16B.1

Verifying Vectors

Randy Bullock* and Tressa L. Fowler

National Center for Atmospheric Research, Boulder, Colorado

1 INTRODUCTION

Predictions of vector fields are more complex to verify than predictions of scalar fields. Verification of vector fields, like wind, typically involves transforming the vector into two scalar values, then applying traditional verification methods to each separately. This paper explores some alternate verification strategies and compares them with more traditional approaches. Appropriate circular statistics are calculated for forecast and observed wind vectors. Vectors are transformed into the scalar measures, divergence and curl, then verified using object-based methods. In vector calculus, divergence fields often give information about regions of inflow or outflow, i.e. sinks and sources, indicating that object-based verification methods might be appropriate. Curl is a measure of rotation, usually referred to by meteorologists as vorticity. The different strategies for verifying wind fields are compared.

Section 2 of this report describes the case studies and domains used for these analyses. In section 3, the translation and evaluation methodologies are discussed. Results for all of the methods are presented in section 4. Finally, section 5 contains the conclusions and a considerable amount of future work.

2 DATA

Surface winds from seven case studies are included in this research. They were chosen purely for convenience, as the authors had the files from a previous study. Fortunately, these seven cases contain many types of wind features and represent different seasons.

2.1 WRF wind forecasts and analyses

Surface winds from the weather research and forecasting (WRF; Skamarock *et al*, 2008) model for seven days in 2005 and 2006 are used as the example forecast data. This model has a 13 km resolution. All forecasts have a 36 hour lead time. The “observations” consist of the WRF analysis (i.e. 0 hour lead time) matching the forecast valid time. The surface winds are specified via two horizontal vectors, *u* and *v*, in meters per second. This paper examines the surface wind field, with upper air winds left for future work. The seven cases are listed in Table 1 with a brief description of the surface wind field from the model analysis.

Table 1: Valid times of WRF forecasts and analyses with a brief description of the surface wind conditions.

Valid Time	Summary of analysis winds
20050712 12Z	Large vortex over southern Illinois
20050817 12Z	Strong winds from E Washington to W Montana and from E Texas to North Dakota
20051021 12Z	Weak vortex over S Illinois, strong winds northwest of front extending from Wisconsin to Kansas.
20051108 12Z	Strong winds on the east coast, over Rocky Mtns, and south of line extending from SE Kansas to Lake Erie.
20060111 12Z	Extremely high winds stretching from the Pacific Northwest into Wyoming.
20060216 12Z	Check mark shaped convergence line from Lake Erie to the Texas panhandle up through the Colorado front range and very high winds over New Mexico.
20060508 12Z	Strong winds over the Rocky Mtns.

2.2 Domains

Analyses are conducted over two domains. The CONUS is used for most analyses. For circular statistics, a smaller domain yields more useful information. The northeastern corner of Colorado (NECO), bounded by 39° and 41° N latitude and 102° and 105° W longitude is used. This area is home to many wind farms, so may be of interest in future wind verification analyses.

3 METHODS

Recent years have seen an explosion in the usability and diagnostic capability of verification methods. However, these have not generally been used to verify fields like wind. The format of the model wind fields does not lend itself well to object-based methods and can make interpretation of traditional statistics unintuitive. Traditional verification methods are included in this study for purposes of completeness and comparison. Transformed forecast and observation fields are used with an existing object based verification method. The utility of circular statistics for forecast verification of winds is examined.

Corresponding author address: Randy Bullock, P.O. Box 3000 Boulder, CO 80307. bullock@ucar.edu

3.1 Traditional methods

For purposes of comparison, some traditional methods of wind verification are included. The error of both wind components *u* and *v*, are examined. The error in both wind speed and direction are accumulated over the entire area and over time.

Calculation of these quantities is done using the Model Evaluation Tools (MET) software (Brown *et al.*, 2009).

3.2 Vector methods

Scalar and vector fields are related as shown in Figure 1. Vectors can be transformed into scalars via a derivative known as divergence. Vectors can be translated to a different vector field via calculation of curl (Baxandall and Liebeck, 2008). Gradient and Laplacian transformations are shown in the figure for reference, though they are not used in this paper.

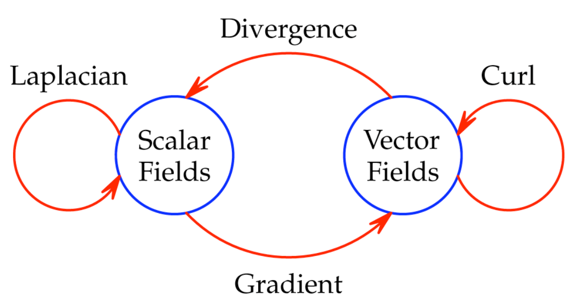

Figure 1: Diagram showing the relationship between scalar fields, vector fields, and their derivatives.

The scalar divergence field can be calculated from any vector field via the following equation:

$$\operatorname{div} \mathbf{F} = \nabla \cdot \mathbf{F} = \frac{\partial \mathbf{F}_u}{\partial u} + \frac{\partial \mathbf{F}_v}{\partial v}$$

Figure 2 shows a zoomed in example of divergence. The wind field is plotted as vectors and the divergence field is represented by colors. Warm colors indicate high absolute values of divergence.

Divergence is a measure of outflow or (if negative) inflow into a region. For a nonzero divergence to exist at some point it is only necessary for there to be a *net* inflow or outflow.

Divergence is a measure of outflow or (if negative) inflow into a region. For a nonzero divergence to exist at some point it is only necessary for there to be a *net* inflow or outflow.

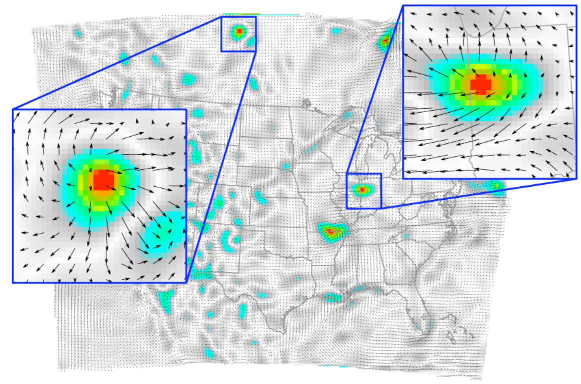

Figure 2: Example of areas of high divergence with wind vectors.

Curl is a vector field derived via partial derivatives of another vector field. It gives the net rotation of the field. For two-dimensional fields, the curl is *one-dimensional* and can thus be treated as a scalar. In two-dimensional Cartesian coordinates, the curl is defined as follows

$$\operatorname{curl} \mathbf{F} = \nabla \times \mathbf{F} = \frac{\partial \mathbf{F}_u}{\partial v} - \frac{\partial \mathbf{F}_v}{\partial u}$$

where *x* and *y* represent the two dimensional coordinates. Meteorologists generally refer to the curl as vorticity.

Probably the classic example of a vector field with curl is one with a spiral vortex pattern. This is not the only way that curl can be present, however. Even wind that flows only in one direction (e.g. easterly) can give rise to a nonzero curl if the speed of flow varies across the flow lines. An easterly wind whose speed increased with latitude over some region would be an example.

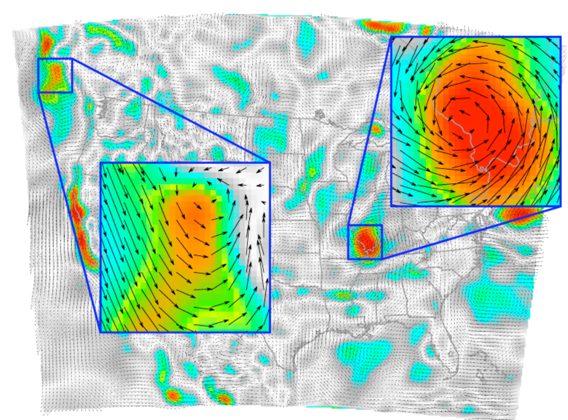

Figure 3: Example showing areas with high values of curl with wind vectors.

It should be noted that in the atmosphere, divergence and curl are often present together in the wind field. For example, a region of low pressure will pull surrounding air into the region, generating a (negative) divergence, but the Coriolis force acting on the inflow will distort the flow pattern into a cyclonic vortex, giving rise to curl.

With data, the divergence and curl must be estimated numerically. For this paper, a straightforward method is used, with more sophisticated estimation left for future work. For each grid location, a 2nd order polynomial is fit via least-squares in two dimensions from the surrounding grid squares. Fifteen grid squares are used in each direction. Since the model has 13km resolution, the divergence and curl is estimated from a nearly 200 km square centered on the location of interest.

The curl and divergence fields have units of meters per second per kilometer, so they represent the change in wind speed over distance. Curl and divergence can be positive or negative. For curl the sign represents the direction of the rotation, while for divergence a negative value actually indicates convergence. For each field, areas with high absolute values are of primary interest. These areas appear to coincide with vortices, fronts, coastal boundaries and other wind features, as shown in Figures 4, 5, and 6.

Figure 4 shows the actual model wind field for 20050712. Note the vortex centered on southern Missouri, the strong winds along the California coast, and the weak boundary stretching from eastern Wisconsin to the northeast corner of Colorado. Figure 5 shows the curl for this same case, with warm colors representing higher absolute values. Note the feature along the California coast and the vortex over southern Missouri.

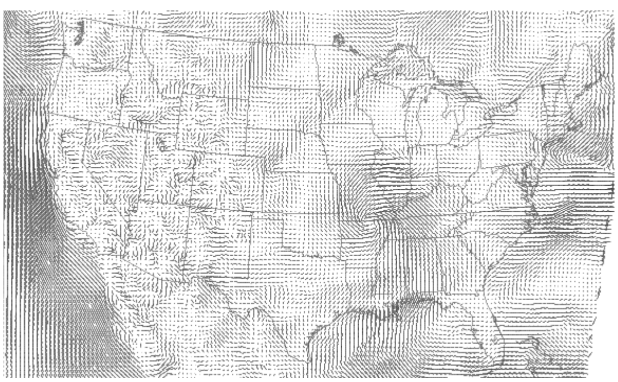

Figure 4: Model surface wind analysis from 20050712 12Z.

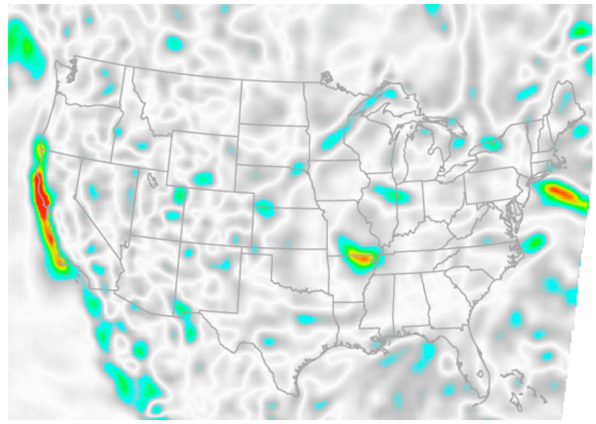

Figure 5: Curl of model surface wind field for 20050712 12Z forecast with a 36 hour lead time.

Figure 6 shows the divergence field for an example model surface wind field from a different day. In this case, note the check mark shaped feature extending over the Colorado front range, through the Texas panhandle out to Lake Erie. The case in Figure 6 also has very high winds over New Mexico. Only the edges of this feature show up as areas of high divergence.

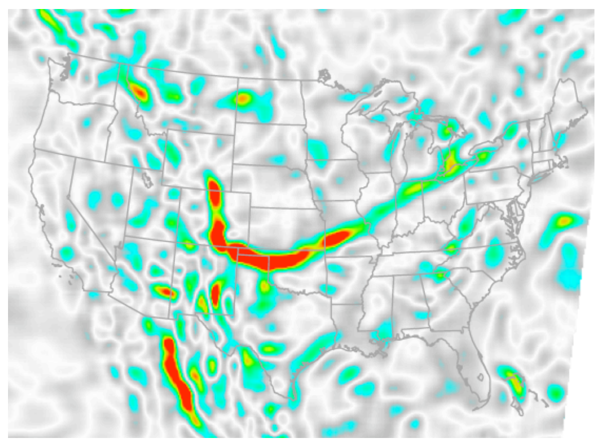

Figure 6: Divergence of model surface wind analysis for 20060216 12Z.

Once scalar fields are derived from the vector winds, they are verified using an object-based verification software package known as method for object-based diagnostic evaluation (MODE, Davis *et al* 2006). The curl and divergence fields appear to lend themselves well to object based analysis as they are somewhat spatially coherent. Also, many wind features have both divergence and curl.

3.3 *Circular statistics*

Wind is represented in many models by its two horizontal components, denoted *u* and *v*. These can be transformed into wind speed and direction, though wind direction is indeterminate when wind speed is very low.

Wind direction, like other orientation data, is not ordered. Thus, the statistics appropriate for wind direction are different in both calculation and interpretation than statistics for scalar values. Many of the concepts remain the same, however. Orientation data can be characterized by a typical value (e.g. mean, median, mode) and spread (e.g. dispersion, variance). Orientations can be uniform, unimodal, or multi-modal and can follow any of several theoretical circular distributions, such as von Mises or wrapped Cauchy (Borradaille, 2003).

The mean circular orientation is generally used to characterize directional data. However, this mean angle is only meaningful for unimodal data. When the orientations are uniform or nearly opposing bi-modal, the calculated mean angle will be zero even though in both cases the mean angle does not exist. A simple example is a set of cases with half of the winds from the east and the other half from the west. The calculated mean orientation of this data would be zero. However, the mean angle is not really north, it does not exist.

The seven cases included in this study do not yield unimodal wind directions when examined over the CONUS. Thus, for this analysis, the wind vectors over northeastern Colorado (NECO) are analyzed. The wind comes from different directions for the seven cases, so even these are not combined. Rather, each case is considered separately. They each appear to be unimodal or nearly so with respect to wind direction.

The mean length and orientation of a set of vectors is found by connecting all the vectors end to end. The vector connecting the beginning to the end is the resultant vector. The angle of this vector is the mean orientation ($\overline{\theta}$). When the length of this vector is divided by the number of vectors used in its calculation (e.g. n), the result is the mean length ($\overline{R}$). Formulas for each are shown below. These formulas represent standard vector addition. The only accommodation required for application to wind vectors is ensuring that the directions are unimodal so the vectors cannot cancel one another out.

$$\overline{\theta} = \mathtt{atan2}\left(\sum_i x_i \sin\theta_i, \sum_i x_i \cos\theta_i\right)$$

$$\overline{R} = \frac{1}{n}\left[\left(\sum_i x_i \cos\theta_i\right)^2 + \left(\sum_i x_i \sin\theta_i\right)^2\right]^{1/2}$$

For interpretability, wind directions are expressed in degrees ranging from -180 to 180. Geographic convention is used rather than geometric, so 0 is “grid” north.

In this research, the mean resultant length and angle are calculated separately for forecasts and observations. These can be compared to each other to see if the typical length and angle of the forecast matches the observations. It is also mathematically possible to calculate these statistics on the difference vector, i.e. the forecast vector minus the observed vector. However, the interpretation of the length and angle of the mean difference vector is unintuitive, so is left for future research.

4 RESULTS

Results are presented for traditional statistics, object based analyses from MODE, and circular statistics. For clarity, each case is considered separately, though traditional and object based methods also lend themselves to cumulative analyses.

4.1 *Traditional statistics*

Table 2 shows the bias and root mean squared error (RMSE) statistics for the u and v components of the surface winds (Wilks, 2006). Table 3 contains the same statistics over the northeastern Colorado (NECO) domain. Over the CONUS, the model tends to underestimate both components of the wind with errors of about 2 to 3 ms^{-1} for both components.

Table 2: Statistics for u and v components of surface winds over CONUS.

	U		V	
	Bias	RMSE	Bias	RMSE
20050712	-0.44	2.6	-0.76	2.5
20050817	-0.51	2.3	-0.43	2.4
20051021	-0.50	2.1	0.48	1.8
20051108	-0.27	2.7	-0.48	2.1
20060111	-1.23	3.3	-0.97	2.9
20060216	-0.31	2.8	-0.14	2.1
20060508	-0.83	2.7	-0.20	2.4
Overall	-0.58	2.7	-0.36	2.6

This research is not really concerned with the quality of the seven forecasts, but in comparisons of different assessments of that quality. Since the different assessments are not directly comparable, it is good to have some idea of which cases are “better” or “worse” according to each assessment to see if they agree. Thus, it is useful to notice that these traditional statistics indicate that over the CONUS, the 20051021 case is pretty good compared with the others (smaller errors, similar bias), while the 20060111 case is somewhat worse (larger errors and bias).

Table 3: Statistics for u and v components of surface winds over NECO.

	U		V	
	Bias	RMSE	Bias	RMSE
20050712	0.07	0.9	-1.78	3.1
20050817	-1.64	2.2	2.68	3.0
20051021	4.13	4.9	-0.06	3.4
20051108	-2.74	2.9	1.45	2.0
20060111	-3.51	3.7	0.38	1.1
20060216	3.44	3.9	-3.39	3.6
20060508	-1.26	1.7	0.87	2.3

Over NECO, the 20051021 and 20060216 cases have the highest RMSE in both components. The 20060216 case also has large bias in both components. The 20050712 and 20060508 cases have the lowest values of RMSE and lowest bias of the u component.

4.2 Object based verification

Divergence and curl fields estimated from wind vectors capture many wind features, such as vortices and fronts. The wind speed field naturally identifies high wind events very well.

For the July 2005 case (20050712 12Z), the divergence, curl, and wind speed fields all indicate an object at the Illinois vortex. The curl field also shows a linear object extending over most of the California coast. The curl and divergence forecast objects are matched by observed objects, but intensity of the objects is too low. In the case of the vortex, the forecast is too small and displaced to the southwest. The wind speed forecast field is too low to have any objects, though the observation objects include the vortex and a small object near the California coast. Thus, for this case, the forecast is able to identify changes in the wind field (i.e. divergence and curl), though it underestimates both their intensity and size. Examination of the wind speed field alone would probably not provide this information. This case illustrates the value of the curl field for detecting features in the wind that are missed by analysis of other fields.

The 20050817 case was characterized primarily by strong winds. The divergence and curl fields picked up some of the edges of the strong winds, but the resulting objects are non-intuitive as they do not contain the center of the strong wind event. The wind speed analysis field picked up the strong wind events. As with the other cases, the forecast winds were much to low.

The 20051021 case has better traditional u and v statistics than the other cases. The most notable features of this case are a front and a weak vortex. The forecast and observed curl fields show these features along with an object along the California coast. The divergence shows the vortex, while the wind speed field picks out the front. The wind speed field also shows an arc shaped area of higher winds north of the vortex, so it does not completely miss this feature.

The high winds characterizing the 20051108 case are well represented in the wind speed objects. The curl and divergence objects for this case seem to detect some edges of these features. The curl represents the east coast winds very well, but the other curl objects and the divergence objects do not clearly represent the features seen in the wind vectors.

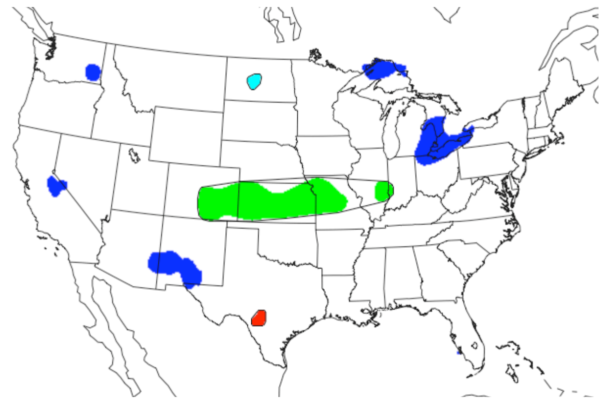

Figure 7: Forecast objects from divergence field for "check mark" case, 20060216.

The “check mark” case from February 2006 is shown in the Figure 4 divergence plot. In all three fields (divergence, curl and wind speed) there are fewer and smaller objects in the forecast field than in the observed field. The forecast divergence field caught some, but not all, of the "check mark", as shown in Figure 7. MODE was able to identify the check mark in the divergence analysis field, as shown in Figure 8.

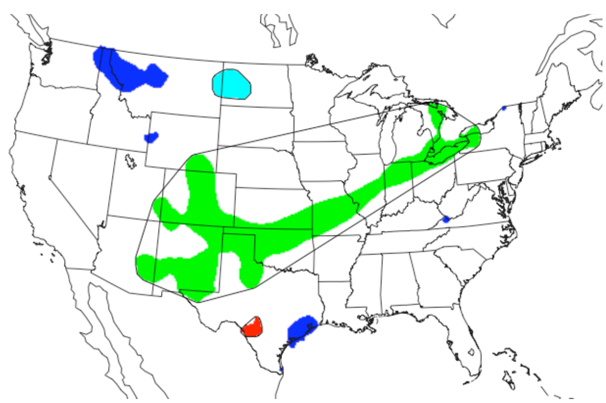

Figure 8: Analysis objects from divergence field for check mark case, 20060216.

The forecast curl field did better, but broke the check mark up into several pieces. There is nothing in the wind speed object field corresponding to the check mark, however the wind speed map shows the check mark in something like a negative sense, since the speed falls to very small values along much of the check mark. There is an object present in the observed wind speed along the Gulf of Mexico that is missed in all three forecast fields. There are objects in the curl fields (forecast and observed) that are missing from the divergence and wind speed fields.

High winds are the primary features of the 20060111 case, which was “worst” in terms of the traditional statistics. Object detection on the wind speed analysis shows these features. Curl has an object over most of Wyoming, though this area only represents a third of the high wind area. The divergence field is very speckled looking with lots of smaller objects near the high winds. It is probably detecting some edges. The forecast vastly underestimated the size and strength of the high wind areas, though it shows some areas with higher winds in the correct places.

For the 20060508 case, there are many objects present in the observed divergence field that are absent from the forecast field. The same holds true for curl. There is a large observed wind speed object over Montana and Wyoming that is not present in the forecast field. There are curl objects along the California coast that are not detected by the wind speed and divergence fields. There is an object on the North Carolina/Virginia coast that is present to some extent in all three fields.

Overall, the wind features were severely underestimated by the forecasts. This assessment was not evident from the bias statistics on the u and v winds. The biases were nearly all negative, but not largely so. Perhaps this is due to smaller or positive forecast bias outside of the features. Most of the wind features were identified by at least one of the curl, divergence, or wind speed fields.

The underestimation of wind speed, divergence and curl in these cases made object based analysis more difficult. With adjustment in the parameters that define and match the objects, this can likely be overcome. That effort is left for future work.

4.3 Circular Statistics

For each of the seven cases over the NECO domain, the mean orientation and mean resultant length are calculated separately for forecast and observation vectors. Results are presented in Table 4.

Table 4: Vector statistics for forecasts and observations for case studies in the NECO domain.

Wind vectors	Mean orientation		Mean resultant length	
	Forecast	Obs	Forecast	Obs
20050712	6	93	1.8	0.2
20050817	-144	-69	2.0	3.0
20051021	-22	42	3.4	4.2
20051108	-134	-98	3.2	5.1
20060111	-115	-100	3.6	6.9
20060216	-8	29	7.7	4.8
20060508	-108	-92	3.4	4.5

The results from the vector statistics tell a different story than the results from the traditional statistics. The two smallest and largest differences are highlighted in Table 4. No cases have both the smallest (largest) orientation differences and the smallest (largest) length differences. The 2050712 case shows a large orientation error. This was the “good” case in terms of bias and RMSE. The “good” bias and RMSE case, 20060508, has a very small average orientation error (16 degrees). The 20051021 case has very close mean forecast and observation vector lengths, with about a 65 degree difference in the orientations. This case is one of two worst cases in terms of RMSE and Bias, though not in terms of differences in mean orientation and length. The 20060216 case is the other worst judging by traditional statistics. The circular statistics confirm this with the large errors in length.

5 CONCLUSIONS AND FUTURE WORK

Together, divergence, curl and wind speed represent many features of a wind field. Certain features lend themselves to identification via one measure over the others while some are detected by all. Object based verification appears to work well on the derived wind fields. Some long linear features are not identified well using the default parameters of MODE verification. The large biases in the selected cases also affect object identification and matching. Adjustment of the MODE settings may alleviate these issues.

The statistics for circular variables may prove useful for verifying forecasts for smaller domains or a single location, such as a wind energy farm instrument site. The information produced by these statistics is somewhat easier to interpret than errors on u and v wind components and the two appear to provide somewhat different information. However, they are uninformative when wind directions are not unimodal, which is often the case when samples cover large spatial or temporal domains. Thus, their use is not recommended for many meteorological applications, particularly for verifying numerical weather prediction models over large domains.

The work included in this paper is very preliminary. Though the results are promising, a considerable amount of further research remains. The methods included here should be tested on a wider variety of forecasts to determine if results are consistent and robust. Extensions and variations of these methods should also be explored.

Translation of scalars to vectors and vice versa can be accomplished by several means, though only some have a physical interpretation. Future work may include translation of traditional scalar fields to vectors and back again or translation of scalars to scalars via the Laplacian transformation. For example, the length of a temperature gradient vector could be derived for forecast and observed grids. Object-based verification of these fields might provide information about how well forecasts identify areas of significant change.

Numerical derivatives can be obtained in many different ways, including higher order polynomials, smaller or larger neighborhoods included in the calculation, and methods of fit other than least squares.

Curl, divergence, and wind speed fields can be verified with any of several other object based methods. Perhaps there is even a way to unify the objects defined by the three fields to create a single wind feature field for verification.

ACKNOWLEDGEMENTS

NCAR is sponsored by the National Science Foundation. This research is in response to requirements and funding by the Air Force Weather Agency (AFWA) and Xcel energy. The views expressed are those of the authors and do not necessarily represent the official policy of AFWA or Xcel.

The authors would like to thank Mike Chapman, Frank McDonough, and Tara Jensen for their advice and interpretations.

REFERENCES

Baxandall, P., and H. Liebeck, 2008: *Vector Calculus*. Dover republication of the edition published by Oxford University Press, New York, 1986. 560pp.

Borradaile, G., 2003: *Statistics Of Earth Science Data: Their Distribution In Time, Space And Orientation*. Springer-Verlag, Berlin And Heidelberg, 280pp.

Brown, B. G., *et al*, 2009: The Model Evaluation Tools (MET): Community Tools For Forecast Evaluation. Preprints, *25th Conference on International Interactive Information and Processing Systems (IIPS),* Phoenix, AZ, Amer. Meteor. Soc., CD-ROM, 9A.6

Davis, C., B. Brown, and R. Bullock, 2006: Object-Based Verification of Precipitation Forecasts. Part II: Application to Convective Rain Systems. Mon. Wea. Rev., 134, 1785–1795.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: A Description of the Advanced Research WRF Version 3. NCAR Technical Note NCAR/TN–475+STR, 113pp.

Wilks, D., 2006: *Statistical Methods In The Atmospheric Sciences*. 2nd Ed. Elsevier, San Diego. 648pp.