Data Manipulator (MDP)
Description	The authors stated that a verification tool should have the following attributes: • Automated and flexible, adaptable to issues concerning forecast parameters/timing/intensity, • Capable of evaluating distortion/timing/amplitude errors, • Able to address large numbers of cases and multiple models rapidly, and • Should be statistically sound and present results that are easy to interpret. The MVT separated the forecast error into amplitude and timing components using the procedure defined by Hoffman (1995, Table C.3). This method used a search technique to find a forecast object on a grid that was similar to an observed object. The authors found that the full grid point-by-grid point search was too slow to be operationally feasible, so they employed accelerated search techniques based on image matching algorithms used by the motion picture industry. The main focus of the paper was a web-based GUI that allows the user to see graphical representations of the model output, and the MDP to be used for selection of dates, initialization times, model domain, verifying field and forecast hour, and other items.
Results	The MVT does not verify model forecasts of any specific phenomena at this time, but research is ongoing. Future plans include more development and testing of the verification method in the MVT. There were no results on the performance of the MVT or MDP since they are not yet fully developed. However, the GUI capabilities are impressive.
Operational Capability	An email conversation with the author indicated that the verification techniques are still being developed and the system is not ready for use. In the conclusion section of the paper that authors state: "The MVT has been tested extensively, but the MDP needs more analysis. MVT has not been proven to be a desired replacement of labor-intensive subjective analysis of model phenomenological verification."

Table C.2.	**Detailed summary of Nachamkin (2004) and Nachamkin et al. (2004), discussed in Section 2.3.2.**
Reference	1) Nachamkin, J. E., 2004: Mesoscale verification using meteorological composites. *Mon. Wea. Rev.*, **132**, 941-955. 2) Nachamkin, J. E., S. Chen, and J. M. Schmidt, 2004: Composite-based verification of precipitation forecasts from a mesoscale model. *Preprint J13.5, Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 5 pp.
Weather Element	1) Mistral events–strong northerly winds that occur over the northern Mediterranean Sea 2) Very heavy precipitation events
Model	Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS)
Data	1) Hourly model forecasts, 0 – 72 hours, from model runs initialized at 0000 and 1200 UTC, Special Sensor Microwave Imager (SSM/I) retrieved winds 2) 24-hour accumulated precipitation: model data valid at 24 and 48 hours initialized at 1200 UTC, NCEP River Forecast Center analyses valid at 1200 UTC
Time Period	1) November 2000 – October 2001 2) 15 April – 7 September 2003
Name of Technique	Composite Verification
Description	1) All contiguous points with wind speeds > 12 m s^{-1} and directions between 270°and 70° were defined as mistrals in the forecasts and observations. The search for such points was limited to a specific region of the Mediterranean where mistrals are known to occur. 2) A heavy rain event was defined as any contiguous area of precipitation over the CONUS containing 25 mm or more of precipitation in a 24-hour period. The model and observed data sets were filtered to include only these heavy rain events. The forecast and observed events were each placed on separate relative grids, at a resolution equal to that of the model, with the center of mass of the events placed at the center grid point; i.e. the forecast relative grid was conditional on the existence of forecast events and the observation relative grid was conditional on the existence of observed events. The observational (forecast) data were then positioned on the forecast (observation) relative grid in the position relative to the forecast (observed) event on the regular grid. The average magnitude of the observed and forecast events (wind speed or precipitation amount) were averaged and the number of events in the observations and forecasts were counted.
Results	The average values were displayed on the relative grids with the forecast values overlaid by the observed values. These maps show differences in the magnitude, size, shape, and location of the forecast and observed events. In general, they answer two questions: what was observed if an event was forecast and what was forecast if an event was observed. Charts of event frequencies also show whether the model over- or under-predicted the number of events.
Operational Capability	This method is based on a fairly simple concept. The statistics are easy to calculate and it can be applied to a variety of phenomena. This method was used on an archive of events to get the overall results of how well the model predicted the occurrence, amounts, and locations of the events. This would be a good technique for climatological verification of specific phenomena, and may be modified for use on individual events.

Table C.3. Detailed summary of Hoffman et al. (1995), discussed in Section 2.3.3.	
Reference	Hoffman, R. N., Z. Liu, J-F Louis, and C. Grassotti, 1995: Distortion Representation of Forecast Errors, *Mon. Wea. Rev.*, **123**, 2758-2770.
Weather Element	Large scale features of precipitable water, 10 m wind field, 500 mb geopotential heights
Model	Operational European Centre for Medium-Range Weather Forecasts (ECMWF), U.S. Air Force (USAF) Global Weather Central global spectral model
Data	SSM/I precipitable water, ERS-1 scatterometer ocean surface wind speed, USAF operational high resolution analysis system
Time Period	29 Dec 1991, 5 March 1992, 4 January 1989
Name of Technique	Object Distortion
Description	An object in this study was considered to be on the synoptic scale, not the mesoscale. The distortion of an object was made up of a spatial displacement error and an amplitude error. Any error not accounted for by the distortion was deemed the residual error. The displacement was defined to be a smooth transformation of a field without modification to the amplitude of the data. The transformation could include translation, stretching, and rotating of the object through movement of individual grid points. Limits were imposed on how far from the original grid point the displaced data could be moved. The values of the data were then multiplied by an amplification factor (positively or negatively) to fit the observed field as closely as possible. The displacement and amplification took place until the RMS error was minimized. The authors discussed another method in which the entire field was displaced and amplified as one object. The entire forecast object was displaced in units of 1° latitude and longitude to match as closely as possible the location of the observed object. Then all magnitudes were multiplied by an amplitude factor that minimized the error between all forecast and observed data point values. As in the previous method, the displacement and amplitude factors were chosen such that the total RMS error in the analysis area was minimized. This was the pre-cursor to the Ebert-McBride CRA technique (Table A.7, Table A.8).
Results	The displacement, amplitude, and residual errors were displayed on a 2-D contour plot once calculated, showing where the largest errors existed in the forecast. In the individual grid point method, vectors showed the direction and magnitude of movement when the individual grid point values were displaced. There was no consideration of temporal displacement.
Operational Capability	This technique was developed as a prototype. It is a relatively simple idea implemented with complex mathematics that could be used in verifying model forecasts of several types of large-scale weather phenomena. It may prove difficult in verifying convective precipitation in east-central Florida given the usual situation of multiple cells forming through boundary interactions. This may also be useful as a post-analysis verification tool.

References

Baldwin, M. E., J. S. Kain, and S. Lakshmivarahan, 2005: Development of an automated classification procedure for rainfall systems. *Mon. Wea. Rev.*, **133**, 844-862.

Bullock, R., B. G. Brown, C. A Davis, M. Chapman, K. W. Manning, and R. Morss, 2004: An object-oriented approach to the verification of quantitative precipitation forecasts: Part I – Methodology. *Preprint J12.4, Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 6 pp.

Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Met. App.,* 11, 141-154.

Case, J. L., J. Manobianco, J. E. Lane, C. D. Immer, and F. J. Merceret, 2004: An objective technique for verifying sea breezes in high-resolution numerical weather prediction models. *Wea. Forecasting*, **19**, 690-705.

Chapman, M., R. Bullock, B. G. Brown, C. A Davis, K. W. Manning, R. Morss, and A. Takacs, 2004: An object oriented approach to the verification of quantitative precipitation forecasts: Part II – Examples. *Preprint J12.5, Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 9 pp.

Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrology*, **239**, 179-202.

Grams, J. S., W. A. Gallus, L. S. Wharton, S. Koch, E. E. Ebert, and A. Loughe, 2004: Use of a modified Ebert-McBride technique to verify IHOP QPF as a function of convective system morphology. *Preprint J13.4, Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 9 pp.

Hoffman, R. N., Z. Liu, J-F Louis, and C. Grassotti, 1995: Distortion Representation of Forecast Errors, *Mon. Wea. Rev.,* **123**, 2758-2770.

Marshall, S. F., Pl. J. Sousounis, and T. A Hutchinson, 2004: Verifying mesoscale model precipitation forecasts using an acuity-fidelity approach. *Preprint J13.3, Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 8 pp.

Nachamkin, J. E., 2004: Mesoscale Verification using meteorological composites. *Mon. Wea. Rev.*, **132**, 941-955.

Nachamkin, J. E., S. Chen, and J. M. Schmidt, 2004: Composite-based verification of precipitation forecasts from a mesoscale model. *Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 5 pp.

Sandgathe, S. A. and L. Heiss, 2004: MVT – An automated mesoscale verification tool. *Joint Session of 20th Conf. on Wea. Anal. and Forecasting / 16th Conf. on Numerical Wea. Pred. / 17th Conf. on Probability and Statistics in the Atmos. Sci.*, Amer. Meteor. Soc., 11 – 15 January, Seattle, WA, 4 pp.

Zepeda-Arce, J, E. Foufoula-Georgiou, and K. K. Droegemeier, 2000: Space-time rainfall organization and its roll in validating quantitative precipitation forecasts. *J. Geophys. Res.*, **105, No. D8,** 10 129-10 146.

List of Acronyms

45 WS	45th Weather Squadron
AMU	Applied Meteorology Unit
ARPS	Advanced Regional Prediction System
AWIPS	Advanced Weather Interactive Processing System
BAMEX	Bow echo And Mcv EXperiment
CCAFS	Cape Canaveral Air Force Station
CEM	Contour Error Map
CONUS	Continental U.S.
CRA	Contiguous Rain Area
GFS	Global Forecast System
GUI	Graphical User Interface
IHOP	International H_2O Project
KSC	Kennedy Space Center
LAPS	Limited Area Precipitation System
MDP	Mesoscale Data Manipulator
MSE	Mean Square Error
MVT	Mesoscale Verification Tool
NAM	North American Mesoscale
NCEP	National Centers for Environmental Prediction
NWP	Numerical Weather Prediction
NWS MLB	National Weather Service in Melbourne, FL
RAMS	Regional Atmospheric Mesoscale System
RMS	Root Mean Square
RUC	Rapid Update Cycle
SMG	Spaceflight Meteorology Group
SREF	Short Range mesoscale Ensemble Forecast
SS	Skill Score
SSM/I	Special Sensor Microwave Imager
TS	Threat Score
USAF	U.S. Air Force
WRF	Weather Research and Forecasting

NOTICE

Mention of a copyrighted, trademarked or proprietary product, service, or document does not constitute endorsement thereof by the author, ENSCO, Inc., the AMU, the National Aeronautics and Space Administration, or the United States Government. Any such mention is solely to inform the reader of the resources used to conduct the work reported herein.