Applications of Bayesian Procrustes Shape Analysis to Ensemble Radar Reflectivity Nowcast Verification

Neil I. Fox^a, Athanasios C. Micheas^b, and Yuqiang Peng^b

^aCorresponding author:

Department of Soil, Environmental and Atmospheric Sciences,

University of Missouri,

332 ABNR Building, Columbia, MO 65211, USA

Email: foxn@missouri.edu

Phone: +1- 573-882-2144

^bDepartment of Statistics,

University of Missouri, Columbia, MO 65211, USA

Abstract

This paper introduces the use of Bayesian full Procrustes shape analysis in object-oriented meteorological applications. In particular, the Procrustes methodology is used to generate mean forecast precipitation fields from a set of ensemble forecasts. This approach has advantages over other ensemble averaging techniques in that it can produce a forecast that retains the morphological features of the precipitation structures and present the range of forecast outcomes represented by the ensemble. The production of the ensemble mean avoids the problems of smoothing that result from simple pixel or cell averaging, while producing credible sets that retain information on ensemble spread. Also in this paper, the full Bayesian Procrustes scheme is used as an object verification tool for precipitation forecasts. This is an extension of a previously presented Procrustes shape analysis based verification approach into a full Bayesian format designed to handle the verification of precipitation forecasts that match objects from an ensemble of forecast fields to a single truth image. The methodology is tested on radar reflectivity nowcasts produced in the Warning Decision Support System – Integrated Information (WDSS-II) by varying parameters in the K-means cluster tracking scheme.

Keywords: Procrustes Shape Analysis. Radar Meteorology. Ensemble Nowcasting. Bayesian. Forecast Verification.

1. INTRODUCTION

As precipitation often occurs in discrete areas, verification of precipitation forecasts can be misleading if simple point-based categorical statistics are employed, but better assessments can be made by using an object-based method, such as that introduced by Ebert and McBride (2000). Over the last decade a number of verification schemes have been proposed and demonstrated which develop the Ebert and McBride (2000) Contiguous Rain Area (CRA) concept to provide further information on a forecast's ability to replicate structural features (e.g., Nachamkin, 2009; Ebert and Gallus, 2009; Keil and Craig, 2009; Davis et al., 2006; Marzban and Sandgathe, 2009; Haiden et al., 2012; Mittermaier et al., 2013; Dorninger and Gorgas, 2013). Reviews of many of these and other spatial verification methods can be found in Gilleland et al. (2009). One of these methods introduced the use of Procrustes shape analysis in the verification of radar reflectivity forecasts of precipitation areas (Micheas et al., 2007; Lack et al., 2010). In that work single precipitation objects from forecast fields were matched to reflectivity objects in a truth image and forecast errors determined from the matching parameters and residuals. The advantage of the Procrustes approach is that it can isolate forecast errors into components due to differences in location, size, orientation and intensity of precipitation areas, allowing assessment of forecast process errors. That methodology produced simple statistics based on the matching of a single object in the forecast to one in the truth (Micheas et al., 2007; Lack et al., 2010).

Two outstanding issues with precipitation forecast verification are how to represent the uncertainty in the verification metric and how to verify ensemble forecasts (Gorgas and Dorninger, 2012; Ebert et al. 2013). This paper extends the application of the classic Procrustes approach by presenting two applications of Bayesian Procrustes shape analysis that specifically address issues associated with ensembles of precipitation forecasts. The first, and main, application is the representation of discrete precipitation objects produced by an ensemble of forecasts in a single average field in a way that retains high intensity features. The transition from the classical approach to the Bayesian Full Procrustes method creates more distributional information that provides information on forecast uncertainty. The second issue concerns the verification of ensemble forecasts. The process presented here, by utilizing a Bayesian approach,

produces distributional information that yields credible sets and makes it a good tool for assessing the accuracy of object based forecasts.

A major advantage of ensemble methodologies is the ability to assess forecast uncertainty and show the range of possible forecast outcomes and the probability of extreme events such that appropriate preparations can be made. It is also desirable to efficiently present that uncertainty while retaining important features of the forecast field. This is especially important when forecasting convective rainfall events in which it is critical to represent the actual storm morphology, peak rainfall rates and areal coverage. However, ensembles of forecasts that contain discrete precipitation ‘objects’ are problematic to convert to a useable precipitation forecast since simple averaging of the multiple fields will tend to smooth and dilate objects.

The majority of nowcast schemes provide a single deterministic forecast reflectivity or precipitation field (e.g. Rinehart, 1981; Dixon and Wiener, 1993; Chornoboy et al., 1994), while others provide a forecast track (e.g., Johnson et al., 1998). Some procedures present a “best-guess” forecast along with some measure of uncertainty (e.g., Seed, 2003; Germann and Zawadzki, 2004; Turner et al. 2004), and others produce a probability field, however this may place unwarranted emphasis on a particular ensemble member. The implication is that one model formulation is superior to the others, and this leads to the questions of whether it is appropriate to use “inferior” models or how one chooses the “best” solution if all model runs are equal. Indeed, one could argue that a true ensemble would contain a multitude of equally probable members. An example of an operational forecast tool that takes this approach is the Short-Term Ensemble Prediction System (STEPS: Bowler et al., 2006; Seed et al., 2013). This system, employed by the UK Met Office for hydrological applications generates a deterministic rainfall forecast field based on the best-guess advection field alongside a probability of precipitation for each grid square. This is an appropriate approach for widespread precipitation systems associated with mid-latitude cyclonic systems, but may be misleading when used with the discrete precipitation areas typical of convective precipitation. In such cases a forecast benefits from the retention of structural information that clearly conveys the areal extent and shape of the storm. Other examples of ensemble forecast methods which take different approaches to the formulation and

presentation of the final precipitation forecast product can be found in Xu et al. (2005), Fox and Wikle (2005), Kober et al. (2012, 2014) and Sloughter et al. (2007).

To address the problem of presenting the output of an ensemble forecast of discrete precipitation areas, in the scheme presented herein, all members of a small ensemble are treated equally, and equivalent objects (cells) from the different members of the ensemble are averaged using a Bayesian Procrustes process that produces optimal mean objects by removing orientation, dilation and translation differences between objects before finding the average shape. By finding the best transformation (rotation; dilation; and translation) of the individual realizations of the objects towards a common mean in the Bayesian framework, a mean object can be defined with associated credible sets that provide information on the uncertainty in position, size, and orientation.

The verification of ensemble forecasts and the representation of uncertainty in the verification measures have previously been explored in papers such as Zacharov and Rezacova (2009) and Santos and Ghelli (2012). In this paper a Bayesian approach to obtain the full set of Procrustes parameters to acquire estimates of the effects of rotation, dilation and location is demonstrated. By using a set of forecast objects produced by a small ensemble of nowcast runs the variability is captured through the assessment of credible sets. An assessment of the Bayesian full Procrustes analysis of a distribution of object shapes by obtaining the Full Procrustes estimate of the mean forecast shape is achieved by verifying that against the truth. These applications are demonstrated using a limited ensemble of nowcasts of radar reflectivity. However, they could equally be applied to model precipitation fields or other forecast products that are usually presented as discrete areas. In this work we have chosen to focus on nowcasts generated by the k-means cluster scheme within the Warning Decision Support System – Integrated Information WDSS-II (Lakshmanan et al. 2007) as this system is designed to handle the more discrete convective rain areas for which the Procrustes method should be best suited, but we also include examples using a spatial scaling nowcast method (Seed 2003) and a stochastic method (Fox and Wikle 2005).

In the following section we describe the Bayesian Procrustes fit that collapses a number of ensemble forecasts to a common mean, and this includes the Bayesian estimation of mean shape. The application to the nowcast ensemble is presented in Section 3 with details of the nowcast production, and presents the results of the demonstration of the ensemble mean and verification process. Section 4 summarizes the findings of this work and discusses some of the additional work that will be required to make this method operational.

2. BAYESIAN FULL PROCRUSTES FIT

In this work ‘cells’ that appear in radar reflectivity images as contiguous areas of pixels that exceed selected thresholds of reflectivity and area are represented as shapes defined by a set of landmark points. These are identified by first obtaining a centroid for the cell (the object), and extending angles from the centroid in fixed directions and observing when the boundary of the cell is hit. The point the radial hits becomes the landmark for that specific angle. The number of angles used, and hence corresponding landmarks, defined by a fixed angular separation as viewed from some centroid location, can be altered in order to achieve the best representation of a cell in an image. In the Procrustes fit two objects are considered to have the same shape if the set of landmark points match one another exactly after one set has been translated, rotated and resized (dilated).

Classic Procrustes analysis provides a straightforward way of estimating shape parameters, but is not capable of accessing the distribution of these. The Bayesian approach does not have this problem as it provides multiple realizations to generate a distribution of parameters from the model. Here we use the Bayesian approach to combine corresponding cells from different forecast runs into a ‘best’ forecast. This requires a one-to-one correspondence between cells across forecasts and may therefore leave cells in one run unmatched. In this case the true image at time T may contain N cells, but all forecasts in the ensemble might contain a different number of cells. If the ensemble member has fewer or more cells than the true image, the forecast is inappropriate and should be penalized appropriately. Even in the case when the forecast has the same number of cells as the truth, the prediction may be unacceptable, depending on the form of the cells involved. For a perfect prediction, we should have the same number of cells in the truth

and any forecast, but this may not occur in real applications. In this case, a forecast with fewer or more identifiable cells from the truth is penalized naturally by the approach we take, namely that a cell from the truth is always forced to match with a cell from the forecast.

This work examines 2-dimensional shapes defined by a set of p landmarks represented by a complex vector $\mathbf{z} = (z_1, \dots, z_p)^T$ where the real part of each z_i is the x-coordinate and the imaginary part is the y-coordinate. There are several ways in which cell correspondence can be accomplished. Herein, we will consider matching cells based on either their location (centroid based), or their shape differences (Procrustes matching). Taking only location into consideration, if the true image contains only one cell S_1 and an ensemble member has two, then we would define them as S_{11} and S_{12} , and find the cell nearest to S_1 , as the corresponding cells between truth and ensemble member. In order to take into consideration the effects of location, dilation and rotation in each cell, we consider a statistical shape analysis framework (e.g., Bookstein, 1991; Stoyan and Stoyan, 1994; Dryden and Mardia, 1998; Micheas and Dey, 2003, 2005; Micheas et al., 2006). For this approach correspondence between cells is accomplished by finding the smallest Procrustes distance between a truth cell and a cell from an ensemble member. For more details on both approaches, we refer to Section 2.2 of Micheas et al. (2007).

2.1 Bayesian Full Procrustes Mean

The first aim is to produce the full Procrustes mean of the distribution of ensemble objects rather than simply averaging each pixel value across all realizations, or even a simple averaging of the properties of each cell from all the ensemble members. This approach will provide a realistic average of the cell properties by treating the effects of translation, rotation and dilation separately. The idea is to find a single shape that best represents the mean of the collection of forecast objects, and this is achieved through an iterative procedure. This is performed in an object-oriented fashion, such that each individual cell is identified in each ensemble member and the shapes of these cells are then matched to a common mean that is initially unknown.

In practice, as each shape is defined by the set of landmarks that define its boundary contained in the vector $\mathbf{z}$, and the match is achieved through the use of a complex regression. Figure 1 shows a set of five ensemble member 30-minute radar reflectivity forecasts. In each case the objects are identified by setting a reflectivity threshold and the boundaries tagged as shown in the second column of images. In this case the reflectivity threshold was set high at 45 dBZ, and this results in a small number of distinct objects. A minimum size limit is then applied in order to filter out small areas that are either the result of noisy data or are too small to be of concern. For this example the size limit was set at 40 pixels where each pixel represents a 1 km by 1 km square. In the limited domain displayed in figure 1, there is one object identified in each forecast and this is delineated using a set of eight landmarks tagged by finding the intersection of radial lines drawn from the centroid of the object to the boundary at intervals of 45° . These objects are shown in the third column of Figure 1.

For the Full Procrustes Fit (FPF) an object is translated, rotated and dilated to produce an exact match with another object. This vector, representing the forecast object, $\mathbf{w} = (w_1, \dots, w_p)^T$, can then be matched to a mean object, $\mathbf{z}$, using

$$\mathbf{w} = c\mathbf{1}_p + be^{i\theta}\mathbf{z} + \boldsymbol{\varepsilon} \quad (1)$$

where c is the translation, b the dilation, and θ is the rotation angle. $\boldsymbol{\varepsilon}$ is a p -dimensional complex error vector assumed to have an isotropic covariance matrix, i.e., any direction for a landmark is equally preferred. In applications where the cells are circular or polygonal with all landmarks about a centroid, this assumption is not unreasonable, and minor deviations from it can be captured through the other parameters involved in the Procrustes regression equation. When isotropy is not appropriate (i.e., the cells must have a certain type of form based on a priori observations), we would need to consider a general covariance structure for the errors. This is a simple modification of the Bayesian hierarchical framework, but will not be considered here.

The idea is to find the FPF of $\mathbf{w}$ onto $\mathbf{z}$ that minimizes the magnitude of the difference in shape between $\mathbf{w}$ and $\mathbf{z}$:

$$D^2(\mathbf{w}, \mathbf{z}) = \|\mathbf{w} - be^{i\theta}\mathbf{z} - c\mathbf{1}_p\|^2 \quad (2)$$

$D^2(\mathbf{w}, \mathbf{z})$ is called the Procrustes distance and shows the magnitude of the difference in shape between the two objects, as was presented in Micheas et al. (2007). In essence the task is to find the optimum combination of b , c and θ that minimizes the sum of the Procrustes distances between the mean object and the corresponding object in each of the ensemble forecast members. So if there are N members of the ensemble, the requirement is to minimize:

$$\sum_{j=1}^N D^2(\mathbf{w}, \mathbf{z}_j) = \sum_{j=1}^N \left\| \mathbf{w} - b_j e^{i\theta_j} \mathbf{z}_j - c_j \mathbf{1}_p \right\|^2 \quad (3)$$

As described in Micheas and Peng (2009), it can then be shown that the observed values of $\mathbf{w}$ represent samples from a model distribution described by

$$f(\mathbf{w}|\cdot) \propto \sigma^{-1} \exp \left\{ -\sigma^{-2} \sum_{i=1}^p \left\| \mathbf{w}_i - (c \mathbf{1}_p + b e^{i\theta} \mathbf{z}_i) \right\|^2 \right\} \quad (4)$$

Initial values of b , c , and θ are chosen arbitrarily and a sequence of iterations is performed using a Gibbs sampler with embedded metropolis-Hastings steps taken from the joint distribution of the parameters. So that for the $(r+1)^{\text{th}}$ iteration $c^{(r+1)}$ is taken from $\pi(c^r | b^r, \theta^r, \mathbf{z}, \mathbf{w})$, $b^{(r+1)}$ is taken from $\pi(b^r | c^r, \theta^r, \mathbf{z}, \mathbf{w})$, and $\theta^{(r+1)}$ is taken from $\pi(\theta^r | b^r, c^r, \mathbf{z}, \mathbf{w})$. For each iteration, r , the Procrustes distance given by (2) is found, and after allowing for a burn in period (the initial series of iterations which may produce wildly varying parameter values) the parameters $(b^{(i)}, c^{(i)}, \theta^{(i)})$ yielding the smallest distance is chosen. For each combination the Procrustes residual of each object is found from equation (2) and the best fit for the set of objects (defined by the minimization of equation (3)) is selected from this set of parameter sets. If we allow there to be r iterations of this process and the best fit is found using the i^{th} iteration, then the estimator of the FPF would be $\mathbf{w}_{(i)}^F = c^{(i)} \mathbf{1}_p + b^{(i)} e^{i\theta^{(i)}} \mathbf{z}$. An estimator of the Procrustes distance is found from the residual of this fit. Explicit details of the computation can be found in Micheas and Peng (2010).

This process can then be repeated k times with different initial values of b , c and θ such that we find k values of each of $b^{(i)}$, $c^{(i)}$ and $\theta^{(i)}$. Using this set of estimated parameters means and credible sets for each parameter are found. In this way the Procrustes mean of the cell (object) is created from the corresponding cells in each ensemble forecast member as a set of landmark coordinates that define the boundary of the mean cell. There are also estimators $(\hat{b}, \hat{\theta}, \text{Re}(\hat{c}), \text{Im}(\hat{c}))$ of the dilation, rotation and translation shifts that express the average transformation required to match the cells in the ensemble members to the mean cell. This

provides a measure of the spread of the ensemble forecasts for a particular cell along with the credible sets that provide information on the confidence of that estimation.

Overall, a single set of cell boundary coordinates is created that minimizes the difference between the observed landmark locations and those of the mean forecast's landmarks. The corresponding cells in each ensemble forecast member as well as the mean of each cell can be compared to the truth. For a perfect forecast, we should have no location ($c = 0$), rotation ($\theta = 0$), and dilation effects ($b = 1$). Deviations from this perfect scenario are penalized appropriately when using the proposed shape analysis measure to assess how well the ensemble is doing in estimating the truth.

An example of the results of this process is shown in figure 2 using three different nowcast schemes. The data is taken from the San Antonio, TX radar (KEWX) on 5 July 2002 for a valid time of 0340 UTC. Three nowcast systems, the WDSS-II k-means cluster scheme (Lakshmanan et al. 2003), the Spectral Prognosis system (SPROG: Seed, 2003), and the Hierarchical Bayesian Model (Xu et al., 2005; Fox and Wikle, 2005), were used to generate three ensembles of six nowcast members apiece for the same forecast time. In this case the three nowcast systems produced ensemble members in different ways: The WDSS nowcasts are produced using a k-means cluster method from the same lead time but with six different reflectivity thresholds to determine the advection; SPROG used six different lead times forecasts from 10 minutes to 60 minutes prior to valid time; and the HBM used six stochastically generated forecasts from the same lead time.

Figure 2 shows the six nowcasts from each of the systems on the left. These nowcasts have a 4 km pixel size such that the 32 x 32 domain shown is 128 km x 128 km in extent. In the center column the Procrustes shapes for the largest and most intense rain object (blue dotted lines) thresholded at 20 dBZ and the mean fitted shape (solid green line) are shown. The left hand image shows a comparison of the three resultant mean forecast objects. It can be seen that the WDSS nowcasts produce a somewhat smaller and differently shaped mean object compared to the other two nowcast schemes, and a greater spread of shapes. However, despite the differences between the nowcast schemes the resulting mean shapes are consistent in this case.

2.2 Full Procrustes Verification

The verification of the forecast by means of the Procrustes fit provides measures of dilation, rotation and translation required to match a forecast object to one in the truth, and the residual of the fit is the Procrustes distance which represents the difference in shape between the two objects. Within the Bayesian framework multiple two-dimensional shapes can be compared to a ‘truth’ image allowing the computation of the Full Procrustes Fit (FPF) that gives a best estimate of the error along with credible sets (confidence intervals). In this way multiple model realization fields can be assessed, either through comparison of individual members (including the ensemble mean field) to the truth, or through distributional information of the entire set of forecast members.

This methodology provides measures of the error in terms of four parameters: translation, dilation, rotation and the residual between the truth shape and the matched forecast shape. The advantage of this approach is that the errors are accompanied by credible sets that provide information on the distribution of the forecasts around the mean and the truth. This information can be used to indicate the confidence in each aspect of the forecast or the spread of ensemble members around both the mean forecast and the truth. Overall this provides additional information on the consistency between ensemble forecast members.

3. APPLICATION TO NOWCAST ENSEMBLE

In order to demonstrate the Bayesian Procrustes mean and verification the method is applied to a small ensemble of nowcasts generated by altering the parameterization of the K-means cluster storm cell advection process within the Warning Decision Support System – Integrated Information (WDSS-II: Lakshmanan et al., 2003, 2007). Rather than ascertain cell motion using individual cell centroid tracks, the K-means clustering estimates the motion of a clustered group of objects in a reflectivity field. The clustering and assessed motion vector can be based on a user-defined band of reflectivity values, such that the motion vector applied to generate the forecast reflectivity image can be based on the motion of sections of the image defined within

that reflectivity band. Therefore, by changing the reflectivity band it is possible to alter the forecast motion of convective storms (Sen Roy et al. 2014; Lakshmanan et al. 2012) and, if a number of different reflectivity bands are used, an ensemble of different forecast fields can be generated that provide a range of solutions representing some realistic indication of the uncertainty in the forecast storm motion (Fox et al., 2007, 2008).

The default reflectivity range upon which the motion forecast is made is 20 dBZ to 60 dBZ. This means that the cluster delineates areas of reflectivity between these two values in consecutive radar volume scans, associates clusters in the sequence of scans and derives motion vectors based on the tracking of clusters from one time to the next. The motion vectors are applied to the full range of reflectivities in the final forecast. Therefore the use of different reflectivity bands will produce slightly different nowcasts as the derived motion will be based upon somewhat different delineated clusters. In this case we chose to use ranges of 20-60 dBZ, 30-60 dBZ, 30-50 dBZ, 20-40 dBZ, and 40-70 dBZ to produce an ensemble of five members, although there is little to prevent using a greater number and variety of ranges. This could also be achieved by altering other parameters in the scheme such as those that control the spatial scale of the clusters or the permitted growth rate of the objects in the forecast.

3.1 Full Procrustes Fit Examples

3.1.1 Case 1: New Braunfels, TX 13 January 2007

To demonstrate the application the procedure is applied to the forecasts for 1200 UTC on 13 January 2007 for the domain of the New Braunfels, TX radar (KEWX) using radar reflectivity images that were processed within WDSS-II to a 1 km x 1 km resolution Cartesian grid. We compare the 10 minute and 60 minute lead-time forecasts for this time. Figure 3 shows the radar image for 1200 UTC, which is assumed to represent the truth, alongside the five 10-minute forecasts for that time made at 1150 UTC. Figure 4 is the same except that in this case the forecasts were 60-minute forecasts generated from 1100 UTC.

Figure 5 shows the fits of the five forecasts for two cells for the 10 minute forecast, and Figure 6 shows the same for the 60 minute forecast, in both cases a 35 dBZ reflectivity threshold was used. Also shown in Figures 5 and 6 are the mean shapes found from the Full Procrustes fit. For these examples we use ten landmarks to define the cells and these are found by locating the centroid of the cell and identifying where reflectivity falls below a threshold of 35 dBZ in ten directions at 36° intervals. The mean fit is shown by the thick line and this represents the best estimate of the location of the cell found from the cells in each of the ensemble members. It can be seen that the mean cell retains the size and definition of the individual member cells rather than being smoothed by the process.

Table 1 shows the mean value of the shift of each of the parameters used to produce the Procrustes mean fit from the cells identified in the ensemble members. This is done for each of the two cells at both 10- and 60-minute lead times. In the table it is seen that the average dilation and rotation shifts needed to match the cells are generally small, however, for both of these parameters the shifts are slightly larger at the 60-minute lead time indicating a greater spread of sizes and orientations of forecast cells at longer lead times. This is more noticeable in the real and imaginary parts of the translation parameter (c). As each pixel has represents a 1 km square, and the calculations are performed by pixels, these values can be interpreted as the translation in x (east-west) and y (north-south) coordinates respectively. The actual coordinates of the landmarks that define the cells are shown in Figures 5 and 6. The 10-minute forecasts shown in Figure 5 illustrate how well the cells from each of the ensemble forecasts are clustered, as one would expect a small average shift in each direction. For the 60-minute forecasts, in Figure 6, there is a much greater range of forecast solutions presented by the ensemble members. Most noticeably the spread of forecast cell locations around the mean of the first cell is significant, and this is reflected in the $Im(c)$ – the mean translation in the y -direction, as seen in Table 2.

In the version of the fitting scheme presented here we have not adjusted any of the parameters. Therefore there is no greater weight given to a fit that minimizes the translation required, for example. There is also no factor related to cell size, and this may be desirable as a small displacement of a large cell would result in overlapping cell areas, whereas for a smaller cell the forecast members might not overlap. On the other hand, one could consider this to be an

objective measure of spread which reduces the perceived poorness of an ensemble forecast of small cells where each member's cells do not overlap. The mean ensemble shifts are calculated over all landmarks and are therefore difficult to interpret as absolute measures. Rather, one can compare two similar forecasts to identify which produces a greater spread of forecast object locations. If one were to change the number of landmarks or the number of ensemble members then the magnitude of the average shift would change. A similar argument could be made when comparing forecasts with different resolutions. In each of these cases it would probably be necessary to normalize the measures in some way to produce reliable comparisons.

3.1.2 Case 2: Melbourne, FL 02 February 2007

For a second case the full Procrustes fits are found for an event that occurred in Florida on 02 February 2007. This case involved a cold frontal passage that triggered severe convection over a widespread area along the front. In this instance five 30 minute lead time forecasts for the time of 1230 UTC are shown in figure 7 along with the observed reflectivity field at that time. The mean fits are found using two reflectivity thresholds, 25 dBZ and 45 dBZ, and are shown in Table 3. For the lower threshold three cells are identified, while the higher threshold only finds one. Figure 8 shows the forecast shape fits superimposed on the observed reflectivity image at the valid time for each reflectivity threshold.

This case illustrates the importance of threshold selection as the lower threshold (25 dBZ) identifies a large cell that encompasses not only the convective line, but also areas of moderate reflectivity closer to the radar. The higher threshold clearly locates a single intense cell, but misses others that may be important. The majority of reflectivity nowcast methods tend to reduce peak intensities with forecast lead time, as does the K-means clustering nowcast scheme used here. This is done as intense storms tend not to maintain the strength of individual cells over long periods and the general reduction in reflectivity reflects the limit of prediction of the behavior of these storms (Seed 2003). It is notable however that although the lower threshold delineated unrealistic areas those areas were consistent across the forecasts and it was possible to fit the objects to a common mean with reasonable parameters.

3.2 Full Procrustes Verification Examples

For the verification of the forecast the shape (either the mean shape or the individual ensemble member shapes) is fitted to the observed shape using the same procedure used to obtain the mean shape. The same parameters are calculated, but they can be interpreted as errors in location, dilation and rotation, and the verification application also uses the Procrustes distance or residual as another measure of the difference between the forecast and the truth. Once the fit has been completed it is possible to compute the residuals that represent the remaining difference between the shape of the forecast and truth object. A dilation of 1 represents a forecast cell that is the same size as the observed cell, while a rotation error of 0 means that the forecast and truth are aligned. In some of the forecasts (especially at the longer lead time) there are unequal numbers of cells in the truth and some forecast members. This problem is not penalized in this version of the verification, but one approach to assessing an appropriate penalty is examined in *Lack et al.* (2010). The same paper also presented a method of adjusting the components of the error either to make them comparable or emphasize the error with which one is most concerned.

3.2.1 Case 1: New Braunfels, TX 13 January 2007

For an illustration of the verification the radar reflectivity data from 1200 UTC on 13 January 2007 is compared to the 10-minute forecast from 1150 UTC, and the 60-minute forecast made at 1100 UTC. In the truth image two cells are diagnosed and, therefore, two cells are identified that can be matched to each identified object in each ensemble member forecast.

In the case presented here the rotation and dilation errors found are small in comparison to the translation errors and the residuals. These are shown as the shape difference in Tables 4a and 4b. For the 10-minute forecast these are small as the forecast cells are similar in shape to the truth. The Procrustes mean cell 1 is very similar in shape to the truth, and this is partly a result of little spread in the ensemble. For the 60-minute forecast one sees differences in shape between the forecast and the truth. This is most notable for cell 1 where the cell is elongated in the north-south direction and there is a range of cell shapes in the ensemble

Table 5 shows the Procrustes distances for each of the two cells in each of five forecast members compared to the truth, along with the Procrustes distance for the ensemble mean. As this distance is a measure of the difference in location between the landmarks of the two objects, a smaller value indicates a better match and forecast. In the case of the 10-minute forecast it can be seen that for cell 1 the ensemble mean found by the Bayesian Procrustes fit method produces the best forecast, while for cell 2 it provides the second best forecast.

In contrast to the 10-minute forecast, the 60-minute lead time forecast mean does not perform as well as the majority of ensemble members. Four of the forecasts are good, but one – the forecast generated by the 20 – 40 dBZ reflectivity range tracking – is poor and influences the mean in a manner that decreases its accuracy in comparison to the other forecasts. Such an influence by a single member of the forecast ensemble is not unlikely given the small size of this ensemble, but in this case a particular problem is caused by this forecast producing a single cell that is located between the two cells that appear in the truth image. This cell is then included in the mean fit for each of the two cells identified in each of the other forecasts and is also verified against each of the two cells in the truth. Ideally one would produce a large enough number of members of an ensemble to diminish the influence of a single run, but in this demonstration a limited number of forecasts provides a good illustration of the methodology.

It could be argued that using a tracking routine that uses the motion of reflectivity areas between 20 and 40 dBZ is not appropriate when one is interested in areas with > 35 dBZ reflectivity and this verification provides evidence that the predictive ability of tracks produced using cells defined by the 20 – 40 dBZ reflectivity band is poor for the type of storm observed in this case. This is not surprising for a severe event during which the identification and tracking of the high intensity storm centers is required. In an operational context one could rapidly ascertain the suitability of including particular members in the ensemble or if, as in this case, the forecast methodology justifies exclusion of that member from the forecast process for the remainder of the event.

3.2.2 Case 3: Kansas City, MO 10 June 2006

Figure 9 shows five 30 minute lead time forecasts compared to the truth for an event that took place in the Kansas City, MO area on 10 June 2006. This event featured a series of fast moving supercells some of which underwent splits and mergers. Figure 10 shows the object fits for the four identified cells for each forecast along with the Procrustes mean fit. It is seen in this instance that some of the forecasts merge the two central cell into a single object and this results in the single object being used twice in the fitting process. This is an inevitable consequence of having forecasts that produce differing numbers of objects and we choose in this scheme to allow the same cell to be used more than once. In this case the result is that two of the mean cells overlap and the northern cell is elongated to the south.

Table 6 shows the overall verification scores for the four cells when compared to the truth. Cell 4, the object that is distorted by forecasts that have the two merged cells identified as a single entity has the lowest score due to this effect.

4. DISCUSSION

The methodology described in this paper can generate a mean forecast field of a parameter that has a discrete areal structure or has a critical threshold value whereby contiguous areas are identified and forecast. In this example we use the most common meteorological parameter of this type, precipitation, and the application to nowcasts seems appropriate as good forecasts of heavy rain and severe weather based on radar reflectivity images are dependent on the retention of areal information and sharp gradients. Whereas in a simple pixelwise averaging scheme objects would become diffuse and lose structural information, this method retains the size and shape of the area of interest. Such straightforward methods can diminish the apparent intensity and structure of a forecast precipitation system and hinder the interpretation of the impact of that system. In addition to the retention of structural information, the Bayesian Procrustes fit provides information on the spread of the ensemble forecast in terms of the mean parameter shifts of the ensemble member objects to match the mean, and also gives the uncertainty in these measures in terms of confidence intervals. While this is only a demonstration of the technique, these results are encouraging. Although one might not expect the mean to provide the single forecast that gives the best verification measures when compared to what actually occurs, it would be hoped

that the averaging process will create a good, reliable and useful forecast, in a situation where it is not known which ensemble member is the best.

There are limitations to the present version; the most obvious of these is that it finds the mean shape in terms of the boundary as defined by a limited number of landmarks within a plane. One future aim would therefore be to improve the efficiency of the computation to handle a larger number of landmarks such that the shape and structure of the objects are better specified. This could then lead to the development of a means to extend the process to a third dimension which could either be vertical extent or intensity. This would allow the inclusion of intensity in the averaging scheme rather than just the areal extent of the object. Another approach that could improve this type of representation of objects would be to use a mixture model. Such models define the object by density functions, usually Gaussian in form, which reduces the number of parameters required to those necessary to define the function while retaining essential characteristics of the object. Such models have been investigated as a verification tool for precipitation forecasts (Deweese 2008; Lakshmanan and Kain 2010), and also for nowcasting (Deweese 2008).

As with any iterative process the computational requirements are great and the time taken to fit objects will grow with the number of cells within the domain, the number of landmarks used, and the number of ensemble members. Therefore, for this approach to become an operational method the efficiency of the computation needs to be optimized. However, even with the computational restrictions the concept of producing composites of ensemble forecasts of precipitation in an object-based manner must be pursued in order to avoid the smoothing effect of simplistic averaging schemes.

The verification methodology presented herein allows an assessment of the quality of the ensemble mean and the individual members either as an overall performance or on an object-by-object basis. However if, as is desirable, there are a greater number of ensemble members then finding the mean is accomplished and assessed in a consistent object-oriented manner that retains the important features of the forecast. Other methods of producing an ensemble mean, such as pixel-by-pixel averaging would usually smooth forecasts, especially in the case of precipitation

forecasts that resolve discrete rainfall areas. Such smoothed forecasts can produce better verification measures, particularly when using common contingency type measures such as critical success index that also work on a pixel basis (Ebert and McBride, 2000). Therefore for an ensemble mean that retains the features of the objects one should use an object-oriented verification scheme (Gilleland et al., 2009).

One issue with object-oriented verification schemes is the dependence on the definition of the object to be verified. When designing a verification scheme one wants to produce an objective method that can be used across multiple forecasts, and that produces measures of accuracy that are readily understood and transferrable. Some problems arise which may explain why such methods are not more commonly used, and these are that error scales that look at residuals or areal differences create open-ended scales in which it is not clear whether a diagnosed value in one circumstance is large or small compared to a different value for a different situation. Coupled to this is the fact that the object definition is dependent upon choices such as threshold level and resolution of the object's perimeter. In the case shown here a choice was made to reduce the number of points that define the shape of the object in order to reduce computation time. The result of this is that information on object structure is lost in the attempt to reduce dimensionality. Therefore the comparison of forecast object to truth is not good, and it is clear that the final residual calculated would change if a different number of landmarks were used. To promote the adoption of object-oriented verification methods the measures need to be normalized in a consistent manner so that events and forecasts with different characteristics can be compared objectively.

Acknowledgements

This research was made possible by National Science Foundation Grant ATM-0434213.

References

- Bookstein, F.L., 1991. *Morphometric tools for landmark data: Geometry and Biology*. Cambridge University Press, Cambridge.
- Bowler, N., Pierce, C.E., Seed, A., 2006. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Quart. J. Royal Met. Soc.*, *132*, 2127-2155.
- Chornoboy, E.S., Matlin, A.M., Morgan, J.P., 1994. Automated storm tracking for terminal air traffic control. *The Lincoln Laboratory Journal*, *7*, 427-448.
- Davis, C., Brown, B., Bullock, R., 2006. Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, *134*, 1772–1784.
- Deweese, T.A., 2009. A hierarchical Bayesian mixture approach for modeling reflectivity fields with application to nowcasting. Ph. D. Dissertation, University of Missouri – Columbia. 201 pp.
- Dixon, M., and G. Wiener, 1993: TITAN: Thunderstorm Identification, Tracking Analysis and Nowcasting – A radar –based methodology. *J. Atmos. Oceanic Technol.*, *10*, 785-797.
- Dorninger, M., Gorgas, T., 2013. Comparison of NWP-model chains by using novel verification methods. *Meteorol. Z.*, *22*, 373-393.
- Dryden, I.L., Mardia, K.V., 1999. *Statistical Shape Analysis*. John Wiley & Sons.
- Ebert, E.E., Gallus Jr., W.A., 2009. Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification. *Wea. Forecasting*, *24*, 1401-1415.
- Ebert, E.E., McBride, J.L., 2000. Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.*, *239*, 179-202.
- Ebert, E., Wilson, L., Weigel, A., Mittermaier, M., Nurmi, P., Gill, P., Göber, M., Joslyn, S., Brown, B., Fowler, T., Watkins, A., 2013. Progress and challenges in forecast verification. *Meteorol. Apps.*, *20*, 130-139.
- Fox, N.I., Limpert, G., Gilmore, W., Miranda, J., Schnetzler, A., Miller, N., Kolehny A., Lack, S., 2007. The sensitivity of object-based nowcasts to object threshold selection. *Preprints 22nd Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction*, Park City, UT, CD-ROM.

Fox, N.I., Lack, S.A., Micheas, A.C., 2008. Applying a Procrustes analysis verification scheme to nowcast ensemble forecast members to determine accuracy of diagnosed convective mode. *Preprints 19th AMS Conference on Probability and Statistics*, New Orleans, LA. CD-ROM.

Fox, N.I. and C.K. Wikle, 2005: A Bayesian quantitative precipitation nowcasting scheme. *Weather and Forecasting*, **20**, 264-275.

Germann, U., and I. Zawadzki, 2004: Scale dependence of the predictability of precipitation from continental radar images. Part II: Probability forecasts. *J. Appl. Meteor.*, **43**, 74-89.

Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B., Ebert, E.E., 2009. Intercomparison of spatial verification methods. *Wea. Forecasting*, **24**, 1416-1430.

Gorgas, T., Dorninger, M. 2012. Quantifying verification uncertainty by reference data variation. *Meteorol. Z.*, **21**, 259–277.

Haiden, T., Rodwell, M.J., Richardson, D.S., Okagaki A., Robinson, T., Hewson, T. 2012. Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Mon. Weather Rev.*, **140**, 2720–2733.

Johnson, J.T., P.L. MacKeen, A. Witt, E.D. Mitchell, G.J. Stumpf, M.D. Eilts, and K.W. Thomas, 1998: The Storm Cell Identification and Tracking Algorithm: An enhanced WSR-88D algorithm. *Wea. Forecasting*, **13**, 263-276.

Keil, C., Craig, G.C., (2009), A displacement and amplitude score employing an optical flow technique. *Wea. Forecasting*, **24**, 1297-1308.

Kober, K., Craig, G.C., Keil, C., Dörnbrack, A., 2012. Blending a probabilistic nowcasting method with a high-resolution numerical weather prediction ensemble for convective precipitation forecasts. *Q. J. R. Meteorol. Soc.* **138**, 755–768.

Kober, K., Craig, G.C., Keil, C., 2014. Aspects of short-term probabilistic blending in different weather regimes. *Q. J. R. Meteorol. Soc.* **140**, 1179–1188.

Lack, S.A., Limpert, G.L., Fox, N.I., 2010. A multiscale object-oriented verification scheme. *Wea. Forecasting*, **25**, 79-92.

Lakshmanan, V., Kain, J.S., 2010. A Gaussian mixture model approach to forecast verification. *Wea. Forecasting*, **25**, 908-920.

Lakshmanan, V., Rabin, R., DeBrunner, V., 2003. Multiscale storm identification and forecast. *Atmos. Res.*, **67-68**, 367-380.

- Lakshmanan, V., Smith, T., Stumpf, G.J., Hondl, K., 2007. The Warning Decision Support System - Integrated Information (WDSS-II). *Wea. Forecasting*, 22, 596-612.
- Lakshmanan V., Crockett, J., Sperrow, K., Ba, M., Xin, L. 2012. Tuning the auto-nowcaster automatically. *Wea. Forecasting*, 27, 1568–1579.
- Marzban, C., Sandgathe, S., 2009. Verification with variograms. *Wea. Forecasting*, 24, 1102-1120.
- Micheas, A.C., Dey, D.K., 2003. Shape Classification procedures with applications. *Proceedings of the Section on Bayesian Statistical Science, American Statistical Association, San Francisco, 2003*.
- Micheas, A.C., Dey, D.K., 2005. Modeling shape distributions and inferences for assessing differences in shapes. *Journal of Multivariate Analysis*, 92, 257-280.
- Micheas, A.C., Dey, D.K., Mardia, K.V., 2006. Complex elliptical distributions with applications to shape analysis. *Journal of Statistical Planning and Inference*, 136, 2961-2982.
- Micheas, A., Fox, N.I., Lack, S.A., Wikle, C.K., 2007. Cell identification and verification of QPF ensembles using shape analysis techniques. *Journal of Hydrology*, 344, 105-116.
- Micheas, A.C., Peng, Y., 2010. Bayesian Procrustes analysis with applications to hydrology. *Journal of Applied Statistics*, 37, 41-55.
- Mittermaier, M., Roberts, N., and Thompson, S.A., 2013. A long-term assessment of precipitation forecast skill using the fractions skill score. *Meteorological Applications*, 20, 176-186.
- Nachamkin, J.E. 2009. Application of the composite method to the Spatial Forecast Verification Methods Intercomparison Dataset. *Wea. Forecasting*, 24, 1390-1400.
- Rinehart, R.E., 1981: A pattern-recognition technique for use with conventional weather radar to determine internal storm motions. Research Atmospheric Technology 13, *Recent Progress in Radar Meteorology*, Richard Carbone, Ed., National Center for Atmospheric Research, Boulder, CO, 105-118.
- Santos, C., Ghelli, A., 2012. Observational probability method to assess ensemble precipitation forecasts. *Q. J. R. Meteorol. Soc.* 138, 209–221.
- Seed, A., 2003: A dynamic and spatial scaling approach to advection forecasting. *J. Appl. Meteor.*, 42, 381-388.

Seed, A., Pierce, C.E., Norman, K., 2013. Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme. *Water Resources Res.*, 49, 6624-6641.

Sen Roy, S., Brata Saha, S., RoyBhowmik, S.K., Kundu, P.K., 2014. Optimization of Nowcast Software WDSS-II for operational application over the Indian region. *Meteorology and Atmospheric Physics*, 124, 143-166.

Sloughter, J., Raftery, A., Gneiting, T., Fraley, C., 2007. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Weather Rev.* 135, 3209–3220.

Stoyan, D., Stoyan, H., 1994. *Fractals, Random Shapes and Point Fields: Methods of Geometric Statistics*. Wiley, Chichester.

Turner 2004

Zacharov P, Rezacova D. 2009. Using the fractions skill score to assess the relationship between an ensemble QPF spread and skill. *Atmos. Res.* 94, 684–693.

Figure captions

Figure 1: The first column shows the 30 minute pseudo-reflectivity forecasts from St. Louis (KLSX) for 0730 UTC on 13th August 2007, as well as the mean forecast reflectivity field (bottom) and the observed reflectivity (top). The second column shows the 45-dBZ thresholded objects from these images. The axes are pixel numbers where each pixel is 1 km x 1 km and objects less than 40 pixels in size are filtered out. The third column shows the fitted shape for the cell identified in each image (red) and the Bayesian Procrustes mean shape (blue).

Figure 2: Example of Bayesian mean Procrustes fits from ensembles of 6 members from three different nowcast schemes from 5 July 2002. The images on the left of the figure show forecasts produced by the Hierarchical Bayesian Model (Xu et al. 2005), the Spectral Prognosis scheme (Seed 2003) and the WDSS K-means cluster method (Lakshmanan et al. 2003). The central images show the Procrustes fits to the largest cell in each image (dotted blue lines) and the Bayesian mean fit (solid green line). The image on the right is a comparison of the mean fits from the three ensembles.

Figure 3: Five 10-minute lead-time forecasts compared to the observed radar reflectivity at 1200 UTC on 13 January 2007 using the San Antonio, TX radar (KEWX).

Figure 4: Five 60-minute lead-time forecasts compared to the observed radar reflectivity at 1200 UTC on 13 January 2007 using the San Antonio, TX radar (KEWX).

Figure 5: Procrustes fits for the two cells from 13 January 2007 at 10-minute lead time. The solid green line shows the truth, the dotted line connecting dotted landmarks show the cells from each forecast realization, and the red shows the common full Procrustes fit.

Figure 6: Procrustes fits for the two cells from 13 January 2007 at 60 minute lead time. The green line shows the truth, the dotted line connecting dotted landmarks show the cells from each forecast realization, and the red line shows the common full Procrustes fit.

Figure 7: Five 30 minute lead time forecasts and the observed reflectivity for 02 February 2007 at 1230 UTC using the Melbourne, FL radar (KMLB).

Figure 8: Shapes identified from the five 30-minute forecasts (black) and Procrustes mean fits (red lines) superimposed on observed radar field at 1230 UTC using a 25 dBZ threshold (left) and a 45 dBZ threshold (right).

Figure 9: Five 30 minute lead time forecasts and the observed reflectivity for 10 June 2006 at 2340 UTC using the Pleasant Hill, MO radar (KEAX)

Figure 10: Identified cells using a 45 dBZ threshold from the 10 June 2006 forecasts shown in Figure 9. Cell 1 is shown in blue, cell 2 in cyan, cell 3 in black and cell 4 in green (which also uses some of the cells delineated in cyan).

Figure 1

Figure 2**HBM Nowcasts** **SPROG Nowcasts** **WDSS Nowcasts** **Bayesian Mean Fits**

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

Figure 8

Figure 9

Figure 10

Table 1: Mean parameter shifts for the fit of the ensemble member cells to the mean Procrustes cell. The values of b represent the average dilation, θ represents the average rotation (in radians), and the real and imaginary parts of c show the average translation in the x - and y - directions respectively required to match the cells to the mean. The figures in parentheses show the 95% credible sets as determined by the Bayesian inference and these indicate the spread of the members around the mean.

	<u>10 minute forecast</u>		<u>60 minute forecast</u>	
	Cell 1	Cell 2	Cell 1	Cell 2
b	1.0003 [1.0001, 1.0005]	0.9999 [0.9996, 1.0001]	1.0032 [1.0030, 1.0034]	0.99921 [0.99899, 0.99942]
θ	-0.000003 [-0.00007, 0.00008]	-0.0000012 [-0.000014, 0.0000070]	-0.0000035 [-0.000108, 0.000000]	0.0005 [0.00077, 0.00015]
Re (c)	0.300 [0.092, 0.529]	-0.679 [-0.923, -0.450]	-7.58 [-7.854, -7.315]	0.077 [-0.084, 0.288]
Im (c)	0.534 [0.349, 0.756]	-1.176 [-1.423, -0.880]	-14.315 [-14.555, -14.073]	1.050 [0.884, 1.214]

Table 2a: Translation fits for T+10 and T+60 for the first identified truth cell from each of the five ensemble forecast members.

	<u>10-minute forecast</u>		<u>60-minute forecast</u>	
	x translation	y translation	x translation	y translation
1	-0.294	-0.466	-7.582	-14.315
2	0.932	0.522	15.062	23.966
3	-0.698	-0.340	-4.220	-5.400
4	-0.197	-0.510	-3.251	-4.706
5	0.300	0.534	-0.036	0.248

Table 2b: Translation fits for T+10 and T+60 for second identified truth cell from each of the five ensemble forecast members.

	<u>10-minute forecast</u>		<u>60-minute forecast</u>	
	x translation	y translation	x translation	y translation
1	-1.336	-2.723	0.077	1.050
2	0.024	1.098	-13.774	-30.754
3	1.895	3.634	4.397	9.908
4	-0.404	-0.576	5.289	10.813
5	-0.679	-1.176	3.710	9.104

Table 3

Table 3: Mean parameter shifts for the fit of the ensemble member cells to the mean Procrustes cell for the case from 2 February 2007 from the Melbourne, FL radar. The first three cells were detected using a 25 dBZ reflectivity threshold, and one cell was identified using a threshold of 45 dBZ.

	<u>25 dBZ Threshold</u>		<u>45 dBZ Threshold</u>	
	Cell 1	Cell 2	Cell 3	Cell 1
b	1.0054 [1.0049, 1.0059]	1.0020 [1.0017, 1.0024]	1.0019 [1.0016, 1.0022]	0.9983 [0.9978, 0.9938]
θ	-0.000034 [-0.000069, 0]	0.000011 [-0.000006, 0.000026]	-0.000022 [-0.000044, -0.000001]	0.000022 [0, 0.000044]
Re (c)	-6.702 [-6.948, -6.456]	2.347 [2.139, 2.556]	-0.452 [-0.646, -0.258]	3.214 [3.066, 3.362]
Im (c)	-0.572 [-0.748, -0.396]	-0.456 [-0.650, -0.263]	-2.167 [-2.342, -1.992]	0.270 [0.018, 0.521]

Table 4a: Estimates of error in the mean shape of the 10-minute ensemble forecast versus the truth.

	Cell 1	Cell 2
Dilation	1.0005	1.0013
Rotation	0.00003	0
Location in x	3.09	0.14
Location in y	1.33	2.41
Shape difference	38.6	78.3

Table 4b: Estimates of error in the mean shape of the 60-minute ensemble forecast versus the truth.

	Cell 1	Cell 2
Dilation	0.9941	1.0005
Rotation	0.000088	0.000155
Location in x	-3.7267	2.2378
Location in y	8.4363	9.4707
Shape difference	967.8	856.5

Table 5a: Procrustes distance between the cell in the truth image and each corresponding forecast cell in the five 10-minute lead time forecast members, as well as the mean forecast cell.

Forecast	Cell 1	Cell 2
1	48.623	63.806
2	47.826	126.94
3	39.684	167.93
4	42.789	132.23
5	52.729	88.984
Mean	38.6	78.3

Table 5b: Procrustes distance between the cell in the truth image and each corresponding forecast cell in the five 60-minute lead time forecast members, as well as the mean forecast cell.

Forecast	Cell 1	Cell 2
1	764.16	847.92
2	1360.2	1019.3
3	678.8	732.94
4	583.87	812.91
5	753.11	861.96
Mean	967.8	856.5

Table 6: Estimates of error in the mean shape of the ensemble forecast versus the truth using the 45 dBZ threshold for the case on 10 June 2006.

	Cell 1	Cell 2	Cell 3	Cell 4
Overall Shape difference	1092	631	1429	2572