9A.5 An Evaluation of a Procrustes Shape Analysis Verification Tool Using Idealized Cases

Steven A. Lack
Neil I. Fox
Anthanasios Christou Micheas
University of Missouri-Columbia, Columbia, Missouri

1. INTRODUCTION

A recently developed verification tool designed by the Department of Statistics at the University of Missouri-Columbia was evaluated utilizing several idealized cases. The verification methodology utilizes a Procrustes fit for shape analysis of individual cells (Dryden and Mardia, 1998). The scheme also includes statistics based on intensity parameters for a complete verification solution. The information on the error based on size, translation, and rotation are combined with error based on intensity values via a penalty function. As the errors are residual sum of squares they are open-ended and this testing procedure allows the assessment of the scale such that (1) different forecast situations can be compared, (2) comparability can be achieved between the different components that make up the total error, and (3) suitable normalization factors can be found that take the previous issues into account to create a robust and practical verification scheme.

The idealized cases are a series of simple geometric objects, such as ellipsoids, and vary in intensity, intensity distribution within the cell, as well as orientation, size, and translation. These idealized cases highlight the usefulness of the new Procrustes verification scheme as it is able to decompose the error contribution of various attributes combined in the penalty function. Results will also be shown for a few real cases for completeness

2. The Procrustes Scheme

The Procrustes verification scheme is an object oriented approach in which shape analysis techniques are employed to identify objects from a forecasting system (Micheas *et al.*, 2007). The scheme is useful in that it utilizes an overall penalty function for forecast skill as well as the ability to break down the error components of the forecast based on size, shape, and intensity for a more robust verification solution.

2.1 *Procrustes Methodology*

The scheme first identifies forecast objects (*i.e.* storm cells of reflectivity) by using a user-defined threshold of size and intensity for the object. Once the object is identified, characteristics of each identified cell are retained such as max and mean intensity along with centroid location. The cells in the forecast field are then matched to the cells in the observed field. Objects can be matched in two ways in the current scheme. First, objects can be matched to minimize the error in their shape parameters given by the Procrustes fit. In other words, those cells with approximately the same shape will be matched. The scheme also uses more of a traditional approach to matching based on minimizing the distance from the forecast to observed object. Work is underway to combine these two matching sequences and possibly add intensity information as well.

Once the matching is accomplished the statistical shape analysis begins. The forecast object is essentially overlaid onto its corresponding observed field and a fit is performed using equation (1).

$$\hat{z}^j = \hat{c}_{jk} + \hat{r}_{jk} e^{i\hat{\varphi}_{jk}} z^{kj} \qquad (1)$$

Equation (1) is known as the full Procrustes fit, the superimposition of z^{kj} onto z^j. Where the first component c is the translation term, r is the dilation term and φ is the rotational component. These terms incorporate the residual sum of squares (RSS) term in the penalty function (2).

$$D = RSS_k + SS_{\text{avg}}^k + SS_{\min}^k + SS_{\max}^k \qquad (2)$$

The other components in the penalty function are the errors based on intensity differences between the forecast and observed object summed for the entire domain. It is easily shown that if the number of objects in the forecast field does not match that in the observed field, the penalty is increased due to the matching of multiple forecast objects to one truth object. The penalty function, in this framework, is only meaningful when multiple forecast products (different models at similar resolutions) are compared to the truth field. The lower the penalty is the better the forecast solution. The penalty function may also be weighted accordingly to the user's specific needs. For example, a hydrologist was using the data may value intensity errors highly and weight the penalty accordingly.

2.2 *Procrustes Output Example*

An example for the shape analysis verification scheme is shown for a sample case from the NCAR Spring 2005 project for an intercomparison of WRF forecast solutions (ICP, 2007). Three versions of the WRF were compared from 13 May 2005 and compared against a Stage II analysis. Figure 1 is an example of one of the three forecast products used; in this case the WRF4NCAR. Figure 2 is the Stage II analysis. For this particular date and example, the NCAR version of the

* *Corresponding author address:* Steven A. Lack, 302 ABNR Building, Univ. of Missouri-Columbia, Dept of Soil, Environmental, and Atmos. Science, Columbia, MO, 65211; email: SALack@mizzou.edu

WRF model outperformed the others (WRF2CAPS and WRF4NCEP). Although the basic skill scores for the date in question favored the NCAR version of the WRF, the skill score was not impressive (CSI ~0.25). The shape analysis scheme for comparison between the models allows some insight into the differences (Tables 1 and 2).

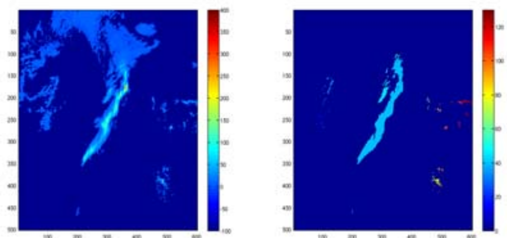

Fig. 1: WRF4NCAR precipitation forecast field (left) in hundredths of an inch. Cell identification as performed by the Procrustes scheme (right) at a 0.10" threshold and a minimum size of 10 km^2.

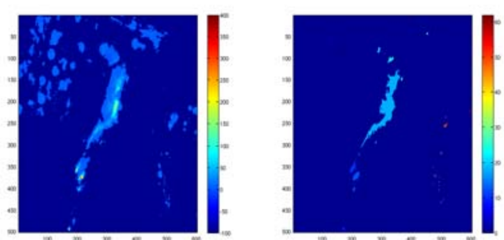

Fig. 2: Stage II precipitation analysis (left) in hundredths of an inch. Cell identification as performed by the Procrustes scheme (right) at a 0.10" threshold and a minimum size of 10 km^2.

	WRF4NCAR	WRF4NCEP	WRF2CAPS
RSS (1*10^3)	5.9151	5.4487	**4.8603**
SStot (1*10^3)	**1.7531**	2.3083	1.9005
# Cells (Truth = 17)	**26**	61	39
Tot Error (1*10^5)	**2.6607**	5.7858	2.8416

Table 1: Shape analysis statistics for the 3 models on 13 May 2005. The bold numbers represent which model is performing the best for the given category. In this situation WRF4NCAR is performing the best in all categories the intensity forecast and the total error, but not the overall shape analysis.

	WRF4NCAR	WRF4NCEP	WRF2CAPS
Min Intensity	**2**	4	4
Mean Intensity	**8926**	26825	9079.9
Max Intensity	**2521220**	5463000	2702100
Dilation	**4.6711**	22.669	11.596
Rotation	0	0	0
Translation	**428290**	7736700	1176000

Table 2: A further breakdown (error per forecast object) of the shape analysis statistics for the 3 models on 13 May 2005. The bold numbers represent which model is performing the best for the given category. In this situation WRF4NCAR is performing the best in all categories.

3. Idealized Cases and Results

Idealized cases were generated to test with the newly developed Procrustes shape analysis scheme. These cases run the gamut of the general types of forecasting flaws. Since this is a relatively new verification scheme it makes sense to test it on idealized cases as well as real cases to get an idea of its usefulness and potential areas for improvement. The idealized cases used in this study range from cases in which there could possibly be a timing error in the model yielding large translation errors to cases that may have no translation error but have significant intensity errors. Most cases used involve simple ellipsoidal objects with some given intensity. There were 14 idealized cases used in this study. The resulting total penalty for all idealized cases is summarized in Table 3.

The first case (case 1) examined dealt with a simplified ellipse with no translation (figure 3). The observed object has a weak normalized intensity about its center, and the forecast object has a higher intensity which is normally distributed within the cell with its peak intensity in the middle. This case yields zero penalty for the dilation, rotation, and translation components with a near zero value for the RSS term of the penalty function, but a large error in terms of max intensity and a slight error in terms of mean intensity. Minimum intensity is negligible in all idealized cases, but could be significant if there are peaks and valleys in the intensity field. This idealized case closely resembles the second case (case 2), in which there is no normalization of intensity over the ellipse. It simply has a higher uniform intensity. This case results in a large overall error due to both the mean and the max intensity producing a high penalty.

Case 3 simply involves two objects of similar reflectivity; however, one object is simply double the size of the original object. Both cells share the same centroid therefore there is no translation error

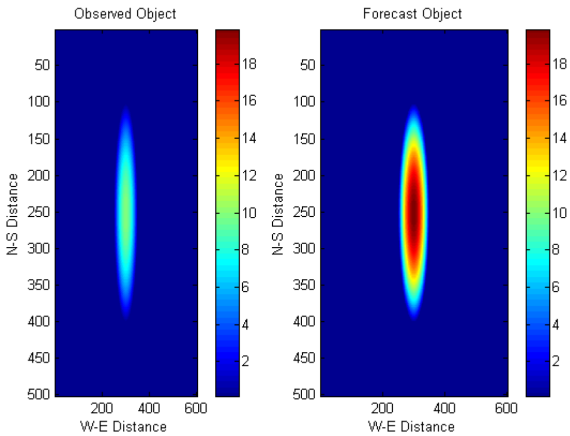

Fig. 3: Idealized case 1 showing the differences in intensity while maintaining similar size and shape.

component. The only error existing is the dilation component from the Procrustes fit.

The fourth idealized case (case 4) is a unique case which deals with the forecast ellipse being rotated about its centroid 90° (figure 4). There is no change in intensity or size or shape of the ellipse. This case ended up yielding the highest error despite having uniform size, shape, and intensity structure. This was because the prior version of the scheme had problems with rotation so that the overall fit was poor giving the high penalty value. This problem has since been corrected so that the values in table 4 reveal more meaningful error characteristics.

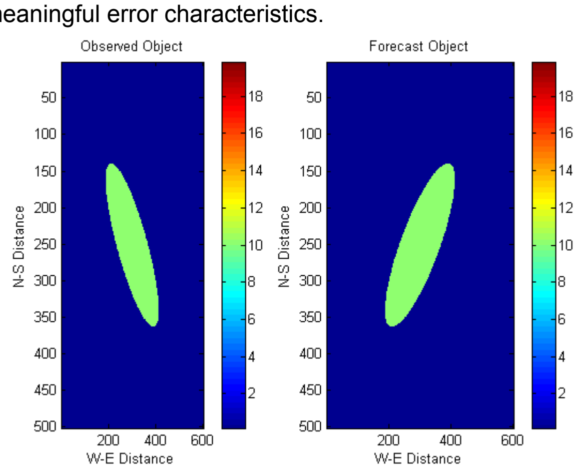

Fig. 4: Idealized case 4 showing the differences in rotating about the centroid 90° with similar size, shape, and intensity parameters.

Case 5 through case 9 represent translation error with a different twist each time. Case 5 represents a small translation with ellipsoids of similar size, shape and intensity. Case 6 simply increases the distance between the ellipsoids. Case 5 and 6 have the lowest combined penalty of all cases as the only thing that is wrong is the location (timing) of the system. Case 7 deals with a translation error with objects of different sizes, while case 8 deals with a translation and rotation component, and case 9 deals with a centroid translation with some overlap in the different sized objects (figure 5). It is no surprise that of all the idealized cases we find these having the highest penalties of all 14 idealized cases.

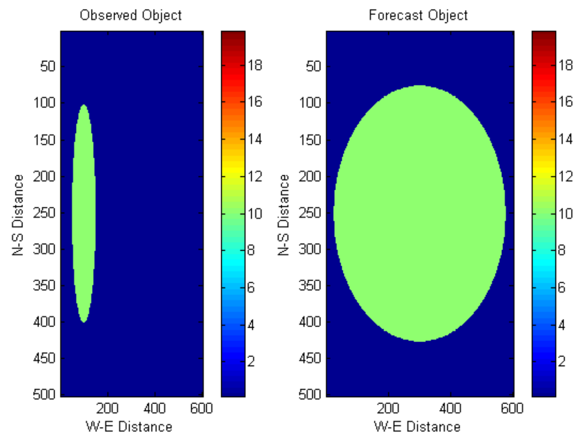

Fig. 5: Idealized case 9 showing the differences in translation from observed object to forecast object while blowing up the forecast object so there is some slight overlap with the observed object. Intensity structure remains the same.

The last of the idealized cases are special cases that do not fit into a generic category, but still test the benefits of using the shape analysis scheme. Case 10 simply concerns objects with the same centroid and intensity structure and even similar size; however, one object is a rectangle and one is an ellipse. Square objects may represent coarse model domains where precipitation realizations do not have smoothed features. Small errors exist in this case. Case 11 utilizes two square objects; however, the forecast object has peaks of reflectivity contained within its outer boundary (figure 6). Case 12 is a smoothed versus a noisy ellipsoid much like case 1 and also yields a moderate-sized penalty. Case 13 uses three objects in the observed field and misplaces though similarly shaped objects in the forecast field (figure 7), which actually leads to one of the smallest penalties in the study based on (2)

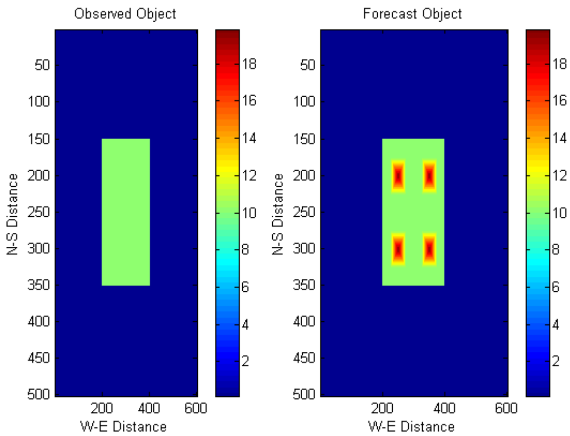

Fig. 6: Idealized case 11 with areas of high intensity peaks in the forecast object rectangle with uniform intensity in the observed rectangle.

Finally case 14 are ellipses that are simply mirrored objects. This type of error could be seen in a real situation in linear convection with either a trailing stratiform region or a leading stratiform region.

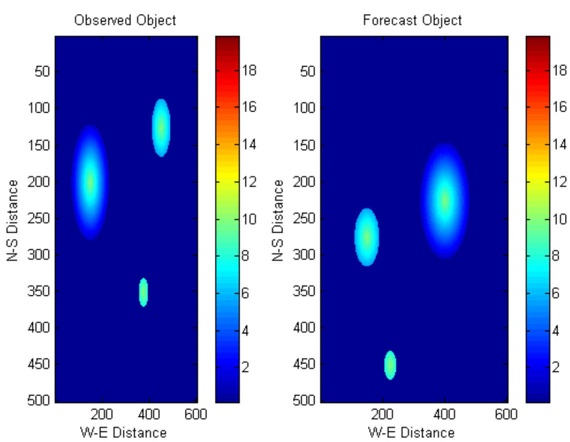

Fig. 7: Idealized case 13 with three observed objects being simply misplaced in the forecast object. There is no change in object size, shape, or intensity.

4. DISCUSSION

Idealized cases were utilized in order to test the applicability of the new Procrustes shape analysis verification tool to a variety of situations. This was done to examine the possible benefits of using such a system which can breakdown error components which include information of size, shape, and intensity. This was also done to see what improvements could be made to the current system or to illustrate how end users may adjust the penalty function to fit their specific needs. Table 3 illustrates the differences in the overall penalty function to each of the idealized cases.

Case #	Total Penalty
5	3.1706E-026
13	9.8632E-025
6	1.0835E-024
14	6.1000E-003
3	1.1475E+000
11	1.0069E+002
1	1.3419E+002
12	1.5497E+002
2	3.0000E+002
10	2.1305E+004
7	4.0677E+004
9	4.3809E+004
8	7.4343E+004
4	7.5281E+004

Table 3: Procrustes shape analysis total penalty for the 14 idealized cases using the penalty function described by (2). The cases are listed in ascending order based on their total penalty.

While a first examination of the penalties associated with various descriptions of the idealized cases may not make intuitive sense at first. A quick description of the breakdown of some of the idealized cases may help. The first question that arises is why idealized case 13 (figure 7) has the second lowest penalty associated with it while the forecast looks totally worthless. The answer lies in the selection of how to match the objects within the verification scheme. This matching is based on minimizing the RSS portion of the penalty function. Since there are three objects that perfectly match in terms of shape, the match is made this way and yields only a translation error component for each cell. If one were to change the matching characteristic to a distance-weighted method a much larger error would result, as now there would be size, shape, and intensity errors to consider. A future application of this verification scheme would to utilize both a distance- and shape-weighted scheme, as well as an intensity-weighted scheme so that cells of similar nature and life cycle are matched appropriated. As with all object-oriented approaches matching can be counterintuitive. With the options this new verification scheme provides, we hope to decrease some of the ambiguity in cell matching.

Another question arises in the form of the translation error cases having lower errors despite being separated by great distances to their counterpart. Again, in the current framework of the penalty function the translation component is not utilized. This can be an instance where the end-user may which to increase the penalty function based on translation if timing is the main concern and not necessarily intensity. The aviation community comes to mind in this instance as a thunderstorm of any intensity is a threat; however, the timing of the event is the main player in keeping takeoff and landing schedules intact.

To illustrate making the initial penalty function (2) more meaningful, consider altering the penalty function into components of similar magnitude and including all possible errors including translation (T), rotation (R), dilation (D), intensity and shape (RSS) errors. An example of such a penalty function is given as (3).

$$D = RSS^{0.5} + SS_{\min} + SS_{\max} + SS_{avg} + (1 - SS_D) * 100 + SS_R * 100 + SS_T^{0.5} \quad (3)$$

The total penalty given each idealized case now makes more intuitive sense and the values based on the penalty function (3) can be seen in table 4. It is now evident that the objects that must be rotated and translated a great deal have the highest total penalty. Cases 7 and 8 respectively deal with situations of the truth and forecast object having the same intensity structure but having signification size, rotation and translation error components. The errors would be magnified to a greater extent had the intensity components exhibited difference as well. Case 12 and 14 exhibit the smallest errors as the objects have similar shape and size with the same centroid and have slightly different intensity structure. Case 14 is a mirrored case, which shows some weakness in the Procrustes scheme. However, if a different threshold was used in the mirrored case the objects would look as if they have different shapes and thus would increase the penalty function. Case 12 is just a random noise intensity field based on the true field so the result has minimal error.

Case #	Total Penalty (from 3)
14	2.43
12	22.63
5	102.56
1	129.00
10	129.46
11	197.48
2	197.99
4	246.74
3	384.86
6	396.35
9	486.15
13	528.83
7	562.95
8	640.74

Table 4: Procrustes shape analysis total penalty for the 14 idealized cases using the penalty function described by (3). The cases are listed in ascending order based on their total penalty.

Another interesting result comes in the examination of case 13 shown in figure 7. In the original version of the Procrustes scheme the objects were matched on similar shape, which minimized the error in intensity which minimizes the total penalty in (2) simply because no translation error was accounted for in (2). An additional run of the Procrustes scheme was used to calculate values in Table 4 using a minimum distance matching approach. When examining the breakdown in error for idealized case 13, translation, dilation, and intensity errors are all greater than zero yielding a high penalty function. When utilizing the Procrustes scheme which minimizes the shape error (RSS), the breakdown in error changes by eliminating the intensity and dilation errors and just leaving a different translation error. When using (3), the total penalty using the minimum distance scheme is 528.83 and the total penalty using the minimum RSS scheme is 263.38. It therefore seems best that the minimum distance version be used as it is makes more intuitive sense until the schemes can be combined.

5. CONCLUSION

Overall, the usefulness of this particular verification scheme is readily apparent. It allows for the breakdown of error into components and the penalty function can be user-defined so that important errors become the dominant player in the end result. The use of these idealized cases also highlights the benefits of testing all verification schemes with some baseline set of cases. It may highlight limitations from one system to the next and provide insight to the developers on which areas need work in the particular scheme. It also allows an assessment of the meaning of the magnitudes of errors produced in cases where the verification scheme uses an open-ended error scale.

6. REFERENCES

Dryden, I.L., and K.V. Mardia, , 1998: Statistical Shape Analysis. John Wiley & Sons.

ICP, 2007: Spatial Forecast Verification Methods Intercomparison Project: http://www.rap.ucar.edu/projects/icp/index.html

Micheas, A., N.I. Fox, S.A. Lack, and C.K. Wikle, 2007: Cell identification and verification of QPF ensembles using shape analysis techniques. *J. Hydrology (*in press).

ACKNOWLEDGEMENTS

The authors would like to thank Paul Kucera at NCAR for the development of the idealized cases used in this study.

This research was funded by the National Science Foundation Award # ATM-0434213.