



available at [www.sciencedirect.com](http://www.sciencedirect.com)



journal homepage: [www.elsevier.com/locate/jhydrol](http://www.elsevier.com/locate/jhydrol)



# Cell identification and verification of QPF ensembles using shape analysis techniques

Athanasios Christou Micheas <sup>a,\*</sup>, Neil I. Fox <sup>b</sup>, Steven A. Lack <sup>b</sup>,  
Christopher K. Wikle <sup>c</sup>

<sup>a</sup> *Department of Statistics, 146 Middlebush Hall, University of Missouri, Columbia, United States*

<sup>b</sup> *Department of Soil, Environmental and Atmospheric Sciences, University of Missouri, Columbia, United States*

<sup>c</sup> *Department of Statistics, University of Missouri, Columbia, United States*

Received 28 February 2006; received in revised form 17 April 2007; accepted 17 May 2007

**Summary** This paper introduces a new verification technique designed for, but not limited to, quantitative precipitation forecasts. It is tested on, and examples are given for, an intercomparison of very short-period nowcasting schemes. One of these nowcasters is a Bayesian scheme that is used in an extensive ensemble formulation, and the verification scheme is uniquely capable of treating both the ensemble members and the mean forecast. The verification scheme uses Procrustes shape analysis methods that are well established in statistics but have not, to date, been applied to meteorological forecast assessment. The Procrustes methodology allows for a decomposition of the forecast error into any number of components such as location (displacement), shape, size, orientation and intensity. Each error component can be afforded a separate weighting such that a cost or value of the forecast can be calculated that accounts for different error types. For example, a forecaster who is concerned with the location of a storm would place greater emphasis on correct location in the forecast than other attributes. This ability to apply weights makes the system particularly suited to real-time verification applications where confidence in the performance of the forecast translates into improved dissemination to users. In addition, the decomposition of the error into parts enables diagnosis of the error sources that can lead to model adjustment and improvement.

© 2007 Elsevier B.V. All rights reserved.

\* Corresponding author. Tel.: +1 573 884 8828.  
E-mail address: [amicheas@stat.missouri.edu](mailto:amicheas@stat.missouri.edu) (A.C. Micheas).

## Introduction

Typical forecast verification schemes, particularly in the field of QPF (Quantitative Precipitation Forecasts), usually employ categorical statistics such as the critical success index (CSI) or Heidke Skill Score. Particularly in the case of spatially dispersed precipitation areas, several authors have criticized such scores as not representing the true value of the forecast to the forecaster (see for example Ebert and McBride, 2000; Grams et al., 2006).

Two issues that have arisen due to current trends in forecasting are the need to verify forecasts in real-time and the need to verify ensemble member forecasts. Some of the newer proposed multiscale verification methods are very useful, but may suffer in the real-time and ensemble situations because of the multi-dimensional nature of the forecast verification problem (e.g. Casati et al., 2005). The solution is to develop a system that is both robust in that it produces useful and objective measures of forecast accuracy (and preferably indications of forecast strengths and weaknesses), while reducing the dimensionality of the verification to an extent whereby multiple realizations of an ensemble member can be verified individually and overall.

Starting with Ebert and McBride (2000), methods have been developed that account for different aspects of forecast errors. In this case, it was noted that QPF may have a value to a forecaster even when the categorical skill scores are very low because the forecast may contain a small error in location (propagation) such that the forecast and actual precipitation areas hardly overlap. On the other hand, the forecast could represent the size, shape and intensity of the precipitation area (cell) accurately, and the forecaster with knowledge of the meteorology or behavior of the forecast model may be able to either adjust the forecast or to interpret it with expert knowledge by including a subjective uncertainty. Therefore, Ebert and McBride (2000) developed a verification methodology that differentiated the location error from all other errors. The work presented herein extends that concept by separating error terms due to propagation, dilation, intensity, rotation and shape.

There are a number of aspects to the rationale behind the approach taken. Firstly, valuable QPFs contain more information than simply the location and intensity of future precipitation. Structural factors such as the shape of the precipitation area, the change of intensity, the alignment and the size all provide information of the type of precipitation, its likely persistence and future development. Therefore, an analysis of model predictions may provide information on model aspects of storm behavior, leading to improvements in user interpretation and value of forecasts and, potentially, improvements to the model itself.

Secondly, following Ebert and McBride (2000), the methodology employs an object-oriented approach with flexibility in the determination of a boundary threshold intensity to define objects. One important point is that there is no restriction on the number of objects in either the forecast or the actual (truth). Therefore, it is critical to penalize a forecast that does not produce the same number of objects (cells) as exist in the "truth". Within this methodology such a penalty is introduced. This allows the verification scheme to reward forecasts that provide good representations of

storm initiation, dissipation, merging and splitting, QPF aspects that are notoriously difficult to parameterize.

Thirdly, in order to fully exploit the benefits of the verification scheme for model interpretation we choose to work in radar reflectivity rather than precipitation rate. This gives a number of advantages of which two are dominant. One is that one avoids the error introduced in the conversion from reflectivity to rainrate. This further retains the structural information contained in reflectivity that a forecaster may use in assessments of storm type and development. However, this means that we are not truly verifying precipitation forecasts, but rather the quality of the forecast radar reflectivity fields compared to actual fields. In each case there is no rainfall (precipitation) truth observation, but this is common with many verifications.

This paper introduces the new verification scheme that uses a shape analysis methodology that reduces the dimensionality of the verification problem by identifying precipitation objects and assigning their characteristics through a number of angular components that specify the shape of the object in the forecast and corresponding true image. In the following section, the ensemble verification scheme is introduced and described. In Section "Nowcast schemes and data", we discuss the case study we consider along with the nowcasting scheme. Section "Application of cell identification and verification methods" presents an example of how the method may be applied using forecasts from a number of different short-period quantitative precipitation forecast systems. This section illustrates the variety of metrics available through the methodology. Based on this example, Section "Concluding remarks" provides a discussion of the observed applicability of the scheme and suggestions for implementation and broader research.

## Ensemble verification methods

The verification of nowcasts is critical for evaluating any proposed forecast, and replacements for the older categorical statistical measures that do not reflect the real value of precipitation forecasts need to be developed. Verification is not best accomplished using a pixel by pixel comparison, as many forecast fields, particularly those that include forecasts of heavy convective rain, contain features (such as pixel clusters or cells) that can easily be identified as separate objects, with grid values in a specific cell being highly correlated. These identifiable features certainly vary from being different in intensity, and a simple way of quantifying uncertainty in nowcasts would be a pixel by pixel comparison (of intensity). Although simplistic, this approach does not account for other effects such as rotation, scale and location changes that alter the intensities of grid points. For an object-based verification of precipitation forecasts, see for example Davis et al. (2006).

In essence, the purpose is to assess differences between the observed image,  $\mathbf{z}_{T+\tau}$ , of size  $K \times L$ , and the  $M$  nowcast realizations from a distribution of predictions that can be represented by  $\mathbf{y}_{T+\tau}^{(1)}, \dots, \mathbf{y}_{T+\tau}^{(M)}$ . That is, the aim is to assess how well each prediction matches the actual image. Typical approaches to this verification problem utilize all intensities to create a penalty function that measures the error. For example, Ebert and McBride (2000), considered the usual

quadratic penalty function between the truth and the  $i$ th forecast as

$$SS_{\text{tot}}^{(i)} = \frac{1}{KL} (\text{vec}(\mathbf{z}_{T+\tau}) - \text{vec}(\mathbf{y}_{T+\tau}^{(i)}))' (\text{vec}(\mathbf{z}_{T+\tau}) - \text{vec}(\mathbf{y}_{T+\tau}^{(i)})), \quad (1)$$

where  $\text{vec}(\mathbf{X})$  denotes the vector with the columns of the matrix  $\mathbf{X}$  stacked. The authors proposed a further decomposition utilizing a shifted forecast that accounts for changes in location and average intensity. Other event-based verification techniques can be found in Ebert et al. (2004).

Herein, we propose a penalty function that accounts for all effects of location, dilation and rotation separately, as well as the difference in levels of intensity, that enjoys a great reduction of dimensionality. The problem with using all intensities (apart from high dimensionality) is that such approaches cannot identify specific cells in an ensemble that may be of particular interest to the forecaster. The fact that from nowcast to nowcast we may have a different number of cells moving across the area in comparison with the observed, dictates that a method of verification should penalize a forecast on a cell-by-cell basis.

Several verification tasks naturally arise, including the development and implementation of cell identification algorithms, which we discuss first.

### Cell identification within nowcasting ensemble members

We first clarify how we may identify corresponding cells between the observed truth and forecasted ensemble members (or between any two ensemble members). This problem has not been addressed before in the hydrometeorological literature, and provides researchers with a novel approach to verification as well as forecasting. Assume that each ensemble member has  $N_i > 0$ , identifiable cells (clusters of intensities), where  $i = 1, 2, \dots, M$ , denotes the ensemble member, such that in a grid of  $K \times L$  pixels, we could have  $N_i$  taking values from 0 (no precipitation) to  $\frac{KL}{4}$ , assuming  $K$  and  $L$  even. We will consider clusters of four or more pixels as valid cells. Also notice that we do not consider the case of forecasts with no cells.

To provide a mathematical model for cells in each ensemble member, for each one of the  $N_i$  identifiable cells, we first obtain a centroid on the grid, denoted by  $(c_x^{ij}, c_y^{ij})$ ,  $j = 1, 2, \dots, N_i$ , the  $j$ th cell within the  $i$ th ensemble member, and  $x$  and  $y$  the horizontal and vertical axis. Thus,  $c_x^{ij}$  takes values in  $\{1, 2, \dots, L\}$  while  $c_y^{ij}$  takes values in  $\{1, 2, \dots, K\}$ . For simplicity, we may assume that the lower left corner of the grid is the point  $(0, 0)$ , for all  $i$  and  $j$ . Once a centroid has been identified, a radial is drawn that starts at the centroid and crosses the cell's edge at some point(s). We consider a set of fixed angles  $\theta_l \in \Theta = \{\theta_l \in [0, 2\pi) : \theta_l = \frac{2\pi l}{A}, l = 0, 1, 2, \dots, A-1\}$ , with  $A$  taking values from 1 to 360, and we obtain the radial distance from the centroid to the edge along the  $l$ th angle, denoted by  $r_{\theta_l}^{ij}$ . Typically, one would choose a value of  $A \ll 360$  to reduce dimensionality. Clearly, for convex cells we will obtain a complete description of the shape of the cell (only one edge point), while for non-convex cells, we choose  $r_{\theta_l}^{ij}$  to be the largest distance among such distances. This means that concave

areas, or holes, may be included in the cell parameterization, but it ensures that no parts of the cell are ever lost.

Since we want to reduce the dimensionality of the problem, without losing vital information, the intensity of a cell should be represented by a small number of summary statistics of intensity. For example, we choose three values that summarize intensity in a cell, namely the average, the lowest and the highest intensities, denoted by  $c_{\text{avg}}^{ij}$ ,  $c_{\text{min}}^{ij}$  and  $c_{\text{max}}^{ij}$ , respectively.

Thus, each ensemble member  $\mathbf{y}_{T+\tau}^{(i)}$ , is described as a collection of  $N_i$  cells, where each cell is described by an  $(A+5)$ -dimensional vector  $S_{ij} = (c_x^{ij}, c_y^{ij}, c_{\text{min}}^{ij}, c_{\text{max}}^{ij}, c_{\text{avg}}^{ij}, r_{\theta_0}^{ij}, \dots, r_{\theta_{A-1}}^{ij})$ ,  $j = 1, 2, \dots, N_i$  (cell number),  $i = 1, 2, \dots, M$  (ensemble member number), which contains a parameterization of the cell in terms of centroid coordinates, three intensity measures and values of centroid-edge distance for  $A$  different angles. Hence, from the original dimension of the realizations being  $M \times K \times L$  we have a dimensionality of  $d = (A+5) \sum_{i=1}^M N_i$ , where  $N_i \leq \frac{KL}{4}$ . Although at first glance, the upper bound for the new dimension of the problem appears to be large, there are usually only a small number of observed cells per ensemble member, and hence there is a great reduction in dimension using this approach.

Some discussion is in order about the actual selection process of the cells we use to describe each ensemble member. Typically, meteorologists have a strong sense of what intensity level is representative of the weather phenomena of interest, and more precisely, they can provide a threshold value. Intensities below this value indicate precipitation that is not significant in terms of severe weather or heavy precipitation. Values above the threshold are of extreme importance in short-term forecasting, as they may indicate cells of imminent danger to public safety. Therefore, by consideration of the physical process underlying the relationship between the observed radar reflectivity and storm characteristics (precipitation rate, hail size) a suitable threshold can be selected that we will call  $F$ .

Now all values below  $F$  are set to the minimum value observed. A boundary algorithm is then applied to the resulting image, that yields the grid coordinates of each cell. Once we have identified each cell's coordinates on the plane, the aforementioned approach is used to obtain the  $(A+5)$ -dimensional vectors that describe each cell in the ensemble.

### Cell correspondence between ensemble members

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).

Assume that the true image at time  $T + \tau$  contains  $N$  cells described by vectors

$$S_j = (c_x^j, c_y^j, c_{\text{min}}^j, c_{\text{max}}^j, c_{\text{avg}}^j, r_{\theta_0}^j, \dots, r_{\theta_{A-1}}^j), \quad (2)$$

$j = 1, 2, \dots, N$ . Since the ensemble member might have fewer ( $N_i < N$ ) or more ( $N_i > N$ ) cells than the true image, the forecast is inappropriate and should be penalized appropriately. Even in the case of  $N_i = N$ , the prediction maybe be unacceptable depending on the values of the vectors of the cells involved. For a perfect prediction, we should have  $N_i = N$ ,

and  $S_{ij} = S_j$ , for all  $j = 1, 2, \dots, N$ , which 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 undertake, namely, a cell from the truth is always forced to match with a cell from the forecast.

In the context of verification, we know exactly the true number  $N$  of cells in the true image. Thus, a natural estimator of the cell  $S_j$  is the mean vector of all the particular cell vectors from each of the ensemble members,  $\hat{S}_j = \frac{1}{M} \sum_{i=1}^M S_{ij}$ ,  $j = 1, 2, \dots, N$ . The obvious drawback with these estimators is that, depending on how the cell indices are attached to a cell, it is possible to compute the wrong cells even in the perfect prediction case. Hence, indices are assigned to cells according to how close, in some sense, the cells are to each other. That is, we order the systems in the ensemble members depending on their proximity to the cells in the true image. For example, taking only location in mind, 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  $S_{12}$ , provided that  $(c_x^1 - c_x^{11})^2 + (c_y^1 - c_y^{11})^2 < (c_x^1 - c_x^{12})^2 + (c_y^1 - c_y^{12})^2$ .

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; Micheas and Dey, 2005; Micheas et al., 2006). To accomplish this, the coordinates of the cells are rewritten using complex coordinates, by considering the grid as being the first quadrant of the complex plane. Thus, since any cell's boundary coordinates can be described by an  $A \times 2$  matrix of  $x$ - $y$  coordinates  $(x_{\theta_l}^j, y_{\theta_l}^j)$ ,  $l = 0, 1, \dots, A-1$ , where  $x_{\theta_l}^j = c_x^j + r_{\theta_l}^j \sin \theta_l$ ,  $y_{\theta_l}^j = c_y^j + r_{\theta_l}^j \cos \theta_l$ , the  $x$  and  $y$  grid coordinates of the boundary point on the axis of angle  $\theta_l$ , of the  $j$ th cell in the true image, and  $x_{\theta_l}^{kj} = c_x^{kj} + r_{\theta_l}^{kj} \sin \theta_l$ ,  $y_{\theta_l}^{kj} = c_y^{kj} + r_{\theta_l}^{kj} \cos \theta_l$ , the corresponding grid coordinates for the  $j$ th cell in the  $k$ th ensemble member, then the  $j$ th true cell is described by the vector of complex numbers

$$\mathbf{z}^j = (x_{\theta_0}^j + iy_{\theta_0}^j, \dots, x_{\theta_{A-1}}^j + iy_{\theta_{A-1}}^j), \quad (3)$$

and similarly for the  $j$ th cell in the  $k$ th ensemble member we define

$$\mathbf{z}^{kj} = (x_{\theta_0}^{kj} + iy_{\theta_0}^{kj}, \dots, x_{\theta_{A-1}}^{kj} + iy_{\theta_{A-1}}^{kj}). \quad (4)$$

We refer to  $\mathbf{z}^j$  and  $\mathbf{z}^{kj}$  as the icons of the storm systems. We define  $\mathbf{z}_H^j = \mathbf{H}\mathbf{z}^j$ , the "Helmertized" landmarks of  $\mathbf{z}^j$ , where  $\mathbf{H}$  stands for the last  $(k-1)$ -rows of the Helmert matrix, a special matrix of independent contrasts (see Dryden and Mardia (1998, p. 34)) that help remove location from  $\mathbf{z}^j$ . These rows of the matrix form an orthonormal basis. In order to remove scaling effects, we simply take the standardized version of  $\mathbf{z}_H^j$ , i.e., we define  $\mathbf{z}_s^j = \frac{\mathbf{z}_H^j}{\|\mathbf{z}_H^j\|}$ , where  $\|\mathbf{z}_H^j\|^2 = \mathbf{z}_H^{j*} \mathbf{z}_H^j$ , and  $\mathbf{z}_H^{j*}$  denotes the conjugate transpose of  $\mathbf{z}_H^j$ . We call  $\mathbf{z}_s^j$  the pre-shape of the shape  $\mathbf{z}^j$ . The pre-shape can be thought of as the standardized shape of the cell, in terms of both location (the average of the landmarks is zero) and scale (size is one).

Hence, let  $\mathbf{z}_s^j$  and  $\mathbf{z}_s^{kj}$  be the corresponding pre-shapes of  $\mathbf{z}^j$  and  $\mathbf{z}^{kj}$ . Then we define an ordering of pre-shapes (cells)

in each ensemble member. Within the  $k$ th ensemble member we will use the  $j$ th cell to estimate the true  $j$ th cell when

$$\mathbf{z}_s^{kj} = \arg \min_{\mathbf{z}_s^{kl}} d(\mathbf{z}_s^j, \mathbf{z}_s^{kl}) \quad (5)$$

for  $l = 1, 2, \dots, N_j$  and where  $d(\mathbf{z}_s^j, \mathbf{z}_s^{kl}) = (\mathbf{z}_s^j - \mathbf{z}_s^{kl})^* (\mathbf{z}_s^j - \mathbf{z}_s^{kl})$  and  $\arg \min$  denotes the argument that minimizes the distance. Notice that for the ordering of the cells in each ensemble member we only take into consideration shape information and not intensity this way. We can easily incorporate intensity using distances between the intensity measures  $c_{\min}^j$ ,  $c_{\max}^j$ , and  $c_{\text{avg}}^j$ , but as far as identifying the corresponding cells between truth and forecast, intensity is not essential. Thus, when it comes to application of the methods, the first task is to apply the cell identification approach above to obtain the proper pairing between systems in realizations and corresponding systems in the true images.

### Shape-analysis techniques in ensemble member verification

Our goal is to extract shape and intensity information from each cell in the true image and its corresponding estimator from the  $k$ th ensemble member, and then penalize the now-cast using these measurements. In essence, we will estimate the amount of the error associated with dilation, rotation and translation of a forecast cell to a cell from the truth. To accomplish this, we first assume that every cell  $S_j$  (or  $\mathbf{z}^j$ ) in the true image, is predicted by the  $j$ th cell from the  $k$ th ensemble member, namely  $S_{kj}$  (or  $\mathbf{z}^{kj}$ ),  $j = 1, 2, \dots, N$ ,  $k = 1, 2, \dots, M$ . In the case where a forecast contains fewer cells than the truth then a cell might be used more than once to estimate a truth cell, and similarly in the opposite case. We treat the problem of estimating intensity and that of shape separately.

In order to capture shape variation (translation, dilation and rotation) we employ a "Procrustes"<sup>1</sup> approach. Dryden and Mardia (1998) provide a good overview of Procrustes methods.

Suppose that there exist a number of dilation parameters  $r_{jk} > 0$ , rotation parameters  $\varphi_{jk} > 0$ , and translation parameter complex vectors  $\mathbf{b}_{jk} = (b_{jk0}, \dots, b_{jkA-1})$ , (often all the coordinates of  $\mathbf{b}_{jk}$  are equal to a single complex parameter  $b_{jk}^0$ ), such that we may dilate, rotate and translate a given storm system from some ensemble member and obtain (recreate) the corresponding true storm system. The transformations required to match a storm in an ensemble member to an actual storm, provide the verification measure. That is,  $\mathbf{z}^j = r_{jk} e^{i\varphi_{jk}} \mathbf{z}^{kj} + \mathbf{b}_{jk} + \varepsilon_j$ , or

<sup>1</sup> In Greek mythology, Damastes (which means 'tamer' in Greek), was a robber who operated on the road from Eleusis to Athens. He used to go by the name Procrustes, and he would offer travellers a bed to rest and sleep. Once asleep, he would fit them to the bed by stretching them if they were too short, or cutting off their limbs if too tall. In our context, we can think of the bed as the shape of an object and the traveller as the shape of the object that needs to be translated, dilated and rotated in order for it to fit as closely as possible to the bed.

$$\begin{aligned} x_{\theta_a}^j + iy_{\theta_a}^j &= r_{jk} e^{i\varphi_{jk}} (x_{\theta_a}^{kj} + iy_{\theta_a}^{kj}) + b_{jka} + \varepsilon_{ja}, \\ a &= 0, 1, \dots, A-1, j = 1, 2, \dots, N, k = 1, 2, \dots, M, \end{aligned} \quad (6)$$

where  $\varepsilon_{ja}$  is an error term from (typically) the complex normal distribution, with mean zero, and variance  $\sigma_j^2 = \sigma_{j1}^2 + \sigma_{j2}^2 i$ . Thus, there are  $N$  complex regression equations that can be used to estimate the parameters of interest for each one of the  $N$  true systems. The least square estimators of  $r_{jk}$ ,  $\varphi_{jk}$ , and  $\mathbf{b}_{jk}$ , can be denoted by  $\widehat{r}_{jk}$ ,  $\widehat{\varphi}_{jk}$ , and  $\widehat{\mathbf{b}}_{jk}$ . In a statistical shape analysis context, Eq. (6) is the regression equation used in Procrustes Analysis, where the two shapes are matched through similarity transformations, and the differences between  $\mathbf{z}^j$  and the fitted  $\mathbf{z}_k^j$ , indicate the magnitude of the difference in shape between  $\mathbf{z}^j$  and  $\mathbf{z}^{kj}$ . The estimator

$$\widehat{\mathbf{z}}_k^j = \widehat{\mathbf{b}}_{jk} + \widehat{r}_{jk} e^{i\widehat{\varphi}_{jk}} \mathbf{z}^{kj}, \quad (7)$$

is often referred to as the Full Procrustes fit (superimposition) of  $\mathbf{z}^{kj}$  onto  $\mathbf{z}^j$ . Now it can be shown that

$$\begin{aligned} \widehat{r}_{jk} &= \frac{|(\mathbf{z}_c^{kj})^* \mathbf{z}_c^j|}{(\mathbf{z}_c^{kj})^* \mathbf{z}_c^{kj}}, \\ \widehat{\varphi}_{jk} &= \arg((\mathbf{z}_c^{kj})^* \mathbf{z}_c^j), \\ \widehat{\mathbf{b}}_{jk} &= \overline{\mathbf{z}^j} - \widehat{r}_{jk} e^{i\widehat{\varphi}_{jk}} \overline{\mathbf{z}^{kj}}, \text{ or} \\ \widehat{\mathbf{b}}_{jk}^0 &= \frac{1}{A-1} \sum_{a=0}^{A-1} (x_{\theta_a}^{kj} + iy_{\theta_a}^{kj}) - \frac{\widehat{r}_{jk} e^{i\widehat{\varphi}_{jk}}}{A-1} \sum_{a=0}^{A-1} (x_{\theta_a}^j + iy_{\theta_a}^j), \end{aligned} \quad (8)$$

where

$$\begin{aligned} \overline{\mathbf{z}^{kj}} &= \frac{1}{A-1} \sum_{a=0}^{A-1} (x_{\theta_a}^{kj} + iy_{\theta_a}^{kj}) \mathbf{1}, \\ \overline{\mathbf{z}^j} &= \frac{1}{A-1} \sum_{a=0}^{A-1} (x_{\theta_a}^j + iy_{\theta_a}^j) \mathbf{1}, \end{aligned} \quad (9)$$

are the natural centroids of each cell,  $\mathbf{1}^T = [1, \dots, 1]$ , and  $\mathbf{z}_c^{kj} = \mathbf{z}^{kj} - \overline{\mathbf{z}^{kj}}$ ,  $\mathbf{z}_c^j = \mathbf{z}^j - \overline{\mathbf{z}^j}$ , the centered cells. Notice that the Procrustes residual sum of squares (RSS) between  $\mathbf{z}^j$  and  $\mathbf{z}^{kj}$  (i.e., when estimating the  $j$ th true cell with the  $j$ th cell from the  $k$ th ensemble member), is defined as

$$\text{RSS}_{jk} = (\mathbf{z}^j - \widehat{\mathbf{z}}_k^j)^* (\mathbf{z}^j - \widehat{\mathbf{z}}_k^j). \quad (10)$$

Now, if the forecast cell fits perfectly the corresponding true cell (i.e.,  $\widehat{r}_{jk} = 1$ ,  $\widehat{\varphi}_{jk} = 0$ , and  $\widehat{\mathbf{b}}_{jk} = 0$ , hence  $\mathbf{z}^j = \mathbf{z}^{kj}$ ), then no mistake was made in estimating this cell by the forecast and thus the Procrustes RSS is zero as expected. The overall Procrustes RSS for the  $k$ th ensemble member becomes

$$\begin{aligned} \text{RSS}_k &= \sum_{j=1}^N (\mathbf{z}^j - \widehat{\mathbf{z}}_k^j)^* (\mathbf{z}^j - \widehat{\mathbf{z}}_k^j) \\ &= \sum_{j=1}^N (\mathbf{z}^j - \widehat{\mathbf{b}}_{jk} - \widehat{r}_{jk} e^{i\widehat{\varphi}_{jk}} \mathbf{z}^{kj})^* (\mathbf{z}^j - \widehat{\mathbf{b}}_{jk} \\ &\quad - \widehat{r}_{jk} e^{i\widehat{\varphi}_{jk}} \mathbf{z}^{kj}). \end{aligned} \quad (11)$$

Moreover, an overall estimator of the  $j$ th true cell based on all  $k$  realizations (i.e., an estimator based on the ensemble) is given by

$$\widetilde{\mathbf{z}}^j = \frac{1}{M} \sum_{k=1}^M \widehat{\mathbf{z}}_k^j. \quad (12)$$

## Penalty function between ensembles

To obtain an objective penalty function we first penalize forecasts for their deviations from the truth, according to the following quadratic function that accounts for location, dilation and rotation effects (together), as well as intensity effects between cells from the truth and ensembles. The penalty function between the truth and the  $k$ th ensemble member is defined as

$$\begin{aligned} D(\mathbf{z}_{T+\tau}, \mathbf{y}_{T+\tau}^{(k)}) &= \sum_{j=1}^N (\mathbf{z}^j - \widehat{r}_{jk} e^{i\widehat{\varphi}_{jk}} \mathbf{z}^{kj} - \widehat{\mathbf{b}}_{jk})^* (\mathbf{z}^j - \widehat{r}_{jk} e^{i\widehat{\varphi}_{jk}} \mathbf{z}^{kj} - \widehat{\mathbf{b}}_{jk}) \\ &\quad + \sum_{j=1}^N (c_{\text{avg}}^j - c_{\text{avg}}^{kj})^2 + \sum_{j=1}^N (c_{\text{min}}^j - c_{\text{min}}^{kj})^2 \\ &\quad + \sum_{j=1}^N (c_{\text{max}}^j - c_{\text{max}}^{kj})^2 = \text{RSS}_k + \text{SS}_{\text{avg}}^{(k)} + \text{SS}_{\text{min}}^{(k)} + \text{SS}_{\text{max}}^{(k)}, \end{aligned} \quad (13)$$

where the components  $\text{RSS}_k$ ,  $\text{SS}_{\text{avg}}^{(k)}$ ,  $\text{SS}_{\text{min}}^{(k)}$  and  $\text{SS}_{\text{max}}^{(k)}$  penalize differences in shape, average, minimum and maximum intensities, respectively. Several similar penalty functions may be created that utilize shape and intensity information from all forecast realizations.

Now, using the estimators  $\widehat{r}_{jk}$ ,  $\widehat{\varphi}_{jk}$ , and  $\widehat{\mathbf{b}}_{jk}^0$ , from the full Procrustes fit on the  $j$ th cell from the  $k$ th ensemble member to the  $j$ th true cell, we can create a penalty function that penalizes an ensemble member according to the individual errors in location, rotation and dilation. Within the  $k$ th ensemble member define

$$\begin{aligned} \text{SS}_{\text{shape}}^{(k)} &= \text{SS}_{\text{loc}}^{(k)} + \text{SS}_{\text{rot}}^{(k)} + \text{SS}_{\text{scale}}^{(k)} \\ &= \sum_{j=1}^N \widehat{\mathbf{b}}_{jk}^0{}^2 + \sum_{j=1}^N \widehat{\varphi}_{jk}^2 + \sum_{j=1}^N \widehat{r}_{jk}^2, \end{aligned} \quad (14)$$

where  $\text{SS}_{\text{loc}}^{(k)}$ ,  $\text{SS}_{\text{rot}}^{(k)}$  and  $\text{SS}_{\text{scale}}^{(k)}$  denote the differences (over all cells within the ensemble member) of location, rotation and dilation, respectively. Then  $\text{MSS}_{\text{shape}}^{(k)} = \frac{1}{N} \text{SS}_{\text{shape}}^{(k)}$ , denotes the average effects of location, rotation and dilation for the  $k$ th ensemble member, and hence a new penalty function is defined by

$$\begin{aligned} D^*(\mathbf{z}_{T+\tau}, \mathbf{y}_{T+\tau}^{(k)}) &= \text{MSS}_{\text{shape}}^{(k)} + \text{SS}_{\text{avg}}^{(k)} + \text{SS}_{\text{min}}^{(k)} + \text{SS}_{\text{max}}^{(k)} \\ &= \frac{1}{N} \text{SS}_{\text{loc}}^{(k)} + \frac{1}{N} \text{SS}_{\text{rot}}^{(k)} + \frac{1}{N} \text{SS}_{\text{scale}}^{(k)} + \text{SS}_{\text{avg}}^{(k)} \\ &\quad + \text{SS}_{\text{min}}^{(k)} + \text{SS}_{\text{max}}^{(k)}. \end{aligned} \quad (15)$$

Both penalty functions can be used to assess how well an ensemble member is performing in estimating the truth. The dimensionality reduction is clearly great, but perhaps the most interesting consequence of these measures is that it allows an experimenter to choose the "best" ensemble member forecast among a set of forecasts being compared, based on several characteristics including location, rotation and scale effects and average, minimum or maximum intensities in cells. Eqs. (13) and (15) provide overall measures of the differences between the truth and a forecast. For a perfect forecast,  $D(\mathbf{z}_{T+\tau}, \mathbf{y}_{T+\tau}^{(k)})$  should be zero, by the very definition of the quantities in (13), since in this case  $\text{RSS}_k$ ,  $\text{SS}_{\text{avg}}^{(k)}$ ,  $\text{SS}_{\text{min}}^{(k)}$  and  $\text{SS}_{\text{max}}^{(k)}$ , would all be zero. Since we aim to break down the error due to differences in shape between truth and forecast, we take for each forecast the average

differences across all cells in location, rotation and scale ( $\frac{1}{N}SS_{loc}^{(k)}$ ,  $\frac{1}{N}SS_{rot}^{(k)}$ , and  $\frac{1}{N}SS_{scale}^{(k)}$ ). A perfect forecast in this case would yield values close to zero for  $\frac{1}{N}SS_{loc}^{(k)}$  and  $\frac{1}{N}SS_{rot}^{(k)}$ , and close to one for  $\frac{1}{N}SS_{scale}^{(k)}$ . Hence,  $D^*(\mathbf{z}_{T+\tau}, \mathbf{y}_{T+\tau}^{(k)})$  should be close to one for a perfect forecast.

## Nowcast schemes and data

Each of the nowcasting schemes used in this study works on a different principle. For this reason they are likely to produce nowcasts that vary in their treatment of precipitation fields containing structures of different shape, spatial scale, and organization and that move in more or less regular ways. In each case, for the verification system to be applied, the nowcast scheme must generate forecast products in the form of a pseudo-reflectivity grid.

The Warning Decision Support System-Integrated Information (WDSS-II) has been developed by the National Severe Storms laboratory (NSSL). Now in its second major version, it contains two cell tracking algorithms: The Storm Cell Identification and Tracking (SCIT) system (Johnson et al., 1993) and the K-Means cluster analysis scheme (Lakshmanan et al., 2003). The first of these is a centroid tracking scheme that produces motion vectors for identified cells. The latter uses a variable cell clustering technique to identify motion of single cells or groups dependent upon reflectivity thresholds set by the user. This latter scheme provides forecasts in the form of pseudo-reflectivity images and was therefore suitable for use in this project.

The spectral prognosis (SPROG) nowcasting system was developed at the Australian Bureau of Meteorology Research Centre (BMRC) and uses a spectral decomposition technique to selectively identify and track features of different spatial scales. Smaller features are decayed in forecast time in keeping with the knowledge of persistence and predictability of rainfall structures in time (Seed, 2003). The resulting forecast spectral components are then recombined into a final nowcast product in the form of a reflectivity image. Notably, the efficiency of this scheme allows it to be run operationally in an ensemble member stochastic format as is currently being tested in the Met Office (Pierce et al., 2005). For this study this was not done.

The University of Missouri Bayesian Hierarchical Model (BHM) uses a discretized integrodifference equation approach to forecast pixel-by-pixel motion and intensity of reflectivity fields (Xu et al., 2005; Fox and Wikle, 2005). The Bayesian nature of the model allows one to obtain prediction distributions, and thus, ensemble members. For this study we extracted 20 ensemble members from the posterior distribution sample. The BHM used for a demonstration in this study runs on a  $32 \times 32$  pixel grid because it is inefficient and currently would not produce a viable ensemble member in a reasonable time if run on a larger grid. We know that this is not adequate as an operational forecast scheme and are currently working on an improved version that will provide nowcasts on a larger grid. This paper deals with the verification issue, for which production of ensemble member forecasts was required. The effectiveness of the verification is not impacted by the size of the grid and would be the same for a larger grid and number of cells. Description of the BHM, including its limitations, can be found in Xu et al. (2005) and Fox and Wikle (2005).

The verification technique explored in this paper was developed to work with ensemble member forecast systems, but is not restricted to these and can be used for any spatial forecast field whether it is an ensemble member or not. The concept is not only to provide objective statistical measures of overall forecast skill, but also to identify individual ensemble member performance such that different treatments of the data or configurations of the forecast model can be assessed. The separation of errors into the various components should provide a tool whereby the cause of the error can be identified and improvements can be made to the nowcast system based on this knowledge. The verification scheme can be run in real time such that assessments of nowcast performance can be made and forecasters could then have confidence in the products or have the ability to choose between different products (even from different systems) depending on relative performance statistics.

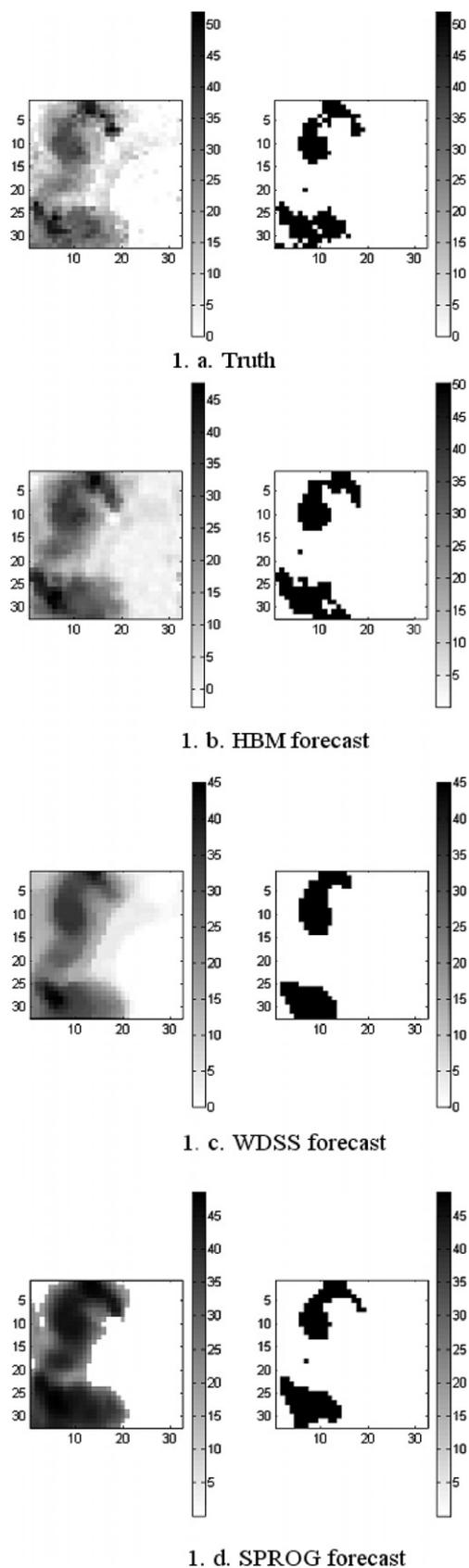
The example case was taken using data from the San Antonio, TX area on 5 July 2002. This day saw a series of intense convective storm cells track over the area resulting in some flash flooding. The individual storms were disorganized and their tracks influenced by the proximity of the coast and complex orography inland. Although the storms were not severe, forecast knowledge of their tracks would have helped in forecasting small catchment response to heavy rainfall.

In order to ensure comparability of the nowcast schemes and, in particular, the performance of the verification, each scheme was run on the same data. In the example case, this includes NEXRAD LEVEL II data reduced to a  $32 \times 32$  grid of 4-km resolution pixels from a mosaic of three radars including Corpus Christi, TX (KCRP), Brownsville, TX (KBRO) and San Antonio, TX (KEWX) to capture the event. Although this resolution does not allow any of the nowcast schemes to perform optimally, it does allow for the BHM scheme to be run in full ensemble mode as a test of the verification scheme's handling of a multirealization product.

## Application of cell identification and verification methods

Initially, we considered nowcasts from the BHM, WDSS, and SPROG models based on observed radar reflectivities at 10-min intervals from 0130 UTC to 0220 UTC. Specifically, we considered nowcasts at 10-min intervals up to 60 min: 0230 UTC to 0320 UTC. All forecasts were based on reflectivity images with dimension 32 by 32 pixels. Note, since the BHM model provides ensemble members (20 in this application), the results presented relate to the posterior mean nowcast, unless stated otherwise.

Fig. 1 shows the truth and forecasts from the three models for a 10-min lead-time nowcast verifying at 0230 UTC. The left side of this figure shows the full reflectivity field for the truth and various nowcasts, and the right side shows the corresponding images after intensities lower than 30 dBZ are removed and the boundary finding algorithm has been applied. The verification methodology then determines the number of "cells" in the truth image, based in this case on 20 angles. For example, the truth image at 0230 UTC is found to contain two cells, with corresponding



**Figure 1** (a) Truth at time 0230 UTC. (b) Posterior mean from the BHM forecast for time 0230 UTC, lead 10 min. (c) WDSS forecast at time 0230 UTC, lead 10 min. (d) SPROG forecast at time 0230 UTC, lead 10 min.

x-, y-coordinates, minimum, average, and maximum intensities:

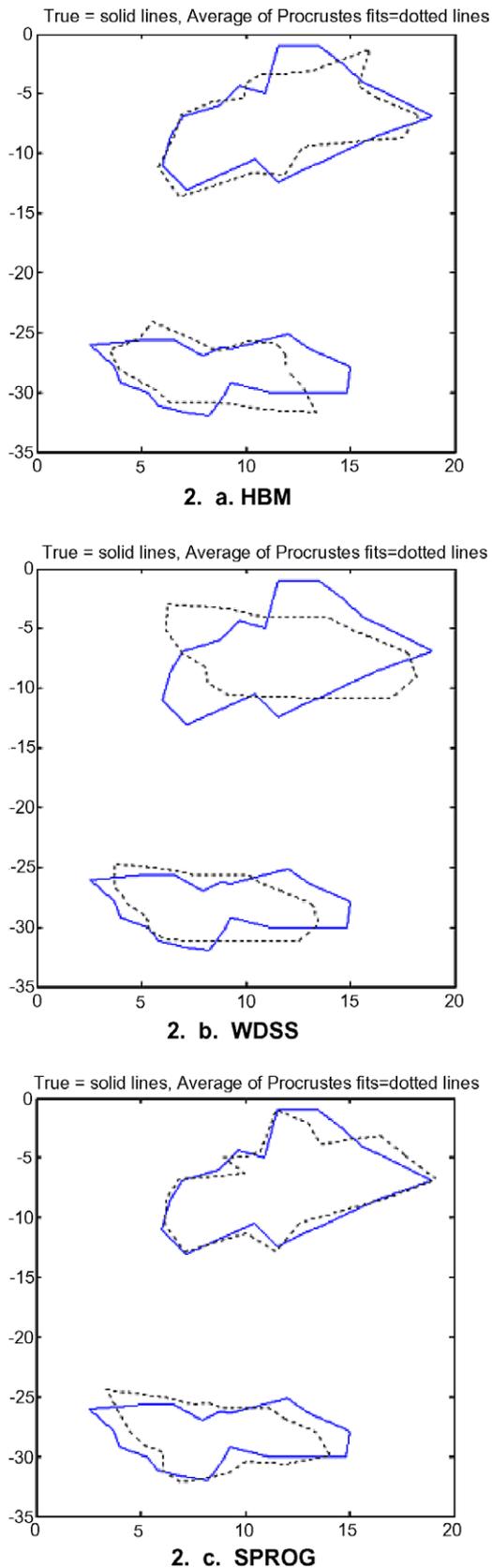
True cell 1 : 8.3, -27.8, 30.4, 36.3, 52.1.

True cell 2 : 11.6, -6.9, 30.0, 37.0, 50.6.

Similarly, cells are also estimated for the corresponding nowcast images. These cells, relative to the truth cells at 0230 UTC, are shown for each nowcast model in Fig. 2. To quantify the relationship between the cells for the truth and nowcasts, we consider verification summaries. Specifically, we consider the total Procrustes residual sum of squares (RSS) and the sum of squares of the intensities ( $SS_{tot}$ ). A summary of these statistics for the nowcasts verifying at 0230 UTC to 0320 UTC (10-min lead time nowcasts ending at the times) is given in Table 1 where only location is used for cell correspondence between truth and nowcast, as discussed in Section “Cell correspondence between ensemble members”. Although the primary purpose here is to demonstrate the methodology, it is interesting to compare the different nowcast methods based on these measures. In terms of the intensities ( $SS_{tot}$ ), up through 40-min lead times, BHM performs best, followed by SPROG and WDSS. For 50- and 60-min lead times, SPROG performs best, followed by BHM and WDSS. Regarding the total Procrustes residual sum of squares, for lead times of 20–50 min, WDSS performs best, followed by SPROG and BHM. At 10-min leads, SPROG is best followed by BHM and WDSS, and at 60 min, BHM is best followed by WDSS and SPROG. Table 2 shows the same RSS statistics but for the case where Procrustes residuals are used to establish correspondence between cells from the truth and nowcasts (as discussed in Section “Shape-analysis techniques in ensemble member verification”). The results are somewhat different, with BHM showing the best performance for medium range lead times.

We now look more closely at a specific verification time, 0340 UTC, for lead times of 10–60 min. Note that these results are based on a threshold intensity of 20 dBZ and 15 angles. This choice, although somewhat subjective, is based on the ability of the cell detection algorithm to identify specific cells in the truth image. The flexibility of choosing different thresholds and number of angles is a strength of the methodology. Table 3 shows the summary statistics for this case when only location is used for cell correspondence. In this example, SPROG typically performs best at shorter lead times, followed by BHM and WDSS. There seems to be some indication that BHM tends to perform better at longer lead times. Table 4 shows the comparable verification results when the Procrustes residuals are used for cell correspondence. Using this measure, the BHM performs the worst at the early lead times while the WDSS is best, followed by SPROG. At the longer lead times, SPROG is best, followed by BHM and WDSS.

Recall that a strength of the Procrustes verification approach is that the overall sum of squares statistics can be further decomposed relative to intensity, shape and scale. Tables 5 and 6 show sum of squares for the minimum, average, and maximum intensity, as well as scale and location for the case when cell correspondence is based on location only, and Procrustes residuals, respectively. These results suggest that each nowcast system has its



**Figure 2** (a) Comparing truth to BHM posterior mean ensemble, time 0230 UTC. (b) Comparing truth to WDSS ensemble member, time 0230 UTC. (c) Comparing truth to SPROG ensemble member, time 0230 UTC.

strengths and weaknesses. Clearly, more case studies should be considered before one can say with confidence that one method is performing consistently well or poor in any one area. However, the Procrustes-based statistics provide a promising suite of measures for such a comparison.

As demonstrated by the above results it is possible to find an over all error for each nowcast ensemble member and time-step. More interestingly, it is possible to express this error through contributions due to location ( $SS_{loc}$ ), dilation ( $SS_{scale}$ ), and intensity ( $SS_{max}$ ,  $SS_{min}$ ,  $SS_{avg}$ ). Not shown in the above tables are errors due to rotation ( $SS_{rot}$ ), as in this case all such errors were negligible. Such a decomposition allows one to determine how each model is performing in a variety of ways. For example, from Tables 5 and 6 it is seen that WDSS outperforms both SPROG and BHM in terms of intensity, but is itself clearly outperformed in terms of location, particularly at longer lead-times. This is not surprising given that SPROG and BHM have in-built tendencies to smooth high intensities, whereas WDSS retains current patterns. However, SPROG and BHM seem to account well for the combination of advection and propagation of precipitation, whereas WDSS has a simpler advection function. There is no evidence that the superior location forecasts come at the expense of dilation, i.e., that location is improved simply by forecasting cell growth, as all the dilation errors are similar.

A further comment is in order about the robustness of the methodology. Sensitivity analysis was performed with respect to both the choice of threshold and the number of angles used. Increasing the number of angles, tends to give smoother estimates of cells, but still in some cases they can be more star shaped. Using a very large number defeats the goal of dimension reduction of the problem. Meanwhile, using only a few, say five, gives crude estimates of the cells in the forecasts. We found that 10–20 angles seem to work quite well in most cases. The choice of threshold is much more subjective. It is expected that the expert meteorologists have a well-developed knowledge of the intensity range that they need to observe, in order to identify a severe storm system. Nonetheless, we investigated both high and lower intensities.

In general, the penalty functions are of similar magnitudes for small changes in threshold and number of angles. We reached the same results, when it came to small or large changes in the number of angles. However, this was not the case with the threshold value. Greatly increasing the threshold, sometimes yields no identifiable cells in some forecasts. Very low thresholds tend to yield very large cells. Changes such as these, greatly affected both Procrustes as well as the intensity sum of squares. Since this is a subjective choice, we advise a careful selection of threshold that satisfies the needs of the experimenter and the analysis.

There is no reason why one particular forecast should produce the best verification when the measures assess different attributes of the forecast. Indeed one major advantage of this verification procedure is that it can identify which forecast (or ensemble member) produces the most accurate representation of the various characteristics of the storm. For ensembles that are made up of members

**Table 1** Results per time period for all cells, using location only for cell correspondence between cells from truth and forecast

Time period	SS <sub>tot</sub>			RSS			# of cells in truth
	WDSS	SPROG	BHM	WDSS	SPROG	BHM	
0230 UTC	25.6253	20.2261	19.459	96.5847	37.9386	48.5222	2
0240 UTC	51.4803	45.3927	43.5694	75.3826	92.3019	139.9885	3
0250 UTC	78.85	66.0539	61.7467	52.929	68.2372	69.6679	3
0300 UTC	93.5189	74.5149	72.0801	33.2543	44.0658	39.3487	3
0310 UTC	98.4979	69.8259	76.5728	51.9803	60.6706	68.8061	4
0320 UTC	111.2577	74.439	88.7457	34.6858	51.8965	29.8134	4

Ten-minute lead time nowcasts ending at the times.

**Table 2** Results per time period for all cells, using Procrustes residuals for cell correspondence between cells from truth and forecast

Time period	RSS			# of cells in truth
	WDSS	SPROG	BHM	
0230 UTC	113.4805	36.5387	71.2615	2
0240 UTC	68.9181	75.3691	89.4667	3
0250 UTC	69.5937	75.2025	61.2549	3
0300 UTC	60.4894	53.9560	44.9139	3
0310 UTC	69.4926	59.6525	47.7261	4
0320 UTC	50.3868	45.7965	54.3680	4

Ten-minute lead time nowcasts ending at the times.

**Table 3** Verification results for the 0340 UTC time period, using location only for cell correspondence between cells from truth and forecast

Verifying 0340 UTC lead time (min)	SS <sub>tot</sub>			RSS		
	WDSS	SPROG	BHM	WDSS	SPROG	BHM
10	32.4169	30.3376	32.9085	42.8283	41.9378	39.1913
20	49.1405	36.3965	41.4114	36.2465	50.5711	48.9398
30	62.829	44.9979	51.2791	49.8125	46.7081	47.9163
40	90.9051	51.7074	56.2427	65.1625	44.8942	54.5747
50	100.567	63.9995	59.6349	93.3849	59.5873	58.2467
60	118.645	64.8223	62.6202	525.2878	59.0587	89.6373

**Table 4** Verification results for the 0340 UTC time period, using Procrustes residuals for cell correspondence between cells from truth and forecast

Verifying 0340 UTC lead in time (min)	RSS		
	WDSS	SPROG	BHM
10	30.4847	34.6459	35.3245
20	27.3061	29.2164	30.5256
30	25.0882	35.4505	41.6773
40	45.1607	33.7056	40.7563
50	70.4598	46.1167	51.2568
60	525.2878	58.6487	66.3705

generated by different models (e.g., models containing different convective parameterization schemes or, in the nowcasting realm, different advection/dispersion formula-

tions) the decomposition of the error may allow a determination of which model is handling the situation best. This would best be achieved in a post-event study rather than

**Table 5** Verification results for the 0340 UTC time period for individual components, using location only for cell correspondence between cells from truth and forecast

Verifying 0340 UTC lead time (min)	Breaking down intensity and shape to specific elements					
		$SS_{min}$	$SS_{avg}$	$SS_{max}$	$\frac{1}{N} SS_{scale}$	$\frac{1}{N} SS_{loc}$
10	WDSS	0.48	49.17	344.99	0.49	61.118
	SPROG	0.46	26.42	396.15	0.83	4.17
	BHM	0.16	18.12	429.01	0.83	10.84
20	WDSS	0.21	37.3	703.66	0.77	26.82
	SPROG	0.17	71.78	805.37	1.36	10.43
	BHM	1.44	64.39	810.4	0.84	8.2
30	WDSS	0.5	82.1	566.9	0.5	97.2
	SPROG	1.0	104.6	991.6	1.4	9.5
	BHM	0.9	106.9	1074.5	0.9	5.0
40	WDSS	1.3	39.8	345.2	0.5	106.9
	SPROG	1.0	149.7	1177.5	0.9	5.7
	BHM	1.0	147.2	1211.4	0.9	5.3
50	WDSS	1.0	54.2	373.2	0.5	115.4
	SPROG	1.1	212.0	1602.7	2.4	86.1
	BHM	1.0	204.7	1573.5	1.8	93.9
60	WDSS	0.9	25.4	185.6	0.2	169.2
	SPROG	0.5	172.0	1401.8	0.7	36.6
	BHM	0.9	155.3	1258.6	0.4	84.0

as a real-time application, and could then be used to assess model applicability to storm type in an analogous way to that described in Grams et al. (2006). However, the verification scheme is efficient enough to provide a variety of real-time verification measures. The 'best' forecast can still be identified from the total error, but this may depend on the judgment of the forecaster as to what particular aspect of the storm is the most important and would require a suitable means of weighting the various components of the total error appropriately.

### Concluding remarks

This paper introduces a new approach to assessing errors in forecast reflectivity fields that is not limited to the short-period forecasts illustrated herein. One can also envision a number of ways in which the verifications can be refined, either systematically, by an individual forecaster, or on an event by event basis. One refinement may be to apply weighting factors to each error term prior to the summation. If one wished to stress storm location as the most important forecast parameter then one could add weight to this term, increasing the penalty for a poor prediction of storm location. Alternatively one may increase the weighting in  $SS_{max}$  to reward forecasts that retain high-

intensity features that may accurately represent areas of severe weather, or the error in shape may reflect the ability of the forecast to maintain or predict structures such as bows and hook-echoes that are indicative of threats of downburst or tornado.

As can be seen in the purely mathematical treatment presented herein, it is apparent that the different error terms can result in numerical values that are orders of magnitude apart. Although this indicates the relative values of the errors it makes intercomparison of the different error terms difficult and it may be desirable to normalize the error values of each source. This could enable one to compare the different components and identify from one model to the next, or one ensemble member to the next, how well each is dealing with the various morphological factors in the storm's behavior.

Finally, there are many avenues of investigation and potential refinements that can be made to the verification methodology. It is planned that further research will be undertaken using the nowcast schemes mentioned here and others, on a number of cases representing a variety of meteorological scenarios. Furthermore, we plan to implement the verification as the standard in our own nowcast scheme developments as we consider detailed performance evaluations of the BHM that reveal how it handles the different aspects of precipitation forecasting

**Table 6** Verification results for the 0340 UTC time period for individual components, using Procrustes residuals for cell correspondence between cells from truth and forecast

Verifying 0340 UTC lead time (min)	Breaking down intensity and shape to specific elements					
		$SS_{\min}$	$SS_{\text{avg}}$	$SS_{\max}$	$\frac{1}{N} SS_{\text{scale}}$	$\frac{1}{N} SS_{\text{loc}}$
10	WDSS	0.64	34.65	483.36	0.38	112.83
	SPROG	0.07	39.16	455.82	0.36	113.61
	BHM	0.08	52.79	477.52	0.38	112.52
20	WDSS	0.26	66.72	874.72	0.62	95.49
	SPROG	1.11	57.88	653.08	0.34	112.78
	BHM	1.09	92.75	765.02	0.36	109.17
30	WDSS	0.83	48.92	575.37	0.44	117.99
	SPROG	1.11	86.66	808.88	0.36	112.89
	BHM	0.99	89.77	987.43	0.73	21.27
40	WDSS	0.5	28.8	479.7	0.5	127.6
	SPROG	1.1	129.6	1004.8	0.4	112.5
	BHM	1.1	118.8	960.7	0.4	108.4
50	WDSS	0.6	25.6	363.1	0.5	128.4
	SPROG	0.9	124.9	1111.7	0.4	109.2
	BHM	1.0	142.1	1171.1	0.4	108.7
60	WDSS	0.9	25.4	185.6	0.2	169.2
	SPROG	1.0	149.6	1246.4	0.5	54.4
	BHM	1.1	96.5	1028.1	0.3	113.3

involved in a good prediction of location structure and intensity.

## Acknowledgements

This research was made possible by National Science Foundation Grant ATM-0434213. We are also grateful to two referees for their constructive comments and suggestions.

## References

- Bookstein, F.L., 1991. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge.
- Casati, B., Ross, G., Stephenson, D.B., 2005. A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorol. Appl.* 11, 141–154.
- 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.* 34, 1772–1784.
- Dryden, I.L., Mardia, K.V., 1998. *Statistical Shape Analysis*. John Wiley & Sons.
- Ebert, E.E., McBride, K., 2000. Verification of precipitation in weather systems: determination of systematic errors. *J. Hydrol.* 239, 179–202.
- Ebert, E.E., Wilson, L.J., Brown, B.G., Nurmi, P., Brooks, H.E., Bally, J., Jaeneke, M., 2004. Verification of nowcasts from the WWRP Sydney 2000 Forecast Demonstration Project. *Wea. Forecasting* 19, 73–96.
- Fox, N.I., Wikle, C.K., 2005. A Bayesian precipitation nowcast scheme. *Wea. Forecasting* 20, 264–275.
- Grams, J.S., Gallus, W.A., Koch, S.E., Wharton, L.S., Loughe, A., Ebert, E.E., 2006. The use of a modified Ebert-McBride technique to evaluate mesoscale model QPF as a function of convective system morphology during IHOP 2002. *Wea. Forecasting* 21, 288–306.
- Johnson, J.T., MacKeen, P.L., Witt, A., Mitchell, E.D., Stumpf, G.J., Eilts, M.D., Thomas, K.W., 1993. The storm cell identification and tracking algorithm: an enhanced WSR-88D algorithm. *Wea. Forecasting* 13, 263–276.
- Lakshmanan, V., Rabin, R., DeBrunner, V., 2003. Multiscale storm identification and forecast. *Atmos. Res.*, 367–380.
- Micheas, A.C., Dey, D.K., 2003. Shape classification procedures with applications. In: *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. *J. Multivariate Anal.* 92, 257–280.
- Micheas, A.C., Dey, D.K., Mardia, K.V., 2006. Complex elliptical distributions with applications to shape analysis. *J. Statist. Planning Inference* 136, 2961–2982.

- Pierce, C.E., Bowler, N., Seed, A., Jones, A., Jones, D., Moore, R., 2005. Use of a stochastic precipitation nowcast scheme for fluvial flood forecasting and warning. *Atmos. Sci. Lett.* 6, 78–83.
- Seed, A.W., 2003. A dynamic and spatial scaling approach to advection forecasting. *J. Appl. Meteor.* 42, 381–388.
- Stoyan, D., Stoyan, H., 1994. *Fractals, Random Shapes and Point Fields: Methods of Geometric Statistics*. Wiley, Chichester.
- Xu, K., Winkle, C.K., Fox, N.I., 2005. A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. *J. Am. Stat. Assoc.* 100, 1133–1144.