A Practical Guide to Wavelet Analysis

Christopher Torrence and Gilbert P. Compo
Program in Atmospheric and Oceanic Sciences, University of Colorado, Boulder, Colorado

ABSTRACT

A practical step-by-step guide to wavelet analysis is given, with examples taken from time series of the El Niño–Southern Oscillation (ENSO). The guide includes a comparison to the windowed Fourier transform, the choice of an appropriate wavelet basis function, edge effects due to finite-length time series, and the relationship between wavelet scale and Fourier frequency. New statistical significance tests for wavelet power spectra are developed by deriving theoretical wavelet spectra for white and red noise processes and using these to establish significance levels and confidence intervals. It is shown that smoothing in time or scale can be used to increase the confidence of the wavelet spectrum. Empirical formulas are given for the effect of smoothing on significance levels and confidence intervals. Extensions to wavelet analysis such as filtering, the power Hovmöller, cross-wavelet spectra, and coherence are described.

The statistical significance tests are used to give a quantitative measure of changes in ENSO variance on interdecadal timescales. Using new datasets that extend back to 1871, the Niño3 sea surface temperature and the Southern Oscillation index show significantly higher power during 1880–1920 and 1960–90, and lower power during 1920–60, as well as a possible 15-yr modulation of variance. The power Hovmöller of sea level pressure shows significant variations in 2–8-yr wavelet power in both longitude and time.

1. Introduction

Wavelet analysis is becoming a common tool for analyzing localized variations of power within a time series. By decomposing a time series into time–frequency space, one is able to determine both the dominant modes of variability and how those modes vary in time. The wavelet transform has been used for numerous studies in geophysics, including tropical convection (Weng and Lau 1994), the El Niño–Southern Oscillation (ENSO; Gu and Philander 1995; Wang and Wang 1996), atmospheric cold fronts (Gamage and Blumen 1993), central England temperature (Baliunas et al. 1997), the dispersion of ocean waves (Meyers et al. 1993), wave growth and breaking (Liu 1994), and coherent structures in turbulent flows (Farge 1992). A complete description of geophysical applications can be found in Foufoula-Georgiou and Kumar (1995), while a theoretical treatment of wavelet analysis is given in Daubechies (1992).

Unfortunately, many studies using wavelet analysis have suffered from an apparent lack of quantitative results. The wavelet transform has been regarded by many as an interesting diversion that produces colorful pictures, yet purely qualitative results. This misconception is in some sense the fault of wavelet analysis itself, as it involves a transform from a one-dimensional time series (or frequency spectrum) to a diffuse two-dimensional time–frequency image. This diffuseness has been exacerbated by the use of arbitrary normalizations and the lack of statistical significance tests.

In Lau and Weng (1995), an excellent introduction to wavelet analysis is provided. Their paper, however, did not provide all of the essential details necessary for wavelet analysis and avoided the issue of statistical significance.

The purpose of this paper is to provide an easy-to-use wavelet analysis toolkit, including statistical significance testing. The consistent use of examples of

Corresponding author address: Dr. Christopher Torrence, Advanced Study Program, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: torrence@ucar.edu
In final form 20 October 1997.

ENSO provides a substantive addition to the ENSO literature. In particular, the statistical significance testing allows greater confidence in the previous wavelet-based ENSO results of Wang and Wang (1996). The use of new datasets with longer time series permits a more robust classification of interdecadal changes in ENSO variance.

The first section describes the datasets used for the examples. Section 3 describes the method of wavelet analysis using discrete notation. This includes a discussion of the inherent limitations of the windowed Fourier transform (WFT), the definition of the wavelet transform, the choice of a wavelet basis function, edge effects due to finite-length time series, the relationship between wavelet scale and Fourier period, and time series reconstruction. Section 4 presents the theoretical wavelet spectra for both white-noise and red-noise processes. These theoretical spectra are compared to Monte Carlo results and are used to establish significance levels and confidence intervals for the wavelet power spectrum. Section 5 describes time or scale averaging to increase significance levels and confidence intervals. Section 6 describes other wavelet applications such as filtering, the power Hovmöller, cross-wavelet spectra, and wavelet coherence. The summary contains a step-by-step guide to wavelet analysis.

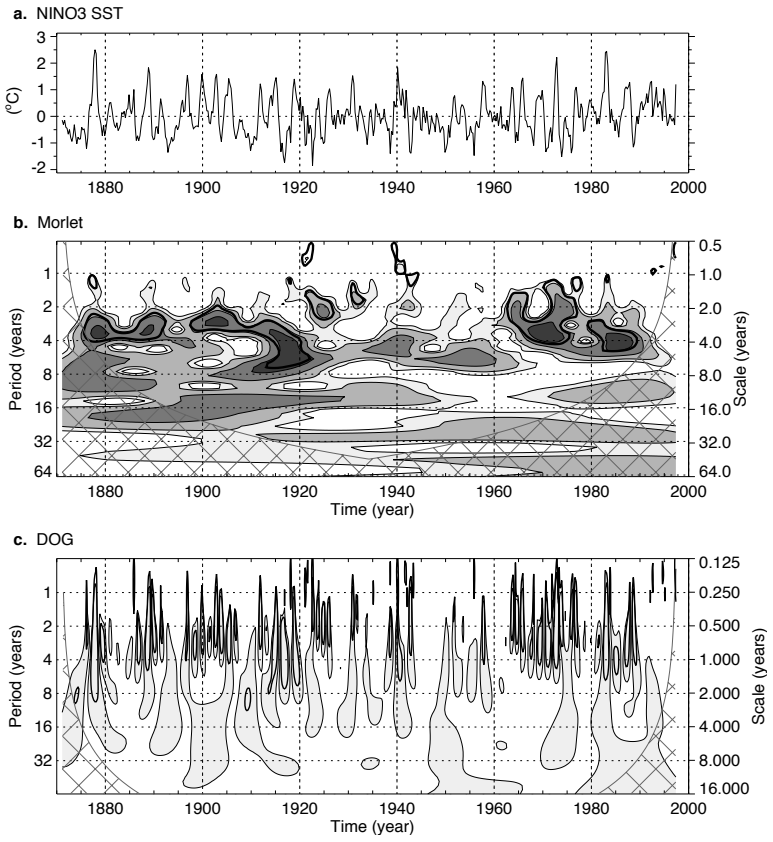

FIG. 1. (a) The Niño3 SST time series used for the wavelet analysis. (b) The local wavelet power spectrum of (a) using the Morlet wavelet, normalized by $1/\sigma^2$ ($\sigma^2 = 0.54°C^2$). The left axis is the Fourier period (in yr) corresponding to the wavelet scale on the right axis. The bottom axis is time (yr). The shaded contours are at normalized variances of 1, 2, 5, and 10. The thick contour encloses regions of greater than 95% confidence for a red-noise process with a lag-1 coefficient of 0.72. Cross-hatched regions on either end indicate the "cone of influence," where edge effects become important. (c) Same as (b) but using the real-valued Mexican hat wavelet (derivative of a Gaussian; DOG $m = 2$). The shaded contour is at normalized variance of 2.0.

2. Data

Several time series will be used for examples of wavelet analysis. These include the Niño3 sea surface temperature (SST) used as a measure of the amplitude of the El Niño–Southern Oscillation (ENSO). The Niño3 SST index is defined as the seasonal SST averaged over the central Pacific (5°S–5°N, 90°–150°W). Data for 1871–1996 are from an area average of the U.K. Meteorological Office GISST2.3 (Rayner et al. 1996), while data for January–June 1997 are from the Climate Prediction Center (CPC) optimally interpolated Niño3 SST index (courtesy of D. Garrett at CPC, NOAA). The seasonal means for the entire record have been removed to define an anomaly time series. The Niño3 SST is shown in the top plot of Fig. 1a.

Gridded sea level pressure (SLP) data is from the UKMO/CSIRO historical GMSLP2.1f (courtesy of D. Parker and T. Basnett, Hadley Centre for Climate Prediction and Research, UKMO). The data is on a 5° global grid, with monthly resolution from January 1871 to December 1994. Anomaly time series have been constructed by removing the first three harmonics of the annual cycle (periods of 365.25, 182.625, and 121.75 days) using a least-squares fit.

The Southern Oscillation index is derived from the GMSLP2.1f and is defined as the seasonally averaged pressure difference between the eastern Pacific (20°S, 150°W) and the western Pacific (10°S, 130°E).

3. Wavelet analysis

This section describes the method of wavelet analysis, includes a discussion of different wavelet functions, and gives details for the analysis of the wavelet power spectrum. Results in this section are adapted to discrete notation from the continuous formulas given in Daubechies (1990). Practical details in applying wavelet analysis are taken from Farge (1992), Weng and Lau (1994), and Meyers et al. (1993). Each section is illustrated with examples using the Niño3 SST.

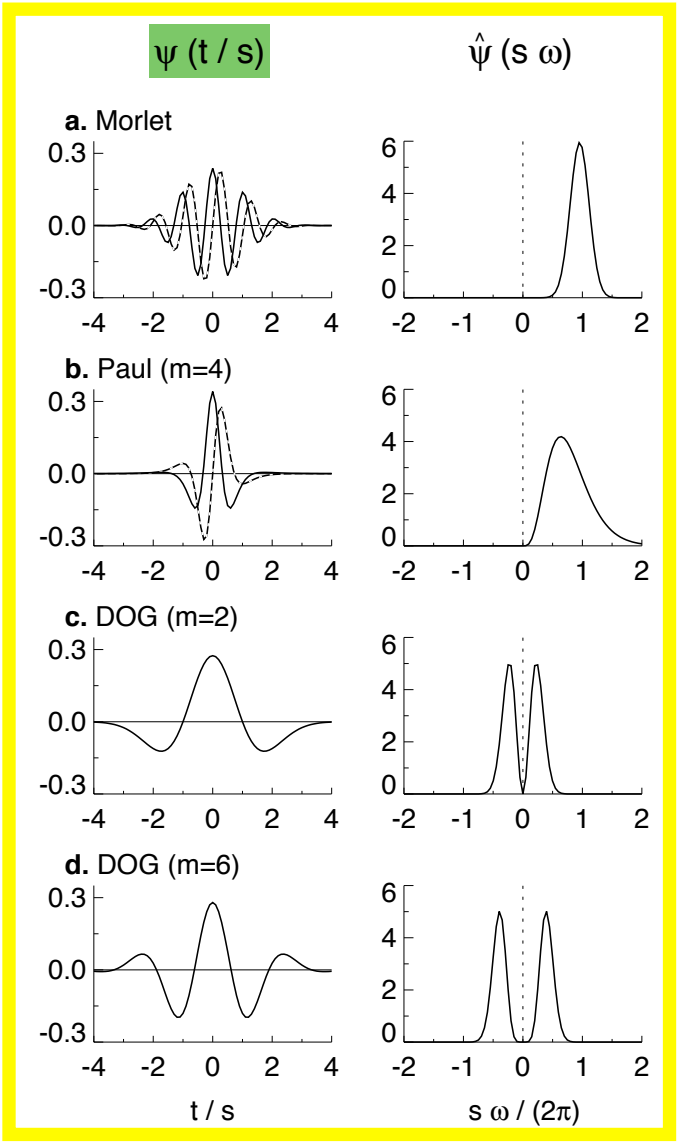

FIG. 2. Four different wavelet bases, from Table 1. The plots on the left give the real part (solid) and imaginary part (dashed) for the wavelets in the time domain. The plots on the right give the corresponding wavelets in the frequency domain. For plotting purposes, the scale was chosen to be $s = 10\delta t$. (a) Morlet, (b) Paul ($m = 4$), (c) Mexican hat (DOG $m = 2$), and (d) DOG ($m = 6$).

a. Windowed Fourier transform

The WFT represents one analysis tool for extracting local-frequency information from a signal. The Fourier transform is performed on a sliding segment of length T from a time series of time step δt and total length $N\delta t$, thus returning frequencies from T^{-1} to $(2\delta t)^{-1}$ at each time step. The segments can be windowed with an arbitrary function such as a boxcar (no smoothing) or a Gaussian window (Kaiser 1994).

As discussed by Kaiser (1994), the WFT represents an inaccurate and inefficient method of time–frequency localization, as it imposes a scale or "response interval" T into the analysis. The inaccuracy arises from the aliasing of high- and low-frequency components that do not fall within the frequency range of the window. The inefficiency comes from the $T/(2\delta t)$ frequencies, which must be analyzed at each time step, regardless of the window size or the dominant frequencies present. In addition, several window lengths must usually be analyzed to determine the most appropriate choice. For analyses where a predetermined scaling may not be appropriate because of a wide range of dominant frequencies, a method of time–frequency localization that is scale independent, such as wavelet analysis, should be employed.

Discuss:
(1) Non-orthogonality
(2) Uncertainty Principle

b. Wavelet transform

The *wavelet transform* can be used to analyze time series that contain nonstationary power at many different frequencies (Daubechies 1990). Assume that one has a time series, x_n, with equal time spacing δt and $n = 0 \ldots N-1$. Also assume that one has a *wavelet function*, $\psi_0(\eta)$, that depends on a nondimensional "time" parameter η. To be "admissible" as a wavelet, this function must have zero mean and be localized in both time and frequency space (Farge 1992). An example is the Morlet wavelet, consisting of a plane wave modulated by a Gaussian:

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2}, \qquad (1)$$

where ω_0 is the nondimensional frequency, here taken to be 6 to satisfy the admissibility condition (Farge 1992). This wavelet is shown in Fig. 2a.

The term "wavelet function" is used generically to refer to either orthogonal or nonorthogonal wavelets. The term "wavelet basis" refers only to an orthogonal set of functions. The use of an orthogonal basis implies the use of the *discrete wavelet transform*, while a nonorthogonal wavelet function can be used

with either the discrete or the *continuous wavelet transform* (Farge 1992). In this paper, only the continuous transform is used, although all of the results for significance testing, smoothing in time and scale, and cross wavelets are applicable to the discrete wavelet transform.

The continuous wavelet transform of a discrete sequence x_n is defined as the convolution of x_n with a scaled and translated version of $\psi_0(\eta)$:

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \psi^* \left[\frac{(n'-n)\delta t}{s} \right], \qquad (2)$$

where the (*) indicates the complex conjugate. By varying the *wavelet scale s* and translating along the *localized time index n*, one can construct a picture showing both the amplitude of any features versus the scale and how this amplitude varies with time. The subscript 0 on ψ has been dropped to indicate that this ψ has also been normalized (see next section). Although it is possible to calculate the wavelet transform using (2), it is considerably faster to do the calculations in Fourier space.

To approximate the continuous wavelet transform, the convolution (2) should be done N times for each scale, where N is the number of points in the time series (Kaiser 1994). (The choice of doing all N convolutions is arbitrary, and one could choose a smaller number, say by skipping every other point in n.) By choosing N points, the convolution theorem allows us do all N convolutions simultaneously in Fourier space using a discrete Fourier transform (DFT). The DFT of x_n is

$$\hat{x}_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / N}, \qquad (3)$$

where $k = 0 \ldots N-1$ is the frequency index. In the continuous limit, the Fourier transform of a function $\psi(t/s)$ is given by $\hat{\psi}(s\omega)$. By the convolution theorem, the wavelet transform is the inverse Fourier transform of the product:

$$W_n(s) = \sum_{k=0}^{N-1} \hat{x}_k \hat{\psi}^*(s\omega_k) e^{i\omega_k n \delta t}, \qquad (4)$$

where the angular frequency is defined as

$$\omega_k = \begin{cases} \dfrac{2\pi k}{N\delta t} & : \quad k \le \dfrac{N}{2} \\ -\dfrac{2\pi k}{N\delta t} & : \quad k > \dfrac{N}{2} \end{cases}. \qquad (5)$$

Using (4) and a standard Fourier transform routine, one can calculate the continuous wavelet transform (for a given s) at all n simultaneously and efficiently.

c. Normalization

To ensure that the wavelet transforms (4) at each scale s are directly comparable to each other and to the transforms of other time series, the wavelet function at each scale s is normalized to have unit energy:

$$\hat{\psi}(s\omega_k) = \left(\frac{2\pi s}{\delta t} \right)^{1/2} \hat{\psi}_0(s\omega_k). \qquad (6)$$

Examples of different wavelet functions are given in Table 1 and illustrated in Fig. 2. Each of the unscaled $\hat{\psi}_0$ are defined in Table 1 to have

$$\int_{-\infty}^{+\infty} |\hat{\psi}_0(\omega')|^2 \, d\omega' = 1;$$

that is, they have been normalized to have unit energy.

Using these normalizations, at each scale s one has

$$\sum_{k=0}^{N-1} |\hat{\psi}(s\omega_k)|^2 = N, \qquad (7)$$

where N is the number of points. Thus, the wavelet transform is weighted only by the amplitude of the Fourier coefficients $\hat{x}_k$ and not by the wavelet function. If one is using the convolution formula (2), the normalization is

$$\psi \left[\frac{(n'-n)\delta t}{s} \right] = \left(\frac{\delta t}{s} \right)^{1/2} \psi_0 \left[\frac{(n'-n)\delta t}{s} \right], \qquad (8)$$

where $\psi_0(\eta)$ is normalized to have unit energy.

d. Wavelet power spectrum

Because the wavelet function $\psi(\eta)$ is in general complex, the wavelet transform $W_n(s)$ is also complex. The transform can then be divided into the real part,

TABLE 1. Three wavelet basis functions and their properties. Constant factors for ψ_0 and $\hat{\psi}_0$ ensure a total energy of unity.

Name	$\psi_0(\eta)$	$\hat{\psi}_0(s\omega)$	*e*-folding time τ_s	Fourier wavelength λ
Morlet (ω_0 = frequency)	$\pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2}$	$\pi^{-1/4} H(\omega) e^{-(s\omega-\omega_0)^2/2}$	$\sqrt{2}s$	$\dfrac{4\pi s}{\omega_0 + \sqrt{2+\omega_0^2}}$
Paul (m = order)	$\dfrac{2^m i^m m!}{\sqrt{\pi(2m)!}}(1-i\eta)^{-(m+1)}$	$\dfrac{2^m}{\sqrt{m(2m-1)!}} H(\omega)(s\omega)^m e^{-s\omega}$	$s/\sqrt{2}$	$\dfrac{4\pi s}{2m+1}$
DOG (m = derivative)	$\dfrac{(-1)^{m+1}}{\sqrt{\Gamma\left(m+\frac{1}{2}\right)}} \dfrac{d^m}{d\eta^m}\left(e^{-\eta^2/2}\right)$	$\dfrac{-i^m}{\sqrt{\Gamma\left(m+\frac{1}{2}\right)}} (s\omega)^m e^{-(s\omega)^2/2}$	$\sqrt{2}s$	$\dfrac{2\pi s}{\sqrt{m+\frac{1}{2}}}$

$H(\omega)$ = Heaviside step function, $H(\omega) = 1$ if $\omega > 0$, $H(\omega) = 0$ otherwise.
DOG = derivative of a Gaussian; $m = 2$ is the Marr or Mexican hat wavelet.

$\Re\{W_n(s)\}$, and imaginary part, $\Im\{W_n(s)\}$, or amplitude, $|W_n(s)|$, and phase, $\tan^{-1}[\Im\{W_n(s)\}/\Re\{W_n(s)\}]$. Finally, one can define the *wavelet power spectrum* as $|W_n(s)|^2$. For real-valued wavelet functions such as the DOGs (derivatives of a Gaussian) the imaginary part is zero and the phase is undefined.

To make it easier to compare different wavelet power spectra, it is desirable to find a common normalization for the wavelet spectrum. Using the normalization in (6), and referring to (4), the *expectation value* for $|W_n(s)|^2$ is equal to N times the expectation value for $|\hat{x}_k|^2$. For a white-noise time series, this expectation value is σ^2/N, where σ^2 is the variance. Thus, for a white-noise process, the expectation value for the wavelet transform is $|W_n(s)|^2 = \sigma^2$ at all n and s.

Figure 1b shows the normalized wavelet power spectrum, $|W_n(s)|^2/\sigma^2$, for the Niño3 SST time series. The normalization by $1/\sigma^2$ gives a measure of the power *relative to* white noise. In Fig. 1b, most of the power is concentrated within the ENSO band of 2–8 yr, although there is appreciable power at longer periods. The 2–8-yr band for ENSO agrees with other studies (Trenberth 1976) and is also seen in the Fourier spectrum in Fig. 3. With wavelet analysis, one can see variations in the frequency of occurrence and amplitude of El Niño (warm) and La Niña (cold) events. During 1875–1920 and 1960–90 there were many warm and cold events of large amplitude, while during 1920–60 there were few events (Torrence and Webster 1997). From 1875–1910, there was a slight shift from a period near 4 yr to a period closer to 2 yr, while from 1960–90 the shift is from shorter to longer periods.

These results are similar to those of Wang and Wang (1996), who used both wavelet and waveform analysis on ENSO indices derived from the Comprehensive Ocean–Atmosphere Data Set (COADS) dataset. Wang and Wang's analysis showed reduced wavelet power before 1950, especially 1875–1920. The reduced power is possibly due to the sparseness and decreased reliability of the pre-1950 COADS data (Folland et al. 1984). With the GISST2.3 data, the wavelet transform of Niño3 SST in Fig. 1b shows that the pre-1920 period has equal power to the post-1960 period.

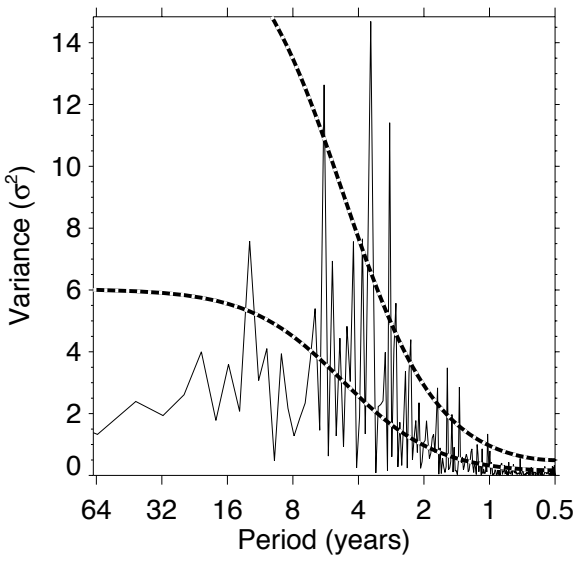

FIG. 3. Fourier power spectrum of Niño3 SST (solid), normalized by $N/(2\sigma^2)$. The lower dashed line is the mean red-noise spectrum from (16) assuming a lag-1 of $\alpha = 0.72$. The upper dashed line is the 95% confidence spectrum.

e. Wavelet functions

One criticism of wavelet analysis is the arbitrary choice of the wavelet function, $\psi_0(\eta)$. (It should be noted that the same arbitrary choice is made in using one of the more traditional transforms such as the Fourier, Bessel, Legendre, etc.) In choosing the wavelet function, there are several factors which should be considered (for more discussion see Farge 1992).

1) *Orthogonal or nonorthogonal*. In orthogonal wavelet analysis, the number of convolutions at each scale is proportional to the width of the wavelet basis at that scale. This produces a wavelet spectrum that contains discrete "blocks" of wavelet power and is useful for signal processing as it gives the most compact representation of the signal. Unfortunately for time series analysis, an aperiodic shift in the time series produces a different wavelet spectrum. Conversely, a nonorthogonal analysis (such as used in this study) is highly redundant at large scales, where the wavelet spectrum at adjacent times is highly correlated. The nonorthogonal transform is useful for time series analysis, where smooth, continuous variations in wavelet amplitude are expected.
2) *Complex or real*. A complex wavelet function will return information about both amplitude and phase and is better adapted for capturing oscillatory behavior. A real wavelet function returns only a single component and can be used to isolate peaks or discontinuities.
3) *Width*. For concreteness, the width of a wavelet function is defined here as the *e*-folding time of the wavelet amplitude. The resolution of a wavelet function is determined by the balance between the width in real space and the width in Fourier space. A narrow (in time) function will have good time resolution but poor frequency resolution, while a broad function will have poor time resolution, yet good frequency resolution.
4) *Shape*. The wavelet function should reflect the type of features present in the time series. For time series with sharp jumps or steps, one would choose a boxcar-like function such as the Harr, while for smoothly varying time series one would choose a smooth function such as a damped cosine. If one is primarily interested in wavelet power spectra, then the choice of wavelet function is not critical, and one function will give the same *qualitative* results as another (see discussion of Fig. 1 below).

Four common nonorthogonal wavelet functions are given in Table 1. The Morlet and Paul wavelets are both complex, while the DOGs are real valued. Pictures of these wavelet in both the time and frequency domain are shown in Fig. 2. Many other types of wavelets exist, such as the Haar and Daubechies, most of which are used for orthogonal wavelet analysis (e.g., Weng and Lau 1994; Mak 1995; Lindsay et al. 1996). For more examples of wavelet bases and functions, see Kaiser (1994).

For comparison, Fig. 1c shows the same analysis as in 1b but using the Mexican hat wavelet (DOG, $m = 2$) rather than the Morlet. The most noticeable difference is the fine scale structure using the Mexican hat. This is because the Mexican hat is real valued and captures both the positive and negative oscillations of the time series as separate peaks in wavelet power. The Morlet wavelet is both complex and contains more oscillations than the Mexican hat, and hence the wavelet power combines both positive and negative peaks into a single broad peak. A plot of the real or imaginary part of $W_n(s)$ using the Morlet would produce a plot similar to Fig. 1c. Overall, the same features appear in both plots, approximately at the same locations, and with the same power. Comparing Figs. 2a and 2c, the Mexican hat is narrower in time-space, yet broader in spectral-space than the Morlet. Thus, in Fig. 1c, the

peaks appear very sharp in the time direction, yet are more elongated in the scale direction. Finally, the relationship between wavelet scale and Fourier period is very different for the two functions (see section 3h).

f. Choice of scales

Once a wavelet function is chosen, it is necessary to choose a set of scales s to use in the wavelet transform (4). For an orthogonal wavelet, one is limited to a discrete set of scales as given by Farge (1992). For nonorthogonal wavelet analysis, one can use an arbitrary set of scales to build up a more complete picture. It is convenient to write the scales as fractional powers of two:

$$s_j = s_0 2^{j\delta j}, \quad j = 0, 1, \ldots, J \tag{9}$$

$$J = \delta j^{-1} \log_2\left(N\delta t / s_0\right), \tag{10}$$

where s_0 is the smallest resolvable scale and J determines the largest scale. The s_0 should be chosen so that the equivalent Fourier period (see section 3h) is approximately $2\delta t$. The choice of a sufficiently small δj depends on the width in spectral-space of the wavelet function. For the Morlet wavelet, a δj of about 0.5 is the largest value that still gives adequate sampling in scale, while for the other wavelet functions, a larger value can be used. Smaller values of δj give finer resolution.

In Fig. 1b, $N = 506$, $\delta t = 1/4$ yr, $s_0 = 2\delta t$, $\delta j = 0.125$, and $J = 56$, giving a total of 57 scales ranging from 0.5 yr up to 64 yr. This value of δj appears adequate to provide a smooth picture of wavelet power.

g. Cone of influence

Because one is dealing with finite-length time series, errors will occur at the beginning and end of the wavelet power spectrum, as the Fourier transform in (4) assumes the data is cyclic. One solution is to pad the end of the time series with zeroes before doing the wavelet transform and then remove them afterward [for other possibilities such as cosine damping, see Meyers et al. (1993)]. In this study, the time series is padded with sufficient zeroes to bring the total length N up to the next-higher power of two, thus limiting the edge effects and speeding up the Fourier transform.

Padding with zeroes introduces discontinuities at the endpoints and, as one goes to larger scales, decreases the amplitude near the edges as more zeroes enter the analysis. The *cone of influence* (COI) is the region of the wavelet spectrum in which edge effects become important and is defined here as the e-folding time for the autocorrelation of wavelet power at each scale (see Table 1). This e-folding time is chosen so that the wavelet power for a discontinuity at the edge drops by a factor e^{-2} and ensures that the edge effects are negligible beyond this point. For cyclic series (such as a longitudinal strip at a fixed latitude), there is no need to pad with zeroes, and there is no COI.

The size of the COI at each scale also gives a measure of the decorrelation time for a single spike in the time series. By comparing the width of a peak in the wavelet power spectrum with this decorrelation time, one can distinguish between a spike in the data (possibly due to random noise) and a harmonic component at the equivalent Fourier frequency.

The COI is indicated in Figs. 1b and 1c by the cross-hatched regions. The peaks within these regions have presumably been reduced in magnitude due to the zero padding. Thus, it is unclear whether the decrease in 2–8-yr power after 1990 is a true decrease in variance or an artifact of the padding. Note that the much narrower Mexican hat wavelet in Fig. 1c has a much smaller COI and is thus less affected by edge effects.

h. Wavelet scale and Fourier frequency

An examination of the wavelets in Fig. 2 shows that the peak in $\hat{\psi}(s\omega)$ does not necessarily occur at a frequency of s^{-1}. Following the method of Meyers et al. (1993), the relationship between the *equivalent Fourier period* and the wavelet scale can be derived analytically for a particular wavelet function by substituting a cosine wave of a known frequency into (4) and computing the scale s at which the wavelet power spec-

TABLE 2. Empirically derived factors for four wavelet bases.

Name	C_δ	γ	δj_0	$\psi_0(0)$
Morlet ($\omega_0 = 6$)	0.776	2.32	0.60	$\pi^{-1/4}$
Paul ($m = 4$)	1.132	1.17	1.5	1.079
Marr (DOG $m = 2$)	3.541	1.43	1.4	0.867
DOG ($m = 6$)	1.966	1.37	0.97	0.884

C_δ = reconstruction factor.
γ = decorrelation factor for time averaging.
δj_0 = factor for scale averaging.

trum reaches its maximum. For the Morlet wavelet with $\omega_0 = 6$, this gives a value of $\lambda = 1.03s$, where λ is the Fourier period, indicating that for the Morlet wavelet the wavelet scale is almost equal to the Fourier period. Formulas for other wavelet functions are given in Table 1, while Fig. 2 gives a graphical representation.

In Figs. 1b,c, the ratio of Fourier period to wavelet scale can be seen by a comparison of the left and right axes. For the Morlet, the two are nearly identical, while for the Mexican hat, the Fourier period is four times larger than the scale. This ratio has no special significance and is due solely to the functional form of each wavelet function. However, one should certainly convert from scale to Fourier period before plotting, as presumably one is interested in equating wavelet power at a certain time and scale with a (possibly short-lived) Fourier mode at the equivalent Fourier period.

i. Reconstruction

Since the wavelet transform is a bandpass filter with a known response function (the wavelet function), it is possible to *reconstruct* the original time series using either deconvolution or the inverse filter. This is straightforward for the orthogonal wavelet transform (which has an orthogonal basis), but for the continuous wavelet transform it is complicated by the redundancy in time and scale. However, this redundancy also makes it possible to reconstruct the time series using a completely different wavelet function, the easiest of which is a delta (δ) function (Farge 1992). In this case, the reconstructed time series is just the sum of the real part of the wavelet transform over all scales:

Q: What if W_n(s_j) = \delta_{ij}? Discuss.

$$x_n = \frac{\delta j \delta t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=0}^{J} \frac{\Re\{W_n(s_j)\}}{s_j^{1/2}}. \qquad (11)$$

The factor $\psi_0(0)$ removes the energy scaling, while the $s_j^{1/2}$ converts the wavelet transform to an energy density. The factor C_δ comes from the reconstruction of a δ function from its wavelet transform using the function $\psi_0(\eta)$. This C_δ is a constant for each wavelet function and is given in Table 2. Note that if the original time series were complex, then the sum of the complex $W_n(s)$ would be used instead.

To derive C_δ for a new wavelet function, first assume a time series with a δ function at time $n = 0$, given by $x_n = \delta_{n0}$. This time series has a Fourier transform $\hat{x}_k = N^{-1}$, constant for all k. Substituting $\hat{x}_k$ into (4), at time $n = 0$ (the peak), the wavelet transform becomes

$$W_\delta(s) = \frac{1}{N} \sum_{k=0}^{N-1} \hat{\psi}^*(s\omega_k). \qquad (12)$$

The reconstruction (11) then gives

$$C_\delta = \frac{\delta j \delta t^{1/2}}{\psi_0(0)} \sum_{j=0}^{J} \frac{\Re\{W_\delta(s_j)\}}{s_j^{1/2}}. \qquad (13)$$

The C_δ is scale independent and is a constant for each wavelet function.

The total energy is conserved under the wavelet transform, and the equivalent of *Parseval's theorem* for wavelet analysis is

$$\sigma^2 = \frac{\delta j \delta t}{C_\delta N} \sum_{n=0}^{N-1} \sum_{j=0}^{J} \frac{|W_n(s_j)|^2}{s_j}, \qquad (14)$$

where σ^2 is the variance and a δ function has been assumed for reconstruction. Both (11) and (14) should be used to check wavelet routines for accuracy and to ensure that sufficiently small values of s_0 and δj have been chosen.

For the Niño3 SST, the reconstruction of the time series from the wavelet transform has a mean square error of 1.4% or 0.087°C.

4. Theoretical spectrum and significance levels

To determine significance levels for either Fourier or wavelet spectra, one first needs to choose an appropriate background spectrum. It is then assumed that different realizations of the geophysical process will be randomly distributed about this mean or expected background, and the actual spectrum can be compared against this random distribution. For many geophysical phenomena, an appropriate background spectrum is either white noise (with a flat Fourier spectrum) or red noise (increasing power with decreasing frequency).

A previous study by Qiu and Er (1995) derived the mean and variance of the local wavelet power spectrum. In this section, the theoretical white- and red-noise wavelet power spectra are derived and compared to Monte Carlo results. These spectra are used to establish a *null hypothesis* for the significance of a peak in the wavelet power spectrum.

a. Fourier red noise spectrum

Many geophysical time series can be modeled as either white noise or red noise. A simple model for red noise is the univariate lag-1 autoregressive [AR(1), or Markov] process:

$$x_n = \alpha x_{n-1} + z_n, \tag{15}$$

where α is the assumed lag-1 autocorrelation, $x_0 = 0$, and z_n is taken from Gaussian white noise. Following Gilman et al. (1963), the discrete Fourier power spectrum of (15), after normalizing, is

$$P_k = \frac{1-\alpha^2}{1+\alpha^2-2\alpha\cos(2\pi k/N)}, \tag{16}$$

where $k = 0 \ldots N/2$ is the frequency index. Thus, by choosing an appropriate lag-1 autocorrelation, one can use (16) to model a red-noise spectrum. Note that $\alpha = 0$ in (16) gives a white-noise spectrum.

The Fourier power spectrum for the Niño3 SST is shown by the thin line in Fig. 3. The spectrum has been normalized by $N/2\sigma^2$, where N is the number of points, and σ^2 is the variance of the time series. Using this normalization, white noise would have an expectation value of 1 at all frequencies. The red-noise background spectrum for $\alpha = 0.72$ is shown by the lower dashed curve in Fig. 3. This red-noise was estimated from $(\alpha_1 + \sqrt{\alpha_2})/2$, where α_1 and α_2 are the lag-1 and lag-2 autocorrelations of the Niño3 SST. One can see the broad set of ENSO peaks between 2 and 8 yr, well above the background spectrum.

b. Wavelet red noise spectrum

The wavelet transform in (4) is a series of bandpass filters of the time series. If this time series can be modeled as a lag-1 AR process, then it seems reasonable that the *local wavelet power spectrum*, defined as a vertical slice through Fig. 1b, is given by (16). To test this hypothesis, 100 000 Gaussian white-noise time series and 100 000 AR(1) time series were constructed, along with their corresponding wavelet power spectra. Examples of these white- and red-noise wavelet spectra are shown in Fig. 4. The local wavelet spectra were constructed by taking vertical slices at time $n = 256$. The lower smooth curves in Figs. 5a and 5b show the theoretical spectra from (16). The dots show the results from the Monte Carlo simulation. On average, the local wavelet power spectrum is identical to the Fourier power spectrum given by (16).

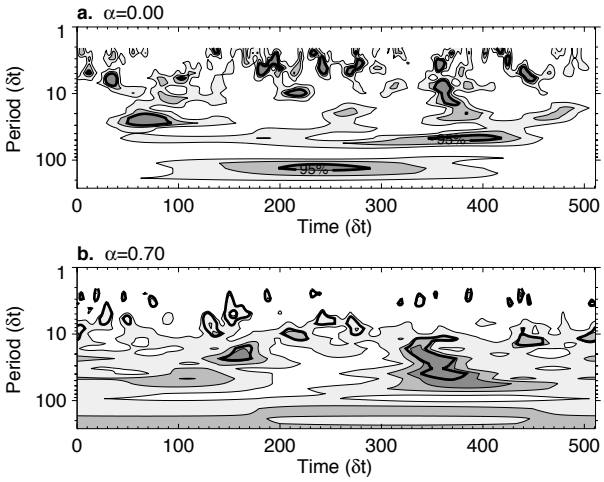

FIG. 4. (a) The local wavelet power spectrum for a Gaussian white noise process of 512 points, one of the 100 000 used for the Monte Carlo simulation. The power is normalized by $1/\sigma^2$, and contours are at 1, 2, and 3. The thick contour is the 95% confidence level for white noise. (b) Same as (a) but for a red-noise AR(1) process with lag-1 of 0.70. The contours are at 1, 5, and 10. The thick contour is the 95% confidence level for the corresponding red-noise spectrum.

Therefore, the lower dashed curve in Fig. 3 also corresponds to the red-noise local wavelet spectrum. A random vertical slice in Fig. 1b would be *expected* to have a spectrum given by (16). As will be shown in section 5a, the average of all the local wavelet spectra tends to approach the (smoothed) Fourier spectrum of the time series.

c. Significance levels

The null hypothesis is defined for the wavelet power spectrum as follows: It is assumed that the time series has a mean power spectrum, possibly given by (16); if a peak in the wavelet power spectrum is significantly above this background spectrum, then it can be assumed to be a true feature with a certain percent confidence. For definitions, "significant at the 5% level" is equivalent to "the 95% confidence level," and implies a test against a certain background level, while the "95% confidence *interval*" refers to the range of confidence about a given value.

The normalized Fourier power spectrum in Fig. 3 is given by $N|\hat{x}_k|^2/2\sigma^2$, where N is the number of points, $\hat{x}_k$ is from (3), and σ^2 is the variance of the time series. If x_n is a normally distributed random variable, then both the real and imaginary parts of $\hat{x}_k$ are normally distributed (Chatfield 1989). Since the square of a normally distributed variable is chi-square distributed with one *degree of freedom* (DOF), then $|\hat{x}_k|^2$ is chi-

square distributed with two DOFs, denoted by χ_2^2 (Jenkins and Watts 1968). To determine the 95% confidence level (significant at 5%), one multiplies the background spectrum (16) by the 95th percentile value for χ_2^2 (Gilman et al. 1963). The 95% Fourier confidence spectrum for the Niño3 SST is the upper dashed curve in Fig. 3. Note that only a few frequencies now have power above the 95% line.

In the previous section, it was shown that the local wavelet spectrum follows the mean Fourier spectrum. If the original Fourier components are normally distributed, then the wavelet coefficients (the bandpassed inverse Fourier components) should also be normally distributed. If this is true, then the wavelet power spectrum, $|W_n(s)|^2$, should be χ_2^2 distributed. The upper curves in Figs. 5a and 5b show the 95% Fourier red-noise confidence level versus the 95% level from the Monte Carlo results of the previous section. Thus, at each point (n, s) in Fig. 1b, assuming a red-noise process, the distribution is χ_2^2. Note that for a wavelet transform using a real-valued function, such as the Mexican hat shown in Fig. 1c, there is only one degree of freedom at each point, and the distribution is χ_1^2.

In summary, assuming a mean background spectrum, possibly red noise [(16)], the distribution for the *Fourier power spectrum* is

$$\frac{N|\hat{x}_k|^2}{\sigma^2} \Rightarrow \frac{1}{2} P_k \chi_2^2 \tag{17}$$

at each frequency index k, and "$\Rightarrow$" indicates "is distributed as." The corresponding distribution for the *local wavelet power spectrum* is

$$\frac{|W_n(s)|^2}{\sigma^2} \Rightarrow \frac{1}{2} P_k \chi_2^2 \tag{18}$$

at each time n and scale s. The 1/2 removes the DOF factor from the χ^2 distribution. (For a real wavelet the distribution on the right-hand side would be $P_k \chi_1^2$.) The value of P_k in (18) is the mean spectrum at the Fourier frequency k that corresponds to the wavelet scale s (see section 3h). Aside from the relation between k and s, (18) is independent of the wavelet function. After finding an appropriate background spectrum and choosing a particular confidence for χ^2 such as 95%, one can then calculate (18) at each scale and construct 95% confidence contour lines.

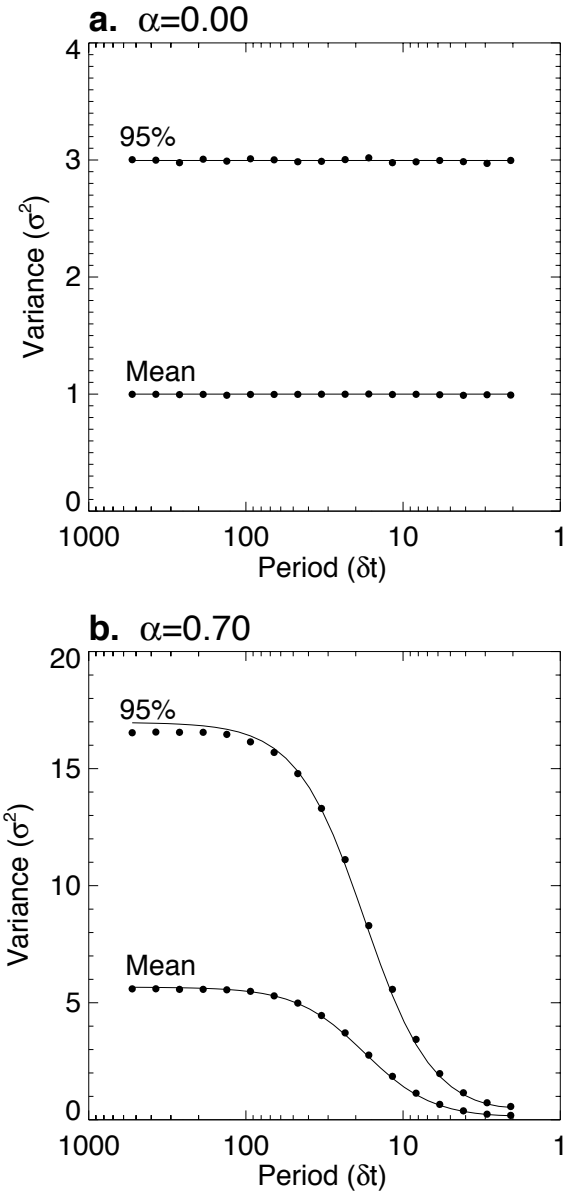

FIG. 5. (a) Monte Carlo results for local wavelet spectra of white noise ($\alpha = 0.0$). The lower thin line is the theoretical mean white-noise spectrum, while the black dots are the mean at each scale of 100 000 local wavelet spectra. The local wavelet spectra were slices taken at time $n = 256$ out of $N = 512$ points. The top thin line is the 95% confidence level, equal to $\chi_2^2(95\%)$ times the mean spectrum. The black dots are the 95% level from the Monte Carlo runs. (b) Same as (a) but for red noise of $\alpha = 0.70$.

As with Fourier analysis, smoothing the wavelet power spectrum can be used to increase the DOF and enhance confidence in regions of significant power. Unlike Fourier, smoothing can be performed in either the time or scale domain. Significance levels and DOFs for smoothing in time or scale are discussed in section 5.

Inside the COI, the distribution is still χ^2, but if the time series has been padded with zeroes, then the mean spectrum is reduced by a factor of $(1 - \frac{1}{2}e^{-2t/\tau_s})$, where τ_s is from Table 1, and t is the distance (in time) from either the beginning or end of the wavelet power spectrum.

The 95% confidence level for the Niño3 SST is shown by the thick contours on Figs. 1b and 1c. During 1875–1910 and 1960–90, the variance in the 2–8-yr band is significantly above the 95% confidence for red noise. During 1920–60, there are a few isolated significant regions, primarily around 2 yr, and at the edge of the usual 2–8 yr ENSO band. The 95% confidence implies that 5% of the wavelet power should be above this level. In Fig. 4b, approximately 5% of the points are contained within the 95% contours. For the Niño3 wavelet spectrum, 4.9% of the points are above 95%, implying that for the Niño3 time series a test of enclosed area cannot distinguish between noise and signal. However, the spatial distribution of variance can also be examined for randomness. In Fig. 4b, the variance shows a gradual increase with period, with random distributions of high and low variance about this mean spectrum. In Figs. 1b and 1c, the significant regions are clustered together in both period and time, indicating less randomness of the underlying process.

d. Confidence interval

The *confidence interval* is defined as the probability that the true wavelet power at a certain time and scale lies within a certain interval about the estimated wavelet power. Rewriting (18) as

$$\frac{|W_n(s)|^2}{\sigma^2 P_k} \Rightarrow \frac{\chi_2^2}{2}, \tag{19}$$

one can then replace the theoretical wavelet power $\sigma^2 P_k$ with the *true wavelet power,* defined as $\mathcal{W}_n^2(s)$. The confidence interval for $\mathcal{W}_n^2(s)$ is then

$$\frac{2}{\chi_2^2(p/2)}|W_n(s)|^2 \le \mathcal{W}_n^2(s) \le \frac{2}{\chi_2^2(1-p/2)}|W_n(s)|^2, \tag{20}$$

where p is the desired significance ($p = 0.05$ for the 95% confidence interval) and $\chi_2^2(p/2)$ represents the value of χ^2 at $p/2$. Note that for real-valued wavelet functions, the right-hand side of (19) becomes χ_1^2, and the factor of 2 is removed from the top of (20). Using (20) one can then find confidence intervals for the peaks in a wavelet power spectrum to compare against either the mean background or against other peaks.

e. Stationarity

It has been argued that wavelet analysis requires the use of nonstationary significance tests (Lau and Weng 1995). In defense of the use of stationary tests such as those given above, the following points are noted.

1) A nonarbitrary test is needed. The assumption of stationary statistics provides a standard by which any nonstationarity can be detected.
2) The test should be robust. It should not depend upon the wavelet function or upon the actual distribution of the time series, other than the assumption of a background spectrum.
3) A non–Monte Carlo method is preferred. In addition to the savings in computation, the chi-square test simplifies comparing one wavelet transform with another.
4) Many wavelet transforms of real data appear similar to transforms of red-noise processes (compare Figs. 1b and 4b). It is therefore difficult to argue that large variations in wavelet power imply nonstationarity.
5) One needs to ask what is being tested. Is it nonstationarity? Or low-variance versus high-variance periods? Or changes in amplitude of Fourier modes? The chi-square test gives a standard measure for any of these possibilities.

In short, it appears wiser to assume stationarity and design the statistical tests accordingly. If the tests show large deviations, such as the changes in ENSO variance seen in Figs. 1b and 1c, then further tests can be devised for the particular time series.

5. Smoothing in time and scale

a. Averaging in time (global wavelet spectrum)

If a vertical slice through a wavelet plot is a measure of the local spectrum, then the *time-averaged wavelet spectrum* over a certain period is

$$\overline{W}_n^2(s) = \frac{1}{n_a}\sum_{n=n_1}^{n_2}|W_n(s)|^2, \tag{21}$$

where the new index n is arbitrarily assigned to the

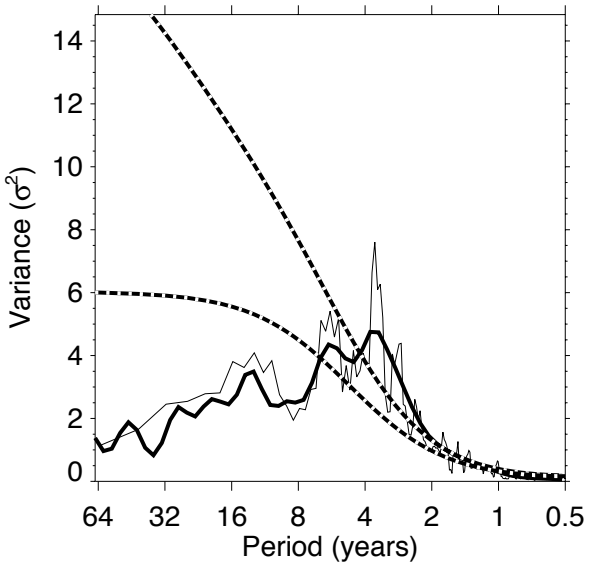

FIG. 6. Fourier power spectrum from Fig. 3, smoothed with a five-point running average (thin solid line). The thick solid line is the global wavelet spectrum for the Niño3 SST. The lower dashed line is the mean red-noise spectrum, while the upper dashed line is the 95% confidence level for the global wavelet spectrum, assuming $\alpha = 0.72$.

midpoint of n_1 and n_2, and $n_a = n_2 - n_1 + 1$ is the number of points averaged over. By repeating (21) at each time step, one creates a wavelet plot smoothed by a certain window.

The extreme case of (21) is when the average is over all the local wavelet spectra, which gives the *global wavelet spectrum*

$$\overline{W}^2(s) = \frac{1}{N}\sum_{n=0}^{N-1}\left|W_n(s)\right|^2 . \qquad (22)$$

In Fig. 6, the thick solid line shows the normalized global wavelet spectrum, $\overline{W}^2(s)/\sigma^2$, for the Niño3 SST. The thin solid line in Fig. 6 shows the same spectrum as in Fig. 3, but smoothed with a five-point running average. Note that as the Fourier spectrum is smoothed, it approaches the global wavelet spectrum more and more closely, with the amount of necessary smoothing decreasing with increasing scale. A comparison of Fourier spectra and wavelet spectra can be found in Hudgins et al. (1993), while a theoretical discussion is given in Perrier et al. (1995). Percival (1995) shows that the global wavelet spectrum provides an unbiased and consistent estimation of the true power spectrum of a time series. Finally, it has been suggested that the global wavelet spectrum could provide a useful measure of the background spectrum, against which peaks in the local wavelet spectra could be tested (Kestin et al. 1998).

By smoothing the wavelet spectrum using (21), one can increase the degrees of freedom of each point and increase the significance of peaks in wavelet power. To determine the DOFs, one needs the number of independent points. For the Fourier spectrum (Fig. 3), the power at each frequency is independent of the others, and the average of the power at M frequencies, each with two DOF, is χ^2 distributed with $2M$ degrees of freedom (Spiegel 1975). For the time-averaged wavelet spectrum, one is also averaging points that are χ^2_2 distributed, yet Figs. 1b and 4 suggest that these points are no longer independent but are correlated in both time and scale. Furthermore, the correlation in time appears to lengthen as scale increases and the wavelet function broadens. Designating ν as the DOFs, one expects $\nu \propto n_a$ and $\nu \propto s^{-1}$. The simplest formula to consider is to define a decorrelation length $\tau = \gamma s$, such that $\nu = 2n_a\delta t/\tau$. However, Monte Carlo results show that this τ is too abrupt at small n_a or large scales; even though one is averaging points that are highly correlated, some additional information is gained.

The Monte Carlo results are given in Fig. 7, which shows the mean and 95% levels for various n_a. These

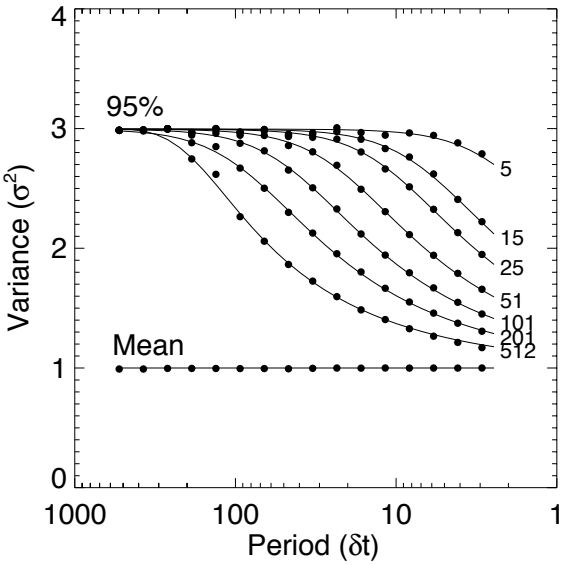

FIG. 7. Monte Carlo results for the time-averaged wavelet spectra (21) of white noise using the Morlet wavelet. The numbers to the right of each curve indicate n_a, the number of times that were averaged, while the black dots are the 95% level for the Monte Carlo runs. The top thin lines are the 95% confidence from (23). The lower thin line is the mean white-noise spectrum, while the black dots are the means of the Monte Carlo runs (all of the means are identical).

curves are best described by the distribution $P_k\chi^2_\nu/\nu$, where P_k is the original assumed background spectrum and χ^2_ν is the chi-square distribution with ν degrees of freedom, where

$$\nu = 2\sqrt{1+\left(\frac{n_a \delta t}{\gamma s}\right)^2}\,. \tag{23}$$

Note that for a real-valued function such as the Mexican hat, each point only has one DOF, and the factor of 2 in (23) is removed. The decorrelation factor γ is determined empirically by an iterative fit of absolute error to the 95% Monte Carlo level and is given in Table 2 for the four wavelet functions. The relative error (or percent difference) between the Monte Carlo and the χ^2_ν/ν distribution was everywhere less than 7% for all scales and n_a values. The thin lines in Fig. 7 show the results of (23) using the Morlet wavelet. Note that even the white noise process has more stringent 95% confidence levels at large scales compared to small. As a final note, if the points going into the average are within the cone of influence, then n_a is reduced by approximately one-half of the number within the COI to reflect the decreased amplitude (and information) within that region.

A different definition of the global wavelet spectrum, involving the discrete wavelet transform and including a discussion of confidence intervals, is given by Percival (1995). An example using Percival's definition can be found in Lindsay et al. (1996).

The 95% confidence line for the Niño3 global wavelet spectrum is the upper dashed line in Fig. 6. Only the broad ENSO peak remains significant, although note that power at other periods can be less than significant globally but still show significant peaks in local wavelet power.

b. Averaging in scale

To examine fluctuations in power over a range of scales (a band), one can define the *scale-averaged wavelet power* as the weighted sum of the wavelet power spectrum over scales s_1 to s_2:

$$\overline{W}_n^2 = \frac{\delta j \delta t}{C_\delta} \sum_{j=j_1}^{j_2} \frac{\left|W_n(s_j)\right|^2}{s_j}\,. \tag{24}$$

Comparing (24) and (14), the scale-averaged wavelet power is a time series of the average variance in a certain band. Thus, the scale-averaged wavelet power can be used to examine modulation of one time series by another, or modulation of one frequency by another within the same time series.

As an example of averaging over scale, Fig. 8 shows the average of Fig. 1b over all scales between 2 and 8 yr (actually 2–7.34 yr), which gives a measure of the average ENSO variance versus time. The variance plot shows a distinct period between 1920 and 1960 when ENSO variance was low. Also shown in Fig. 8 is the variance in the Southern Oscillation index (SOI), which correlates well with the changes in Niño3 SST variance (0.72 correlation). Both time series show consistent interdecadal changes, including a possible modulation in ENSO variance with a 15-yr period. To examine more closely the relation between Niño3 SST and the SOI, one could use the cross-wavelet spectrum (see section 6c).

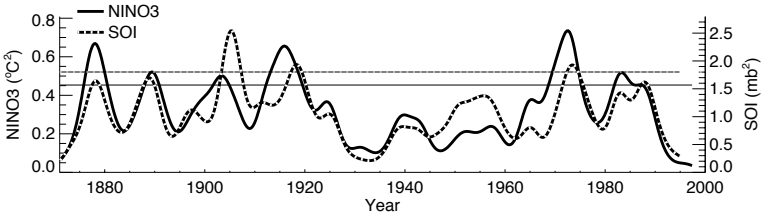

FIG. 8. Scale-averaged wavelet power (24) over the 2–8-yr band for the Niño3 SST (solid) and the SOI (dashed). The thin solid line is the 95% confidence level from (26) for Niño3 SST (assuming red noise $\alpha = 0.72$), while the thin dashed line is the 95% level for the SOI (red noise $\alpha = 0.61$).

As with time-averaged wavelet spectrum, the DOFs are increased by smoothing in scale, and an analytical relationship for significance levels for the scale-averaged wavelet power is desirable. Again, it is convenient to normalize the wavelet power by the expectation value for a white-noise time series. From (24), this expectation value is $(\delta j\, \delta t\, \sigma^2)/(C_\delta S_{\text{avg}})$, where σ^2 is the time-series variance and S_{avg} is defined as

$$S_{\text{avg}} = \left(\sum_{j=j_1}^{j_2} \frac{1}{s_j}\right)^{-1}\,. \tag{25}$$

The black dots in Fig. 9 show the Monte Carlo results for both the mean and the 95% level of scale-averaged wavelet power as a function of various n_a, where $n_a = j_2 - j_1 + 1$ is the number of scales averaged.

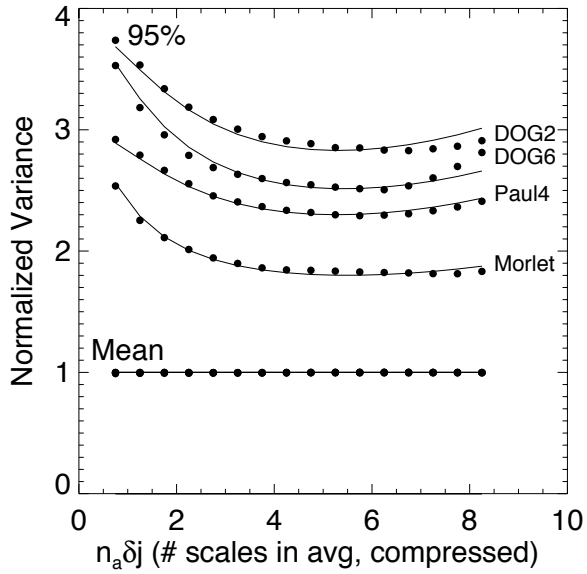

FIG. 9. Monte Carlo results for the wavelet spectra averaged over n_a scales from (24), using white noise. The average from (24) is centered on scale $s = 16\delta t$ for convenience, but the results are independent of the center scale. To make the graph independent of the choice for δj, the x axis has been compressed by the Monte Carlo δj of 0.25. The top black dots are the 95% level for the Monte Carlo runs, while the lower black dots are the means. The means for all four wavelet bases are all the same, while the 95% level depends on the width of the basis in Fourier space, with the Morlet being the most narrow. The top thin lines are the 95% confidence from (28). The lower thin line is the mean white-noise spectrum.

Using the normalization factor for white noise, the distribution can be modeled as

$$\frac{C_\delta S_{\text{avg}}}{\delta j \delta t \sigma^2} \overline{W}_n^2 \Rightarrow \overline{P} \frac{\chi_\nu^2}{\nu}, \tag{26}$$

where the scale-averaged theoretical spectrum is now given by

$$\overline{P} = S_{\text{avg}} \sum_{j=j_1}^{j_2} \frac{P_j}{s_j}. \tag{27}$$

Note that for white noise this spectrum is still unity (due to the normalization). The degrees of freedom ν in (26) is modeled as

$$\nu = \frac{2 n_a S_{\text{avg}}}{S_{\text{mid}}} \sqrt{1 + \left(\frac{n_a \delta j}{\delta j_0} \right)^2}, \tag{28}$$

where $S_{\text{mid}} = s_0 2^{0.5(j_1+j_2)\delta j}$. The factor $S_{\text{avg}}/S_{\text{mid}}$ corrects for the loss of DOF that arises from dividing the wavelet power spectrum by scale in (24) and is observed in the Monte Carlo results. Note that for a real-valued function such as the Mexican hat, each point only has one DOF, and the factor of 2 in (28) is removed. The decorrelation distance δj_0 is determined empirically by an iterative fit of absolute error between (28) and the 95% level of the Monte Carlo results and is given in Table 2. The thin lines in Fig. 9 show the results of (28) for the Morlet, Paul ($m = 4$), DOG2, and DOG6 wavelet functions. For these wavelets, the relative error between the χ_ν^2 distribution using (28) and the Monte Carlo results is less than 1.5%. It should be noted that (28) is valid only for confidences of 95% or less. At higher confidence levels, the distribution begins to deviate significantly from χ^2, and (28) is no longer valid.

In Fig. 8, the thin solid and dashed lines show the 95% confidence levels for the Niño3 SST and the SOI using (25)–(28). In this case, $\delta j = 0.125$, the sum was between Fourier periods 2 and 8 yr (actually 2.1–7.6 yr), $n_a = 16$, $S_{\text{avg}} = 0.221$ yr, $S_{\text{mid}} = 3.83$ yr, $\delta j_0 = 0.60$, and $\nu = 6.44$. Since the two time series do not have the same variance or the same red-noise background, the 95% lines are not equal.

6. Extensions to wavelet analysis

a. Filtering

As discussed in section 3i, the wavelet transform (4) is essentially a bandpass filter of uniform shape and varying location and width. By summing over a subset of the scales in (11), one can construct a *wavelet-filtered* time series:

$$x'_n = \frac{\delta j \delta t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=j_1}^{j_2} \frac{\Re\{W_n(s_j)\}}{s_j^{1/2}}. \tag{29}$$

It is also possible to filter simultaneously in j and n.

This filter has a response function given by the sum of the wavelet functions between scales j_1 and j_2.

This filtering can also be done on both the scale and time simultaneously by defining a threshold of wavelet power. This "denoising" removes any low-amplitude regions of the wavelet transform, which are presumably due to noise. This technique has the advantage over traditional filtering in that it removes noise at all frequencies and can be used to isolate single events that have a broad power spectrum or multiple

events that have varying frequency. A more complete description including examples is given in Donoho and Johnstone (1994).

Another filtering technique involves the use of the two-dimensional wavelet transform. An example can be found in Farge et al. (1992), where two-dimensional turbulent flows are "compressed" using an orthonormal wavelet packet. This compression removes the low-amplitude "passive" components of the flow, while retaining the high-amplitude "dynamically active" components.

b. Power Hovmöller

By scale-averaging the wavelet power spectra at multiple locations, one can assess the spatial and temporal variability of a field of data. Figure 10a shows a *power Hovmöller* (time–longitude diagram) of the wavelet variance for sea level pressure (SLP) anomalies in the 2–8-yr band at each longitude. The original time series at each longitude is the average SLP between 5° and 15°S. At each longitude, the wavelet power spectrum is computed using the Morlet wavelet, and the scale-averaged wavelet power over the 2–8-yr band is calculated from (24). The average wavelet-power time series are combined into a two-dimensional contour plot as shown in Fig. 10a. The 95% confidence level is computed using the lag-1 autocorrelation at each longitude and (26).

Several features of Fig. 10 demonstrate the usefulness of wavelet analysis. Fig. 10c shows the time-averaged 2–8-yr power as a function of longitude. Broad local maxima at 130°E and 130°W reflect the power associated with the Southern Oscillation. This longitudinal distribution of power is also observed in the 2–8-yr band for Fourier spectra at each longitude (not shown). The zonal average of the power Hovmöller (Fig. 10b) gives a measure of global 2–8-yr variance in this latitude band. Comparing this to Fig. 8, one can see that the peaks in zonal-average power are associated with the peaks in Niño3 SST variance, and, hence, the 2–8-yr power is dominated in this latitude band by ENSO.

With the power Hovmöller in Fig. 10a, the temporal variations in ENSO-associated SLP fluctuations can be seen. While the low power near the date line region is apparent throughout the record, the high power regions fluctuate on interdecadal timescales. From the 1870s to the 1920s, strong decadal fluctuations in the 2–8-yr power are observed in the Australian region. In contrast, the eastern Pacific fluctuations are strong only through 1910 and appear to have little

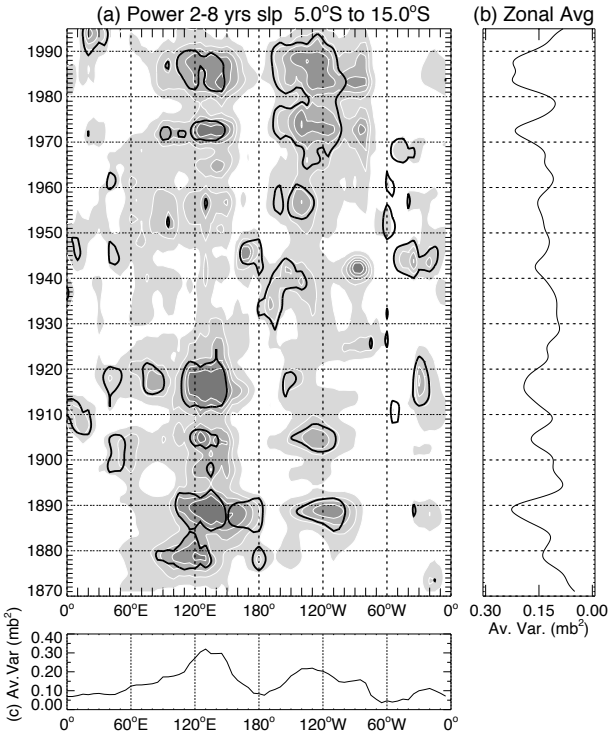

FIG. 10. (a) Power Hovmöller of 2–8-yr averaged wavelet power in SLP. The original time series at each longitude is the average SLP between 5° and 15°S. The contour interval is 0.1 mb². The thick contour is the 95% confidence level, using the corresponding red-noise spectrum at each longitude; (b) the average of (a) over all longitudes; (c) the average of (a) over all times.

power in the 1915–30 period. The generally low power observed in Figs. 1 and 8 between 1930 and 1950 mainly reflects a lack of power in the Australian region, with the eastern Pacific having some significant fluctuations in the 1940s. The large zonal-scale fluctuations in both regions return in the 1950s, with the strongest amplitudes after 1970. The diminished power after 1990 is within the COI, yet may reflect the changes in ENSO structure and evolution seen in recent years (Wang 1995).

c. Cross-wavelet spectrum

Given two time series X and Y, with wavelet transforms $W_n^X(s)$ and $W_n^Y(s)$, one can define the *cross-wavelet spectrum* as $W_n^{XY}(s) = W_n^X(s)W_n^{Y*}(s)$, where $W_n^{Y*}(s)$ is the complex conjugate of $W_n^Y(s)$. The cross-wavelet spectrum is complex, and hence one can define the *cross-wavelet power* as $|W_n^{XY}(s)|$.

Confidence levels for the cross-wavelet power can be derived from the square root of the product of two chi-square distributions (Jenkins and Watts 1968).

Assuming both wavelet spectra are χ^2 distributed with ν DOFs, the probability distribution is given by

$$f_\nu(z) = \frac{2^{2-\nu}}{\Gamma^2\left(\frac{\nu}{2}\right)} z^{\nu-1} \mathrm{K}_0(z), \qquad (30)$$

where z is the random variable, Γ is the Gamma function, and $K_0(z)$ is the modified Bessel function of order zero. The cumulative distribution function is given by the integral $p = \int_0^{Z_\nu(p)} f_\nu(z)\, dz$, where $Z_\nu(p)$ is the confidence level associated with probability p. Given a probability p, this integral can be inverted to find the confidence level $Z_\nu(p)$.

If the two time series have theoretical Fourier spectra P_k^X and P_k^Y, say from (16), then the cross-wavelet distribution is

$$\frac{\left|W_n^X(s) W_n^{Y*}(s)\right|}{\sigma_X \sigma_Y} \Rightarrow \frac{Z_\nu(p)}{\nu} \sqrt{P_k^X P_k^Y}, \qquad (31)$$

where σ_X and σ_Y are the respective standard deviations. For $\nu = 1$ (real wavelets), Z_1 (95%) = 2.182, while for $\nu = 2$ (complex wavelets), Z_2 (95%) = 3.999.

Figure 11a shows the wavelet power spectrum of Niño3 SST using the Paul ($m = 4$) wavelet, while Fig. 11b shows the wavelet power for the SOI. Note that the narrow width in time of the Paul gives better time localization than the Morlet but poorer frequency localization. Finally, Fig. 11c shows the cross-wavelet power for the Niño3 SST and the SOI and indicates large covariance between the time series at all scales between 2 and 8 yr. The 95% confidence level was derived using (31) and assuming red-noise spectra (16) with $\alpha = 0.72$ for Niño3 SST and $\alpha = 0.61$ for the SOI.

d. Wavelet coherence and phase

Another useful quantity from Fourier analysis is the *coherence*, defined as the square of the cross-spectrum normalized by the individual power spectra. This gives a quantity between 0 and 1, and measures the cross-correlation between two time series as a function of frequency. Unfortunately, as noted by Liu (1994), this coherence is identically one at all times and scales. In Fourier analysis, this problem is circumvented by smoothing the cross-spectrum before normalizing. For wavelet analysis, it is unclear what sort of smoothing (presumably in time) should be done to give a useful measure of coherence. This smoothing would also seem to defeat the purpose of wavelet analysis by decreasing the localization in time. Liu (1994) suggests plotting the real and imaginary parts (the co- and quadrature-wavelet spectra) separately, and also plotting the *coherence phase*, defined as $\tan^{-1}\left[\Im\{W_n^{XY}(s)\}/\Re\{W_n^{XY}(s)\}\right]$.

The co- and quadrature-wavelet spectra for the Niño3 SST and the SOI (not shown) do not appear to give any additional information, especially in conjunction with the coherence phase shown in Fig. 11d. The shaded region in Fig. 11d shows where the phase difference between Niño3 SST and the SOI is between 160° and 200°. It is well known that the Niño3 SST and the SOI are out of phase, yet this shows that the time series are within ±20° of being 180° out of phase over all periods between 2 and 8 yr. Furthermore, this out-of-phase behavior is consistent with changes in the cross-wavelet power, with periods of low variance, say between 1920 and 1960, associated with more random phase differences.

7. Summary

Wavelet analysis is a useful tool for analyzing time series with many different timescales or changes in variance. The steps involved in using wavelet analysis are as follows:[1]

1) Find the Fourier transform of the (possibly padded) time series.
2) Choose a wavelet function and a set of scales to analyze.
3) For each scale, construct the normalized wavelet function using (6).
4) Find the wavelet transform at that scale using (4);
5) Determine the cone of influence and the Fourier wavelength at that scale.
6) After repeating steps 3–5 for all scales, remove any padding and contour plot the wavelet power spectrum.
7) Assume a background Fourier power spectrum (e.g., white or red noise) at each scale, then use the chi-squared distribution to find the 95% confidence (5% significance) contour.

For other methods of wavelet analysis, such as the

[1]Software and examples are available from the authors at URL: http://paos.colorado.edu/research/wavelets/.

orthogonal wavelet transform, see Farge (1992). The results presented in section 4 on statistical significance testing are presumably valid for higher-dimensional wavelet analysis (assuming an appropriate background spectrum can be chosen), but this has not been tested and is left to future research. More research is also needed on the properties and usefulness of the cross-wavelet, wavelet coherence, and co- and quadrature spectra.

In the wavelet analysis of Niño3 sea surface temperature, the Southern Oscillation index, and the sea level pressure, it was found that the variance of the El Niño–Southern Oscillation changed on interdecadal timescales, with a period of low variance from 1920 to 1960. Using both the filtered 2–8-yr variance and the cross-wavelet power, the changes in Niño3 SST variance appear to be well correlated with changes in the SOI. The SLP power Hovmöller suggests that these changes are planetary in scale, while Torrence and Webster (1997) use wavelet analysis to show that interdecadal changes in ENSO are also related to changes in Indian monsoon variance. Further studies are necessary to determine the extent and possible causes of these interdecadal changes.

It is hoped that the analysis presented here will prove useful in studies of nonstationarity in time series, and the addition of statistical significance tests will improve the quantitative nature of wavelet analysis. Future studies using wavelet analysis can then concentrate on the results rather than simply the method.

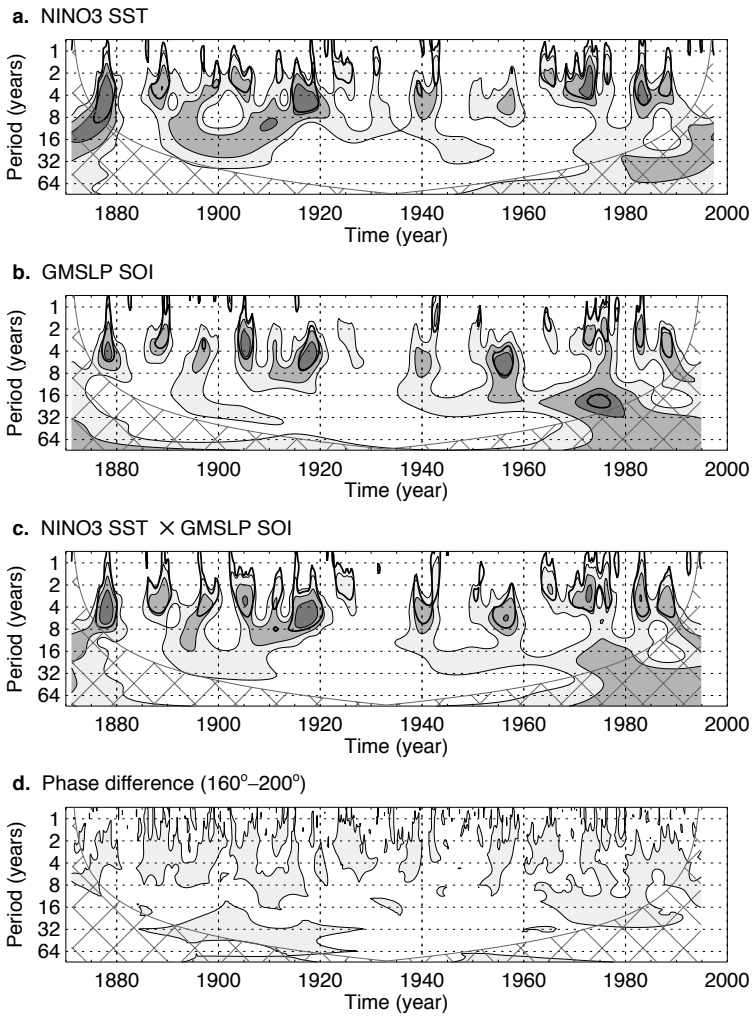

FIG. 11. (a) The wavelet spectrum for the Niño3 SST using the Paul ($m = 4$) wavelet. The contours are at normalized variances of 2, 5, and 10, while the thick contour is the 95% confidence level (red noise $\alpha = 0.72$). (b) same as (a) but for the GMSLP SOI (red noise $\alpha = 0.61$); (c) the cross-wavelet power for the Niño3 SST and the SOI. Contours are at 2, 5, and 10, while the thick contour is the 95% confidence from (31), with the red noise given in (a) and (b); (d) the phase difference between Niño3 SST and the SOI, with the filled contour enclosing regions between 160° and 200°.

Acknowledgments. Thanks to C. Penland, A. Siegal, R. Tomas, and M. Wheeler for many useful discussions on statistics and wavelet analysis. The UKMO GISST and UKMO/CSIRO GMSLP data were kindly provided by D. Rowell and the U.K. Hadley Centre Meteorological Office. The manuscript was much improved through the help of the anonymous reviewers. Special thanks to P. Webster and G. Kiladis for insightful discussions and for providing the impetus and freedom to pursue this research. Computing was done at the PAOS Computer Facility. This work was funded by NOAA Office of Global Programs Grant NA56GP0230 and National Science Foundation Grants NSF ATM9526030 and NSF ATM9525860.

References

Baliunas, S., P. Frick, D. Sokoloff, and W. Soon, 1997: Time scales and trends in the central England temperature data (1659–1990): A wavelet analysis. *Geophys. Res. Lett.,* **24,** 1351–54.

Chatfield, C., 1989: *The Analysis of Time Series: An Introduction*. 4th Ed. Chapman and Hall, 241 pp.

Daubechies, I., 1990: The wavelet transform time-frequency localization and signal analysis. *IEEE Trans. Inform. Theory,* **36,** 961–1004.

——, 1992: *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 357 pp.

Donoho, D. L., and I. M. Johnstone, 1994: Ideal spatial adaptation by wavelet shrinkage. *Biometrika,* **81,** 425–455.

Farge, M., 1992: Wavelet transforms and their applications to turbulence. *Annu. Rev. Fluid Mech.,* **24,** 395–457.

——, E. Goirand, Y. Meyer, F. Pascal, and M. V. Wickerhauser, 1992: Improved predictability of two-dimensional turbulent flows using wavelet packet compression. *Fluid Dyn. Res.,* **10,** 229–250.

Folland, C. K., D. E. Parker, and F. E. Kates, 1984: Worldwide marine temperature fluctuations 1856–1981. *Nature,* **310,** 670–673.

Foufoula-Georgiou, E., and P. Kumar, Eds., 1995: *Wavelets in Geophysics*. Academic Press, 373 pp.

Gamage, N., and W. Blumen, 1993: Comparative analysis of low-level cold fronts: Wavelet, Fourier, and empirical orthogonal function decompositions. *Mon. Wea. Rev.,* **121,** 2867–2878.

Gilman, D. L., F. J. Fuglister, and J. M. Mitchell Jr., 1963: On the power spectrum of "red noise." *J. Atmos. Sci.,* **20,** 182–184.

Gu, D., and S. G. H. Philander, 1995: Secular changes of annual and interannual variability in the Tropics during the past century. *J. Climate,* **8,** 864–876.

Hudgins, L., C. A. Friehe, and M. E. Mayer, 1993: Wavelet transforms and atmospheric turbulence. *Phys. Rev. Lett.,* **71,** 3279–3282.

Jenkins, G. M., and D. G. Watts, 1968: *Spectral Analysis and Its Applications*. Holden-Day, 525 pp.

Kaiser, G., 1994: *A Friendly Guide to Wavelets*. Birkhäuser, 300 pp.

Kestin, T. A., D. J. Karoly, J.-I. Yano, and N. Rayner, 1998: Time–frequency variability of ENSO and stochastic simulations. *J. Climate,* in press.

Lau, K.-M., and H.-Y. Weng, 1995: Climate signal detection using wavelet transform: How to make a time series sing. *Bull. Amer. Meteor. Soc.,* **76,** 2391–2402.

Lindsay, R. W., D. B. Percival, and D. A. Rothrock, 1996: The discrete wavelet transform and the scale analysis of the surface properties of sea ice. *IEEE Trans. Geosci. Remote Sens.,* **34,** 771–787.

Liu, P. C., 1994: Wavelet spectrum analysis and ocean wind waves. *Wavelets in Geophysics,* E. Foufoula-Georgiou and P. Kumar, Eds., Academic Press, 151–166.

Mak, M., 1995: Orthogonal wavelet analysis: Interannual variability in the sea surface temperature. *Bull. Amer. Meteor. Soc.,* **76,** 2179–2186.

Meyers, S. D., B. G. Kelly, and J. J. O'Brien, 1993: An introduction to wavelet analysis in oceanography and meteorology: With application to the dispersion of Yanai waves. *Mon. Wea. Rev.,* **121,** 2858–2866.

Percival, D. P., 1995: On estimation of the wavelet variance. *Biometrika,* **82,** 619–631.

Perrier, V., T. Philipovitch, and C. Basdevant, 1995: Wavelet spectra compared to Fourier spectra. *J. Math. Phys.,* **36,** 1506–1519.

Qiu, L.-J., and M.-H. Er, 1995: Wavelet spectrogram of noisy signals. *Int. J. Elec.,* **79,** 665–677.

Rayner, N. A., E. B. Horton, D. E. Parker, C. K. Folland, and R. B. Hackett, 1996: Version 2.2 of the global sea-ice and sea surface temperature data set, 1903–1994. Tech. Rep., CRTN 74, Hadley Centre for Climate Prediction and Research, 21 pp. [Available from Meteorological Office, London Road, Bracknell, Berkshire RG12 2SY, United Kingdom.]

Spiegel, M. R., 1975: *Schaum's Outline of Theory and Problems of Probability and Statistics*. McGraw-Hill, 372 pp.

Torrence, C., and P. J. Webster, 1998: The annual cycle of persistence in the El Niño–Southern Oscillation. *Quart. J. Roy. Meteor. Soc.,* in press.

Trenberth, K. E., 1976: Spatial and temporal variations of the Southern Oscillation. *Quart. J. Roy. Meteor. Soc.,* **102,** 639–653.

Wang, B., 1995: Interdecadal changes in El Niño onset in the last four decades. *J. Climate,* **8,** 267–285.

——, and Y. Wang, 1996: Temporal structure of the Southern Oscillation as revealed by waveform and wavelet analysis. *J. Climate,* **9,** 1586–1598.

Weng, H., and K.-M. Lau, 1994: Wavelets, period doubling, and time-frequency localization with application to organization of convection over the tropical western Pacific. *J. Atmos. Sci.,* **51,** 2523–2541.