The Discrete Wavelet Transform and the Scale Analysis of the Surface Properties of Sea Ice

Ronald W. Lindsay, Donald B. Percival, and D. Andrew Rothrock

Abstract—The formalism of the one-dimensional discrete wavelet transform (DWT) based on Daubechies wavelet filters is outlined in terms of finite vectors and matrices. Both the scale-dependent wavelet variance and wavelet covariance are considered and confidence intervals for each are determined. The variance estimates are more accurately determined with a maximal-overlap version of the wavelet transform. The properties of several Daubechies wavelet filters and the associated basis vectors are discussed. Both the Mallat orthogonal-pyramid algorithm for determining the DWT and a pyramid algorithm for determining the maximal-overlap version of the transform are presented in terms of finite vectors. As an example, we investigate the scales of variability of the surface temperature and albedo of spring pack ice in the Beaufort Sea. The data analyzed are from individual lines of a Landsat TM image (25-m sample interval) and include both reflective (channel 3, 30-m resolution) and thermal (channel 6, 120-m resolution) data. The wavelet variance and covariance estimates are presented and more than half of the variance is accounted for by scales of less than 800 m. A wavelet-based technique for enhancing the lower-resolution thermal data using the reflected data is introduced. The simulated effects of poor instrument resolution on the estimated lead number density and the mean lead width are investigated using a wavelet-based smooth of the observations.

I. INTRODUCTION

The wavelet transform is a relatively new tool for analyzing geophysical data. As originally formulated by Grossman and Morlet [1], the wavelet transform maps a function y(x) of continuous location x into a function W(x, s) of continuous location x and scale s, where ∞ < s < ∞ and 0 < s < ∞. Given a finite set of N equally spaced observations y = {y(x0 + iΔx), i = 1, · · · , N}, where x0 is an offset and Δx is the sample interval, we can approximate the continuous wavelet transform on a digital computer by discretizing s also. The usual choice in earlier applications of the wavelet transform was a linear discretization, i.e., iΔx. More recently, Daubechies [2] and others have formulated an orthogonal discrete wavelet transform (DWT) specifically designed for analyzing a finite set of observations over the set of scales sj = 2j−1Δx, i.e., a dyadic discretization. An advantage of the orthogonal DWT is its ability to partition the variance of the elements of y on a scale by scale basis. This partitioning leads to the notion of the scale-dependent wavelet variance, which in many ways is analogous to the more familiar frequency-dependent Fourier power spectrum.

The wavelet variance is thus a natural tool for investigating the spatial scales of variability in geophysical data. Not only is the scale-dependent variance estimated, but the locations of events contributing to the variance at each scale are also determined.

The intent of this paper is to provide a clear self-contained exposition on the orthogonal DWT and on the wavelet variance and to demonstrate the usefulness of these tools to the remote sensing community by analyzing the surface properties of sea ice as an example. Our exposition of the DWT is formulated entirely in terms of finite matrices and vectors and hence should be readily accessible by practitioners. Our formulation of the wavelet variance and the related concept of wavelet covariance discusses both how to efficiently estimate these quantities and also how to generate confidence intervals for them.

The pack ice of the Arctic Ocean is crossed by a broken network of leads that introduce a large amount of narrowly concentrated variability in the surface properties of the ice. Between the leads, on the ice floes, the surface of spring pack ice is otherwise highly uniform for measurements averaged over several meters. Many physical processes in the leads are radically different from those found on the floes. The processes of most concern to geophysicists include the heat flux from the ice, the ice formation rate, the wind stress on the ice, and the strength of the ice. All of these processes depend on the area covered by the leads, the lead widths, the lead temperature, and the lead albedo. The accurate measurement of these lead properties is, in turn, dependent on the size of the leads and the accuracy and measurement resolution of the instruments. A widely available satellite instrument often used in the Arctic is the advanced very high resolution radiometer (AVHRR) which has a nominal resolution of 1.1 km at nadir. Images obtained with this instrument show a rich assortment of leads which has a nominal resolution of 1.1 km at nadir. Images obtained with this instrument show a rich assortment of leads in almost any cloud-free image of pack ice, but the question remains regarding how much variability associated with leads is missed (averaged out) by the large area sampled in each pixel. We demonstrate that this question can be addressed through the one-dimensional wavelet analysis of transacts from a high-resolution image.

Our example comes from individual lines from a single Landsat image of sea ice that was taken April 16, 1992 in the Beaufort Sea. The image covers approximately 200 km with a pixel spacing of 25 m, which is much smaller than the resolution of the AVHRR. In this scene the dominant scale of both the albedo and the temperature variability is found at between 200–800 m.

Manuscript received March 23, 1995. This work was supported by the Office of Naval Research under Grant N00014-90-J-1074.

The authors are with the Applied Physics Laboratory, University of Washington, Seattle, WA 98105 USA.

Publisher Item Identifier S 0196-2892/96/02854-9.

0196-2892(96)02854-9 © 1996 IEEE
The remainder of the paper is organized as follows. In the first part of Section II we discuss the general properties of the DWT and compare the wavelet transform to windowed Fourier transform techniques. In Section II-A we describe the orthogonal DWT in a self-contained exposition based upon simple ideas from matrix theory. The orthogonal DWT can be computed efficiently using a recursive procedure known as the pyramid algorithm, which we discuss in Section II-B. The pyramid algorithm for the orthogonal DWT involves a subsampling operation that leads to inefficient estimates of the wavelet variance. Therefore, in Section II-C we describe the maximal overlap pyramid algorithm, which is a variation of the pyramid algorithm that eliminates subsampling and leads to efficient estimates of the wavelet variance. The DWT depends on the choice of a mother wavelet filter \( H \), which is a vector with \( m \) elements where \( m \) is a small, even number. We provide some guidance in Section II-D on the choice of \( H \). In Section III we define the wavelet variance and covariance and describe ways of obtaining confidence intervals for these quantities. We then apply these analysis techniques in Section IV where we look at the wavelet variance and covariance of albedo and surface temperature series, discuss how to enhance the low-resolution surface temperature series using the high-resolution albedo and a wavelet-based model, and demonstrate how the analysis of lead number density and mean lead widths is affected by instrument resolution using a wavelet-derived simulation of a low-resolution instrument.

II. WAVELETS

Why wavelets? A spatial series \( y(x) \) tells us everything about the value of a function at a specific location \( x \) but nothing about the spatial scales of the variability. A Fourier transform of the entire series tells us everything about the scales of variability (wavenumber content) of the function, but nothing about the locations of the variability. The wavelet decomposition is a compromise between the two forms of the data (observations and Fourier transform) that identifies both scales and the location of events in the spatial series.

The wavelet transform decomposes a signal into sets of coefficients: each set is associated with a spatial scale and each element in a set is associated with a particular location. The scales start with the sample interval \( \Delta x \); each subsequent scale is twice as large as the previous one. As we shall see, the transform is orthogonal so that the variability at the various scales is separable. The transform is reversible so that an exact reconstruction of the series is possible. The transform is commonly accomplished by the successive application of one highpass and one lowpass filter corresponding to a single wavelet family. There are a number of different wavelet families that are distinguished by different characteristics of the transfer functions of the filters.

A traditional method for attempting to define a localized measure of the series variance at a given frequency is the windowed Fourier transform. This analysis technique applies a moving window, commonly of fixed width and taper, over the data to determine localized spectra. The wavelet technique can be said to use windows that adjust according to the scale, so that small-scale variability is highly localized in the analysis and larger-scale variability is poorly localized: the differences between the two techniques are illustrated in Fig. 1, which shows schematic frequency/space diagrams for both the Fourier analysis with large and small windows and for the wavelet analysis. The interpretation of the Fourier analysis is particularly problematic in cases where the series \( y(x) \) has transient features. It is in these cases, which are characteristic of sea ice surface temperature and albedo observations, that wavelet analysis is often most useful.

The similarity between the two approaches is greatest when the variance of an entire series is analyzed without regard to the locations of specific events. A Fourier analysis would decompose the variance of a series of length \( N \) across approximately \( N/2 \) wavenumbers, where a wavelet analysis would decompose the variance across approximately \( \log_2 N \) scales. A number of decisions must be made in performing an analysis of variance with each approach. Using the Fourier analysis one must decide whether or not to detrend the data, the amount and shape of data tapering that should be applied, as well as the type of smoothing to apply to the spectral estimates. Using the wavelet analysis, polynomial trends of a certain degree can be automatically removed, but one must decide which wavelet family to use and how boundary conditions at the ends of the series are to be handled.

Wavelets were introduced by Grossman and Morlet [1] as functions whose translations and dilations can be used as a basis for expansions of other functions. Their work has since been expanded upon considerably, and wavelets are now being used in an increasing number of applications, such as in turbulence studies, image analysis, sonar signal interpretation, electromagnetic propagation studies, and data compression. A recent compilation of geophysical applications is found in [3].

A. Discrete Wavelet Transform

The discrete wavelet transform (DWT) takes a series of \( N \) observations and produces \( N \) new values called wavelet coefficients. The DWT is structured around a set of filtering operations that determine these coefficients. The wavelet coef-
coefficients contain information about the magnitude and location of events of different scales, or, in filtering parlance, different pass-bands. The decomposition is performed sequentially, starting with the smallest scales and progressing to the larger scales, with the scale doubling in size with each iteration.

We initially assume that the sample size is a power of two, \( N = 2^L \), but later on we can relax this condition. When \( N \) is a power of two, the final stage of the decomposition yields a single wavelet coefficient for the largest possible scale, but in practice the decomposition is often taken only to an intermediate level. There are significant complications from end-effects, which we ignore in this initial presentation but address subsequently.

The discrete wavelet transform decomposes a vector of observations \( \mathbf{y} = \{y(x_i), i = 1 \text{ to } N\} \), where \( x_i = x_0 + i \Delta x \), \( x_0 \) is an offset, and \( \Delta x \) is the sample interval. The decomposition uses orthonormal wavelet basis vectors \( \psi_{j,k} \), \( j = 1 \text{ to } L \) that can be organized into sets distinguished by a scale index \( j \). Each vector in the \( j \)th set is distinguished by a location index \( k \). The scale index \( j \) runs from 1 to \( L \), and the physical scale associated with the \( j \)th set is \( s_j = \Delta x 2^{-j} \). These orthonormal vectors can be used to decompose the series much like the Fourier analysis, but unlike the Fourier sines and cosines, the vectors \( \psi_{j,k} \) are localized in space in the sense that only a portion of each vector is nonzero; that is, each wavelet basis vector is typically sparse (except those associated with the largest scales). The location index \( k \) typically indicates where the nonzero portion of each basis vector occurs. Half as many locations are needed with each step up in scale, and hence \( k \) runs from 1 to \( 2^{-j} N \). In addition, the length of the nonzero portions of these vectors roughly doubles in length with each step up in scale.

As we shall see, the wavelet basis vectors are formed by translating and convolving the elements of a highpass mother wavelet filter \( \mathbf{H} \) and a closely related lowpass scaling filter \( \mathbf{G} \), so that a range of scales and locations are represented. The filters \( \mathbf{H} \) and \( \mathbf{G} \) can be represented as vectors of length \( m \), where \( m \) is commonly an even integer. The filtering is accomplished by a convolution operation defined below. The smallest-scale wavelet basis vectors \( (j = 1) \) are based on shifted versions of the mother wavelet filter, so only \( m \) elements of the vector \( \psi_{1,k} \) are nonzero.

The simplest mother wavelet filter is the Haar, which is based on lag-one differences of the series. The Haar wavelet filter is of length \( m = 2 \). The mother wavelet filter is the two element vector

\[
\mathbf{H} = \left\{ \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right\}
\]

and the scaling filter is

\[
\mathbf{G} = \left\{ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\}.
\]

For \( N = 8 \) observations the \( j = 1 \) wavelet basis vectors are

\[
\psi_{1,1} = \left\{ -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, 0, 0, 0 \right\}
\]

\[
\psi_{1,2} = \left\{ 0, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, 0 \right\}
\]

\[
\psi_{1,3} = \left\{ 0, 0, 0, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0 \right\}
\]

\[
\psi_{1,4} = \left\{ 0, 0, 0, 0, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\}.
\]

The \( j = 2 \) wavelet basis vectors are

\[
\psi_{2,1} = \left\{ -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0 \right\}
\]

\[
\psi_{2,2} = \left\{ 0, 0, 0, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0 \right\}.
\]

The \( j = 3 \) wavelet basis vector is

\[
\psi_{3,1} = \left\{ -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right\}.
\]

For the discrete case in general, the larger-scale wavelet basis vectors are not easily written in closed form and are determined by an iterative process of reconstruction outlined below. In practice, all of the information about the observation vector is contained in the wavelet coefficients, and the wavelet basis vectors themselves rarely have to be determined. A discussion of the properties of various other wavelet families is delayed until Section II-D.

The orthonormal characteristics of the wavelet basis vectors are expressed with the inner product as

\[
\langle \psi_{j,k}, \psi_{j,k} \rangle = \sum_{i=1}^{N} \psi_{j,k}(x_i) \psi_{j,k}(x_i) = \begin{cases} 1 & \text{if } j = J \text{ and } k = K \text{ and } \\
0 & \text{otherwise} \end{cases}
\]

indicating that the wavelet basis vectors are orthonormal across both locations and scales. The observation vector \( \mathbf{y} \) is decomposed using \( L \) sets of wavelet basis vectors \( \psi_{j,k} \) and wavelet coefficients \( D_{j,k} \), and the mean of the series \( \bar{y} \). The decomposition is expressed as

\[
\mathbf{y} = \sum_{j=1}^{L} \sum_{k=1}^{n_j} D_{j,k} \psi_{j,k} + \bar{y} \mathbf{1}
\]

where \( n_j = 2^{-j} N \) is the number of coefficients (locations) for scale \( j \) and \( \mathbf{1} \) is a vector of ones. The wavelet coefficients can be determined by the inner product of the observations and the basis vectors

\[
D_{j,k} = \sum_{i=1}^{N} y(x_i) \psi_{j,k}(x_i) = \langle \mathbf{y}, \psi_{j,k} \rangle.
\]

However, we will shortly indicate a much more efficient method of determining the coefficients using the filters \( \mathbf{H} \) and \( \mathbf{G} \). The sets of coefficients \( D_{j,k} \) and \( \bar{y} \) constitute the DWT. The total number of values in the DWT is the same as in the observation vector.

In our example with the Haar wavelet, the first coefficients at each scale are

\[
D_{1,1} = \frac{y(x_2) - y(x_1)}{\sqrt{2}}
\]
The first scale coefficients are proportional to differences on a scale of \( \Delta x \), the second scale coefficients are proportional to differences on a scale of \( 2\Delta x \), and the third scale coefficients are proportional to differences on a scale of \( 4\Delta x \). The idea that the wavelet coefficients are related to differences (of various orders) of (weighted) average values of portions of \( y \) concentrated in space is not special to the Haar wavelet, but is fundamental to all wavelet transforms discussed in this paper.

Because of the orthogonal characteristics of the wavelet basis vectors, the vector \( y \) is easily separated into components according to scale. The vector \( y \) is seen to be the sum of the mean and a set of \( L \) details (each a vector of length \( N \))

\[
V_j = \sum_{k=1}^{n_j} D_{j,k} \psi_{j,k}
\]

which contain that portion of the series associated with scale \( s_j = \Delta x 2^{-j} \). We can also define a smooth for each intermediate scale index \( j = J \) and with physical scale \( 2^J \Delta x \); the smooth is the portion of the series not accounted for by the smaller-scale details

\[
S_J = y - \sum_{j=1}^{J} V_j.
\]

The smooth at any scale is the sum of the smooth and the detail from the next larger scale

\[
S_J = V_{J+1} + S_{J+1}.
\]

We can also define the rough at scale \( J \) as the difference between the original series and the scale \( J \) smooth; the rough is also the sum of all the lower scale details

\[
R_J = y - S_J = \sum_{j=1}^{J} V_J.
\]

The smooths are lowpass filtered versions of \( y \), the details are bandpass versions, and the roughs are highpass versions. The separation of the series into smooths, roughs, and details is called a multiresolution analysis [4].

In Fig. 2, we show the wavelet decomposition of a sample derived from a Landsat TM image of sea ice. (The decomposition in the figure is based on a mother wavelet filter of length \( m = 8 \) which will be introduced shortly.) The spikes correspond to leads (cracks in the ice). The observation vector \( y \) is shown at the top and the details, smooths, and roughs are shown for six levels of decomposition. Note how leads are reduced in magnitude and expand in width in the smooths, which appear as low-resolution versions of the original series. The largest-scale rough shows the leads as spikes, with the large-scale variations seen in the smooth removed. The details show the most variability at a scale of about \( j = 4 \) and include excursions in both directions.

With the Haar wavelet, the DWT is based on nonoverlapping differences. More information on the variability of the vector of observations is obtained by also considering the differences that overlap. For example with the \( j = 1 \) scale coefficients we used only \( y(x_2) - y(x_1) \) and \( y(x_4) - y(x_3) \) but additional information is obtained if we also include the \( y(x_3) - y(x_2) \) difference. This leads to what we call the maximal-overlap (MO) algorithm. It determines all possible differences at each scale and results in \( N \) coefficients for each scale. We will designate these coefficients as \( d^{ij,k} \), \( j = 1 \) to \( L \) and \( k = 1 \) to \( N \) (ignoring end effects). The DWT coefficients are a subset of the MO coefficients. Both sets of coefficients are efficiently computed with pyramid algorithms, both using the compact filters \( H \) and \( G \).

**B. Orthogonal-Pyramid Algorithm**

The orthogonal-pyramid algorithm for calculating the DWT was introduced by Mallat [5]. It is structured around the operations of highpass and lowpass filtering and subsampling-by-two. The algorithm uses the pair of filters \( H \) and \( G \) (the highpass mother wavelet filter and the lowpass scaling filter). The filters are quadrature mirror filters, in that \( G \) is the reverse
of $H$ with every second element multiplied by $-1$:

$$g_p = (-1)^p h_{m-p+1},$$

$$p = 1 \text{ to } m.$$  \hfill (8)

The orthogonal-pyramid algorithm consists of an iterative scheme in which at each step the wavelet and scaling coefficients are computed from the scaling coefficients of the previous step. The level 0 scaling coefficients for the first step are the elements of the observation vector, $A_0, k = y(x_k), k = 1 \text{ to } N$. The scaling coefficients are found by convolving $G$ with the smaller-scale coefficients $A_{j-1,k}$ and subsampling by two (decimation)

$$A_{j,k} = \sum_{p=1}^{m} g_p A_{j-1,(2k+m-1-p)};$$

$$k = 1 \text{ to } 2^{-j} N \hfill \hfill (9)$$

where again we have ignored complications from end effects. The two in the index $2k + m - 1 - p$ for the $j-1$ scaling coefficients accomplishes the subsample-by-two operation. The wavelet coefficients $D_{j,k}$ are found by convolving $H$ with the smaller-scale scaling coefficients $A_{j-1,k}$ again with decimation

$$D_{j,k} = \sum_{p=1}^{m} h_p A_{j-1,(2k+m-1-p)};$$

$$k = 1 \text{ to } 2^{-j} N.$$.  \hfill (10)

The discrete filters $H$ and $G$ have only a finite number $m$ of nonzero elements, so the convolutions are easily calculated. The vectors $A_j = \{A_{j,k}, k = 1 \text{ to } 2^{-j} N\}$ and $D_j = \{D_{j,k}, k = 1 \text{ to } 2^{-j} N\}$ are made up of weighted sums of the smaller-scale coefficients $A_{j-1,k}$; the set of weights ($H$ and $G$) are the same for all scales. Fig. 3 shows which elements of the observation vector contribute to each of the scaling coefficients using the Haar filter and how the number of scaling coefficients is reduced by roughly a factor of two with each step up in scale.

At each scale $j$ then, starting at the smallest scale, the signal is decomposed into a set of wavelet coefficients $D_{j,k}$ and a set of scaling coefficients $A_{j,k}$. The first decomposition has $N/2$ wavelet coefficients $D_{1,k}$ and $N/2$ scaling coefficients $A_{1,k}$ as well. Thus the total number of values retained is still $N$. The second decomposition is applied to $A_1$. It has $N/4$ wavelet coefficients $D_{2,k}$ and $N/4$ scaling coefficients $A_{2,k}$; the $A_{1,k}$ coefficients are no longer needed and the total number of values retained is still $N = N/2 + N/4 + N/4$. This process may be repeated until there is but one scaling coefficient $A_{L,1} = \sqrt{N} \hat{y}$.

For a complete decomposition, the number of observations $N$ must be a power of two because of the successive subsampling. In addition, if an accurate reconstruction is to be accomplished and the wavelet filter has a length greater than 2, boundary conditions must be assumed in order to calculate coefficients near the end of the series. Two common assumptions are: 1) that the series may be extended in a periodic fashion (e.g., $y(x_{N+i}) = y(x_i)$), where the series is assumed to be periodic with period $N$; or 2) in a symmetric fashion (e.g., $y(x_{N+i}) = y(x_{-i+1})$), where the series is reflected symmetrically about $i = N/2$. One can also extend the series with a constant, such as $y(x_{N+i}) = y(x_N)$. In an analysis context, such as presented in Section IV where data compression and regeneration are not the goals, a “brick wall” boundary condition may be used in which any convolution that extends beyond the end of the data series is not permitted. This boundary condition is appropriate in an analysis when there is no compelling reason to assume that the data are periodic or symmetric in structure.

The original signal can be exactly reconstructed, if all the coefficients are retained, by working back down the pyramid, successively computing the approximation at each smaller scale. At each scale, the set of wavelet coefficients and the set of scaling coefficients are both doubled in length by inserting one zero after each sample. The insert-one-zero (ioz) operation places a zero after each element in a vector, for example

$$\text{ioz} \{y_1, y_2, y_3, y_4\} = \{y_1, 0, y_2, 0, y_3, 0, y_4, 0\}.$$ \hfill (11)

The scaling coefficients at each smaller scale are built on convolutions of the larger-scale scaling and wavelet coefficients.
with the appropriate filters:

\[ A_{j, k} = \sum_{p=1}^{m} g_{m-p+1} 10^{j} [A_{j+1, (2k-p)}] \]

\[ + \sum_{p=1}^{m} h_{m-p+1} 10^{j} [D_{j+1, (2k-p)}] \]  \hspace{1cm} (12)

The procedure begins with the coefficients from the highest level decomposed, \( A_{L,k} \) and \( D_{L,k} \), and continues in an iterative fashion to the original series \( A_0 = y \). Note that the filters must be used in reverse order in the reconstruction.

The smooth \( S_j \) for scale \( j \) is found by following this procedure starting with the vector \( A_j \) and setting all lower-scale wavelet coefficients to zero; similarly the detail \( V_j \) is found by setting all the \( A_j \) to zero and all the wavelet coefficients to zero except those for the selected scale, the \( D_j \). In each case the reconstruction is carried out in the same iterative fashion until the series is of the original length \( N \). The wavelet basis vectors can also be obtained with the reconstruction procedure. The wavelet basis vector \( \psi_{j, k} \) is obtained by setting the single wavelet coefficient \( D_{j, k} = 1 \); all other wavelet and scaling coefficients are set to zero and the reconstruction procedure is followed.

The decomposition and reconstruction operations can be expressed as matrix operators. In the case of an \( m = 4 \) filter, the first decomposition would use two \( N/2 \times N \) matrices \( \mathcal{H}_1 \) and \( \mathcal{G}_1 \) operating on a vector of length \( N \), yielding a vector of length \( N/2 \). The matrix \( \mathcal{H}_1 \) is made up of the highpass filter elements

\[ \mathcal{H}_1 = \begin{bmatrix} h_4 & h_3 & h_2 & h_1 & 0 & 0 & \cdots \\ 0 & 0 & h_4 & h_3 & h_2 & h_1 & \cdots \\ 0 & 0 & 0 & 0 & h_4 & h_3 & \cdots \\ & & & & & & \end{bmatrix} \]  \hspace{1cm} (13)

while the matrix \( \mathcal{G}_1 \) is made up of the lowpass filter elements

\[ \mathcal{G}_1 = \begin{bmatrix} g_4 & g_3 & g_2 & g_1 & 0 & 0 & \cdots \\ 0 & 0 & g_4 & g_3 & g_2 & g_1 & \cdots \\ 0 & 0 & 0 & 0 & g_4 & g_3 & \cdots \\ & & & & & & \end{bmatrix} = \begin{bmatrix} h_1 & -h_2 & h_3 & -h_4 & 0 & 0 & \cdots \\ 0 & h_1 & -h_2 & h_3 & -h_4 & 0 & \cdots \\ 0 & 0 & 0 & h_1 & -h_2 & h_3 & \cdots \\ & & & & & & \end{bmatrix} \]  \hspace{1cm} (14)

The matrices for subsequent decompositions, \( \mathcal{G}_j \) and \( \mathcal{H}_j \), would be similar in structure and have dimensions \( 2^{-j+1} N \times 2^{-j+1} N \). The first scale scaling coefficients and wavelet coefficients are

\[ A_1 = \mathcal{G}_1 y \]  \hspace{1cm} (15)

\[ D_1 = \mathcal{H}_1 y \]  \hspace{1cm} (16)

while subsequent scale coefficients are

\[ A_j = \mathcal{G}_j A_{j-1} \]  \hspace{1cm} (17)

\[ D_j = \mathcal{H}_j A_{j-1} \]  \hspace{1cm} (18)

The matrix operators can also be used to construct the lowpass and band-pass filters that could be used to calculate the smooths and the details directly from \( y \) for scales greater than \( j = 1 \). The coefficients for scale \( j = 2 \) are

\[ A_2 = \mathcal{G}_2 G_1 y \]  \hspace{1cm} (19)

\[ D_2 = \mathcal{H}_2 G_1 y \]  \hspace{1cm} (20)

while the subsequent scale coefficients are

\[ A_j = (\mathcal{G}_j \mathcal{G}_{j-1} \mathcal{G}_{j-2} \cdots \mathcal{G}_1) y \]  \hspace{1cm} (21)

\[ D_j = (\mathcal{H}_j \mathcal{G}_{j-1} \mathcal{G}_{j-2} \cdots \mathcal{G}_1) y \]  \hspace{1cm} (22)

Each row in the matrix operators \( (\mathcal{G}_j, \mathcal{G}_{j-1}, \mathcal{G}_{j-2}, \cdots, \mathcal{G}_1) \) and \( (\mathcal{H}_j, \mathcal{G}_{j-1}, \mathcal{G}_{j-2}, \cdots, \mathcal{G}_1) \) has identical nonzero elements just as we saw for \( \mathcal{G}_1 \) and \( \mathcal{H}_1 \) and these elements constitute, respectively the lowpass filter \( \mathcal{G}_j \) and the band-pass filter \( \mathcal{H}_j \). The scale \( j \) smooth could then be obtained by convolving \( \mathcal{G}_j \) with \( y \) and the scale \( j \) detail by convolving \( \mathcal{H}_j \) with \( y \).

The matrix formulation of the reconstruction is based on the orthogonality of the matrices \( \mathcal{G}_j \) and \( \mathcal{H}_j \). The first level decomposition may be written as

\[ \begin{bmatrix} A_1 \\ D_1 \end{bmatrix} = \begin{bmatrix} \mathcal{G}_1 \\ \mathcal{H}_1 \end{bmatrix} y \]  \hspace{1cm} (23)

and since the filters are orthonormal,

\[ [\mathcal{G}_1^T \mathcal{H}_1^T] \begin{bmatrix} \mathcal{G}_1 \\ \mathcal{H}_1 \end{bmatrix} = I. \]  \hspace{1cm} (24)

The final step in the reconstruction can then be expressed as

\[ y = \mathcal{G}_1^T A_1 + \mathcal{H}_1^T D_1. \]  \hspace{1cm} (25)

In a reconstruction starting from a larger scale, the procedure steps down the scales

\[ A_{j-1} = \mathcal{G}_j^T A_j + \mathcal{H}_j^T D_j. \]  \hspace{1cm} (26)

C. Maximal-Overlap Pyramid Algorithm

The maximal-overlap decomposition can also be calculated with a pyramid scheme introduced by Percival and Guttrop [6]. This method uses the constant-length filters \( H \) and \( G \) but applies them \( 2^j-1 \) times at each level of decomposition, each time to a subsampled-by-\( 2^j-1 \)-and-shifted version of the approximation from the previous scale. For example, if we start with

\[ y = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, \cdots \} \]  \hspace{1cm} (27)

then the series is convolved with \( H \) and \( G \) (but not subsampled) to give

\[ a_1 = \{a_{1,1}, a_{1,2}, a_{1,3}, a_{1,4}, a_{1,5}, a_{1,6}, a_{1,7}, a_{1,8}, \cdots \} \]  \hspace{1cm} (28)

\[ d_1 = \{d_{1,1}, d_{1,2}, d_{1,3}, d_{1,4}, d_{1,5}, d_{1,6}, d_{1,7}, d_{1,8}, \cdots \} \]  \hspace{1cm} (29)

(again ignoring end effects). The vector \( a_1 \) is then separated into two series, each a subsampled and shifted-by-zero or -one version of \( a_1 \),

\[ \{a_{1,1}, a_{1,3}, a_{1,5}, a_{1,7}, \cdots \} \]  \hspace{1cm} (30)

\[ \{a_{1,2}, a_{1,4}, a_{1,6}, a_{1,8}, \cdots \} \]  \hspace{1cm} (31)
each is filtered separately with \( H \) and \( G \), and then interleaved to form the full sets of scaling coefficients and wavelet coefficients

\[
a_2 = \{a_{2,1}, a_{2,2}, a_{2,3}, a_{2,4}, a_{2,5}, a_{2,6}, a_{2,7}, a_{2,8}, \ldots \} \quad (31)
\]
\[
d_2 = \{d_{2,1}, d_{2,2}, d_{2,3}, d_{2,4}, d_{2,5}, d_{2,6}, d_{2,7}, d_{2,8}, \ldots \}. \quad (32)
\]

For the third scale, \( a_2 \) is divided into four series, each shifted by zero, one, two, or three, and again filtered by \( H \) and \( G \) and interleaved. Fig. 3 shows how the scale 1 coefficients are found with a single convolution of the observation vector, the scale 2 coefficients are found with two interleaved convolutions of the scale 1 coefficients, and the scale 3 coefficients are found from four interleaved convolutions of the scale 2 coefficients. In addition, the heavy solid lines indicate which of the maximal-overlap coefficients constitute the set of DWT coefficients.

In general, let \( q \) be the series of indices of the original data structure that are subsampled and shifted into \( 2^{j-1} \) shorter series: \( q = r, r + 2^{j-1}, r + 2(2^{j-1}), r + 3(2^{j-1}), \ldots \), where \( r = 1 \) to \( 2^{j-1} \). Then the scaling coefficients for each \( r \) at scale \( j \) are found by convolving \( G \) with the smaller-scale coefficients \( a_{j-1,k} \)

\[
a_{j,q} = \sum_{p=1}^{m} g_p a_{j-1,q-2^{j-1}p}. \quad (33)
\]

This convolution is performed \( 2^{j-1} \) times, once for each value of \( r \), and the results interleaved to make a new series \( a_{j,k} \) of length \( N \). The wavelet coefficients are found in a similar manner by convolving \( H \) with the smaller-scale scaling coefficients

\[
d_{j,q} = \sum_{p=1}^{m} h_p a_{j-1,q-2^{j-1}p} \quad (34)
\]

and interleaving the results. At each stage the sets of scaling coefficients and the wavelet coefficients both retain \( N \) values, although points near the edges depend on how boundary conditions are handled. This algorithm requires \( O(N \log_2 N) \) multiplications (the same as the FFT) while the orthogonal-pyramid algorithm requires only \( O(N) \).

**D. Wavelet Families**

There are a variety of wavelet and scaling filters that satisfy the crucial orthonormality requirements. Here we are only interested in discrete wavelet filters with compact support; in other words those in which the mother wavelet filters and scaling filters have finite length. A wavelet family consists of all the wavelet basis vectors for all scales and locations derived from a single mother wavelet filter. Different wavelet families make different trade-offs between the degree of localization and the degree to which ideal highpass filters are approximated, which always adhere to the orthonormality conditions outlined above. In Fig. 4 we have plotted the nonzero elements of the wavelet basis vectors for scales \( j = 1 \) and \( j = 2 \) for four different Daubechies wavelet families. Note how the number of nonzero elements increases as the scale increases and how the vectors appear less jagged as well. The most localized wavelet is the Haar, with but two elements; it is the worst highpass approximation.

Daubechies [2] has developed the theory for obtaining higher order mother wavelet filters with compact support and identified two sets of filters, the extremal-phase and the least-asymmetric. These filters have even length \( m \) between 2 and 20, but are identified not by the length but by the number of vanishing moments, \( n = m/2 \). The moment of order \( \lambda \) of a wavelet filter is defined as

\[
M_{\lambda} = \sum_{p=1}^{m} p^{\lambda} h_p. \quad (35)
\]

As the number of vanishing moments increases the wavelet filter become longer and its approximation to an ideal highpass filter improves. The order of the filter is equal to the number of vanishing moments and in the case of the Daubechies filters it is half the length of the filter.

The wavelet basis vectors derived from filters of order \( n \) are orthogonal to polynomials of order \( n - 1 \). This means that trends in the series of order \( n - 1 \) are not represented in the wavelet coefficients, again ignoring end effects. For example, the Haar wavelet (order \( n = 1 \)) is orthogonal to a constant (zeroth order polynomial) but not orthogonal to a linear trend (first-order polynomial), so a linear trend in \( y(x) \) will offset the wavelet coefficients by a constant related to the magnitude of the slope. The Daubechies second-order wavelet basis vectors are orthogonal to such a trend so that the wavelet coefficients would not be offset due to the presence of the linear trend. However, a second-order trend would again offset the wavelet coefficients.
The degree of symmetry in a wavelet is important in reducing the phase shift, or shift in position, of features during the decomposition. The Daubechies least-asymmetric family was designed to have smaller shifts than the extremal-phase family. The phase shift of the Haar wavelet filter is very large and can lead to distortions in the location of features in the transform coefficients. Filters of the same length in the two sets of families have transfer functions whose squared moduli are identical but whose phase shifts are different.

In Fig. 5 the squared modulus of the transfer function for the wavelet filters \( (H_1 \text{ and } H_2) \) and scaling filters \( (G_1 \text{ and } G_2) \) are shown for the same four families and two scales as in Fig. 4. Ideally, the wavelet filter would be a perfect highpass \((j = 1)\) or band-pass \((j > 1)\) filter, yet the Haar shows significant leakage from both lower and higher frequencies (cross-hatched areas), with a lesser amount of leakage seen in the other wavelet families. Percival and Guttorp [6] have illustrated that this leakage can lead to a significant bias in estimates of the power spectrum computed via the Haar wavelet variance.

### III. WAVELET VARIANCE AND COVARIANCE

#### A. Wavelet Variance

The orthonormal wavelet decomposition of an observation vector leads to a natural partition of the variance by scale. This is best seen in examining the sample variance of the vector \( y \)

\[
\sigma_y^2 = \frac{1}{N} \sum_{i=1}^{N} [y(x_i) - \bar{y}]^2 = \frac{1}{N} \sum_{i=1}^{N} [y(x_i)]^2 - \bar{y}^2
\]

where \( \bar{y} \) is the sample mean. Because the wavelet basis vectors are orthonormal, the sum of the squares of the series can be expressed as the sum of the squares of the wavelet coefficients

\[
\sum_{i=1}^{N} [y(x_i)]^2 = \sum_{j=1}^{L} \sum_{k=1}^{n_j} D_{j,k}^2 + Ny^2
\]

so the variance may be expressed in terms of the wavelet coefficients

\[
\sigma_y^2 = \frac{1}{N} \sum_{j=1}^{L} \left\{ \sum_{k=1}^{n_j} D_{j,k}^2 \right\}.
\]

The contribution to the variance of \( y \) associated with each scale \( j \) is then

\[
\sigma_{y,j}^2 = \frac{n_j}{N} \left( \frac{1}{n_j} \sum_{k=1}^{n_j} D_{j,k}^2 \right) = \frac{\sigma_D_{j,j}^2}{2^j}
\]

since \( n_j = N/2^j \), where \( \sigma_D_{j,j}^2 \) is the sample variance of the wavelet coefficients \( D_{j,k} \) at scale \( j \) (assuming their sample mean is zero). We call \( \sigma_{y,j}^2 \) the sample wavelet variance; it is also equal to the sample variance of the detail \( V_j \).

End effects are sometimes handled in an analysis context by establishing a “brick wall” boundary condition, which prohibits convolutions that extend beyond the ends of the series. If such a condition is imposed, then the number of coefficients is reduced and the DWT sample estimate of the wavelet variance is

\[
\tilde{\sigma}_{y,j}^2 = \frac{1}{2^j \tilde{n}_j} \sum_{k=1}^{n_j} D_{j,k}^2
\]

where the number of DWT coefficients at scale \( j \) is obtained from the formulas

\[
\tilde{n}_1 = \frac{N}{2} - \left( \frac{m-2}{2} \right)
\]

\[
\tilde{n}_2 = \frac{N}{4} - \left( \frac{m-2}{2} + \frac{m}{4} \right)
\]

\[
\tilde{n}_3 = \frac{N}{8} - \left( \frac{m-2}{2} + \frac{m+2}{4} \right)
\]

\[
\tilde{n}_j = \frac{N}{2^j} - (m-2), \quad j \geq 4.
\]
LINDSAY et al.: DISCRETE WAVELET TRANSFORM AND SCALE ANALYSIS

Here \( \lfloor x \rfloor \) refers to the greatest integer less than or equal to \( x \), the \(^\sim\) hat refers to DWT sample estimates with a "brick wall" boundary condition. The hat used below refers to the MO sample estimates with a "brick wall" boundary condition.

With the MO method, the variance associated with each scale can be more accurately determined by including more information about the variations of the signal over distances equal to the scale \( s_j = \Delta x 2^{-j} \). As we saw before for the Haar wavelet, the wavelet coefficients in the DWT at scale \( j = 1 \) are the differences in \( y(x_i) \) between points \( i = 1 \) and 2, 3 and 4, 5 and 6, and so on, while with the MO method we also include the differences between points 2 and 3, 4 and 5, and so on. The MO sample estimate of the wavelet variance at scale \( j \) is given by

\[
\hat{\sigma}_{y,j}^2 = \frac{1}{2N} \sum_{k=1}^{n_j} d_{j,k}^2.
\]

This is the sample variance of the MO wavelet coefficients (assuming their sample mean is zero), scaled by \( 2^j \). For the "brick wall" boundary condition, the quantity

\[
n_j = N - m_j + 1
\]

is the number of MO coefficients at scale \( j \), where

\[
m_j = (2^j - 1)(m - 1) + 1
\]

is the length of the filters \( H_j \) and \( G_j \).

B. Wavelet Covariance

In many geophysical applications the spatial or temporal structure of the covariance between two signals is of interest. This spatial structure and a scale analysis of the covariance can also be illuminated with the wavelet formalism. The wavelet covariance was compared to the Fourier cross spectra by Hugdins et al. [7] using atmospheric surface-layer measurements of the horizontal and vertical velocities (stress) and the vertical velocity and temperature (heat flux). They used nonorthogonal cubic-spline wavelets and found good agreement between the methods but concluded that the wavelet covariance provided a better analysis of the low frequency variations.

The sample covariance of two series \( y(x_i) \) and \( z(x_i) \) is

\[
\hat{\sigma}_{yz} = \frac{1}{N} \sum_{i=1}^{N} \left[ y(x_i) - \bar{y} \right] \left[ z(x_i) - \bar{z} \right]
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} y(x_i) z(x_i) - \bar{y} \bar{z}
\]

\[
= \frac{\langle y, z \rangle}{N} - \bar{y} \bar{z}.
\]

The inner product of two vectors can be expressed in terms of their wavelet decompositions as

\[
\langle y, z \rangle = \left( \sum_{j=1}^{L} \sum_{k=1}^{n_j} D_{j,k}^y \psi_{j,k} + \bar{y} \right) \left( \sum_{j=1}^{L} \sum_{k=1}^{n_j} D_{j,k}^z \psi_{j,k} + \bar{z} \right).
\]

B. Wavelet Covariance

In many geophysical applications the spatial or temporal structure of the covariance between two signals is of interest. This spatial structure and a scale analysis of the covariance can also be illuminated with the wavelet formalism. The wavelet covariance was compared to the Fourier cross spectra by Hugdins et al. [7] using atmospheric surface-layer measurements of the horizontal and vertical velocities (stress) and the vertical velocity and temperature (heat flux). They used nonorthogonal cubic-spline wavelets and found good agreement between the methods but concluded that the wavelet covariance provided a better analysis of the low frequency variations.

The sample covariance of two series \( y(x_i) \) and \( z(x_i) \) is

\[
\hat{\sigma}_{yz} = \frac{1}{N} \sum_{i=1}^{N} \left[ y(x_i) - \bar{y} \right] \left[ z(x_i) - \bar{z} \right]
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} y(x_i) z(x_i) - \bar{y} \bar{z}
\]

\[
= \frac{\langle y, z \rangle}{N} - \bar{y} \bar{z}.
\]

The inner product of two vectors can be expressed in terms of their wavelet decompositions as

\[
\langle y, z \rangle = \left( \sum_{j=1}^{L} \sum_{k=1}^{n_j} D_{j,k}^y \psi_{j,k} + \bar{y} \right) \left( \sum_{j=1}^{L} \sum_{k=1}^{n_j} D_{j,k}^z \psi_{j,k} + \bar{z} \right).
\]

Hence, in light of the orthonormality properties of the vectors \( \psi_{j,k} \), the sample covariance of the series may be written in terms of the wavelet coefficients

\[
\hat{\sigma}_{yz} = \sum_{j=1}^{L} \left\{ \frac{1}{N} \sum_{k=1}^{n_j} D_{j,k}^y D_{j,k}^z \right\}
\]

where the contribution to the covariance associated with each scale \( j \) is then

\[
\hat{\sigma}_{yz,j} = \frac{1}{N} \sum_{k=1}^{n_j} D_{j,k}^y D_{j,k}^z
\]

\[
= \frac{1}{N} \langle D_j^y, D_j^z \rangle.
\]

If a "brick wall" boundary condition is imposed the DWT estimate is

\[
\hat{\sigma}_{yz,j} = \frac{1}{2N} \sum_{k=1}^{n_j} D_{j,k}^y D_{j,k}^z
\]

Just as with the variance, the MO method allows a more accurate determination of the covariance associated with each scale. The MO covariance estimate at scale \( j \) is given by

\[
\hat{\sigma}_{yz,j} = \frac{1}{2N} \sum_{k=1}^{n_j} d_{j,k}^y d_{j,k}^z
\]

where \( d_{j,k}^y \) and \( d_{j,k}^z \) are the MO wavelet coefficients. The estimate of the scale-dependent wavelet cross correlation can then be written as

\[
R_j = \frac{\hat{\sigma}_{yz,j}}{\hat{\sigma}_{y,j} \hat{\sigma}_{z,j}}.
\]

The wavelet correlation is analogous to the coherence determined in Fourier cross-spectral analysis.

C. Confidence Intervals

If \( y \) is considered to be a realization of a stationary process with expected value \( \mu = E[y] \) and variance \( \sigma_y^2 = E[(y - \mu)^2] \), the wavelet coefficients can also be considered to be realizations of \( L \) different stationary processes, one for each scale, each with an expected value of zero and a variance \( 2^j \sigma_y^2 \). The confidence intervals for \( \hat{\sigma}_{y,j}^2 \) can then be determined based on the sample estimates \( \hat{\sigma}_{y,j}^2 \) and \( \hat{\sigma}_{y,j}' \). Percival [8] has developed the theory for determining the uncertainty in the wavelet variance estimate for wavelet filters of various lengths under a Gaussian assumption. The estimates given above of \( \hat{\sigma}_{y,j}^2 \) and \( \hat{\sigma}_{y,j}' \) are unbiased and asymptotically normally distributed. The variance of the variance estimates is proportional to the integral of the spectral density squared \( P_{D,j} \) of the wavelet coefficients. For the DWT estimate it is approximately

\[
\text{var}(\hat{\sigma}_{y,j}^2) = \frac{P_{D,j}}{2^{2j-1} n_j}
\]

and for the MO method it is

\[
\text{var}(\hat{\sigma}_{y,j}^2) = \frac{P_{D,j}}{2^{2j-1} n_j}.
\]
The integral of the spectral density squared can be estimated from the autocovariance functions \( a_{r,D} \) or \( a_{r,d} \) of the wavelet coefficients

\[
\hat{P}_{D,j} = \frac{1}{2} \int_{-1/2}^{1/2} |S_{D_1}(f)|^2 \, df
\]

or

\[
\hat{P}_{d,j} = \frac{1}{2} \int_{-1/2}^{1/2} |S_{d_1}(f)|^2 \, df
\]

(57)

The error variance of the two methods can be compared with the asymptotic relative efficiency of the MO estimator relative to the DWT estimator

\[
\varepsilon_j = \lim_{n \to \infty} \frac{\text{var}(\tilde{\sigma}_{y,s,j}^2)}{\text{var}(\tilde{\sigma}_{y,s,j}^2)}
\]

(59)

Percival [8] determined that \( \varepsilon_j \) is always less than unity and can, in fact, approach one half for processes typical for those encountered in geophysics. We will now concentrate only on the more accurate MO estimator.

Under the restrictive assumption that the estimate of \( \hat{P}_{d,j} \) is close to its true value, then \( \tilde{\sigma}_{y,s,j}^2 \) approximately follows a chi-squared distribution and an approximate 100(1 - 2\( \eta \))% confidence interval for \( \tilde{\sigma}_{y,s,j}^2 \) is

\[
\left[ \frac{\eta \tilde{\sigma}_{y,s,j}^2}{Q_{\eta}(1-p)}, \frac{\eta \tilde{\sigma}_{y,s,j}^2}{Q_{\eta}(-p)} \right]
\]

(60)

where \( Q_{\eta} \) is the \( \eta \times 100\% \) point for the \( \chi^2 \) distribution with \( \eta \) degrees of freedom. The equivalent number of degrees of freedom is given by

\[
\eta = \frac{2E(\tilde{\sigma}_{y,s,j}^2)}{\text{var}(\tilde{\sigma}_{y,s,j}^2)}
\]

\[
= \frac{\tilde{\sigma}_{y,s,j}^2 / n_j}{\hat{P}_{d,j}}
\]

(61)

for \( \eta \) greater than about 30. For a smaller \( \eta \), a more accurate estimate is given by

\[
\eta \approx \frac{n_j}{2j^2}
\]

(62)

As we saw above, the wavelet decomposition of the covariance is determined by the product of the coefficients from the two decompositions performed separately. The MO estimator \( \tilde{\sigma}_{y,s,j} \) is asymptotically normally distributed with mean \( \sigma_{y,s,j} \equiv 2^{-j} E\{d_{j,k}^y \} \) and variance

\[
\text{var}(\tilde{\sigma}_{y,s,j}) = \frac{1}{2^{2j+1} n_j} \left[ \int_{-1/2}^{1/2} |S_{d_1,y}(f)|^2 \, df \right]
\]

\[
+ \int_{-1/2}^{1/2} |S_{d_1}(f)S_{d_1}(f)| \, df
\]

(63)

where \( S_{d_1,y}(f) \) is the cross spectrum of the coefficients \( S_{d_1}(f) \) and \( S_{d_1}(f) \) are the autospectra. The integral of the cross spectrum squared is determined from the covariance function \( a_{r,d_1}a_{r,d_1} \) of the wavelet coefficients \( d_{j,k}^y \) and \( d_{j,k} \)

\[
\int_{-1/2}^{1/2} |S_{d_1,y}(f)|^2 \, df = \sum_{r=-n_j+1}^{n_j-1} a_{r,d_1}a_{r,d_1}^*
\]

(64)

and the integral of the product of the spectra from the autocovariance

\[
\int_{-1/2}^{1/2} |S_{d_1}(f)S_{d_1}(f)| \, df = \sum_{r=1}^{n_j-1} a_{r,d_1}a_{r,d_1}^* + a_{0,d_1}a_{0,d_1}^*
\]

(65)

Again, under the restrictive assumption that the spectral estimates are close to the true values, an approximate 100(1 - 2\( \eta \))% confidence interval for \( \tilde{\sigma}_{y,s,j} \) is based in this case on a normal distribution

\[
\left[ \tilde{\sigma}_{y,s,j} - \Phi^{-1}(1-p) \sqrt{\text{var}(\tilde{\sigma}_{y,s,j})}, \tilde{\sigma}_{y,s,j} + \Phi^{-1}(1-p) \sqrt{\text{var}(\tilde{\sigma}_{y,s,j})} \right]
\]

(66)

where \( \Phi^{-1}(p) \) is the \( p \times 100\% \) point for the standard normal distribution for a 95% confidence interval \( \Phi^{-1}(0.05) = 1.95 \).

IV. ANALYSIS OF SEA-ICE SURFACE PROPERTIES

A. Albedo and Temperature Variance and Covariance

Examples of the wavelet decomposition and estimation of the wavelet variance are taken from a Landsat TM image of pack ice. The data are from a single line from a Landsat TM image, channels 3 and 6, which was obtained on April 16, 1992 in the Beaufort Sea. The pixel spacing \( \Delta x \) is 25 m, and there are 8428 points in the line. The first example (Fig. 6) is of the surface albedo as measured by channel 3, for which the instrument resolution is a nominal 30 m and the spectral interval is 0.63-0.69 \( \mu \)m. The first panel shows the measured albedo; the digital counts have been scaled to give a mean albedo of 0.70. The spikes of low brightness correspond to leads. Leads are narrow, irregular cracks in the thick ice and consist of thin ice and sometimes some open water. The second panel shows the multiresolution analysis of the series with details for eight levels of decomposition, \( V_1 \) through \( V_8 \), and the level 8 smooth, \( S_8 \). The observation vector is the sum of all eight of the details and the smooth. Note that the leads correspond to areas of high variability in the detail signals. As we will see below, the greatest variability is at a scale of 400 m, \( j = 5 \). In this and all subsequent wavelet decompositions we have used the Daubechies least asymmetric wavelet filter no. 4, DLA4.

The second example (Fig. 7) is of the thermal channel (channel 6, 10.4-12.5 \( \mu \)m) from the same Landsat TM image line. The point separation is still 25 m but the nominal resolution of the thermal channel is only 120 m. Here, not
only is the spatial resolution reduced but the digital resolution of the temperature estimate is large (0.7 K). No adjustments for atmospheric interference have been applied. Leads are manifested as warm spikes, all of which have corresponding dark spikes in the channel 3 line. The low spatial resolution is reflected in the much lower variability at small scales compared to that seen in the larger scales. At the smallest scale detail, leads are not seen at all. The variability at the smallest scale is dominated by jumps in the temperature associated with the discretization of the temperature in the digital signal. In the temperature series the greatest variability is at a scale of 800 m, \( j = 6 \).

In the top panels of Figs. 8 and 9 the MO wavelet variance is shown for the two series with 95% confidence intervals indicated. In each case the variance is normalized by the total variance of the series. Note how the variance-versus-scale curves both show a broad peak at scales between about 200 m and 800 m, but that there is much less variability at the smallest scales in the temperature record due to the larger averaging area of the temperature sensor. The sample standard deviation of the albedo series is 0.042, and 54% of the total variance is accounted for by scales \( j = 4 \) to 6 (200–800 m). The sample standard deviation for the temperature series is 1.1°C and only 45% of the variance is accounted for by scales \( j = 4 \) to 6 due in part to the lower spatial resolution of the thermal sensor and in part to the greater temperature variance seen at very large scales. Also shown in these two figures is the cumulative sum of the squares over distance of the wavelet coefficients (times 2^\( -j \)) for scales \( j = 3 \) to \( j = 8 \). In this type of presentation it is clear that the total variance is dominated by the two leads near kilometers 110 and 120, that just these two leads contribute more than half of the variance at all scales above 400 m, and that the proportion of the variance they contribute increases with scale. Note how the larger scales are truncated at the ends due to the “brick wall” boundary conditions imposed in the analysis.

In Fig. 10, an example is seen of the wavelet covariance between the albedo and temperature records from the Landsat image line in Figs. 6 and 7. The two channels are anticorrelated, with much of the variance found in the contrast between warm, dark leads and cold, bright ice, so for convenience we plot the covariance of the albedo and the negative of the temperature. Shown is the wavelet covariance \( \sigma_{T_a,T_t} \) with 95% confidence intervals and the cumulative sums of the product of the albedo and temperature wavelet coefficients (times 2^\( -j \)). Again we can see the broad peak in the covariance at scales of 400–800 m and the domination of the covariance by the two leads near the middle of the line.

With so much of the variability in a single line contributed by just two leads, we can expect a great deal of change in the wavelet variance as we move to other transects in the image. Indeed this is the case. We have determined the variance for

![Fig. 6. Wavelet decomposition of the surface albedo of sea ice for a single line of a Landsat image derived from channel 3: (top) The albedo and (bottom) the level 8 smooth and the details for levels \( j = 1 \) to 8. The vertical scales in all of the plots are the same.](image)

![Fig. 7. Wavelet decomposition of the surface temperature of sea ice for a single line of a Landsat image derived from channel 6: (top) The observed temperature and (bottom) the level 8 smooth and the details for levels \( j = 1 \) to 8.](image)
five separate lines from the same Landsat TM image. The parallel lines are separated by 25 km, span a distance of 100 km, and contain between 7014–8428 points per line. Fig. 11 shows the variance, covariance, and correlation for channels 3 and 6 for each of the lines. Here we have plotted the variance, not the normalized variance, in order to highlight the range of total variability which changes dramatically from one portion of the image to another. Although they all show a maximum in the variance at midsize scales, the position of the maximum and the shape of the curves vary considerably. The correlation curves fall into two categories: one with high correlation at scales larger than about 0.2 km (three lines) and another that shows lower correlation at scales from 0.1–0.4 km and the same high correlation at larger scales. The lines with low correlation at the intermediate scales also show low variance in both the albedo and the temperature. These are portions of the image with fewer leads. In all of the lines half of the total variance is found in scales at and below 800 m in both channels.

B. Surface Temperature Enhancement

Calculations of various lead properties, such as the number of leads, the lead widths, and the sensible heat flux from leads, depend critically on the spatial resolution of the measurements. While lead number density and lead widths can be estimated from the albedo measurements, the sensible heat flux requires the thermal measurements. The relatively low resolution of the Landsat thermal sensor seriously degrades these estimates. However, the wavelet formalism allows us to make a scale-dependent enhancement of the channel 6 surface temperature measurements based on the high correlation between the albedo and the temperature at large scales. This enhancement thus extends the effective resolution of the temperature mea-
LINDSAY et al.: DISCRETE WAVELET TRANSFORM AND SCALE ANALYSIS

Fig. 10. (a) The wavelet covariance of the surface albedo and the negative of the temperature for a single line of the Landsat image with 95% confidence intervals indicated. (b) The cumulative sum of the product of the wavelet coefficients for scales \( j = 3 \) to 8.

Measurements to smaller scales. The fall-off of the correlation between the albedo and the temperature at small scales seen in Fig. 11 is a result of the low resolution of the thermal infrared sensor compared to that of the shortwave sensor. A higher-resolution estimate of the surface temperature may be obtained by assuming the relationship between albedo and temperature at larger scales is maintained in the small scales. The temperature is considered to be determined by the smooth of the measured temperature at some level \( J \) and a weighted sum of the details of the temperature and the renormalized albedo at the smaller scales. The model is

\[
T = S^T_J + \sum_{j=1}^{J} [w_j V^T_j + (1 - w_j)\hat{b} \nu^\alpha_j]
\]

where

\[
\hat{D}^T_{j,k} = w_j D^T_{j,k} + (1 - w_j)\hat{b} D^\alpha_{j,k}.
\]

Using the smooth of the measured temperature insures that the large scale thermal structure is maintained, a structure which is well measured by the thermal sensor. The normalization factor \( \hat{b} \) is determined with a least-squares linear regression between the wavelet coefficients for temperature and albedo at a reference scale \( J + 1 \)

\[
D^T_{j+1,k} = \hat{b} D^\alpha_{j+1,k}.
\]

No intercept is needed because the sample means of the coefficients are close to zero. The set of weights is determined from the observed scale-dependent cross correlation \( R^2_j \) via

\[
w_j = \frac{R^2_j}{1 + R^2_j}.
\]

Thus, if the correlation between \( D^T_{j,k} \) and \( D^\alpha_{j,k} \) is unity (indicating that there is no noise in the temperature coefficients) then \( w_j = 0.5 \); on the other hand if the correlation is zero (indicating that \( D^T_{j,k} \) is totally dominated by noise), then \( w_j = 0 \). A simpler procedure that gives essentially the same results is to set each \( w_j \) to zero and reconstruct the temperature from the scale \( J \) temperature smooth and the smaller scale albedo details, i.e.,

\[
T = S^T_J + \sum_{j=1}^{J} \hat{b} V^\alpha_j.
\]

The reference scale \( J = 5 \) (400 m) is selected to correspond to the scale of maximum covariance. The technique is based on the assumptions that the temperature and the albedo are linearly related and that this relationship is the same for all scales below the reference scale. The complex relationship between temperature and albedo is affected by ice thickness, snow depth, wind speed, cloud cover, and sun angle, all of which change over time; however, the Landsat data (and other observations) clearly show a strong relationship, and a linear approximation of this relationship can be used to obtain an approximate estimate of the fine resolution temperature signal.

Fig. 12 shows the reconstructed enhanced temperature for the line of the Landsat data shown in Fig. 7. The use of the channel 3 wavelet coefficients in determining the small scales for channel 6 forces the surface temperature wavelet variance to have a similar shape at small scales to that of the albedo. The sample variance is increased 7% in the enhanced series (the sample standard deviation increased from 1.10-1.14 °C). As we will show below, when we use the enhanced temperature the estimated number density of leads is increased, the estimated mean width is decreased, while simultaneously the estimated lead concentration remains nearly the same. In addition, estimates of the regional average sensible heat flux from leads for this image is increased by 17%.
 resolution instruments may be simulated using a wavelet-derived smooth of the data, which is a lowpass version of the observations. This simulation does not account for the point spread functions of individual instruments. The basic idea is to decompose the observations on a scale-by-scale basis using the wavelet decomposition and to assume that data from a low-resolution instrument can be simulated by setting the small-scale wavelet coefficients to zero—this procedure amounts to substituting the wavelet smooth at a certain scale for the observation vector. Using the smooths gives us insight on the scales of variability that exist in the observations that are important for derived quantities of interest. For example, if a derived quantity does not change with a step up in smooth scale, the quantity is insensitive to the variability of the observations at the smaller smooth scale. These results are summarized in what we call a scale dependency curve showing the value of the derived quantity as a function of the scale of the smooth used as input.

We show some examples of lead statistics for the five sample lines. Leads are identified as regions of relatively warm ice.
They are defined in terms of the potential open water \[ T_{\text{floes}}(22) \]

\[ T_{\text{floes}}(22) = C \left( T_{\text{floes}}(22) - T_{\text{floes}}(22) \right) \]

We use a span \( \zeta = 750 \text{ m} \), which is more than twice as big as the largest leads. Leads are taken to be regions in which \( \delta \)

\[ \delta(x_i) = \frac{T(x_i) - T_{\text{floes}}(x_i)}{T_{\text{floes}}(x_i)} \]  (72)

where \( T_{\text{floes}} = -1.8 \text{ °C} \) is the temperature of open water. The floe temperature is obtained from a running median filter

\[ T_{\text{floes}}(x_i) = \text{median}_\zeta \left[ T(x_i) \right]. \]  (73)

The mean lead width can be computed from the density distribution of lead widths (number weighted) or weighted by the area of the leads (area weighted). The scale dependency curves for the two methods are shown in Fig. 13(b). The number density computed with the observed temperature is insensitive to the smooth used, because the instrument resolution is 130 m; however, the number density computed with the enhanced temperature continues to increase at the smallest scales. We also see by extrapolation that the number density drops to near zero for smooth scales above 1 km.

The mean lead width is the number of leads per 100 km of transect, while the width \( w \) of a lead is found from the length of a run of points above the threshold. These two quantities depend strongly on the level of the smooth used. Fig. 13(a) shows the scale dependency curve for the number density, which is estimated from both the observed and the enhanced temperatures. The lead number density using the finest scale (25 m, no smoothing) is twice as large using the enhanced temperature instead of the observed temperature. Using larger-scale smooths of the two signals give the same result, because the enhancement only affects the smallest scales. At scales below 100 m the number density computed with the observed temperature continues to increase at the smallest scales. We also see by extrapolation that the number density drops to near zero for smooth scales above 1 km.

The best mother wavelet filter to use for a particular application depends on a number of trade-offs. The length of the filter affects the size of the largest scale that can be analyzed. Longer filters reduce the accuracy of the wavelet variance estimate because they allow for fewer independent samples. The filters used to establish the wavelet coefficients are not perfect band pass filters. For this reason, they allow energy in the signal from higher and lower frequencies to be included in the energy represented in the coefficients for a particular scale. The transfer function of the filter indicates how much spectral leakage is expected and, as a consequence, if a possible bias in the estimated wavelet variance relative to Fourier estimates may occur. In some applications, the phase shift introduced by the filter is important when the exact location of features is significant to the analysis. The more

V. CONCLUSION

We have outlined the formalism of the one-dimensional discrete wavelet transform based on Daubechies wavelet filters in terms of finite vectors and matrices. We have shown how both the scale-dependent wavelet variance and wavelet covariance are easily determined from the DWT and how the confidence intervals for each are determined. The variance estimates are more accurately determined with a maximal-overlap version of the wavelet transform. We discussed the properties of several Daubechies wavelet filters and the associated basis vectors. Both the Mallat orthogonal-pyramid algorithm for determining the DWT and a pyramid algorithm for determining the maximal-overlap version of the transform were presented in terms of finite vectors.

The transfer function of the filter indicates how much spectral leakage is expected and, as a consequence, if a possible bias in the estimated wavelet variance relative to Fourier estimates may occur. In some applications, the phase shift introduced by the filter is important when the exact location of features is significant to the analysis. The more
symmetric filters, such as the Daubechies least asymmetric, reduce the amount of the shift. For many applications, the results of the scale analysis are nearly identical for any of the wavelet families (excluding the Haar) and the shortest family is appropriate. In our analysis of Landsat sea-ice data we have used the DLA4 wavelet filter as a compromise between the different criteria: it is the shortest of the least asymmetric wavelet filters, there is only a modest amount of leakage, and the phase shift is small.

A simple scale analysis of the variability of a field, as performed in this study, does not require a two-dimensional data set unless the field is significantly anisotropic. If this is the case, the data slices along different directions could be analyzed independently or the two-dimensional field could be analyzed as a whole. The wavelet decomposition can be extended to two dimensions. In the two-dimensional procedure the orthogonal two-dimensional wavelet basis matrices are each associated with a scale, a location, and one of three orientations: horizontal, vertical, and diagonal. An efficient decomposition algorithm based on one-dimensional wavelet filters was introduced by Mallet [5]. Applications of the two-dimensional decomposition include texture analysis, image smoothing, image compression, edge detection, orientation analysis, and computer vision.

In addition to the analysis of variance and the compression of data, the wavelet decomposition offers a unique method of spatial-adaptive smoothing of a signal that contains large discontinuities, such as those obtained over sea ice with leads. The method is called wavelet shrinkage and was developed by Donoho and Johnstone [11]. Wavelet shrinkage is based on the principle that noise in the signal is reflected in the wavelet decomposition as noise in the coefficients and that discontinuities in the signal are represented by a few large coefficients. All coefficients less than a threshold are set to zero and those larger than the threshold are shrunk toward zero. Donoho and Johnstone developed specific optimal thresholds. Tests with the Landsat data show very encouraging results; the smoothed signal appears very smooth over floes, the jumps at leads remain sharp, the magnitude of the signal within leads remain high, and the apparent widths of the leads remain the same.

In our example of the use of the DWT for analyzing the spatial scales of variability of sea ice, the variance of both the albedo and the surface temperature show a maximum at scales near 500 m with a rapid decline at smaller scales. The albedo variability drops continuously at small scales (25 to 100 m) while the temperature variability levels off to a flatter rate more characteristic of white noise at smaller scales. The more rapid fall-off of the temperature variance at intermediate scales is a reflection of the lower resolution of the thermal sensor. The uncertainties in the variance estimates increase as the scale increases. About 50% of the variance occurs at scales less than 800 m, so the AVHRR data, which has a resolution of 1.1-5 km, samples less than half of the total variance for this scene. In other scenes, with much larger variances, the AVHRR data may capture substantially more or less of the total variance. The fraction of the total variance found at small scales depends on both the lead concentration (and the lead brightness or temperature distribution) and on the large-scale variability that may exist independent of the leads. The covariance between channels 3 and 4 is small at the smallest scales due to the relatively poor resolution of the thermal sensor. The maximum covariance is seen at scales of about 500 m. We expect the covariance to be large in areas with a lot of leads, because leads are both much warmer and much darker than the surrounding ice. Over ice floes there is little reason to expect any strong correlation between the two signals, and that is what we observe in the plots of the wavelet coefficients of the covariance.

The wavelet decomposition allows for a particularly revealing form of scale analysis in which the value of a derived quantity is determined using data smoothed to increasingly larger scales. The scale dependency curves show the value of a derived quantity versus the scale of the wavelet-based smooth used for the observed quantities. It depends on: 1) the spatial structure of the observed quantities; 2) the filtering process used to produce the smooths; and 3) the spatial structure of the model process. The curves show what scales are most important for the quantity of interest. By determining the scale below which the derived quantity no longer changes, we can establish the minimum resolution requirements for measuring the observed quantities. Alternatively, if enough high-resolution data is examined, it may be possible to determine the typical shapes of the scale dependency curves and determine a correction to values derived from low-resolution data.

Wavelet decomposition is a new and rapidly evolving field. Wavelet analysis can now be easily implemented in a wide variety of applications using the estimates of the smooths and the roughs of the data and the partition of variance to address specific geophysical problems. Wavelet analysis offers a rich, new way of thinking about data. Here we have tried to outline the technique in enough detail to allow others to easily adapt it to other geophysical problems. Our specific interest is sea-ice surface properties and the technique has proved to be a flexible and productive analysis tool.

ACKNOWLEDGMENT

The Landsat data was provided through the courtesy of K. Steffen, University of Colorado, Boulder.

REFERENCES

Ronald W. Lindsay received the B.S. degree in physics from the University of California, Davis, in 1968, and the Master’s degree from the Department of Atmospheric Sciences, University of Washington, Seattle, in 1976.

He has been active in polar research for most of his career. He has worked with air-sea interaction modeling, air pollution studies, collection and analysis of arctic radiosonde observations, and the design and installation of automatic arctic weather stations. His recent interests and publications include work on the remote sensing of sea ice using visible and thermal satellite imagery.

Mr. Lindsay is a member of the American Meteorological Society and the American Geophysical Union.

Donald B. Percival received the B.A. degree in astronomy from the University of Pennsylvania, Philadelphia, in 1968, the M.A. degree in mathematical statistics from George Washington University, Washington, DC, in 1976, and the Ph.D. degree in statistics from the University of Washington, Seattle, in 1982.

From 1968 to 1978, he was an Astronomer with the Time Service Division of the U.S. Naval Observatory in Washington, DC, where he worked on the generation of atomic clock time scales and on the analysis of the frequency stability of high-performance oscillators. Since 1982, he has been a Senior Mathematician at the Applied Physics Laboratory, University of Washington, with a joint appointment in the Department of Statistics. His current research interests include time series and spectral analysis, simulation of stochastic processes, computational environments for interactive time series and signal analysis, and statistical analysis of underwater turbulence. He is coauthor with A. Walden of Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques (Cambridge: Cambridge Univ. Press, 1993).

Dr. Percival is a member of the American Statistical Association.