Introduction to Orthonormal Wavelet Analysis with Shift Invariance: Application to Observed Atmospheric Blocking Spatial Structure

AIME FOURNIER

Department of Physics and Department of Geology and Geophysics, Yale University, New Haven, Connecticut

(Manuscript received 25 February 1999, in final form 29 March 2000)

ABSTRACT

Orthonormal wavelet analysis (OWA) is a special form of wavelet analysis, especially suitable for analyzing spatial structures, such as atmospheric fields. For this purpose, OWA is much more efficient and accurate than the nonorthogonal wavelet transform (WT), which was introduced to the meteorological community recently and which is more suitable for time series analysis. Whereas the continuous WT is strictly correct only for infinite domains, OWA is derived from periodizing and discretizing the infinite-domain case and so is correct for periodic boundary conditions.

Unlike Fourier spectra, OWA is not shift invariant. Nor is it equivariant like the WT; that is, the OWA output does not shift as its input shifts. Two remedies are to combine all possible shifts, known as the overcomplete, nonorthogonal shift equivariant WT, or else to use a “best shift,” known as best shift wavelet analysis. Although shift invariant and orthonormal w.r.t. arbitrary inputs, the latter’s optimization generally depends nonlinearly on a reference structure.

OWA is a generalization of Fourier series, and the associated multiresolution analysis (MRA) generalizes Reynolds averaging. Like these, OWA/MRA on discrete and continuous domains satisfy analogous identities arithmetically exactly, unlike the WT. OWA/MRA is much more efficient than Fourier series for analyzing several examples, including nearly singular synthetic structures, and a 90-day Hovmöller diagram of observed geopotential height, containing five atmospheric blocking events. OWA’s variance compression is comparable to that of EOF analysis, but is much faster and less data-dependent.

Implementing the basic OWA operations is documented with WaveLab, a freely available package requiring MATLAB.

1. Introduction

Understanding atmospheric structure and dynamics requires both functional and statistical analysis. This paper is about relatively new approaches to these analyses and the situations in which they are preferable. Historically there have mainly been two approaches. One option is to use the observable field values at discrete locations. For example, Onsager (1949) introduced a statistical hydrodynamics based on the approximation of a continuous vorticity field by a set of discrete “point vortices” of constant circulation and variable location. Of course, another “spatial” representation is by values on a fixed grid, as in finite-difference numerical modeling. The alternative technique has been Fourier series, or similar orthogonal function series. Lorenz (1960) showed that analogs of the fields’ physical conservation laws were obeyed by truncations of Fourier series. Charney (1963) observed this to be the case also for general orthogonal function series and for point vortices, but not for fixed-grid discretization. He called the point-vortex and Fourier representations “dual” in the sense of the Heisenberg uncertainty principle, or wave particle duality. These terms usually refer to the reciprocal uncertainty of location and wavenumber information in quantum mechanics. Actually, an uncertainty principle applies to the statistical variances of spatial and wavenumber representations of any distribution (Chui 1992; Wickerhauser 1994; Prestin and Quak 1999). The implied trade-off of information can be a basis for choosing between the analysis methods that we will discuss.

The point-vortex technique is limited to relatively simple kinds of dynamics. Thus to maintain conservation laws, some kind of orthogonal function expansion is needed, in order to treat realistic dynamics. Because actual atmospheric fields may be spatially localized and can have large gradients, Fourier series and related glob-
analyses are not ideally suited; some kind of compromise representation is called for.

a. Spatial versus wavenumber compromises

1) Before wavelets

The uncertainty principle suggests that there can be a mutual exchange of information content, or resolution, between the “extreme” cases of points or waves. In recent decades there have been various imperfect attempts to formulate compromise representations, which somehow combine localization in space and scale. Sig-gia (1977) introduced the use of “wave packets” to partially decouple position and scale in turbulent dynamics. Zimin (1981) introduced a disjoint wavenumber band representation, later identified by Zimin and Hussain (1995) as a nonorthogonal wavelet basis. Both these approaches required very complicated approximations. Besides these contributions from turbulence investigators, there was also an application to meteorology of bandpass filtering (BPF) designed for limited spatial domains, which had nonnegligible domain boundary effects (Bettge and Baumhefner 1980). These investigations, and those in other sciences, set the stage for the

Fig. 1. (a) Geostationary satellite earth image. (b) Box-averaged earth image. (c) Fourier-truncated earth image. (d) Wavelet-truncated earth image.
there has been little applied turbulence (Meneveau 1991; Farge et al. 1992; 1995; Torrence and Compo 1998; Park and Mann 2000), and has been applied in the space domain to observed 1995; Lau and Weng 1995; Mak 1995). There are already several reviews available; for example, Kumar and Foufoula-Georgiou (1997) review wavelet analysis for geophysical applications. Examples throughout broader areas of physics are collected by van den Berg (1999).

2) Relevant Wavelet Applications

OWA is especially relevant when dominant spatially “local” structures are present in a synoptic field, for example, during a blocking situation. [Following Hansen and Sutera (1984), we define blocking, in the 55°–80°N average, by slowly propagating ridges in the 50-kPa height, which exceed the zonal mean by at least 250 m for at least seven days.] The present paper is the first in a sequence that argues that in the case of compactly organized structures such as blocking, wavelet-based analysis techniques are more appropriate than Fourier analysis, to evaluate the relative importance of different-scale effects. [This outstanding issue has been investigated by numerous authors, e.g., Tsou and Smith (1990), using part of the present dataset.] Although wavelet analysis has been applied in the time domain to atmospheric boundary layer turbulence (e.g., Foufoula-Georgiou and Kumar 1994) and climatic time series (e.g., Bolton et al. 1995; Lau and Weng 1995; Mak 1995; Torrence and Compo 1998; Park and Mann 2000), and has been applied in the space domain to observed (Zubair et al. 1992; Zubair 1993) and numerically simulated turbulence (Meneveau 1991; Farge et al. 1992; Fournier and Stevens 1996), there has been little application to observed global synoptic meteorological data. The present author has endeavored to remedy this situation in Fournier (1996, 1999) and Fournier (1999, three manuscripts submitted to J. Atmos. Sci., hereafter F1, F2, and F3).

This paper is also designed as a complement to earlier tutorials by Lau and Weng (1995) and Torrence and Compo (1998), whose methods are suitable for analysis of time series, not spatial structures. The latter authors presented several formulas that should be identified as approximations to appropriate integrals, to lowest order in the time increment $\Delta t$ and/or the scale increment $\Delta j$; they correctly emphasized the importance of choosing sufficiently small increments. [Shensa (1992) discusses exact methods for such integrals, in the course of his thorough discussion of the relationship of OWA to the continuous wavelet transform (CWT) and other related transforms.] This is adequate for sufficiently smooth signals but may lead to difficulties in reconstructing, filtering, decomposing the variance of, or otherwise analyzing less-smooth signals. By contrast, the methods described in the present paper are arithmetically exact, that is, accurate to computer precision, for the analysis of a uniform sampling of any spatial field. The analyses of the field’s samples arithmetically exactly satisfy equations analogous to those of the continuous field’s analyses.

b. A motivating example

The following further motivates the hypothesis that OWA should be advantageous for atmospheric problems. Figure 1a shows an image of the earth. Cloud patterns suggest the complicated dynamics and structure
Table 2. Fourier vs OWA formulas comparison. Column 1 describes the row contents. Columns 2–4 contain formulas for the Fourier and wavelet analyses described in the text.

<table>
<thead>
<tr>
<th>Fourier</th>
<th>Wavelet curtailed at level</th>
<th>Wavelet curtailed at level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series</td>
<td>$f(\lambda) = \sum_{n=-\infty}^{\infty} \hat{f}_n F_n(\lambda)$</td>
<td>$f(\lambda) = {f}<em>{a} + \sum</em>{j=0}^{\infty} f^j(\lambda)$</td>
</tr>
<tr>
<td>Coefficient</td>
<td>$\hat{f}_n = \langle F_n f \rangle$</td>
<td>$\hat{f}<em>{j,\lambda} = \langle F_n f \rangle</em>{\lambda}$</td>
</tr>
<tr>
<td>Basis function</td>
<td>$F_n(\lambda) = e^{ik\pi n}$</td>
<td>$W_{n}(\lambda) = e^{i\lambda n}$</td>
</tr>
<tr>
<td>Orthonormality</td>
<td>$\langle F_n F_m \rangle, = \delta_{nm}$</td>
<td>$\langle W_{n} W_{p} \rangle, = \delta_{np}$</td>
</tr>
<tr>
<td>Parseval identity</td>
<td>$\langle f^* f \rangle = \sum_{n=1}^{2^{l}} (\hat{f}_n g^* + \hat{g}_n f^d)$</td>
<td>$\langle f^* f \rangle = \sum_{n=1}^{2^{l}} \sum_{p=1}^{2^{l}} \hat{f}<em>{n} \hat{g}</em>{p}$</td>
</tr>
</tbody>
</table>

of the underlying fields, and exhibit a broad range of scales, anisotropy, and organized structures.

Often in image processing or geoscience one needs to use a relatively small number of values to represent such an earth image. Say we want to reduce the $2^{18}$ pixel values to only $2^{13}$ values of some kind. One method, corresponding to a purely spatial viewpoint, would be to simply average the pixel values in $2^3 \times 2^3$ squares, yielding the piecewise uniform image in Fig. 1b. Obviously, much important detail structure has been lost.

A method corresponding to a purely wavenumber viewpoint could be to truncate wavenumbers $\geq 2^3$. The image reconstructed from the remaining $(2^3 \times 2^3)^3 = 2^{13}$ Fourier–sine–transform coefficients is shown in Fig. 1c. Although the image is much smoother and more realistic, there is still very little detail left.

The alternative method is to combine some of the spatial and wavenumber information, using coefficients of expansion in a wavelet basis, which is simultaneously localized in both space and wavenumber. The reconstruction from the $2^{13}$ largest-magnitude coefficients is shown in Fig. 1d. Much of the structure lost by the previous two methods has been regained, such as the ITCZ and the vortices west of South America. This observation has been the main motivation to the investigations of Fournier (1995, 1996, 1999), F1, F2, F3, and Fournier and Stevens (1996) into how to apply OWA to atmospheric spatial structure and dynamics.

c. Outline

An introduction to the basic elements of OWA follows a review of simple statistical ideas familiar to meteorologists, in section 2. Unlike most popular introductions, which were geared toward time series analysis, this includes a review of periodization, appropriate for analysis on latitude circles. Also briefly reviewed are extensions initiated by Beylkin (1992) and Shensa (1992) to compensate for the lack of shift invariance of the basic analysis. This important objection can be satisfactorily dealt with by either using all possible shifts simultaneously (with minor additional computation) or choosing a “best shift.” There are at this time few references to this subject available, and mostly they are aimed at an audience specializing in statistical mathematics or signal processing. The representation of blocking structures by best shift is investigated in F3 and summarized in section 2h. Section 3 contains a detailed survey of examples, directly comparing the accuracy of equally truncated Fourier and wavelet series, for analyzing various test signals. This tutorial is to familiarize the reader with OWA enough to understand its suit-

Table 3. Pyramidal indexes $\ell_n$ for $j = 3$. The horizontal spacing around $\ell_{3,1}$ indicates the support of $W_{3,1}$, while the 2D tableau arrangement illustrates the inheritance of scale/location information under the computations (2.5a)–(2.5e).

<table>
<thead>
<tr>
<th>$j$</th>
<th>$\ell_{2,1}$</th>
<th>$\ell_{2,2}$</th>
<th>$\ell_{2,3}$</th>
<th>$\ell_{2,4}$</th>
<th>$\ell_{3,1}$</th>
<th>$\ell_{3,2}$</th>
<th>$\ell_{3,3}$</th>
<th>$\ell_{3,4}$</th>
<th>$\ell_{4,1}$</th>
<th>$\ell_{4,2}$</th>
<th>$\ell_{4,3}$</th>
<th>$\ell_{4,4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$\ell_{2,1}$</td>
<td>$\ell_{2,2}$</td>
<td>$\ell_{2,3}$</td>
<td>$\ell_{2,4}$</td>
<td>$\ell_{3,1}$</td>
<td>$\ell_{3,2}$</td>
<td>$\ell_{3,3}$</td>
<td>$\ell_{3,4}$</td>
<td>$\ell_{4,1}$</td>
<td>$\ell_{4,2}$</td>
<td>$\ell_{4,3}$</td>
<td>$\ell_{4,4}$</td>
</tr>
<tr>
<td>3</td>
<td>$\ell_{3,1}$</td>
<td>$\ell_{3,2}$</td>
<td>$\ell_{3,3}$</td>
<td>$\ell_{3,4}$</td>
<td>$\ell_{4,1}$</td>
<td>$\ell_{4,2}$</td>
<td>$\ell_{4,3}$</td>
<td>$\ell_{4,4}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unauthenticated | Downloaded 11/24/23 04:09 AM UTC
FIG. 2. (a) Three possible orthonormal bases plotted (vertically offset) as a function of $\lambda$. Left: The basis corresponding to gridpoint values. Center: The Fourier basis. Ordinate is wavenumber $n$. Right: The Daubechies-4 wavelet basis and $\perp W_{0,1}^{2j}$. The ordinate indicates $\perp W_{0,1}^{2j}$, $j = 0, \ldots, 4$. (b) Different Daubechies-4 orthonormal bases possible for values (left to right) 0 to 4 of level of curtailing $R$ (Table 2), as a function of $\lambda$.

FIG. 3. (a) The location (abscissa $\lambda$)–wavenumber (ordinate $n$) plane; (b) space representation; (c) wavenumber representation; (d)–(f) OWA representations curtailed at levels $R = 0, 1, 2$, respectively.

2. Average and detailed components of structures
a. Basic MRA concepts
Since the work of Reynolds, it has been a common practice in turbulence (large eddy simulation) and meteorological modeling to decompose a field $f(x)$ into its average and deviation parts: $f = W_\perp f + Wf$. Here $W_\perp f$ is the result of some kind of averaging projection operation, $W_\perp$ is a smoothing operator, and the residual $Wf$ defines a detailing operator, $W = 1 - W_\perp$. The label $\perp$ will appear in different ways here, to signify some kind of orthogonality. For functions and operators...
these orthogonalities are expressed by \( \langle \mathcal{W}^j f \mathcal{W}^j f \rangle = 0 \) [where \( \langle f \rangle = \int f(x) \, dx \) and \( \mathcal{W}^j \mathcal{W}^j = 0 \), respectively. For example, low-pass filtering (LPF) (Blackmon 1976; Bettge and Baumhefner 1980; Saltzman 1990) gives a \( \mathcal{W}^j \) in the “ideal” LPF case, \((\mathcal{W}^j)^2 = \mathcal{W}^j\), but the commonly used running average is not a projection: each application of running average further smoothes the field. For concreteness, henceforth \( \mathcal{W}^j f \) and \( \mathcal{W}^j f \) are taken to be \( \langle f \rangle \), and \( f^* = f - \langle f \rangle \), respectively.

This is generalized to MRA as follows. Begin with a trivial-looking “telescoping sum” decomposition:

\[
\bar{f}_j = \bar{f}_0 + (\bar{f}_1 - \bar{f}_0) + \cdots + (\bar{f}_j - \bar{f}_{j-1})
\]

\[
= \bar{f}_0 + \sum_{j=0}^{J-1} (\bar{f}_{j+1} - \bar{f}_j) = \mathcal{W}^j f.
\]

Note that this expansion is true regardless of the symbols’ definitions. If \( \bar{f}_j \) stands for a “fine approximation” of \( f \) at some “finescale” \( j = J \), then one can always derive it from a “gross approximation” \( \bar{f}_0 \) at some “coarse scale” \( j = 0 \) by adding a series of “details,”

\[
f^*_j = \bar{f}_{j+1} - \bar{f}_j = \mathcal{W}^j f, \quad (j = 0, \ldots, J - 1),
\]

between successive approximations. Now let us combine these simple ideas with OWA.

We are interested in \( f \)'s that depend on periodic coordinates such as

\[-\pi \leq \lambda = (2\pi - 1) \pi < \pi,
\]

so in the following sections we survey periodic OWA; see, for example, Perrier and Basdevant (1989) for de-
tails. (Henceforth “periodic” shall be subsumed into OWA, etc.) First we review Fourier analysis. For ease of discussion, all formulas and notation are summarized in Table 2, arranged in columns by analysis method and in rows by the ingredients of the analyses. Additional notations are \((f)_\lambda = (2\pi)^{-1} \int_{-\pi}^{\pi} f(\lambda) \, d\lambda\) and \((a)_\alpha = \sum_{n=-\infty}^{\infty} a_n\).

b. Fourier analysis

If \(f\) is 2\(\pi\) periodic and either absolutely integrable and has bounded \(\partial_x f\) at \(\lambda\), or else square integrable and continuous at \(\lambda\), then \(f(\lambda)\) may be expanded in the Fourier basis. In practice, the limit \(\infty\) in the Table 2, column 1 formulas is replaced by the number \(2^J\) of uniform function samples available; thus we get discrete analogs that are arithmetically exact.

c. Wavelet analysis

More generally than for Fourier series, any 2\(\pi\) periodic, continuous or absolutely integrable function \(f(\lambda)\) may be expanded in a periodic orthonormal wavelet basis. The wavelet coefficients \(f_{j,k}\) are defined using \(\langle \cdot, \cdot \rangle\), or, in Fourier space, \(\langle \cdot, \cdot \rangle_\lambda\). Again, in practice \(\infty\) is replaced by \(J - 1\) so that the coefficients (including \((f)_\lambda\)) total \(1 + \sum_{j=0}^{J-1} 2^j\) in number and are computed as described in section 2f. All the wavelet equations in Table 2 follow from section 9.3 of Daubechies (1992).

1) DIFFERENT BASES AND THE MEANING OF THE LABELS

Another way to see how the OWA basis compromises between point and wave representations is given by Fig. 2a. The left panel depicts basis functions taking the value 1 at each of \(2^J\) longitudes,

\[\lambda_{j,k} = \left[2^{j-1}(\ell - 1) - 1\right]\pi,\]  

and zero elsewhere. This basis underlies the standard uniform discretization of \(f(\lambda)\). It is entirely localized in \(\lambda\). The middle panel depicts the sines and cosines with wavenumbers \(n\) from 0 to \(2^J-1\), as used in the FFT. The Fourier basis is perfectly localized in wavenumber, sometimes equated with scale, as defined \(\propto n^{-1}\). In the right panel, obviously the basis functions are somewhat localized in space. The basis functions come in octaves or levels \(j = 0, \ldots, J - 1\), with \(2^J\) functions at each level. The functions at lower \(j\) are clearly less localized in \(\lambda\) so \(j\) stands for resolution. It turns out that they are
more localized in scale (properly defined as intervals of \( n \)), as we will see in Fig. 5b. In some sense, some scale information has been traded off to regain lost location information. Such trade-offs must always satisfy the uncertainty principle w.r.t. scale and location. We can visualize this with a unit box, Fig. 3a, symbolizing a 2D location–wavenumber plane to be covered by \( 2^j = 8 \) basis elements. The space representation (no scale or wavenumber information), the wavenumber representation (no location information), and the wavelet representation (some of both kinds of information) are symbolized by the tilings in Figs. 3b–d, respectively. The uncertainty principle is expressed by each tile in every tiling having the same area, \( 2^{-j} = \frac{1}{8} \). As one tile dimension is reduced, to increase the resolution in that aspect, the other dimension must be increased, which reduces its resolution aspect.

2) Wavelets on the real line

The periodic OWA basis functions \( W_{j,k} \) are obtained from those on the real line, \( \psi_{j,k} \), by periodization (see appendix A). The \( \psi_{j,k} \) are generated by \( 2^j \) dilation and \((k - 1)\) shift:

\[
\psi_{j,k}(t) = 2^{j/2} W(2^j t - k + 1), \quad -\infty < t < \infty. \tag{2.2}
\]

There are many possible “mother wavelets” (Meyer 1993), \( W \), which generate such a representation. Most \( W \)s cannot be written in terms of explicit functions, but many can be defined by

\[
W(t) = 2^{j/2} \sum_{l=-\infty}^{\infty} w_l W(2^j t - l + 1), \tag{2.3a}
\]

where the scaling function \( W^j \) (Daubechies 1992), also called “wavelet father” (Meyer 1993), is the solution of the dilation equation (e.g., Kaiser 1994; Strang and Nguyen 1995):

\[
W^j(t) = 2^{j/2} \sum_{l=-\infty}^{\infty} w_l^j W(2^j t - l + 1). \tag{2.3b}
\]

Sometimes (2.3b) is called the “scaling identity” (e.g., Strichartz 1993), or the “two-scale equation” (e.g., Wickerhauser 1994), or (2.3a) and (2.3b) together are called the “two-scale relations” (e.g., Chui 1992). Ultimately, everything is determined by the two filter sequences \( w \) and \( w^j_l = (-1)^{l-j} w_{j-l}^j \). For any \( w_l \), (2.3c) (even with any odd integer replacing 3) implies \( \sum_{l=-\infty}^{\infty} w_l^j W_{j+1,l} = 0 \) for any integer \( l \). This in turn implies \( \int_{-\infty}^{\infty} W^j(t) W(t) dt = 0 \) as long as \( \int_{-\infty}^{\infty} W^j(t) W^j(t - \ell) dt = \delta_{\ell,0} \). Equations (2.2), (2.3a)–(2.3c) are derived from (1.3.4), (5.1.32), (5.1.15), (5.1.35), respectively, of Daubechies (1992). The significance of the cascade interpretation of (2.3a), (2.3b) to turbulence is discussed in appendix B.

There is one case for which \( W^j \) and \( W \) may be expressed explicitly. For \( w = 2^{-1/2} (-1, 1) \) the solutions to (2.3a), (2.3b) are

\[
W^j(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{otherwise}, \end{cases}
\]

\[
W(t) = \begin{cases} -1, & 0 \leq t < 2^{-1} \\ 1, & 2^{-1} \leq t < 1 \\ 0, & \text{otherwise}. \end{cases} \tag{2.4}
\]

(Daubechies 1992). These relations are schematically depicted in Figs. 4a,b. The orthonormal basis generated by these solutions via (A.1) and (2.2) was invented by Haar (1910).

Note that Haar’s solutions are compactly supported (CS, i.e., they vanish outside a closed, bounded set of \( t \)), discontinuous, and zero mean: \( \int_{-\infty}^{\infty} W(t) dt = 0 \). Daubechies (1988, 1992) generalized the solutions, showing that, with certain requirements on \( w \) (reviewed by F3), solutions \( W^j \) exist that are CS and have smoothness and a number of vanishing moments that all increase roughly linearly with the support length \( 2N \) of \( W \). Figures 4c,d show the simplest (\( N = 2 \)) Daubechies solutions, for

\[
w^j = 2^{-5j/2}[(1, 3, 3, 1) + 3^{3j/2}(1, 1, -1, -1)].
\]

In Fig. 4 the wavelet and scaling-function graphs are horizontally aligned to accurately show how features (such as peaks, dips, and cusps) combine to construct the larger structure. The size (inversion) of the \( f \)th smaller scaling function indicates the magnitude (sign) of the coefficient \( w^j_l \) of \( W_{j-2N+2} \) that multiplies it.

The investigator should choose a particular wavelet type according to the problem to be addressed. For example, F1 uses the Daubechies-14 wavelet, with only \( 2N = 14 \) nonzero \( w_l \). For the purpose of analyzing a partial differential equation solution, this wavelet was chosen to provide the most CS basis available that had continuous second derivatives. In the present section 4, Fournier (1999), and F2 and F3, the Daubechies-20 wavelet is used, with only \( 2N = 20 \) nonzero \( w_l \). This was the smoothest CS wavelet available, selected on the assumption that atmospheric structures are at least twice differentiable on the scale observed.

3) Wavelet location and support

Figure 5a shows the \( W_{j,k} \)s (for \( \lambda_j = 0 \) in Table 2, and \( N = 10 \)) in an arrangement that sets up Fig. 9. The \( W_{j,k} \)s are centered approximately at the pyramidal set of longitudes \( \lambda_{j,k} = \lambda_{j,k} \). Each \( \lambda_{j,k} \) is related to the grid (2.1) by the pyramidal index set

\[
\ell_{j,k} = 2^{j/2}(k - 1) + 1,
\]

and constructs a grid from \(-\pi\) to \(\pi\) with step \(2^{j/2}\pi\). Pyramidal simply describes the constraint \( k = 1, \ldots, 2^j \) as exemplified in Table 3. The support of \( W_{j,k} \) has length equal to \( 2\pi \min[1, 2^{j/2}(2N - 1)] \).
d. Generalization to arbitrary curtailing

The functions $W_{j,k}^+$ defined by (A.1), replacing $W$ by $W^+$ in (2.2), satisfy the limited orthogonalities in Table 2. The $W_{j,k}$ are averaging functions, associated with LPF, and $f_0 = f_{0,1} = \langle f \rangle$, since $W_{0,1}(\lambda) = 1$. This is called OWA “curtailed at level $R$” (Nason and Silverman 1995). Its physical interpretation is that the scaling coefficients $\tilde{f}_{R,k}$ now contain all information corresponding to a certain low-wavenumber interval, bounded $\lambda \leq 2^R$.

Figure 2b shows $J = 5$ possible curtailing. The leftmost panel copies the lower half ($j < J - 1$) of the right panel of Fig. 2a. Level $j = J - 1$ is unaffected by curtailing at levels $R < J$. Curtailing at level $R = J$ corresponds to the identity transformation. From left to right, as $R$ increases from 0 to $J - 1$ the first $2^R$ basis functions shown in the leftmost panel are replaced by the corresponding scaling functions $W_{k,1}^R$. The $2^J - 2^R$ wavelets above are unaffected. Each of these panels (taken together with the $j = J - 1$ half not shown) illustrates another possible orthonormal basis. For $J = 3$ the tilings corresponding to $R = 1$ and 2 are in Figs. 3e,f, respectively.

e. Fourier picture of OWA

OWA can also be visualized in the Fourier domain; the $\tilde{W}_{j,k}^R$ are shown in Fig. 5b. From the $n$th Fourier coefficient of (A.1) we see that the magnitude $|\tilde{W}_{j,k}^R|$ is independent of location index $k$ for a given resolution index $j$, so it is only plotted once, for $k = 1$ (black curve). Also, the phase $\arg \tilde{W}_{j,k}^R$ is just a shift by $2\pi(1$
1 December 2000

FOURNIER

3865

Fig. 5. (Continued)

- $k$)2$^{-j}n$ from the continuous Fourier transform of $W$
at frequency $2^{-j}n$ (gray curve for each $n$). Thus for $n = 2^j$
the phase shift is zero, while for $n = 2^{-j}$ every change of $k$ by 2
sweeps a whole cycle.

Since both $W$ and $\hat{W}$ are localized in their respective
Ôspaces, $f_{j,k}$ selects the contribution to $f$ from location
$2^{-j}k$ and scale $2^{-j}$. The set $\{f_{j,k}\}$ defines
an MRA of $f$. It is essentially a sophisticated kind of
BPF technique, to use a term more familiar to climatologists (Blackmon 1976; Saltzman 1990). The $\hat{W}_{j,k,n}$
are for BPF, while the $W_{j,k,n}$ are for LPF. The BPF band-
width and band center are seen to increase with $j$ as the
spatial resolution increases (Fig. 5a), in accordance with
the uncertainty principle (Meyer 1993; Wickerhauser
1994). Fourier and wavelet series are further compared
in appendix C.

f. Practical OWA

In practice, $2^j$ discrete uniform samples $f(\lambda_{jk}) = 2^{nj}\hat{f}_{j,k}$ of $f$ are given. Then discrete analogs of the $2^j$
scaling coefficients $\overline{f}_{j,k+1}$, $k = 0, \ldots, 2^j - 1$, and the
$2^{-j} - 2^j$ wavelet coefficients $\tilde{f}_{j,k+1}$, $k = 0, \ldots, 2^{-j} - 1$,
j = $R$, \ldots, $J - 1$, are computed in $O(2^j)$ convolution-
decimation calculations:

$$\overline{f}_{j,k+1} = \sum_{l=-\infty}^{\infty} w^{l}_{j+1} \overline{f}_{j+1,k+2^{-j}l},$$

$$\tilde{f}_{j,k+1} = \sum_{l=-\infty}^{\infty} w_{j+1} \tilde{f}_{j+1,k+2^{-j}l}$$

(2.5a)

(Beylkin 1992; Nason and Silverman 1995, flipping the sign of $\epsilon$, to conform with WaveLab). (The decimation
of $a_{j}$ is $a_{2j}$, $\epsilon = 0, 1$.) For $2\pi$ periodic $f$ and $w_{\epsilon_{j}+1} = w_{\epsilon_{j}+2j} = 0$ (the periodic CS case)
these become

$$\overline{f}_{j,k+1} = \sum_{l=0}^{2^j-1} w^{l}_{j+1} \overline{f}_{j+1,k+2^{-j}l+2\pi}$$

(2.5b)

$$\tilde{f}_{j,k+1} = \sum_{l=-2j}^{j} w_{l+1} \tilde{f}_{j+1,k+2^{-j}l+2\pi}$$

(2.5c)
Equation (2.5) involves just $w$, so $W_{j,k}$ never needs to be evaluated. This efficient computation is due to the “self-similarity” (2.3a), (2.3b). Note that the corresponding FFT requires $O(J^2)$ calculations and is therefore slower. The reconstruction is the inverse of this orthogonal transformation:

$$\tilde{f}_{j+1,l-\epsilon_l} = \sum_{k=-\infty}^{\infty} (w_{l-2k} \tilde{f}_{j,k+1} + w_{l-2k} \tilde{f}_{j,k+1})$$

$$= \sum_{k=-\infty}^{\infty} \sum_{l'=k}^{\infty} (w_{l'-2k} \tilde{f}_{j,(l'-k)/2+1} + w_{l'-2k} \tilde{f}_{j,(l'-k)/2+1}),$$

$$l - \epsilon_l = 1, \ldots, 2^{j+1}. \quad (2.5d)$$

In the periodic CS case this becomes

$$\tilde{f}_{j+1,l-\epsilon_l} = \sum_{k=-\infty}^{\infty} \sum_{l'=k}^{\infty} w_{l'-2k} \tilde{f}_{j,(l'-k)/2+1}$$

$$+ \sum_{k=-\infty}^{\infty} \sum_{l'=k}^{\infty} w_{l'-2k} \tilde{f}_{j,(l'-k)/2+1}.$$  \hspace{1cm} (2.5e)

A visual pattern corresponding to (2.5a)–(2.5c) corresponds to ascending a pyramid resembling Table 3; likewise (2.5d), (2.5e) correspond to descending. This discrete wavelet transform (DWT) satisfies the discrete an-

The binary sequence $\epsilon_l = 0, 1$ appearing in (2.5) determines whether even- or odd-indexed samples are used at level $j$ of convolution–decimation, as introduced independently by Shensa (1992) and (in more algorithmic detail) by Beylkin (1992). The reference longitude $\lambda_0$ is related to $\epsilon$ in appendix D. Including all possible $\epsilon$, the “overcomplete” set of $\tilde{f}$ corresponds to using all $2^j$ possible periodized shifts of 2-diluted $W$, instead of the orthonormal set of $2^j$ shifts indexed by $k = 1, \ldots, 2^j$, but still only requires $O(J^2)$ calculations (Beylkin 1992), the same cost as the FFT. Appropriately ordered, this larger set of analysis functions generates the nonorthogonal shift-equivariant WT (SEWT), sometimes called “undecimated WT” (Shensa 1992), “stationary WT” (Nason and Silverman 1995), “shift-invariant WT” (Lang et al. 1995), “translation-invariant WT” (Coifman and Donoho 1995; Del Marco and Weiss 1997) or “maximal-overlap DWT” (MODWT; Percival and Mofjeld 1997). The most physical description is equivariant (e.g., von Sachs et al. 1998), meaning a shift of the function produces the same shift of the SEWT [as opposed either to the relations (D.3), wherein the “detail bits” of the shift are truncated, or to (D.4)]. This term is distinguished from “shift invariant,” describing the Fourier spectrum $|\tilde{f}_n|^2$ but not the Fourier transform itself, which merely diagonalizes the shift operator.

The unrestricted domain of location index $k$ “fills in the gaps” left from the orthonormal basis. Figure 6 illustrates this overcomplete, or redundant, representation. The standard OWA (SOWA, defined by $\lambda_0 = 0$) basis functions are shown in heavy outline. The other $2^j - 1$ possible orthonormal sets are all contained in the redundant set.

Among the first SEWT applications were denoising (Coifman and Donoho 1995; Lang et al. 1995) and estimating noisy transients and time delays (Pesquet 1996). Further applications of SEWTs and their wavelet-packet generalizations (Pesquet 1996; Del Marco and Weiss 1997) include examining magnetic field shock waves (Walden and Contreras Cristan 1998), assessing formula errors with observed turbulent heat fluxes (Qi and Neumann 1997), synthesizing multifractal anisotropic 2D cloud profiles (Davis et al. 1999), and analyzing river water levels (B. Whitcher et al. 1999, manuscript submitted to Water Resour. Res.). In particular, Percival and Mofjeld (1997) demonstrated the power of their MODWT for detecting and meaningfully aligning MRA features with the original data, and efficiently decomposing data variance, and explained MODWT’s advantages over both short-time Fourier transform and complex demodulation.

With a minor abuse of notation, the coefficients may be denoted $\tilde{f}_{j,2^{j-l}}$, and $\tilde{f}_{j,2^{j-l}}$, where the fractional $k$ index accurately reflects the location defined by $\lambda_{jk} = 2^{j-l}$ and allows $l = 1, \ldots, 2^j$. Or, generalizing Shensa’s (1992) Eq. (3.7), use (D.1) to define
Fig. 7. Comparison of truncated wavelet vs Fourier analyses of (a)–(e) Ramp, Table 4, row a, with best-shift wavelet; (f) as in (e), with standard wavelet; (g)–(k) as in (a)–(e), for the singular function, Table 4, row c; (l)–(p) the bumps function, Table 4, row e; (q)–(u) the block function, Table 4, row f; (v)–(z) the discontinuous sine function, Table 4, row i.

\[
\overline{f}_{j,2^{-l-1}} = \overline{2}^{j-2^{-l-1}} f_{j,2^{-l}} \quad \text{and} \quad \overline{f}_{j,2^{-l}} = \overline{2}^{j-2^{-l}} f_{j,2^{-l}},
\]

which are easily shown to be shift equivariant, and by (D.4) reduce to tautologies of the form \( A = A \) as \( l \to 2^{-j}/k \). Although not orthonormal, the SEWT still satisfies a Parseval identity,

\[
\sum_{i=1}^{2^j} f(\lambda_{ij})g(\lambda_{ij}) = \sum_{l=1}^{2^j} \left( 2^j \overline{f}_{j,2^{-l-1}} \overline{g}_{j,2^{-l-1}} + \sum_{k=1}^{l-1} \overline{f}_{j,2^{-l-1}} \overline{g}_{j,2^{-l-1}} \right),
\]

which is essential for any kind of analysis of variance, for example, the wavelet energetics of F1, F2, and F3. The factors \( 2^j \) compensate for an emphasis of the larger scales in this representation, in a sense, the “price” of spreading the large-scale representation over a finer location grid. In a scale analysis of covariance, omitting the factors \( 2^j \) would cause larger-scale contributions to be emphasized, relative to the orthonormal analysis.

g. Best shift of the wavelet basis

A particular orthogonal representation, corresponding to a particular

\( \xi = \xi^* \quad \text{or} \quad \lambda = \lambda^* \),

may be extracted from the overcomplete set by maximizing some measure of the representation’s efficiency, w.r.t. \( \xi \). Such a best-shift wavelet analysis (BSWA) was developed roughly contemporaneously by several investigators; see Liang and Parks (1996), Pesquet (1996), and their references. The \( \lambda^* \) computation is shift equivariant (but generally depends nonlinearly on one reference input. For the appropriate \( \lambda^* \), the BSWA is shift invariant and orthonormal w.r.t. arbitrary inputs.

One may find \( \lambda^* \) in \( O(J^2 \delta^2) \) calculations by minimizing a cost functional such as \( \mathcal{E}(\overline{f}, \overline{f}) \) [the analysis entropy of Coifman and Wickerhauser (1992)] w.r.t. \( \xi \), where

\[
\mathcal{E}(\xi) = \sum_{i=1}^{2^j} \delta \mathcal{E}(\|a_i\|^2 \lambda^*), \quad \delta \mathcal{E}(\xi) = -x^2 \log_2 x^2.
\]
As shown by Wickerhauser (1994), $E$ is useful for at least two reasons. It is convex w.r.t. possible analyses $(\tilde{f}, \tilde{g})$, that is,

$$E[xa + (1-x)b] \geq xE(a) + (1-x)E(b)$$

for all $0 \leq x \leq 1$.

Hence any weighted averaging of analyses increases $E$, reducing information, so conversely reducing $E$ increases information, pointing to the most efficient representation. In the extreme cases, $0 \leq E(a) \leq J$, where $E(a) = 0$ if and only if $a$ has at most one nonzero term (maximum representation efficiency), and $E(a) = J$ if and only if $a$ is independent of $l$ (worst case: no information). Wickerhauser (1994) also shows that OWAs optimized in similar ways to this approach the optimal EOF analysis at substantially reduced computational cost. Liang and Parks (1996) classify several alternative cost functionals and evaluate their suitability for fast algorithms and effectiveness for compression. Del Mar and Weiss (1997) extended the method from base-2 to base-$M$ in (2.3a), (2.3b), and to extra flexibility in minimizing $E$ by simultaneously combining multiple $l$s.

### h. Application of BSWA

Let us briefly review some recent applications of BSWA that are closely related to problems investigated by traditional techniques. BSWA is applied by F3 to observed 50-kPa geopotential height data. In four out of five cases of time- and latitude-averaged atmospheric blocking, only two optimally shifted wavelets contribute more variance (at least 89%) than four or more SOWA wavelets. For certain $j = 1$ type blocks, the observed
height distribution remarkably resembles the wavelet $W_{1,k}(\lambda)$ itself. The study of F3 discusses physical reasons why this might be the case and other topics, for example, how to interpret physically the signal-processing motivations underlying wavelet design. Also, F3 derives a minimal Fourier coefficient model of blocking from $W_{1,k}$, relates it to wavelet design considerations, and interprets the minimal model with regard to spectral dynamics models of Platzman (1962) and Wiin Christensen and Wiin-Nielsen (1996).

3. Examples of wavelet analysis

Now we shall compare examples of various $N = 2$ analyses of several test functions $f$ accompanying the free WaveLab software (Buckheit et al. 1995). Each $f$ from Table 4 was first centered and normalized by $f - \langle f \rangle \rightarrow f$, $\|f\|^{-1} \rightarrow f$.

**a. Fourier analysis versus SOWA and BSWA**

Except where noted, both the BSWA and Fourier series were truncated at the number of SOWA coefficients exceeding the magnitude shown, to illustrate the relative merits of the methods. At the left of Figs. 7e,f,k,z, the wavelets are vertically sorted by the variance percentage (VP) each contributes to the reconstruction, cumulatively labeled for each $(j, k)$. After summing some (usually small) number of terms, for example, 3(8) for Fig. 7e (f), the eventual inferiority of the Fourier components (not graphed) is indicated by the right-hand labels’ $(n)$ lower scores in the VP measure. The VP measure also indicates analysis of variance efficiency.

Figure 7a shows the “Ramp” function, such as might model a breaking wave. [In fact, $r^{-1}f(x)$ solves the nonlinear advection equation $U_t = -UU_x$; see F1.] The truncated Fourier series exhibits the Gibbs phenomenon.
(Fig. 7b), while the truncated BSWA yields a much better approximation (Fig. 7c). For most $x$, the BSWA errors are much smaller than the Fourier series errors (Fig. 7d). Note in Fig. 7e how the smaller-scale (higher $j$) wavelets appear only at the discontinuity, and how the cusps of the larger-scale (lower $j$) wavelets align and sum to a smooth line. Figures 7e,f show the improvement of BSWA over SOWA by the VP measure. Note that VP compression does not necessarily follow visually perceived accuracy; pointwise SOWA errors (not shown) are less than BSWA's shown in Fig. 7d, as indicated in row a of Table 4.

How well do the methods reconstruct a point singularity? Table 4, row b, shows the truncation errors for the Kronecker (smoothed Dirac) function. Graphically this is similar to the less singular $f$ analyzed in row c and Figs. 7g–k. With only 23 coefficients the Fourier series has barely started to build up this singularity (Fig. 7h) and performs terribly w.r.t. VP (Fig. 7k). The OWA errors are significantly smaller (Fig. 7j). The Fourier series is better for Table 4, row c, than row b, reflecting the weaker localization in $x$, but in neither case does it perform nearly as well as OWA.

Further delocalizing in $x$, consider Table 4, row d. Again the Fourier series improves but is still inferior to OWA, especially for describing the cusp (not shown). In this case the OWA actually yields two more coefficients larger than the threshold (0.01), relative to the Fourier analysis.

Functions a–d might represent various kinds of vortices, examples of localized phenomena occurring in geophysical fluid dynamics. This suggests that wavelets are useful for representing vortices, as has been demonstrated by Farge (1992), Farge et al. (1992, 1999), and Schneider et al. (1997). Wavelets are also advantageous for representing topography-like structures such
as shown in Figs. 7l–p. In order to capture the very localized “bumps,” Fourier analysis must include high wavenumbers, which contaminate the level “plains” (Fig. 7m) and increase the overall truncation error (Fig. 7o). The small-scale wavelet components only occur where needed in Fig. 7p. Similar behavior appears for the piecewise uniform function shown in Figs. 7q–u.

We now consider functions in which Fourier analysis has some advantage. Clearly for the purely harmonic cases, Table 4, rows g and h, the Fourier analysis does a somewhat better job. In these two analyses there were many more significant OWA coefficients compared to Fourier coefficients, so for the sake of comparison, truncation was made on the magnitude of the latter (as not in the other examples).

Suppose we add a little discontinuity to a sine wave. Figures 7w,x show the reconstructions using 35 coefficients for Fourier or BSWA analysis, determined from the BSWA coefficient magnitudes. Figure 7z shows better VP compression by the Fourier coefficients. As seen in Table 4, row i, Fourier analysis also has slightly smaller mean absolute error, because the wavelets in this case are visibly not smooth enough and so introduce some roughness (Fig. 7x). However, OWA reproduces the discontinuities more sharply and so has smaller rms and max errors. Note: in this case BSWA does not improve over SOWA, since the two discontinuities are separated in $x$ by 0.42, which is nondyadic and so not amenable to a single best shift. We speculate that the multiple-Å approach of Del Marco and Weiss (1997) may help here.

We conclude this survey with two cases related to harmonic functions, yet in a sense more amenable to OWA. Table 4, row j, shows the truncation errors for spatially varying wavenumber, $n_0(x) = 21[x(20x + 1)]^{-1}$ (away from $x = 0, 1$). For the low-wavenumber

![Fig. 7. (Continued)](image-url)
part Fourier analysis does better, but at high wave-numbers the truncation greatly increases the Fourier series error, and OWA does better. OWA also does better with the linearly varying wavenumber “chirp” function, Table 4, row k. OWA outperforms Fourier analysis in this case at both the extreme large and small wave-numbers.

b. MRAs and SEWTs versus ideal BPF

Now we revisit some of the wavelet coefficients just discussed, combined scale-by-scale for \( \lambda_0 = 0 \) (SOWA case) to form MRAs, or sorted by each \( \lambda_0 \) to form SEWTs. Figure 8 shows three analyses of the functions in Table 4, rows e and j. The sum of the \( f_j^* \) (or the
Fig. 9. (a) Hovmöller diagrams of $Z(\lambda, 65^\circ N, 70$ kPa, $t$) (Dm), vs $\lambda$ ($^\circ$) and $t$ (day). White bars indicate $t$ intervals $T$ and $t$ average crest longitudes. The horizontal bar graph at right is $\pi/((\lambda/\partial t))$, for each $T$. Dark (light) curves and lower-left (right) to upper-right (left) diagonal structures typify eastward (westward) progressing synoptic weather patterns. Vertical (horizontal) alignment indicates quasi steadiness (rapid motion). Lines parallel to the main diagonal correspond to the speed (90 day) $212p\sin 65^\circ$ km h$^{-1}$. (b) As in (a), arranged by $j$ as in Fig. 5. (h) As in (b), except here each panel labeled “BPF $j$” corresponds to Fourier BPF (3.1). (n) $Z(\lambda, 70$ kPa, 44.5 day) and topography (Dm) at 65$^\circ N$ (ordinate) vs $\lambda$ ($^\circ$).

For comparison with traditional techniques, attempt the same effect with the ideal BPF achieving roughly the same dyadic partitioning in $n$: Compare $f^\psi$ with

$$f^\psi(\lambda') = 2\text{Re} \sum_{n=-j}^{j-1} \hat{f}_n F_n(\lambda')$$

$$= \langle f \psi_n(\lambda' - j) \rangle$$

(3.1)

$\{\lfloor r \rfloor$ being the integer part of $r$), using nonorthogonal wavelet

$$\psi_n(\lambda) = \csc(2^{-1}\lambda) \{ \sin[(2n - 2^{-1})\lambda] - \sin[(n - 2^{-1})\lambda] \},$$

whose $n$'th Fourier coefficient $\hat{\psi}_n = 1$ ($n \leq |n| \leq 2n - 1$), 0 (otherwise). This filter is wavenumber orthogonal ($\langle \psi_n \psi_m \rangle = \delta_{n,m}$) and zero mean ($\langle \psi_n \rangle = 0$), and decomposes the zonal deviation $f^\psi = \sum_{n=0}^{\infty} f^\psi_n$ and zonal variance $\langle f^\psi \rangle = \sum_{n=0}^{\infty} \langle f^\psi_n \rangle^2$. The results of this BPF of the cases from Table 4, rows e and j, are shown in Figs. 8c,f. In every case some of the features revealed by SOWA–MRA and SEWT still appear, but with much poorer spatial resolution, especially at large scales.
There are contaminating oscillations where the graphs should be flat. The $f_{j}^{0}$ are smoother, but the $f_{j}^{*}$ may be made smoother by using smoother wavelets. In terms of analysis of variance (label to right of $j$th curve indicating contribution from $j$th component), SOWA–MRA achieves the most compression for both functions (Figs. 8a,d). SEWT has some variance compression advantage over BPF for the variable wavenumber function (Figs. 8e,f) but, as noted before, not for the function containing discontinuities with nondyadic separation (Figs. 8b,c).

4. Time-dependent analyses of geopotential height

a. MRA

In these sections five blocking periods are analyzed, of 7–13-day duration (Hansen and Sutera 1984; Fournier 1996, 1999; F2; F3). After F2, $\varphi = 65^\circ$N was judged to be a representative latitude to discriminate blocking from nonblocking, since it cuts through the block structure. It is standard to depict blocking event sequences with a Hovmöller (1949) diagram. This is shown in Fig. 9a for $Z(\lambda, t)$ at 70 kPa, with $t = 0 \pm 0000$ UTC 1
December 1978. The light regions of this figure clearly depict four blocking events in the time intervals $T = [0, 6.5], [19, 26.5], [44, 56.5], [77, 83]$ days near the Atlantic and one in $T = [28.5, 40]$ days near the Pacific. The third Atlantic blocking event is very strong and clearly moves in a retrograde (westward) manner. Apart from these features it is difficult to extract other insight from the complicated patterns. The time and zonal average phase velocity $\langle \partial \lambda / \partial t \rangle_z$ does not consistently show the slowing commonly associated with blocking. There is a range of scales displayed, since there are both regions with many close contour curves and broad, more uniform areas.

A standard way to proceed is to filter the Hovmöller diagram in Fourier space for each $t$ (Wiin-Nielsen and Chen 1993). For comparison, take the $N = 10$ MRA

$$Z(\lambda, t) = \sum_{j=0}^{J-1} Z_j^*(\lambda, t), \quad Z_j^*(\lambda, t) = \sum_{k=1}^{K} \hat{Z}_{jk}(t)W_j(\lambda),$$

shown in Figs. 9b–g. The sum of these six panels reproduces Fig. 9a. The $j = 1$ by itself clearly shows the

---

**Fig. 9. (Continued)**

h. BPF 1 (42.32%)

i. BPF 2 (30.74%)

j. BPF 3 (19.56%)

k. BPF 4 (6.30%)

l. BPF 5 (1.04%)

m. BPF 6 (0.03%)
same blocking ridge pattern as Fig. 9a. This suggests that information encoded in just the two numbers $Z_{1,1}$ and $Z_{1,2}$ may instantaneously describe the presence of blocking, at least for these data. Although $j = 2$ evidently partly contributes to the block, generally the intermediate scales $2^{−j} = \frac{1}{8}, \frac{1}{16}$ capture patterns associated with synoptic cyclone waves, from Atlantic size ($j = 2$) down to Hudson Bay size ($j = 4$). Generally speaking, there are more prograde (eastward tending, dark curves) structures evident during nonblocking. For example, for $|\lambda| = 90$ during the first nonblocking interval, $[7, 18.5]$ days, and $Z_2$ for $0 \leq \lambda \leq 100^\circ$ (over Europe and Russia) during the fourth nonblocking interval, $[57, 76.5]$ days; $Z_1$ at these times and also during the third nonblocking interval, $[40.5, 43.5]$ days. The observable tendencies are usually reflected by

It should be noted that there are exceptions to these observations. That is, occasionally $Z_1$ for $|\lambda| \leq 90$ during the first nonblocking interval, $[7, 18.5]$ days, and $Z_2$ for $0 \leq \lambda \leq 100^\circ$ (over Europe and Russia) during the fourth nonblocking interval, $[57, 76.5]$ days; $Z_1$ at these times and also during the third nonblocking interval, $[40.5, 43.5]$ days. The observable tendencies are usually reflected by

b. Comparison with Fourier BPF

The apparent low-amplitude (a few meters), small-scale ($j > 4$) steady structures over the eastern United States and Europe in Figs. 9f,g seem unphysical. These are probably artifacts but not due to this analysis; almost identical patterns are seen in high-$j$ Fourier BPF (3.1) of these data, shown in Figs. 9h–m. This figure also shows how blocks, belonging to the large-scale (low $j$) band, are more poorly represented by simple BPF than by MRA. For example, although $Z_1$ exhibits greater amplitude during some blocking events, it cannot indicate in which hemisphere the block lies. Figure 9n shows $Z(\lambda, 44.5$ day) and the surface topography at $65^\circ$N. This time was chosen from the high-$j$ panels of previous figures as exhibiting relatively high-amplitude small-scale structure. It is clear that the high-$j$ steady structures must be associated with orography, especially mountain ranges. The perturbations are probably smaller than the original measurement accuracy and so probably indicate mountain-generated artifacts from the National Meteorological Center (now the National Centers for Environmental Prediction) analysis procedure (K. E. Trenberth 1997, personal communication).

c. Wavelet coefficient phase space

To investigate whether the presence of blocking might be described by just the two coefficients $Z_{1,1}$ and $Z_{1,2}$, their time series are presented in Fig. 10a. For the most part, the curves tend to stay at large magnitudes, apart from each other during blocking, and to approach each other at small magnitudes during nonblocking. As might be expected, the Eastern and Western Hemisphere wave-
let modes exchange signs during the Pacific blocking relative to the Atlantic blocking.

The evolution of $\hat{Z}_{k}(t), k = 1, 2$, can also be described in a dynamical phase space; see Fig. 10b. Ordering the quadrants in the standard way indicated (Shapiro 1977), it may be seen that Atlantic blocking is roughly characterized by clockwise orbits, $\hat{Z}_{1,1} > 0$ leading $\hat{Z}_{1,2} < 0$, mostly in quadrant II, while Pacific blocking is characterized by counterclockwise orbits, $\hat{Z}_{1,1} < 0$ lagging $\hat{Z}_{1,2}$, in quadrants III and IV. Nonblocking points tend to cluster nearer to the origin in all quadrants. The empty regions enclosed by the orbits suggest centers of instability, although either longer time series or projection onto higher-dimensional phase space, or both, would be required to further support this. If such diagnostic qualities were robust for more data, they could be useful in predicting blocking, an outstanding problem in extended-range weather prediction.

5. Conclusions

Orthonormal wavelet analysis (OWA) yields the scale separation, or bandpass filtering, of Fourier analysis, while retaining spatial information. The combined accuracy of retained information is limited by the uncertainty principle. The examples in this paper, whether from images, explicit functions, or observed geopotential height, illustrate the greater efficiency of OWA over Fourier analysis, in many cases.

The author offers this tutorial and examples, including the guide (Table 1) to the use of freely available software, with three purposes in mind:

1) to supplement the well-established material reviewing wavelet analysis of time series (e.g., Lau and Weng 1995; Torrence and Compo 1998) with techniques that are suitable for analysis of spatial structures;

2) to help a broader part of the meteorological community to go beyond the limitations of pure location or wavenumber representations; and

3) to illustrate, using atmospheric data, the advantages and complexities that await the user of these OWA techniques, in regard to physical interpretation.

Acknowledgments. I thank A. B. Davis, A. R. Lupo, and an anonymous reviewer for much useful input; B. Saltzman, R. B. Smith, K. R. Sreenivasan, R. R. Coifman, A. R. Hansen, P. D. M. Parker, J. C. van den Berg, and J. D. Berner for comments on earlier versions; and the National Center for Atmospheric Research for providing the observational data. This material is based upon work supported by the National Science Foundation under Grant 9420011 and through NCAR Project 36211014.
APPENDIX A
Periodization

The periodization (e.g., Wickerhauser 1994) of the orthonormal wavelet basis functions \( \psi_{jk} \) on the real line is given in Table 2, where the periodization operator \( (\cdot)^{\text{per}} \) acts on a function \( A \) defined on the real line as

\[
A^{\text{per}}(x) = \sum_{l=-\infty}^{\infty} A(x + l), \quad (A.1)
\]

and \( \lambda \) is a reference origin longitude that may be chosen to optimally align the OWA to the analyzed function (cf. section 2g).

APPENDIX B
Significance of OWA for Turbulence Modelers

All of the graphical depictions of (2.3a), (2.3b) in section 2c(2) illustrate another property of OWA that has attracted researchers in turbulence: a fractal-like self-similarity. Notice in Fig. 4c how the scaling function reproduces itself as a linear combination of smaller copies of itself (2.3b). A related linear combination generates the mother wavelet in Fig. 4d (2.3a). Turbulence cascade modelers have invoked their application to turbulence studies, such as that wavelets are generated by such a process has encouraged their application to turbulence studies, such as the CWT studies of Arneodo et al. (1999). Wavelets are an expansion in amplitudes \( |\tilde{f}_j| \) times scalings by a factor \( n \) of special shifts by \( -\arg \hat{f}_j \) of the mother wave \( \cos(2\pi x) \) (Chui 1992); the bottom form is an expansion in amplitudes \( 2^{j/2}\tilde{f}_j \) times scalings by a factor \( 2^j \) of general shifts by \( 2^j k - 1 \) of the mother wavelet \( W(x) \). Thus in both cases scale is introduced as a multiplier of the argument \( x \), and clearly the linear index \( n \) affords more scale resolution than the exponential index \( j \). However, because of the global structure of cos, the individual shifts (phases) \( \arg \hat{f}_j \) do not localize structures in \( x \), but must be taken all together to do so. Because \( W \) is CS, each individual location index \( k \) does localize structures in \( \lambda \), although it can be seen in Fig. 5a that for low \( j \) the periodization (A.1) slightly diffuses this localization.

APPENDIX D
Relation of the \( \epsilon_j \) to the Choice of \( \lambda_* \)

The \( \epsilon_j \) are related to the choice of \( \lambda_* \), and the shift equivarience (limited to a given resolution) of the basis functions \( W_{jk} \) as follows. Let

\[
\gamma'(\lambda) = f(\lambda - \lambda_*) \quad (D.1)
\]

be a shift by \( \lambda_* \), where \( \epsilon \) is used to represent \( \lambda_0 \) on the grid (2.1) as

\[
\lambda_* = \lambda_{j,\epsilon+1}, \quad 0 \leq \epsilon = \sum_{j=0}^{j-1} \epsilon_j 2^{-j-1} \leq 2' - 1. \quad (D.2)
\]

The limiting cases correspond to \( \epsilon = (0, \ldots, 0) \) and \( \epsilon = (1, \ldots, 1) \), respectively. It has been shown (e.g., Nason and Silverman 1995) that the coefficients \( \tilde{f}_{jk} \) and \( \tilde{F}_{jk} \) are shifted in \( k \) by \( \lambda_* \) with the “detail bits” \( \epsilon_{j,\lambda} \) set to zero, that is, by

\[
0 \leq \sum_{j=0}^{j-1} \epsilon_j 2^{-j-1} = |2^{-j}\epsilon_*| \leq 2^j - 1, \quad (D.3)
\]

from the coefficients \( \gamma'(\tilde{f}_{jk}) \) and \( \gamma'(\tilde{F}_{jk}) \) computed with \( \epsilon = (0, \ldots, 0) \). (Examples are provided by F1.) A corollary generalizes Shensa’s (1992) Eq. (3.6): for any \( \epsilon \),

\[
\tilde{f}_{j,k^{\text{mod}2}(k+1)} = \gamma(2^{j/2}\tilde{f}_{j,k+1}), \quad \tilde{F}_{j,k^{\text{mod}2}(k+1)} = \gamma(2^{j/2}\tilde{F}_{j,k+1}), \quad (D.4)
\]

precisely expressing the shift variance of OWA. Ultimately this implies that there are only \( (J-R)^2 \) different \( f_{jk} \) and \( 2^{j\text{mod}R} \) different \( \tilde{f}_{jk} \) for all possible \( \epsilon \), that is, all possible \( \lambda_* \).
REFERENCES


Schneider, K., N. K.-R. Kevlahan, and M. Farge, 1997: Comparison


