

Wavelets and Field Forecast Verification

WILLIAM M. BRIGGS

Department of Soil, Crop, and Atmospheric Sciences, Cornell University, Ithaca, New York

RICHARD A. LEVINE

Division of Statistics, University of California at Davis, Davis, California

(Manuscript received 28 March 1996, in final form 1 October 1996)

ABSTRACT

Current field forecast verification measures are inadequate, primarily because they compress the comparison between two complex spatial field processes into one number. Discrete wavelet transforms (DWTs) applied to analysis and contemporaneous forecast fields prove to be an insightful approach to verification problems. DWTs allow both filtering and compact physically interpretable partitioning of fields. These techniques are used to reduce or eliminate noise in the verification process and develop multivariate measures of field forecasting performance that are shown to improve upon existing verification procedures.

1. Introduction

The use of wavelets in data analysis is a rapidly growing field that is only recently being exploited in the atmospheric sciences. The power of wavelet analyses is that frequency *and* time decomposition of data (e.g., time series data) is possible. In contrast, traditional Fourier analysis supplies only frequency decomposition. Meteorologists may be most familiar with wavelets in turbulence studies (e.g., Gao and Li 1993; Hagelberg and Gamage 1994; Katul and Vidakovic 1995), time series analysis (e.g., Lau and Weng 1995; Weng and Lau 1994; Meyers et al. 1993), and other miscellaneous geophysical data analyses (e.g., Chao and Naito 1995; Kumar and Foufoula-Georgiou 1993; Serrano et al. 1992). One of the largest (nonmeteorological) applications of wavelets has been in the field of image analysis (Laurent et al. 1993), to remove noise or compress images for data transmission [see Press et al. (1992) for a good introduction]. Excellent mathematical introductions to the theory of wavelets may be found in Chui et al. (1994) and Daubechies (1992). An introduction with examples catering to geoscientists has been presented by Kumar and Foufoula-Georgiou (1994). We additionally recommend the very readable paper by Vidakovic and Müller (1994), nonthreateningly entitled "Wavelets for Kids."

Wavelets have also found their way into statistics,

particularly in nonparametric regression (e.g., Nason 1994) and nonparametric function and density estimation (e.g., Donoho and Johnstone 1994; Donoho et al. 1995). It is with these recent statistical insights that we hope to demonstrate how wavelet analysis can be used in providing measures to compare images and scores to assist in field forecast verification. The measure of closeness between images is a general problem that is encountered in areas other than meteorology; for example, in the quantification of electronic transmission error, image compression algorithms, or pattern recognition studies. In this article we use the terms image, grid, and field interchangeably.

The primary purpose of field comparison for meteorologists is in forecast diagnostics. Typically, we are presented with an analysis field, which is given as "truth," and a corresponding forecast field. However, analysis fields contain error, or random noise, as a result of the statistical methodology used to create them (Daley 1993). Forecast fields also contain such error. We will give a method that uses wavelets for "cleaning up," in the sense of removing noise, both fields after which common univariate scores can be recalculated. This procedure will reduce the effects of sampling variability in the verification process and result in forecast performance scores that, hopefully, provide more information about the true underlying processes. This idea is analogous to techniques that have been used in the past, for example, spectrally filtering fields (a discussion appears in Van den Dool and Rukhovets 1994) or filtering fields with empirical orthogonal functions (EOFs; e.g., Liv- ezey et al. 1995). Once two fields have been filtered, either to remove noise or to remove data at wavelengths

Corresponding author address: Dr. William M. Briggs, Dept. of Soil, Crop, and Atmospheric Sciences, Cornell University, 1123 Bradford Hall, Ithaca, NY 14853-1901.
E-mail: wmb2@cornell.edu

of no interest, point scores can be recomputed to give insight as to closeness of the filtered fields. Wavelets can also be utilized to filter fields, and new research shows how this may be done in a statistically optimal manner.

Our goal is to rate how “close” the forecast field is to the analysis field. This “closeness” gives us a sense of the predictive power of the forecast model that generated the field forecast. Various objective statistical measures have been developed to measure closeness. Among these are the anomaly correlation coefficient (ACC, Miyakoda et al. 1972), the familiar Pearson product moment correlation, and the root-mean-square error (rmse) and its variants (e.g., normalized rmse, mean absolute deviation, etc.). However, these measures have all been judged inadequate. Murphy and Epstein (1989) and Radok and Brown (1993) detail the interpretation and statistical problems with the ACC. We should also note the ACC cannot even be computed when there is no climatological field available, a nontrivial concern in measuring field closeness. Taylor (1991) shows that other measures in use (rmse and mean-absolute deviation) are also poor, in the sense of giving incomplete information, when used in image comparison.

One of the major difficulties with these traditional image comparison/field closeness measures is that they are single numbers (point or univariate scores). Undoubtedly the application of point scores is popular due to the ease and quickness of use. But the goal of these scores is to compare the closeness of two complex fields. Any one number that attempts to squeeze all the information about intricate spatial and distributional aspects of the closeness will almost certainly be inadequate (Murphy and Winkler 1987; Murphy 1991). In this paper we are not presenting a complete solution to the absolute field forecast verification problem. Instead, we hope to persuade the reader that reliance on single scores is misguided and that some form of a multivalued score is better. Later, we will show that an objective multivariate measure can be developed using wavelets.

A complete distributional description of closeness is unlikely, unless significant simplifying assumptions are made, due to the complexity and dimensionality of the problem (Murphy 1991). We can do better than point scores, however. For example, Hoffman et al. (1995) present a multivariate score based on what they term the distortion representation of forecast error. Their approach “breaks” the score into pieces that explain different aspects of the error for use in a specific application. As we will see, the wavelet approach is similar in that wavelet transformations decompose fields into scales representing different detail levels of the field. We exploit this characteristic to develop intuitive multivariate measures of closeness to compare the forecast and analysis field at the multiple scales of detail.

This paper thus has two goals. The first is to show how wavelet techniques can reduce or eliminate noise in the verification process. Second, multivariate mea-

asures of field closeness are developed via wavelet transformations. We claim these multivariate scores allow for a more thorough evaluation of forecast performance than the currently utilized point scores. To this end, the paper unfolds as follows. In section 2 we discuss the traditional approach of field forecast verification. We provide further impetus for improving closeness measures via wavelets. To describe the application of wavelets to forecast verification, we need to familiarize the reader with wavelet mathematics. Therefore, we first delve, briefly, into the theory of discrete wavelet transforms in section 3. We also present recent advances in wavelet statistics that will be used in the latter part of the paper. We emphasize that these statistical methods are applicable in any data analytic setting. In section 4 we show how wavelets can be used in statistical image comparison. We also carry out simulations using actual and synthetic (but realistic) forecast fields and (an actual) analysis field to illustrate the filtering of univariate closeness measures. We then construct multivariate closeness measures based on wavelet decomposition. Section 5 contains concluding remarks.

2. Traditional statistical approaches

There are two problems in field forecast verification. The first is to quantify or define what is meant by verification and the second is to develop and interpret scores based on this quantification. We first outline the goal of absolute field forecast verification and then examine what common point scores are actually measuring in the field forecast framework.

The use of a single measure (point score) in evaluating any forecast, whether for a field or for a univariate variable (e.g., probability of precipitation forecast), is seriously deficient because no single measure can capture all relevant information regarding quality. Murphy and Winkler (1987) and Murphy (1991) describe this consequence as the problem of absolute forecast verification. They show that any complete scoring system must include enough information to allow for reconstruction of the joint distribution of the forecast and analysis values. The familiar example of (point location) probability of precipitation (PoP) forecasts highlights this fact. In scoring PoP forecasts we typically examine the Brier score (see, e.g., Wilks 1995). This measure alone is insufficient for fully explaining the overall performance of the forecasts. Likewise, we cannot hope to reconstruct the complex spatial information in field forecasts from a point score. We need a more nearly complete measure of the joint distribution of the (usually nonprobabilistic) forecast and analysis fields. However, to accomplish this task, we must find a way to overcome the complexity and overwhelming dimensionality of the problem.

Following a similar definition advanced by Murphy (1991) for point forecasts, we define the dimensionality of forecast verification to be the minimum number of

specifications needed to fully describe the joint distribution of forecasts and observations. Numerically, dimensionality is defined as the number of distinct forecasts times the number of distinct observations minus one. The dimensionality of the PoP verification example, when there are 11 distinct probabilities allowed (e.g., 0.0, 0.1, 0.2, etc.), is 21. For field forecasts, this number is essentially the number of grid points squared. The dimensionality of forecast verification for even a small grid is huge (since fields contain continuous variables, it may be possible to introduce parametric models for the variables under consideration to reduce the dimensionality).

As mentioned in section 1, common point scores used in field forecast verification are the rmse and ACC (resembling ordinary sample linear correlation). In using these (or any) scores it is important to understand what they are attempting to measure.

It is convenient to present the rmse and ACC in matrix norm form. In particular, we define the two scores as

$$\text{rmse} = \frac{\|\mathbf{A} - \mathbf{F}\|_f}{n^{1/2}}$$

and

$$\text{ACC} = \frac{\text{Tr}(\mathbf{A} - \mathbf{C})^T(\mathbf{F} - \mathbf{C})}{\|\mathbf{A} - \mathbf{C}\|_f \|\mathbf{F} - \mathbf{C}\|_f}, \quad (2.1)$$

where \mathbf{A} is the analysis field, \mathbf{F} is the forecast field, \mathbf{C} is the climate field, $\|\cdot\|_f$ indicates the Frobenius (or Euclidean) norms of the argument, respectively, Tr denotes the trace of the matrix and is defined as the sum of the diagonal elements of the corresponding matrix, and n is the number of matrix elements. Note that ACC can be used only when \mathbf{C} is meaningfully defined (as a replacement, ordinary correlation might be used in the absence of \mathbf{C}). Writing the scores in matrix form emphasizes the fact that we are examining a relationship between two fields not simply a collection of independent observations; that is, the values at each grid point will almost certainly not be statistically independent. In practice, however, the arguments are treated as vectors not as matrices and essential information is lost (Briggs and Levine 1996; the authors will argue in subsequent work that the more common L_2 matrix norm should be used instead of the Frobenius, or Euclidean).

It is important to recognize that these two scores are measuring different aspects of closeness. ACC is a good measure of linear associativity between two fields but ignores bias. The rmse ignores linear associativity but is a good measure of accuracy (Murphy 1995). Additionally, it is possible to express ACC as a component of MSE (Murphy and Epstein 1989).

If two sets of random variables are independent and identically distributed (iid), then the exact form of the sampling distributions of the rmse is known (that of the ACC is not known). We can then use these sampling distributions to assess the statistical significance for any

particular score received between two sets of data. The problem with field forecasts is that the iid assumption is violated because meteorological fields have complex covariance structures between different spatial grid points (e.g., Perrie and Toulany 1989). Adjacent, and even far-removed, grid points are highly correlated with one another. Therefore, in principal, we must treat the field forecast as one sample from an n -dimensional multivariate distribution, where n is the number of grid points. This assumption is in contrast to n samples from a univariate distribution as is commonly assumed. There exist multivariate techniques for statistically comparing two sets of variables (or two fields) with complex correlation patterns. However, these standard multivariate approaches fail in our case due to the lack of replications in the observations (e.g., see Anderson 1984, chapter 10; recall, we are trying to develop a closeness measure for a single forecast and analysis field). If we could somehow remove the covariance between observations on our grid—that is, make the grid points statistically independent—we could use and correctly interpret the standard scores (although this would not remove the inherent problems with using point scores as a measure of multifaceted forecast quality). In a certain sense, discrete wavelets provide a method to orthogonally transform (i.e., make statistically independent) data. This transform allows us to partition the original data in an objective manner. We can then compute standard scores for each of these partitions resulting in a multivalued score for field forecasts.

Various authors have sought methods in which to improve upon the standard scores, namely the ACC, for a particular forecast-analysis field pair in the sense of examining the score for only a part of the forecast. Typically, this refinement has been accomplished by spectral filtering (e.g., Branstator et al. 1993; Van den Dool and Rukhovets 1994) or through projections with empirical orthogonal functions (EOFs; e.g., Livezey et al. 1995). In the first method, wavelengths considered “too small” are filtered out of both the analysis and forecast fields and the standard scores are recomputed. This filtering tends to eliminate small-scale variation so that the filtered ACC is higher and the filtered rmse is lower. This result is also true for EOF filtering. Both analysis and forecast fields are transformed in the standard way and only the EOFs above a predetermined significance level are kept, the rest are set to zero. Again, small-scale variation from the fields is removed and the standard scores improve (if only one unique forecast-analysis pair is available, singular value decomposition, instead of EOF decomposition, can be performed keeping only the “largest” singular values corresponding to a predetermined percentage of variance). Wavelet transforms are an improvement over EOF filtering. The EOF transformation produces a large number of components making interpretation difficult and unclear. On the other hand, interpretation of wavelet scales is physically justifiable [for an explanation of EOFs, see, e.g., Preis-

dorfer (1988)]. More will be said about this subject in section 4.

Both ideas, EOFs and wavelet transforms, have merit and can be used for intensive field forecast diagnostics. Also, examination of forecast performance by wavelength via spectral analysis can point to certain model flaws needing correction, or to suggest reliability of certain scales of forecasts (Branstator et al. 1993). Similarly, it is possible to partition forecast error into orthogonal blocks by EOF method provided a history of forecasts and analyses are available. There exist formal statistics that allow one to identify the noise floor in EOF analysis (see, e.g., Wilks 1995), but it is not clear from the literature that these methods are used with any consistency. More often, an arbitrary cutoff is chosen, say those EOFs that explain at least 90% of the variance.

Discrete wavelet transforms allow us to objectively identify and eliminate insignificant contributions to closeness scores that can be attributed to noise, under certain assumptions on the noise component (given in section 4). If these assumptions are violated, then using wavelet filtering can give misleading results by filtering out relevant information that “resembles” the assumed noise (in a statistical sense). If the noise model assumptions are believed to be valid, it will be shown, through simulations and examples, that the standard scores tend to improve when using wavelet filtering techniques (in the sense of removing the contribution of noise to the scores).

3. The discrete wavelet transform

This section will introduce the terminology and algorithms used in performing the discrete wavelet transform, but will skip detailed mathematical theory and justification. For more specifics, see the sources mentioned in section 1. We use discrete as opposed to continuous wavelet transforms because of the orthogonality property of the former. This quality ensures that the wavelet coefficients (to be defined) are statistically independent from one scale to the next. More common continuous wavelet transforms do not provide the same statistical advantages. Those interested in other applications besides verification can benefit from reading this section as the statistical results presented are widely applicable to a broad range of problems.

We first outline the mechanics of the discrete transform for images. The methodology developed is quite general and may be used for either one- or two-dimensional data. To transform, we must first select a so-called mother wavelet. We present statistical methods based on entropy principles that choose this mother wavelet from a library of potential choices in an objective manner. After the data is transformed into the wavelet space, we can eliminate insignificant elements by a process known as thresholding. We detail recent advances in wavelet statistics that show how to threshold in an objective manner. Thresholding can also be done on one-

or two-dimensional data. For example, thresholding can be employed in nonparametric regression problems or time series analysis (Nason 1994).

a. Mechanics of the discrete transform $n_r \times n_c$

Let an image (or grid, or field) be written as the real-valued matrix \mathbf{G} consisting of n_r rows and n_c columns. It is possible to construct an orthogonal operator Ψ , called the mother wavelet, such that the discrete wavelet transform (WT) is given by

$$\mathbf{W} = \Psi\mathbf{G}, \quad (3.1)$$

where \mathbf{W} is the wavelet transform matrix of \mathbf{G} whose elements $w_{i,j}$ are called wavelet coefficients (Donoho and Johnstone 1994). In other words, we project \mathbf{G} onto the orthonormal basis Ψ yielding a transformed field \mathbf{W} . A discussion of how the elements of Ψ are found such that it is an orthogonal basis can be found in, for example, Daubechies (1992). Certainly Ψ is not unique, and finding new representations for this matrix is an active area of research (e.g., Chui et al. 1994). If Ψ is well chosen, the transformed field or matrix \mathbf{W} can be sparse; that is, many, even the vast majority, of elements may be at or near zero, with only a few elements being relatively large. This characteristic has allowed WTs to be used in image compression with great success (Laurent et al. 1993). These small or zero elements can be eliminated by thresholding and thus less information must be stored. Wavelets have the ability to represent \mathbf{G} by retaining only a fraction of the original number of data values.

Equation (3.1) is readily invertible—that is, $\Psi^{-1}\mathbf{W} = \mathbf{G}$. Typically (3.1) is only performed for matrices of the size $n_r = 2^{k+1}$ and $n_c = 2^{l+1}$ (Donoho and Johnstone 1994; Press et al. 1992). These dyadic or “power of 2” matrices will be used in the present work. Specifying matrices in this form greatly aids in reducing the number of calculations necessary to perform the transform [of $O(n)$, Vidakovic and Müller 1996]. It is possible to form a dyadic matrix from any nondyadic matrix simply by “padding” the end rows and columns with dummy values, say, zeros. If this is done, however, it must be kept in mind when performing any subsequent analyses. DWT like discrete Fourier transforms assume that the data is periodic (e.g., Press et al. 1992) on its domain. Because most data is not periodic, some edge distortion in the transform coefficients is to be expected. Though the impact of boundary problems will be slight, it cannot be completely ignored (e.g., Wickerhauser 1994).

The approach taken here for wavelet transforming an image is the sequential method as outlined in Press et al. (1992) and Saito (1994). We first WT each row of \mathbf{G} , and then WT the columns of the transformed rows to produce one matrix of wavelet coefficients. Another method to WT an image breaks the original image matrix into separate detail and “mean” matrices, all of less dimension than the original matrix (Saito 1994; Dau-

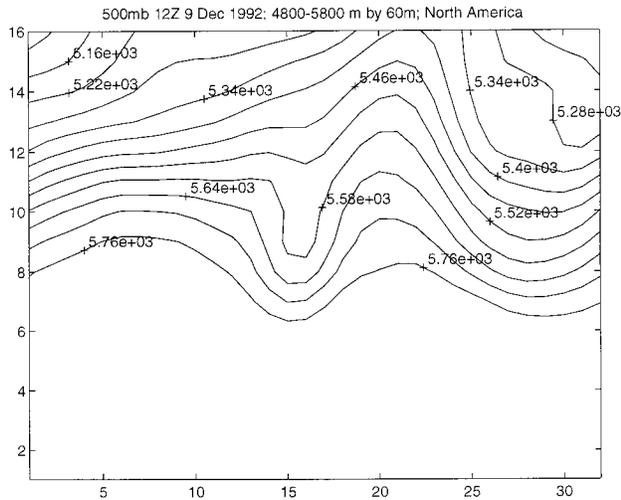


FIG. 1. The 500-mb heights analysis of 1200 UTC 9 December 1992 from ECMWF over North America. The contour interval is from 4800 to 5800 m by 60 m. The axis numbers indicate the corresponding row and column numbers of the matrix.

bechies 1992; Vidakovic and Müller 1996). In our context we prefer the first method because we desire to work with the block scale coefficients, which are more easily interpretable. Both are equivalent in the mathematical sense.

To be more explicit, \mathbf{W} will contain $l + 1$ (from $n_c = 2^{l+1}$) orthogonal blocks. Each block has $n_r = 2^{k+1}$ rows. The number of columns in the first block is 2. The number of columns in the remaining l blocks is 2^h , $h = 1, 2, \dots, l$. As can be seen, the number of columns in each successive block increases by a power of two indicating wavelet representation at finer and finer scales of the original data. The first block of two columns corresponds to, what may be thought of, as the “mean” of the data with respect to the structure of Ψ . The second block cuts the data in half and the columns represent how each of these halves resemble the structure of Ψ . This interpretation can be, admittedly, confusing at first glance but we urge readers to stop and appreciate this process as we use the final WT matrix \mathbf{W} in the following sections.

Throughout this paper we will be using the data presented in Fig. 1. Figure 1 is an $n_r = 2^{3+1} = 16$ by $n_c = 2^{4+1} = 32$ field of 500-mb heights over (roughly) North America (1200 UTC 9 December 1992, data taken from the National Center for Atmospheric Research’s archive of a European Centre for Medium-Range Weather Forecasts analysis from the World Climate Research Programme). A background map has been left out of the figure to emphasize detail (and to emphasize that this data could just as likely have been an image, for example, an uncorrupted computer graphic).

b. Choice of the mother wavelet

There exist many different mother wavelets (orthogonal bases). The question of which mother wavelet is

best to use for transforming a particular field or vector of data is an important and sometimes neglected one. The ideal WT reduces the data to the greatest degree—that is, produces the most coefficients near zero. To aid in this choice, we will study the L_2 matrix norm of the WT matrix \mathbf{W} (Vidakovic and Müller 1996). The L_2 matrix norm (as opposed to the Frobenius or Euclidean norm) is the most commonly used matrix norm (Golub and van Loan 1989). In this work, as in much of the current statistical literature, we also define the norm as the total “energy” of the matrix. The procedures developed in this section are equally applicable to one-dimensional datasets (e.g., time series or regression problems).

We adopt the procedure developed by Goel and Vidakovic (1995) and Katul and Vidakovic (1995) used in choosing the best mother wavelet. It is based on minimizing the entropy of the wavelet transformed matrix. The idea is that the WT disbalances the energy of an image and that the most disbalanced transform is best. This approach makes intuitive sense in that the minimum entropy will be with the transform that produces the greatest ratio of few large coefficients to many small coefficients. We start with a library of mother wavelets, compute an entropy score for each transform, and pick the mother wavelet that produces the best score. We will have more to say on the subject of this library in a moment, but for now we use the library of mother wavelets based on the increasing coefficients of Daubechies and Symmlet wavelets along with the Haar wavelet (Daubechies 1992; Press et al. 1992). Producing a picture of these various bases is probably not instructive as knowledge of their form will not be of much use in the next sections. Readers interested in such forms are referred to Daubechies (1992).

The measure found to be the most resilient by Goel and Vidakovic (1995) is the Shannon entropy measure given by

$$\phi(\mathbf{W}) = -\sum_{ij} w'_{ij} \log w'_{ij}, \tag{3.2}$$

where w'_{ij} are the nonnegative normalized wavelet coefficients—that is, $w'_{ij} = |w_{ij}|/\sum |w_{ij}|$ and $0 \log 0 = 0$ by definition. A similar argument was used by Saito (1994) in his algorithm to select a best basis from a library while simultaneously thresholding (thresholding will be discussed below). The best WT will minimize $\phi(\mathbf{W})$ from the library of possible mother wavelets.

Recall there are a very large number of orthogonal bases Ψ . To make practical use of Shannon’s entropy measure, we must limit ourselves to a finite few whose value is proven by experiment. Clearly, work needs to continue in this area. But whatever library is eventually selected for a task, the user must make explicit the eventual mother wavelet choice. We have had good success with both the Haar and the Daubechies mother wavelets and recommend these as starting points (although there are many others, such as Symmlets, Coiflets, etc.).

TABLE 1. Shannon entropy scores received for each of the listed mother wavelets. The Db are Daubechies wavelets of increasing order.

	Haar	Db 4	Db 6	Db 8	Db 10	Db 12	Db 16	Db 20	Db 30
Score	1.947	1.874	1.861	1.803	1.892	1.857	1.847	1.825	1.920

An example of Shannon’s entropy measure applied to the image of Fig. 3.1 with our library of mother wavelets is found in Table 1. Here we present the Haar wavelet and an increasing series of Daubechies wavelets as our library with the corresponding entropy scores. Table 1 indicates that the Daubechies 8 mother wavelet is the optimal choice from our library.

c. Wavelet thresholding

Thresholding refers to the process of shrinking the coefficients of **W**—that is, setting to zero or shrinking toward zero certain coefficients, in an effort to remove insignificant information. Generally, there are two manners of thresholding, hard and soft. Hard thresholding is of the form

$$w_{ij} = \begin{cases} 0, & |w_{ij}| < \lambda \\ w_{ij}, & |w_{ij}| \geq \lambda, \end{cases} \quad (3.3)$$

where all $|w_{ij}| < \lambda$ are set to zero and the rest are kept. Soft thresholding is of the form

$$w_{ij} = \text{sign}(w_{ij})(|w_{ij}| - \lambda)_+, \quad (3.4)$$

where all $|w_{ij}| < \lambda$ for some fixed constant λ are set to zero and the rest are shrunk toward zero by an amount λ (Donoho et al. 1995). Donoho and Johnstone (1994) maintain that the largest scale’s wavelet coefficients should be left unthresholded regardless of their size. Experiments with meteorological fields suggest this is a good practice and that hard thresholding is ideal for data compression, while statistical arguments suggest soft thresholding be used in analysis settings.

The best choice for λ depends on the assumed model for the data. Donoho and Johnstone (1994) and Donoho (1992) provide a method to find λ for one-dimensional data. They first suppose data of the form $y_i = f(x_i) + e_i$ is observed, where y_i are the data, $f(x_i)$ is the function generating the data, and e_i is noise distributed as $N(0, \sigma^2)$. It is a remarkable property of the WT that it has no effect on noise (Nason 1994); that is, if a pure noise dataset is presented to wavelet transform, the resulting transformed dataset will be indistinguishable (in a statistical sense) from the original untransformed image. This fact should not be surprising as the entire purpose of the WT is to capture real processes in the data, not noise.

The above authors proved that, if this form of the data is correct, then the λ that is near minimax (under certain regularity conditions) is of the form

$$\hat{\lambda}_U = \frac{(2 \log n)^{1/2} \hat{\sigma}}{n^{1/2}}, \quad (3.5)$$

where $\hat{\sigma}$ is also estimated from the data as the median absolute deviation of the wavelet coefficients at the smallest scale (Donoho et al. 1995; they also divide $\hat{\sigma}$ by 0.6745, which slightly increases $\hat{\lambda}_U$). This $\hat{\lambda}_U$ is known as the “universal” thresholder. We note that this estimator may be ideal for many one-dimensional datasets.

Johnstone and Silverman (1995) sought to generalize the universal thresholder $\hat{\lambda}_U$ by allowing for data with correlated noise, which will certainly be present in images. Their version of $\hat{\lambda}_U$ is very similar to the universal one and is given by

$$\hat{\lambda}_{Uh} = \frac{(2 \log n)^{1/2} \hat{\sigma}_h}{n^{1/2}}, \quad (3.6)$$

where a different $\hat{\lambda}_{Uh}$ is computed for each scale or block h and $\hat{\sigma}_h$ is the standard deviation of the wavelet coefficients at that scale.

A final thresholding estimate is due to Goel and Vidakovic (1995) and is used in an atmospheric turbulence study by Katul and Vidakovic (1995). The methodology is based on the Lorentz curve and is appealing since no distributional form is prescribed for the noise. The Lorentz curve threshold is derived from entropy arguments as was the Shannon entropy measure for mother wavelet selection. It attempts to find the place in the data where the signal first rises above the noise and may be written as

$$\hat{\lambda}_E = \frac{\|\mathbf{W}\|_2}{n^{1/2}}. \quad (3.7)$$

Since atmospheric data is certainly highly correlated, we propose a combination of (3.6) and (3.7) as a threshold

$$\hat{\lambda}_{Eh} = \frac{\|\mathbf{W}_h\|_2}{n_h^{1/2}}. \quad (3.8)$$

This threshold gives us the nice properties of the entropy threshold applied to each scale or block h , for all $h > 2$. [The first two blocks are not thresholded because of statistical optimality considerations, which are discussed in Donoho and Johnstone (1995)].

Figure 2 presents an example of wavelet compression. The solid contour is identical to the analysis field of Fig. 1. The dashed lines are the result of WTing the analysis field with the Daubechies 8 mother wavelet, applying hard threshold (3.8), and inverse WTing. Only 131 of the original 512 data points were kept after thresholding, for a reduction to 26% of the original. DeVore (1993) shows that, for some images, this can be improved to as little as 5%!

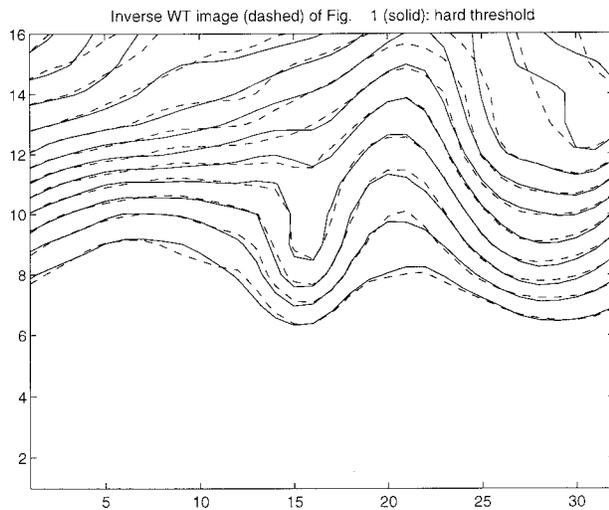


FIG. 2. The solid line is again Fig. 1. The dashed lines are inverse WT images using the threshold estimates as detailed in the text.

4. Use of wavelets with common scores

As mentioned in section 2, we have two goals. The first is to present a method to objectively improve the standard forecast performance scores between fields by removing insignificant information from the relevant data fields. We will do this by WTing the analysis and forecast fields, soft thresholding, inverse WTing, then recomputing the scores. For examples we will present scores based on simulations of realistic forecast fields using the analysis field described in section 3. Furthermore, multiple pseudoforecasts decreasing in prediction accuracy (or increasing in forecast *badness*) will be simulated, and the original and modified scores will be compared. We choose to study simulations over real forecast fields because we are able to specify the exact characteristics of the forecast field. Thus, we have one less unknown to consider when examining the behavior of the WT scoring system.

The second and more important objective is to develop a multivariate closeness score that better captures information about field closeness. An actual numerical forecast will be used to demonstrate how WTing can be employed to construct a multivariate closeness score. The components of this measure are the standard scores computed for each scale of the WTed data. We can thus obtain, for example, an indication of how each scale contributes to the overall rmse, or we can compute an rmse for each scale.

a. Removing noise from point scores

Assume that we can write the observed analysis and forecast fields as $\mathbf{A} = \mathbf{A}' + \boldsymbol{\epsilon}$, and $\mathbf{F} = \mathbf{F}' + \boldsymbol{\eta}$, where \mathbf{A}' and \mathbf{F}' are the true fields and $\boldsymbol{\epsilon}$ and $\boldsymbol{\eta}$ are fields of correlated Gaussian random error. Writing the fields in this manner has some justification [the objective anal-

ysis used in producing the analysis fields contains error of this sort, see Daley (1993)]. In this sense, each data field contains a certain amount of information attributed to error that is not helpful or is insignificant for comparison purposes. Using WTs, we can objectively identify and remove (with soft thresholding), at each scale, the insignificant portions of the data that are solely attributable to this noise. These WT fields produce estimates that are closer to the true fields \mathbf{A}' and \mathbf{F}' than \mathbf{A} and \mathbf{F} , respectively. We can use these estimates to recompute or modify closeness scores thus giving us better or tighter information. We caution that if the error does not follow the assumptions as given, inconsistencies can arise. In particular, it is possible to threshold out true information that has been incorrectly identified as noise. Therefore, in order to utilize the results of this section, the noise assumptions must be taken into consideration.

To illustrate and study the changes in point scores via WT, we will consider a simulation study. Simulations allow control of the forecast–analysis pair. Thus, interpretations are not clouded by the uncertainty that would be present if actual fields were to be used. We simulate the pseudoforecast field \mathbf{F} by assuming $\mathbf{F} \sim N(\mathbf{A}, s\boldsymbol{\Sigma})$ —that is, a normal field with mean \mathbf{A} (the analysis field) and covariance $s\boldsymbol{\Sigma}$ (a 512×512 matrix). The scale parameter s allows simulation of various levels of forecast badness. For example, $s = 1$, say, allows simulation of forecasts that are “good” in that we are only adding a small amount of noise. A value of $s = 400$ simulates forecasts that are considered “bad.” This technique is arguably subjective, but visual inspection of these pseudoforecasts at various values of s indicated the model we have chosen is adequate for an initial study of the performance of the scores. The matrix $\boldsymbol{\Sigma}$ is chosen such that correlations decrease (from grid point to grid point) by 0.9^ρ , where ρ is a function returning the Euclidean distance between grid points. Adjacent grid points are defined as being one unit apart, etc. This choice leads to a smooth decrease in the correlations as grid point distance increases. Other choices could be made for the function ρ , such as an exponential function (Daley 1993, chapter 4). Still more complex relationships describing correlation could be utilized in future studies, such as that described in Hancock and Wallis (1994). The normal field choice, however, will serve as a good starting point for illustration.

As an example, Fig. 3 shows one realization of \mathbf{F} with $s = 400$. The influence of \mathbf{A} can be readily seen, although differences due to the variance $s\boldsymbol{\Sigma}$ are also visible. We argue that, as a first cut, these differences are adequate to simulate “real” forecasts.

The simulation to demonstrate the removal of noise from point scores proceeds as follows. For each fixed s , simulate 50 pseudoforecasts \mathbf{F} . For each of these \mathbf{F} , compute the $\text{ACC}(\mathbf{A}, \mathbf{F})$ and $\text{rmse}(\mathbf{A}, \mathbf{F})$. Now WT \mathbf{A} and \mathbf{F} , apply soft threshold (3.8), inverse WT both fields and recompute the ACC and rmse. The resultant 50

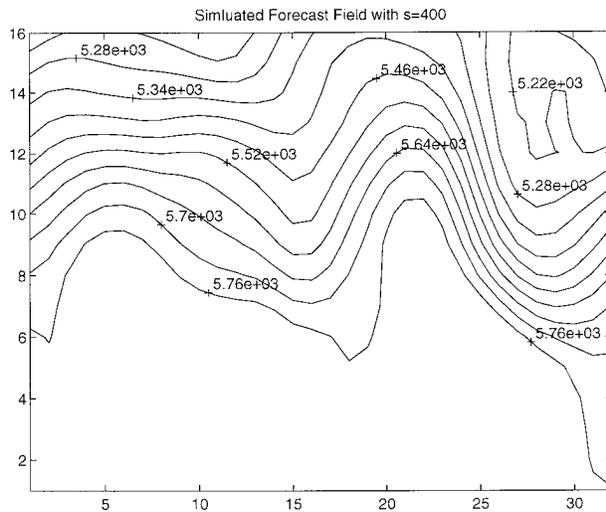


FIG. 3. Simulated forecast field with $s = 400$.

scores, both before WTing and after, are averaged. A new value of s is fixed and the process repeated. Fifty averaged simulations at each fixed s were used to control uncertainty due to sampling variability.

Figure 4 presents the results for s ranging from 1 to 400. Figure 4a shows the ACC, and Fig. 4b displays the rmse. For each, the solid line is the standard score while the dotted line is the WT modified score. As can be clearly seen, both scores show improvement after WTing in that the ACC increased and the rmse decreased. More specifically at a value of $s = 400$, perhaps not untypical for extended range forecasts, the ACC improved by about 5% while the rmse improved by about 18%, that is, a significant portion of the noise was removed. Increased ACC and decreased rmse indicate an improvement as a consequence of the removal, on average, of small-scale variation as discussed in section 2. There is no guarantee that any *individual* simulated forecast would have improved scores, as some form of added noise which could improve any given forecast can be easily imagined. But we would expect, due to the conditions of this experiment, an *average* improvement because we start by degrading the analysis field and then comparing it with itself.

We reran the simulation, this time including an EOF analysis (actually singular value decomposition, SVD, as we have unique fields). For each simulated \mathbf{F} and \mathbf{A} we performed the SVD and kept only the leading three singular values (typically about 80%–90% of the total variance), and set the remaining ones to zero. The data was then transformed back and the standard scores calculated. With pseudoforecasts of the type we are using, we found there to be almost no noticeable improvement in the score's values. This shows that SVD is not as effective as DWT in removing this type of noise. Of course, a change in the correlation function ρ , or other simulation assumptions, or the use of actual forecasts may lead to different results in the scores.

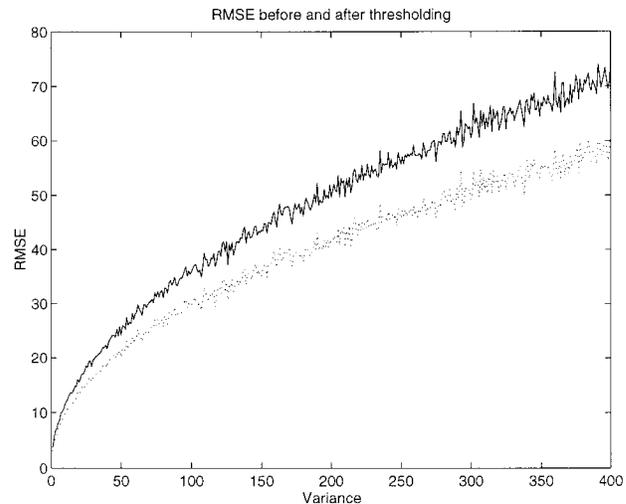
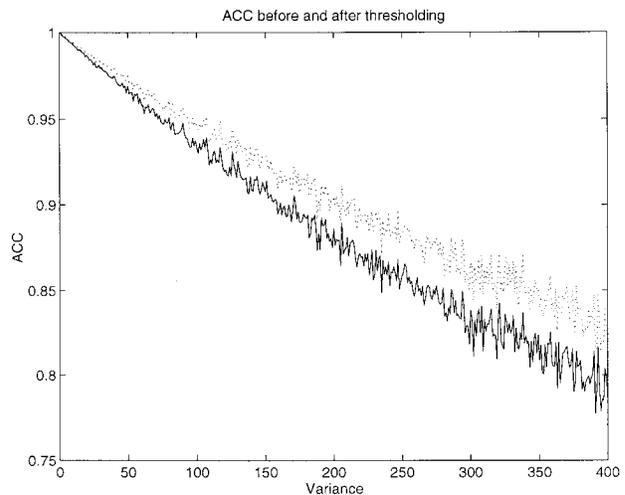


FIG. 4. Mean results of 50 simulations for each value of s of (a) ACC and (b) rmse. The vertical axis is the unit of the score. In each, the horizontal axis is the variance parameter s . Both scores improve after WTing and thresholding.

Overall, the WT modified scores behaved as expected in that, as more and more noise was added to the analysis, and the noise corrupted analysis was compared with its uncorrupted self, the modified scores adjusted more strongly. These simulations have nothing to say about any particular bad forecast, that is, we would not necessarily expect a true forecast with an error variance $s = 400$ to show as drastic an improvement in the scores. This is because a true forecast will be apart (or not close) from the analysis in ways other than noise. Rather, we are taking a known field and seeing how modified closeness scores perform when the noise level is specified (and noise that follows our assumptions).

b. Example of a multivariate score

We now present a detailed example of a multivariate score based on statistics computed in the wavelet space.

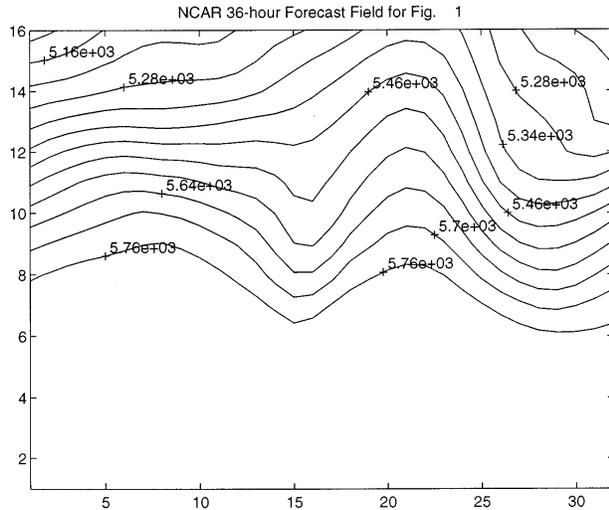


FIG. 5. The 36-h NCAR CCM-2 forecast corresponding to Fig. 1. The contour levels are the same as Fig. 1.

Recall that for our particular dataset **A**, there are $l + 1$ blocks representing different orthogonal dyadic scales of the data. It is possible to both compute the degree to which each scale/block contributes to the overall score(s) and to compute scores for each dyadic scale. For the former we compute the percent of each score explained by scale after thresholding. For the latter, we first inverse WT each scale (setting the others to zero) and then compute scores for these inverse WT fields. Here, we use hard threshold (3.8) on the data to present cleaner images, although we stress that it is *not* necessary to threshold for this method. These concepts can be made clear with an example.

Figure 5 details a 36-h forecast made by the National Center for Atmospheric Research (NCAR) Community Climate Model 2 (CCM-2) for the analysis field of Fig. 1. The ACC between the two fields is 0.974. The rmse is 21.83 m. Note that an ACC of 0.974 is typical, and in this case, interpreted as a very good overall agreement between the forecast and analysis fields, whereas in this case an rmse of 21.83 m indicates a fairly poor forecast. This can be explained upon examination of Figs. 1 and 5. The forecast has captured the overall flow of the field (good ACC) but has missed the details, like the depth of the trough in the center (bad rmse). To illustrate the concepts of the previous section with a real forecast, after WTing, thresholding, and inverse WTing, the modified ACC is 0.979, and the modified rmse is 18.23 m. This represents a change of 0.6% and 4.9%, respectively. However, the discrepancy in interpretation between ACC and rmse still remains (recall that, as mentioned in section 2, the ACC and rmse are measuring different aspects of field closeness).

Notice that WTs allow us to examine and compare the analysis and forecast fields at different geographic scales. Thus, a multivariate score utilizing and distinguishing information from and between these scale/

TABLE 2. A comparison of scores ACC and rmse by percent at each scale, and a comparison of correlation and rmse calculated separately for each scale.

Scale	1	2	3	4	5
%ACC _{<i>h</i>}	0.80	0.09	0.10	0.01	0.00
%MSE _{<i>h</i>}	0.48	0.16	0.20	0.08	0.08
<i>r_h</i>	0.998	0.968	0.978	0.808	0.409
rmse _{<i>h</i>}	11.56	6.98	7.87	4.85	5.68
ER _{<i>h</i>}	0	0.01	0.25	0.03	0.51

blocks of detail may help explain where the two fields differ. Let us first study how the information at each scale contributes to the overall scores. The percent each scale contributes to the ACC can be calculated by the following:

$$\%ACC_h = \frac{\text{Tr}(\mathbf{A}_h - \mathbf{C}_h)^T(\mathbf{F}_h - \mathbf{C}_h)}{\text{Tr}(\mathbf{A} - \mathbf{C})^T(\mathbf{F} - \mathbf{C})}, \quad (4.1)$$

where the subscript h represents the scale under consideration. Because of the linearity of the transform, the score %ACC_{*h*} can be calculated in either the wavelet space or in the data space. If the percentage is calculated in the data space, the wavelet coefficients at all other scales except h should first be set to zero before inverse WTing. We will illustrate these techniques graphically later.

The percent each scale contributes to MSE (MSE, as opposed to rmse, is used because the presence of the radical in the definition of rmse prohibits calculation of a percent score) can be calculated by

$$\%MSE_h = \frac{\|\mathbf{A}_h - \mathbf{F}_h\|_f^2}{\|\mathbf{A} - \mathbf{F}\|_f^2}. \quad (4.2)$$

The score %MSE_{*h*} can also be calculated in the wavelet space or in the data space (with the same recommendations about inverse transforming as before). Table 2 shows the values calculated for %ACC_{*h*} and %MSE_{*h*}.

Examination of %ACC_{*h*} shows that about 80% of ACC is explained by scale $h = 1$. Simulations, like those carried out in section 4a, reveal this trait to hold in general. Even fields that are widely different from each other produce large values for %ACC₁. This phenomena occurs because, in a sense, %ACC₁ is attempting to measure the correlation between the means of the two fields which will, in meteorological forecast fields, always be somewhat close. Because %ACC₁ is so large, little is left to explain with the other %ACC_{*h*}. Thus, we should judge the remaining %ACC_{*h*} relative to each other, not necessarily to %ACC₁. Scales 2 and 3 explain the bulk of the remaining ACC after scale 1 is taken into consideration, while scales 4 and 5 add little. The same kinds of interpretations can be given for %MSE_{*h*}. Note that %MSE₂ and %MSE₃ provide a larger percentage of the MSE than the corresponding ACC scores. This recognition might help explain the discrepancy in interpretation between rmse and ACC. We will explore

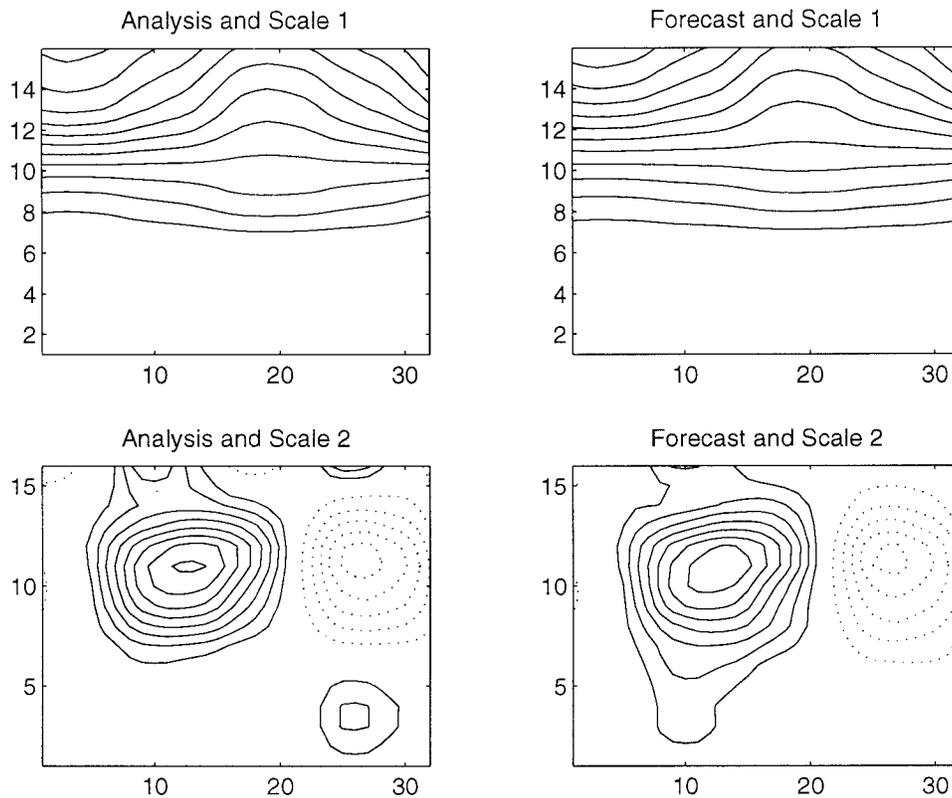


FIG. 6. A series of images in the data space comparing the analysis and forecast fields at each of the five successively finer orthogonal wavelet scales.

this possibility further when developing and interpreting the multivariate score later.

As we can see, though, the scores $\%ACC_h$ and $\%MSE_h$ are helpful in only a limited sense. They do give increased information about forecast performance by scale with respect to the common measures, but we can do better. The ordinary linear correlation r , or ACC, can be geometrically interpreted as the cosine of the angle between two vectors (Briggs and Levine 1996; Persson 1996). Likewise, rmse can be interpreted as a function of a distance between these two vectors. In our case these vectors are our forecast and analysis matrices. Both of these measures are important to understand aspects of field closeness, therefore both should be examined *simultaneously* for any given field pair. Consideration of either independent of the other ignores available information. A third measure, given below, is introduced that will complete the geometric picture. The geometric (and energy) interpretation will be emphasized in the construction of a closeness measure of the forecast-analysis fields by scale. Below, the mean of each matrix \mathbf{A}_h and \mathbf{F}_h is subtracted before any calculations are performed.

Each score, r and rmse, can be calculated in the usual manner for each scale h . Designate these scores as r_h and $rmse_h$. Because of linearity both can be calculated in the wavelet or data space. Values for $rmse_h$ calculated

in the wavelet space by dyadic scale, however, will not sum to rmse. This lack of additivity follows because, in the wavelet space, each matrix \mathbf{A}_h and \mathbf{F}_h consist of only n_h grid points where $\sum_h n_h = n$. In the data space each matrix \mathbf{A}_h and \mathbf{F}_h consist of n data points. Recall in section 2 that calculation of rmse included $n^{1/2}$ in the denominator. If we were to calculate the rmse at each scale h in the wavelet space, $rmse_h$ would have a different denominator for every $h(n_h^{1/2})$. When we calculate rmse by scale in the data space the denominator equals $n^{1/2}$. The two calculations are related by

$$\frac{n^{1/2}}{n^{1/2}_h} rmse_h(\text{data}) = rmse_h(\text{wavelet}). \quad (4.3)$$

Define the energy ratio (ER) by scale as

$$ER_h = |1 - \xi|, \quad (4.4)$$

where

$$\xi = \min \left\{ \frac{\|\mathbf{A}_h\|_2}{\|\mathbf{F}_h\|_2}, \frac{\|\mathbf{F}_h\|_2}{\|\mathbf{A}_h\|_2} \right\}.$$

Defining ξ in this manner ensures that ER is bounded between 0 and 1. The ER can be interpreted through energy arguments as discussed in section 3. Each scale has a certain amount of energy, and it is helpful to ask whether \mathbf{F} has the same amount of energy for each scale

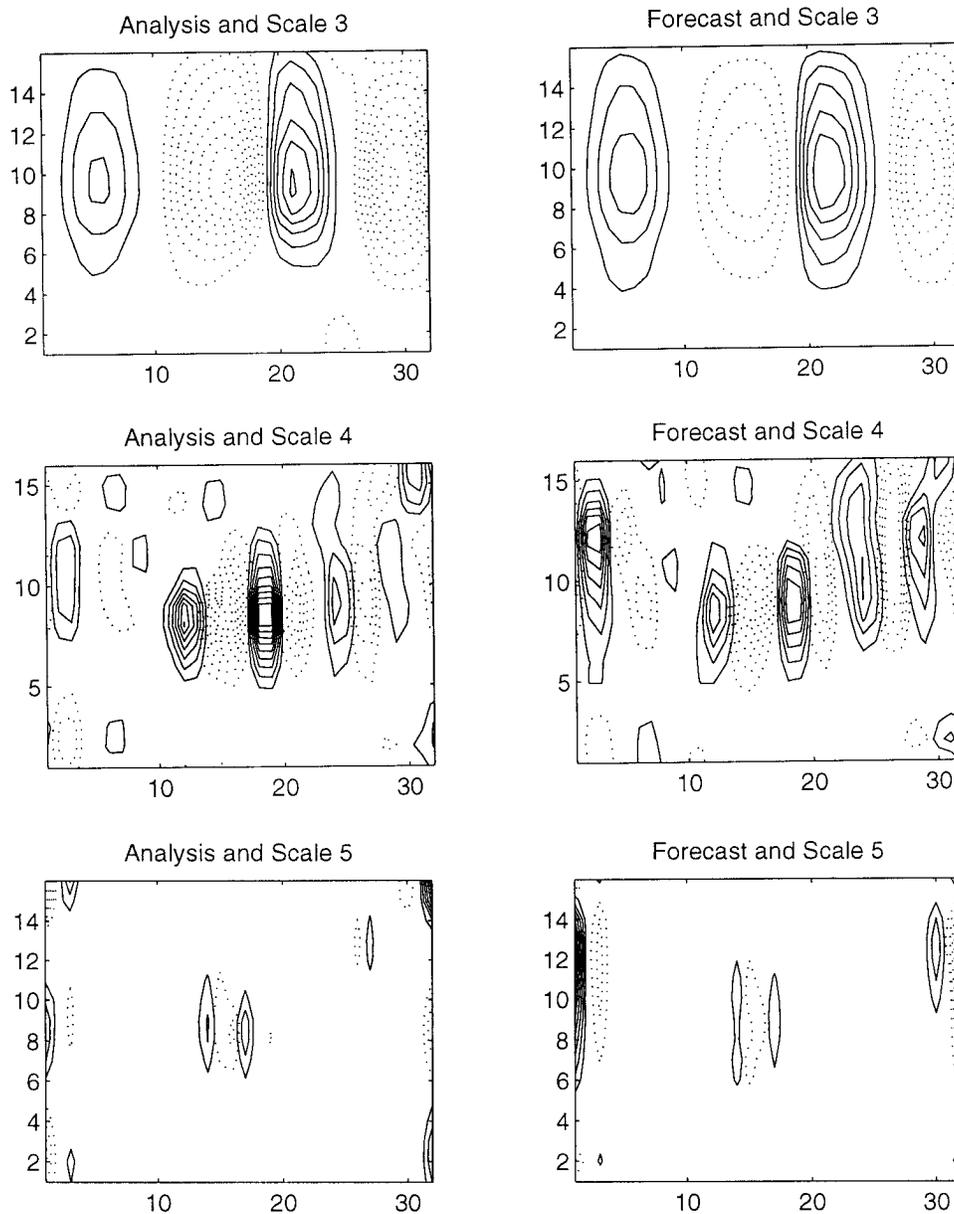


FIG. 6. (Continued)

h as does **A**. If we write the matrices **A** and **F** as two vectors, stringing the columns for each, respectively, then ER geometrically represents the ratio of the lengths of these vectors. If the length of each vector at scale h is identical, then $ER_h = 0$. Of course it is possible to express ER as a function of ordinary linear correlation and MSE, but examining it independently can lead to important insights of field closeness. Table 2 lists the scores received for r_h , $rmse_h$, and ER_h .

Recall from section 3 that in the wavelet space, each block h of wavelet coefficients represents how each dyadic portion of the data resembles the structure of the mother wavelet. Upon transforming back to the data

space, then, the different scales provide an interpretation of the data at various levels of resolution. Hence, in a sense, they provide information for comparison of the features of the forecast and analysis fields at 2^{l+1} different geographical scales of detail. As a side note, the wavelet decomposition of the data into dyadic scales is a convenient mechanism to provide compact representation of the fields. As mentioned in section 2, alternative methods such as SVD break down the data in an analogous manner, but SVD transforms produce a separate field or detail level for each column of data. For fields of even modest size, this is a detriment as interpretation is very cumbersome and difficult. In contrast,

the number of dyadic fields under a wavelet decomposition are limited by powers of two, thus allowing for easier manageability and interpretation.

This breakdown of the comparison of fields into three scores at the, here, five levels of detail further highlights the discrepancy in interpretation between the overall ACC and rmse. As can be seen by examining Table 2, the fields differ at scales 2 and 3. But the difference is best explained as a difference in distance (rmse) and not angle (correlation). Furthermore, the energy in the fields (ER) differs appreciably at scale 3. The overall rmse picks up these differences and declares the forecast to be poor. However, at the first dyadic scale, the ACC and ER scores indicate a good forecast (in fact the ER shows the fields **A** and **F** to have the same length or hold equivalent amounts of energy). The overall ACC interprets these commonalities as showing a good agreement between analysis and forecast fields. Of course, scales 4 and 5 may be interpreted similarly. The importance of analyzing more than one measure simultaneously is apparent in this example.

To further study the analysis and forecast fields, let us examine the contour plots of the different dyadic scales in Fig. 6. These plots are perhaps most revealing. They provide point by point geographical comparison of the analysis and forecast fields at each of the five dyadic scales. Figure 6a (contour interval of about 100 m) represents the overall data structure of the fields. As can be seen, the analysis and forecast fields agree quite well. Figure 6b (contour interval of about 100 m) shows the data at the second dyadic level such that only features that are found in both halves of the overall field are present. The remaining dyadic scales are filtered out (set to zero). The analysis field in this figure has a slightly tighter positive contour gradient than the forecast. It also (in the lower right-hand corner) has additional features not picked up by the forecast. Overall, the fields are close. Similarly, Figs. 6c–e (contour intervals of about 5–10 m) display features of the data that span successive halvings of the space (with all other features filtered out). In each, the forecast field captures the general features of the analysis field in that the contours are in the same geographic location, however, the gradients do not always match. These differences in the gradients are picked up by the rmse and $rmse_h$, but not necessarily by the ACC or r_h .

Examination of Fig. 6 will also confirm, in a subjective way, most of the results of the numerical scores (Table 2). Additionally, because of time constraints, one would not wish to graphically examine images of forecast and analysis pairs by scale operationally. Instead, the closeness measures r_h , $rmse_h$, and ER_h could be kept track of, and particular exemplary or poor performance could be flagged and examined in detail (perhaps visually or with scores like $\%ACC_h$ and $\%MSE_h$). Once again, it must be emphasized that use of a multivariate score will naturally lead to greater insight of forecast performance than will use of univariate scores.

One final note about the number of dyadic scales. We have presented these graphics at five dyadic scales representing $2^5 = 32$ columns of the original data in order to give the largest number of orthogonal scales to analyze. Alternatively, we could have produced four dyadic scales representing $2^4 = 16$ rows of the original data. Slightly different graphics than those in Fig. 6 would have emerged. In the general case, these differing graphics may be insightful and give additional information on field closeness.

5. Conclusions

Field forecast verification is a difficult problem. Typical users are forced to rely on familiar point scores, even with the knowledge that these scores are insufficient and not completely statistically justifiable. The high level of covariance between different grid points violates the assumption of statistical independence so that traditional interpretations for scores such as rmse and correlation are suspect. Further, sampling distributions for univariate scores like the rmse and ACC are unknown. For example, without knowledge of a sampling distribution, it is impossible to gauge the significance of any particular score received and impossible to rate the difference between two competing forecasts.

Two methods utilizing wavelet transforms in the problem of field forecast verification are given. The first assumes that both the analysis and forecast fields are corrupted by noise of a known parametric form. This noise is removed through WTing and thresholding using procedures that are statistically optimal. Standard scores are then recomputed on the “noise-free” fields, giving greater information about the closeness of the true fields. The second scheme uses wavelets to orthogonally decompose both analysis and forecast fields into dyadic scales so that field closeness can be examined at different levels of detail. This procedure is accomplished by computing standard scores and percent contribution to the overall value of these scores at each of the dyadic scales. Visual comparison is also possible with this method.

This article should not be viewed as an answer to the development of a robust sufficient statistic to measure field–image closeness. Rather, it is an attempt to incrementally improve interpretation of forecast verification measures that exist. Attempts to “break up” the scoring system, such that we can examine closeness over a number of different scales, are helpful. The use of wavelets in partitioning a field is especially intriguing because the partitions can be made orthogonal, that is, statistically independent. This work is only the first step in utilizing wavelets in this context and no doubt further improvements will suggest themselves in the future.

Acknowledgments. We would like to thank David Ruppert and Wayne Bresky for their helpful comments along the way. We also thank two anonymous reviewers

who greatly helped the exposition of this paper. Partial support for this research was provided by NIH training grant ES07261. A great portion of work was done with the assistance of the WaveLab MATLAB toolbox written by D. Donoho, I. Johnstone, J. Buckheit, and S. Chen freely available from <http://playfair.stanford.edu>.

REFERENCES

- Anderson, T. W., 1984: *An Introduction to Multivariate Statistics*. 2d ed. Wiley, 675 pp.
- Branstator, G., A. Mai, and D. Baumhefner, 1993: Identification of highly predictable flow element for spatial filtering of medium- and extended-range numerical forecasts. *Mon. Wea. Rev.*, **121**, 1786–1802.
- Briggs, W. M., and R. A. Levine, 1996: Wavelets and field forecast verification. Cornell University Biometrics Unit Tech. Rep. BU 1318M, 461 pp. [Available from Biometrics Unit, Ithaca, NY 14853.]
- Chao, B. F., and I. Naito, 1995: Wavelet analysis provides a new tool for studying Earth's rotation. *Eos*, **76**, 161–165.
- Chui, C. K., L. Montefusco, and L. Puccio, 1994: *Wavelets: Theory, Algorithms, and Applications*. Academic Press, 627 pp.
- Daley, R., 1993: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- Daubechies, I., 1992: *Ten Lectures on Wavelets*. SIAM, 357 pp.
- DeVore, R. A., 1993: Adaptive wavelet bases for image compression. *Wavelets, Images, and Surface Fitting*, P. J. Laurent, A. Le Méhauté, and L. L. Shumaker, Eds., A. K. Peters, 197–219.
- Donoho, D. L., 1992: De-noising by soft-thresholding. Tech. Rep. 409, 37 pp. [Available from Dept. of Statistics, Stanford University, Stanford, CA 94305.]
- , and I. M. Johnstone, 1994: Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.
- , and —, 1995: Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Stat. Assoc.*, **90**, 1200–1224.
- , —, G. Kerkycharian, and D. Picard, 1995: Wavelet shrinkage: Asymptopia? *J. Roy. Stat. Soc.*, **57**, 301–369.
- Gao, W., and B. L. Li, 1993: Wavelet analysis of coherent structures at the atmosphere–forest interface. *J. Appl. Meteor.*, **32**, 1717–1725.
- Goel, P., and B. Vidakovic, 1995: Wavelet transformations as diversity enhancers. Discussion Paper 95-04, 21 pp. [Available from ISDS, Duke University, Durham, NC 27708.]
- Golub, G. H., and C. F. van Loan, 1989: *Matrix Computations*. Johns Hopkins Press, 642 pp.
- Hagelberg, C. R., and N. K. K. Gamage, 1994: Applications of structure preserving wavelet decompositions to intermittent turbulence: A case study. *Wavelets in Geophysics*, E. Foufoula-Georgiou and P. Kumar, Eds., Academic Press, 45–80.
- Hancock, M. S., and J. R. Wallis, 1994: An approach to statistical spatial-temporal modeling of meteorological fields. *J. Amer. Stat. Assoc.*, **89**, 368–390.
- Hoffman, R. N., Z. Liu, J. F. Louis, and C. Grassotti, 1995: Distortion representation of forecast errors. *Mon. Wea. Rev.*, **123**, 2758–2770.
- Johnstone, I. M., and B. W. Silverman, 1995: Wavelet threshold estimators for data with correlated noise. Tech. Rep. 37 pp. [Available from Dept. of Statistics, Stanford University, Stanford, CA 94305.]
- Katul, G., and B. Vidakovic, 1995: The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering. Discussion Paper 95-05, 31 pp. [Available from ISDS, Duke University, Durham, NC 27708.]
- Kumar, P., and E. Foufoula-Georgiou, 1993: A new look at rainfall fluctuations and scaling properties of spatial rainfall using orthogonal wavelets. *J. Appl. Meteor.*, **32**, 209–222.
- , and —, 1994: Wavelet analysis in geophysics: An introduction. *Wavelets In Geophysics*, E. Foufoula-Georgiou and P. Kumar, Eds., Academic Press, 1–43.
- Lau, K. M., and H. Weng, 1995: Climate signal detection using wavelet transform: How to make a time series sing. *Bull. Amer. Meteor. Soc.*, **76**, 2391–2402.
- Laurent, P. J., A. Le Méhauté, and L. L. Shumaker, 1993: *Wavelets, Images, and Surface Fitting*. A. K. Peters, 528 pp.
- Livezey, R. E., J. D. Hoopingarner, and J. Huang, 1995: Verification of official monthly mean 700-hPa height forecasts: An update. *Wea. Forecasting*, **10**, 512–527.
- Meyers, S. D., B. G. Kelly, and J. J. O'Brien, 1993: An introduction to wavelet analysis in oceanography and meteorology: With application to the dispersion of Yanai waves. *Mon. Wea. Rev.*, **121**, 2858–2866.
- Miyakoda, K., G. D. Hembree, R. F. Strickler, and I. Shulman, 1972: Cumulative results of extended forecast experiments. Part I: Model performance for winter cases. *Mon. Wea. Rev.*, **100**, 836–854.
- Murphy, A. H., 1991: Forecast verification: Its complexity and dimensionality. *Mon. Wea. Rev.*, **119**, 1590–1601.
- , 1995: The coefficients of correlation and determination as measures of performance in forecast verification. *Wea. Forecasting*, **10**, 681–688.
- , and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338.
- , and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, **117**, 572–581.
- Nason, G. P., 1994: Wavelet regression by cross-validation. Tech. Rep. 447, 45 pp. [Available from Dept. of Statistics, Stanford University, Stanford, CA 94305.]
- Perrie, W., and B. Toulany, 1989: Correlations of sea level pressure fields for objective analysis. *Mon. Wea. Rev.*, **117**, 572–581.
- Persson, A., 1996: Forecast error and inconsistency in medium range weather prediction. Preprints, *13th Conf. on Probability and Statistics in the Atmospheric Sciences*, San Francisco, CA, Amer. Meteor. Soc., 253–259.
- Preisendorfer, R. W., 1988: *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, 425 pp.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in C*. 2d ed. Cambridge University Press, 994 pp.
- Radok, U., and T. J. Brown, 1993: Anomaly correlation and an alternative: Partial correlation. *Mon. Wea. Rev.*, **121**, 1269–1271.
- Saito, N., 1994: Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length comparison. *Wavelets in Geophysics*, E. Foufoula-Georgiou and P. Kumar, Eds., Academic Press, 299–324.
- Serrano, E., R. Compagnucci, and M. Fabio, 1992: The use of wavelet transform for climatic estimates. *Proc. Fifth Int. Meeting on Statistical Climatology*, Toronto, Canada, AES, Environment Canada, 259–262.
- Taylor, C. C., 1991: Measure of similarity between two images. *Spatial Statistics and Imaging*, A. Possolo, Ed., IMS Lecture Notes, Vol. 20, IMS, 382–391.
- Van den Dool, H. M., and L. Rukhovets, 1994: On the weights for an ensemble-averaged 6–10-day forecast. *Wea. Forecasting*, **9**, 457–465.
- Vidakovic, B., and P. Müller, 1994: Wavelets for kids. ISDS Discussion Paper 94-13, 26 pp. [Available from ISDS, Duke University, Durham, NC 27708.]
- Weng, H., and K. M. Lau, 1994: Wavelets, period doubling, and time-frequency localization with application to organization of convection over the tropical Western Pacific. *J. Atmos. Sci.*, **51**, 2523–2541.
- Wickerhauser, M. L., 1994: Comparison of picture compression methods: Wavelet, wavelet packet, and local cosine transform coding. *Wavelets: Theory, Algorithms, and Applications*, C. K. Chui, L. Montefusco, and L. Puccio, Eds., Academic Press, 585–621.
- Wilks, D. S., 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.