Maximum Likelihood Spectral Fitting: The Batchelor Spectrum

BARRY RUDDICK

Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada

AYAL ANIS

Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada, and Israel Oceanographic and Limnological Research, Haifa, Israel

KEITH THOMPSON

Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada

(Manuscript received 19 May 1999, in final form 14 January 2000)

ABSTRACT

A simple technique for fitting spectra that is applicable to any problem of adjusting a theoretical spectral form to fit observations is described. All one needs is a functional form for the theoretical spectrum and an estimate for the instrumental noise spectrum. The method, based on direct application of the maximum likelihood approach, has several advantages over other fitting techniques. 1) It is unbiased in comparison with other least squares or cost function–based approaches. 2) It is insensitive to dips and wiggles in the spectrum, due to the small number of fitted parameters. It is also robust because the range of wavenumbers used in the fit is held fixed, and the built-in noise model forces the routine to ignore the spectrum as it gets down toward the noise level. 3) The method provides a theoretical estimate for error bars on the fitted Batchelor wavenumber, based on how broad or narrow the likelihood function is in the vicinity of its peak. 4) Statistical quantities that indicate how well the observed spectrum fits the theoretical form are calculated. This is extremely useful in automating analysis software, to get the computer to automatically flag ''bad'' fits.

The method is demonstrated using data from the Self-Contained Autonomous Microstructure Profiler (SCAMP), a free-falling temperature microstructure profiler. Maximum likelihood fits to the Batchelor spectrum are compared to the SCAMP-generated fits and other least squares techniques, and also tested against pseudodata generated by Monte Carlo techniques.

Pseudocode outlines for the spectral fit routines are given.

1. Introduction: Turbulent microstructure—what are χ and ε?

When a thermally stratified fluid is stirred, fluid parcels are moved from warm to cold regions (and vice versa), where they become thermal anomalies. As a result, the mean-square thermal perturbations $\overline{(T')^2}$ grow. The budget equation for the variance of these thermal anomalies is, neglecting a divergence term (Osborn and Cox 1972),

$$\frac{\partial}{\partial t}\overline{(T')^2} = -2\overline{(w'T')}\frac{\partial \overline{T}}{\partial z} - \chi_\theta, \qquad (1)$$

where primes denote turbulent quantities, κ_T the thermal diffusivity, and

$$\chi_\theta \equiv +2\kappa_T\overline{(\nabla T')^2}.$$

The first term on the right-hand side of (1), the product of turbulent heat flux and average thermal gradient, describes the production of thermal anomalies by stirring.

As soon as thermal anomalies are formed, they begin to blend into the background by molecular heat diffusion. The rate of decrease of thermal variance due to molecular diffusion is represented by the second term on the right-hand side of (1), $-\chi_\theta$. If the production and dissipation terms balance on average, the turbulent heat flux $\overline{w'T'}$ can be estimated by measuring χ_θ and the mean temperature gradient (Osborn and Cox 1972). Microstructure profiles of a single component of temperature gradient (i.e., T'/z) on centimeter or smaller scales are often used to infer χ_θ and hence heat flux.

Similarly, stirring imparts kinetic energy [TKE $= (\rho/2)(\overline{u'^2} + \overline{v'^2} + \overline{w'^2})$] to the turbulent motions [first term on the right-hand side of Eq. (2) below], and this

Corresponding author address: Dr. Barry R. Ruddick, Department of Oceanography, Dalhousie University, Halifax, NS B3H 4J1, Canada.
E-mail: barry.ruddick@dal.ca

energy is partially expended in increasing the potential energy of the system [second term on the right-hand side of Eq. (2)], with the rest being dissipated by molecular viscosity [ε, third term on the right-hand side of Eq. (2)]. The TKE budget equation expressing this balance is, approximately (Osborn 1980),

$$\frac{\partial}{\partial t}\overline{\mathrm{TKE}} = -\overline{\rho}\,\overline{(u'w')}\frac{\partial \overline{u}}{\partial z} - g\overline{(\rho' w')} - \overline{\rho}\varepsilon,$$

where

$$\varepsilon = -\nu\overline{\frac{\partial u_i'}{\partial x_j}\left(\frac{\partial u_i'}{\partial x_j} + \frac{\partial u_j'}{\partial x_i}\right)} \qquad (2)$$

is the rate of viscous dissipation and is proportional to the mean-square straining of the turbulent eddies. It is usually measured by direct observation of microscale shears. The rhs expresses, respectively, the increase of TKE by stirring, the conversion to mean potential energy by turbulent buoyancy flux, and energy loss by viscous dissipation.

Observations of ε, usually obtained directly via observations of microscale shear, can be used to infer an upper bound for the turbulent buoyancy flux (Osborn 1980). When both χ and ε are observed, then the combination can be used to infer mixing efficiency (Oakey 1982; Moum 1996) and have been used to test for unequal mixing of heat and salt (Ruddick et al. 1997; St. Laurent and Schmitt 1999).

Dillon and Caldwell (1980) and Oakey (1982) showed that it is possible to estimate χ_θ directly plus ε indirectly by fitting observed temperature gradient spectra to the theoretical "Batchelor" form, whose cutoff wavenumber depends on ε and thermal diffusivity (section 2). An instrument used for this purpose is described in section 3.

The main content of this paper, section 4, is a description of a relatively simple and useful technique for fitting theoretical spectra to observations. It was developed in the context of Batchelor spectral fitting, but should be useful in the context of fitting any theoretical spectrum to observations. It is based on the direct application of the maximum likelihood estimation (MLE) principle, resulting in an efficient estimator. The technique has three additional advantages over more conventional least squares techniques:

- least squares techniques are often biased estimators when the errors are non-Gaussian, while MLE estimators are unbiased;
- error bars—a theoretical estimate for the standard error of the fit parameters (in this case, the Batchelor wavenumber and ε); and
- rejection criteria—a statistical measure to indicate when the observed and theoretical spectra are too poor a match to be accepted.

Monte Carlo tests of the method, in comparison with least squares and other cost function techniques, are shown in section 5. Examples of the application of the technique to Batchelor spectral fitting are given in section 6. Discussion is in section 7, and the MLE metacode is given in the appendix.

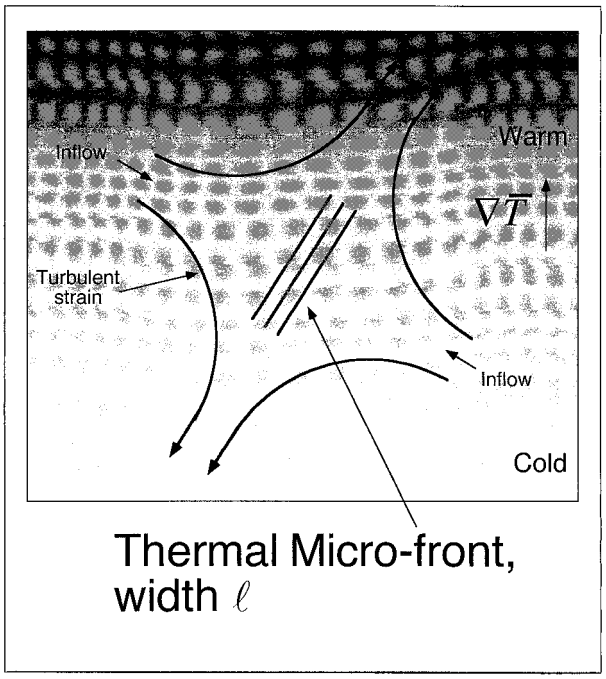

FIG. 1. Straining field in a mean temperature gradient, and microfront produced by the convergence.

2. Batchelor spectrum

Batchelor (1959) discussed the small-scale behavior of passively convected, diffusive scalars like temperature and salinity in the presence of turbulence. He noted that for scales smaller than the Kolmogorov viscous cutoff scale, $(\nu^3/\varepsilon)^{(1/4)}$, the flow appears as a randomly varying straining motion with root-mean-square strain rate (and corresponding inverse timescale) $\tau^{-1} = 0.5(\varepsilon/\nu)^{(1/2)}$. The strain concentrates the large-scale temperature gradient to form a "microfront," and molecular diffusion then acts to keep the front from becoming infinitely sharp (see Fig. 1). Strain and diffusion balance when the frontal width is the diffusion scale, $l = (\kappa_T\tau)^{1/2}$, and eliminating τ in favor of ε gives

$$l^{-1} \equiv k_B = \left(\frac{\varepsilon}{\nu\kappa_T^2}\right)^{1/4}, \qquad (3)$$

the "Batchelor cutoff wavenumber." Note that the units of k_B are in radians per meter; the conversion factor 2π radians per cycle must be applied to convert to cpm.

Note that $\tau N = 2E^{-1/2}$, where $E = \varepsilon/\nu N^2$ and N is the buoyancy frequency, which must be larger than about 15 for the turbulent flow to be actively overturning (Ivey and Imberger 1991, their Table 1). Hence the timescale for production of microfronts is less than half of

the buoyancy timescale, the timescale for evolution of the flow.

Batchelor (1959) was then able to solve the advection–diffusion equation for temperature, driven by turbulent strain and a large-scale temperature gradient, to derive an analytic expression for the high-wavenumber part of the temperature gradient spectrum:

$$S_B(k;\, k_B,\, \chi_\theta) = (q/2)^{1/2}\chi_\theta k_B^{-1}\kappa_T^{-1} f(\alpha),$$

where

$$f(\alpha) = \alpha\left(e^{-\alpha^2/2} - \alpha\int_\alpha^\infty e^{-x^2/2}\,dx\right)$$

$$\alpha = k k_B^{-1}\sqrt{2q}. \tag{4}$$

The universal constant q is in the range 3.4–4.1 (Dillon and Caldwell 1980; Oakey 1982), and the value 3.4 is used here in order to match that used in the Self-Contained Autonomous Microstructure Profiler (SCAMP) software. Dillon and Caldwell (1980) note that a percentage error in q leads to twice the percentage error in ε.

This spectrum is a self-similar form that scales with the Batchelor wavenumber, k_B, rising slowly to the Batchelor cutoff, and falling exponentially beyond it. Changing χ_θ shifts the spectrum vertically, and changing ε with χ_θ fixed shifts the spectrum without change of shape along a line of slope -1 (Fig. 2).

Dillon and Caldwell (1980) made well-resolved observations of temperature gradient spectra and tested these against the hypothesized Batchelor scaling, forming ensemble-averaged dimensionless spectra. They found the observed gradient spectra matched the Batchelor form exceptionally well for intense turbulent events {Cox numbers $\{\chi_\theta/[6\kappa_T)(\overline{T}_2)^2]\} > 2500$} and deviated at low wavenumbers for Cox numbers less than 500. In all cases the peak and cutoff agreed well with theory.

Oakey (1982) made direct velocity-based observations of ε simultaneously with temperature gradient measurements and compared the ε found from Batchelor fits to selected, well-resolved records with the velocity-based ε. He found the agreement to be excellent, proving that ε can be estimated from well-resolved temperature gradient spectra. Kocsis et al. (1998) have more recently compared direct and indirect field observations of dissipation, and found excellent agreement.

3. SCAMP

The SCAMP is a small, hand-held profiling instrument manufactured by Precision Measurement Engineering of Encinitas, California. It free falls slowly through the water column at a nominal late of 0.1 m s^{-1} and measures T at 100 Hz in order to resolve the Batchelor spectral peak. The instrument also can be fitted with microscale conductivity, accurate T, conductivity, and dissolved oxygen sensors, and has a pressure sensor for instrument depth.

Processing sofware (HOST) provided with the system estimates χ_θ and ε by fitting the Batchelor form to the measured temperature gradient spectrum. The HOST software uses a weighted least squares algorithm, adjusting *both* χ_θ and ε to minimize

$$\text{C1} = \sum_{k_i}[(S_{\text{obs}} - S_B)/S_B]^2, \tag{5}$$

which is the square of the difference between the Batchelor spectrum (S_B) and observed (S_{obs}), weighted by the squared Batchelor spectrum. The wavenumber range used for the fit is selected to avoid regions dominated by instrumental noise. Figure 3 illustrates the HOST fit and compares it to the corresponding MLE fit.

Poor convergence of the HOST fit algorithm (version 1.08, since improved) led us to develop improved fit techniques. An additional goal was to develop fitting algorithms with rejection criteria that could be completely automated, making it unnecessary to manually edit the results for poor fits. These should be useful in a variety of spectral fitting contexts, and we describe them in section 4.

We began by constructing an instrumental noise model spectrum and including that in the theoretical fitted spectrum. We also constrained χ_θ match the integrated temperature gradient spectrum less modeled noise, leaving a single parameter to be varied, k_B. Both these improvements are described in more detail in section 4a. The remainder of the fitting procedure involved generating a family of Batchelor spectra plus noise curves for all values of k_B and selecting the one curve that provides the ''best'' match to the observed spectrum (Fig. 2).

4. Maximum likelihood spectral fitting

Fitting a theoretical spectral form to observations is a common problem and is usually approached by a least squares technique. Most time series textbooks discuss the estimation of autoregressive moving average models from time series of observations via least squares estimation; however, those (linear) techniques are difficult to adapt to the fitting of a parametrically defined spectrum such as the Batchelor form. Luketina and Imberger (2001) discuss fitting the Batchelor spectrum to an observed spectrum and describe their experiences with a variety of least squares techniques. They also clearly discuss issues relating to instrument noise, finestructure contamination, and the range of wavenumbers to be used in the fit. The spectral fitting techniques we describe in this section (allowing for instrument noise in the fitted spectrum, reduction to a single-parameter fit, and direct application of the maximum likelihood method) yield a stable, robust, and accurate algorithm that seems to be an improvement over least squares. Although the techniques are not new, they have not been described in the context of the common and important task of fitting a spectrum, and so are described below. We describe general improvements in section 4a, then we give a ''textbook style'' introduction to estimating the mean by the

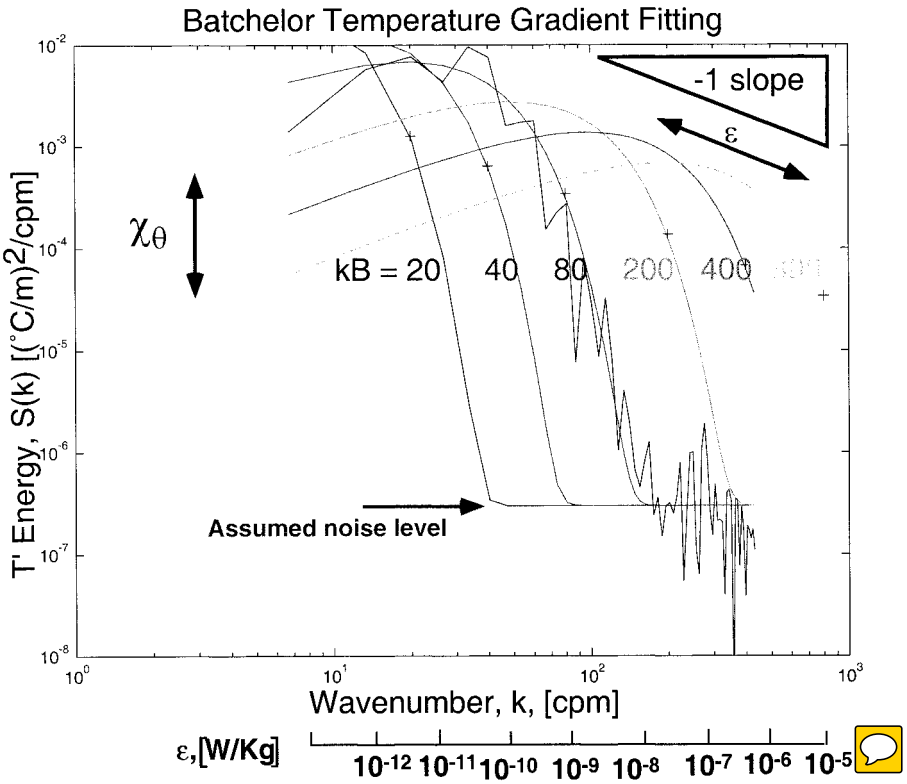

FIG. 2. Batchelor spectrum for various k_B, including estimated SCAMP instrumental noise level [3×10^{-7} (°C m^{-1})^2 (cpm)$^{-1}$], with χ_θ constrained according to Eq. (9). The k_B value corresponding to each curve is indicated with a plus, and the approximate corresponding dissipation level, ε, is shown by the second logarithmic scale below the k axis. The effects of changing χ_θ and ε on the spectrum are indicated by arrows. An observed spectrum is shown.

method of maximum likelihood in section 4b and discuss spectral fitting by direct application of MLE in section 4c. A significant advantage of MLE is that, when the model is valid, the shape of the likelihood function yields an estimate for error bars on the fitted parameters. This is discussed in section 4d.

Another aspect of fitting a spectrum to observations is that of deciding when a fit is too poor to be accepted: sometimes the theoretical spectrum cannot be made to fit the observations acceptably well for any value of the parameters. In section 4e we discuss various measures of spectral misfit and our experiences with them, and recommend a set of criteria that should allow automatic rejection of poor fits.

Cost function approach. Prior to implementing the MLE algorithm (section 4c), we experimented with a variety of least squares–based cost functions, the first being weighted least squares, C1, described in the previous section. (A cost function is simply a measure of misfit.) We experimented with several other cost functions with the arbitrary names C2–C10, which amounted to least squares with different kinds of weighting. The only one of these that offered significant improvement over C1 was

$$\mathrm{C3} = \sum_k [\ln(S_{\mathrm{obs}}) - \ln(S_{\mathrm{th}})]^2, \tag{6}$$

where S_{th} is defined in Eq. (7) below. This adjusts parameters to minimize the mean squared instance in log space between the observed and theoretical spectra. Luketina and Imberger (1999) minimize a similar quantity, involving the mean squared distance in log space along a slope of +1. Luketina and Imberger also give an excellent discussion of the issues relating to instrument drop speed, fine structure contamination, instrumental noise, and the range of wavenumbers to be used in the fit.

We found that C3 works reasonably well, except that Monte Carlo tests in section 5 found it to be biased low. We eventually developed the MLE algorithm because we realized that experimenting with cost functions was too ad hoc, with little theoretical guidance.

a. General improvements

1) ALLOW FOR INSTRUMENTAL NOISE SPECTRUM $S_n(k)$ AS A COMPONENT OF THE THEORETICAL SPECTRUM

It is important to recognize that instrumental noise will act as a lower limit to any measured spectrum.

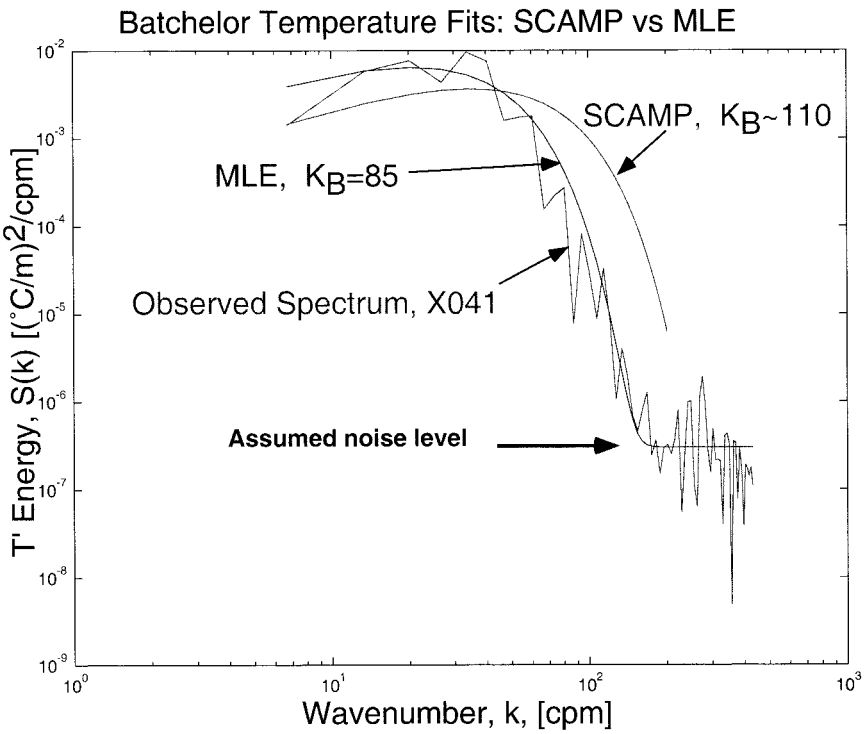

FIG. 3. SCAMP fit example, using C1, Eq. (5). MLE fit (section 4) is shown for comparison. Note the cutoff for the SCAMP fit (C1) to avoid the noise-dominated high-wavenumber region. Same observational spectrum is used in Figs. 2 and 3.

Luketina and Imberger (2001) limit the range of their fit to wavenumbers not dominated by noise. However, if the fitted wavenumber range is varied with the trial k_B, then the cost function has local minima due to the statistical dips and wiggles in the observed spectrum, making it more difficult to locate the global minimum, or to decide which minimum to select (cf. Luketina and Imberger, their Fig. 6).

If the instrumental noise spectrum $S_n(k)$ can be estimated from bench tests, from quiet sections of field records, or any other way, then it is simple to include it in the theoretical spectrum that is fit to observations:

$$S_{\text{th}}(k) = S_B(k) + S_n(k). \tag{7}$$

The simple modification (7) has the benefit that the fitted spectrum only changes in the regions where the theoretical spectrum exceeds the noise level, and remains constant where it lies below (Fig. 2). Thus, any cost function does not change at locations where the signal is less than the noise, and the fit ''ignores'' those regions. For example, the derivative of the cost function C3 with respect to k_B can be easily calculated as

$$\frac{\partial \text{C3}}{\partial k_B} = -2\left(\frac{1}{1 + S_n/S_B}\right)\frac{\partial}{\partial k_B}\ln(S_B). \tag{8}$$

The factor on the right-hand side of Eq. (8) in parentheses goes from 1 to 0 where the signal becomes less than the noise, giving a smoothly varying weighting factor that depends on the expected signal–noise ratio.

In our experience, it is better to slightly overestimate the noise than underestimate it. Otherwise, most fit techniques will spend effort inappropriately trying to fit to noise. The MLE fit in Fig. 3 was done using the simplest of noise models—white noise—and in this case the fit appears to work well and is not sensitive to the chosen noise level. A more detailed and accurate noise model, as is used in all later examples, allows successful fits in situations with a much lower signal–noise ratio. That noise model was made from a parametric fit based on the SCAMP thermistor circuit and theoretical/electronics considerations (M. Head 1999, personal communication).

2) REDUCE THE NUMBER OF FREE PARAMETERS ADJUSTED

Dillon and Caldwell (1980) and Oakey (1982) showed how χ_θ is constrained by the integrated temperature gradient spectrum, after allowance for instrument noise:

$$\chi_\theta = 6\kappa_T \int_0^\infty (S_{\text{obs}} - S_n)\, dk. \tag{9}$$

Kocsis et al. (1998) demonstrated that χ_θ estimated by

fitting it as a free parameter correlated closely with that estimated from the integral (9). In the case of fitting the Batchelor spectrum, only a single parameter, k_B, remains to be adjusted. This amounts graphically to sliding the Batchelor curve along a slope of -1, and selecting the k_B and corresponding ε values that give the best fit (Fig. 2).

The advantages of reducing the dimension of the parameter space searched are many. First, valuable observations are not "used" in adjusting parameters that can be determined in other ways. Second, the adjusted value of a parameter might be inconsistent with other constraints, making it difficult to decide what to believe. Third, searching in multiple dimensions to minimize a computed quantity can be difficult and time consuming, especially if that quantity is not a quadratic. The SCAMP software fit algorithm searched in two dimensions using the Marquardt–Levy algorithm, and did not converge rapidly enough in the k_B direction, leading to significant inaccuracies in that parameter.

b. *A simple example: Maximum likelihood, Gaussian errors, and least squares*

The material in this section can be found in several statistics texts, such as Priestley (1981), and is included to remind the reader how MLE leads to the familiar least squares minimization approach in the special case of Gaussian errors. The MLE concepts are then applied to the specific case of spectral fitting in section 4c.

Suppose we measure the length of an object many times, each time with an error. For example, if the error is distributed as a Gaussian, then the probability of obtaining a measured length y is given by

$$f(y) = C \exp\left[-\frac{(y - \tilde{a})^2}{2\sigma^2}\right], \tag{10}$$

where $\tilde{a}$ is the true length, σ is the standard deviation of the error (here assumed constant), and C is a normalizing factor to make the integral of f equal 1. If we make a histogram of a large number of measurements, it will look something like Fig. 4a. If the error has a different distribution, then f will have a correspondingly different shape, but the concepts and equations below [Eqs. (11)–(13)] still hold.

What should we do if we do not know the true length, $\tilde{a}$, and wish to find the best estimate of it from our observations? The expected probability density function (pdf) of the measured length is shown for $a \sim 1.7$ in Fig. 4b. The MLE procedure is analogous to considering the family of pdf's for all values of a, and picking the one that agrees best with the observed histogram, that is, by sliding the Gaussian curve (Fig. 4b) along the y axis until it agrees best with the histogram (Fig. 4a). For MLE, best agreement is the value of a for which the particular set of observations *is most probable.*

Consider a single measurement—the probability of finding it in the ith bin of Fig. 4a, centered on y_i is the pdf times the bin width:

$$p_i = f(y_i)\Delta y. \tag{11}$$

If we make N measurements, $y_i = y_1, y_2, \ldots, y_N$, the combined probability P of obtaining that particular set of measurements is the product of the probability for each measurement:

$$P = \prod_{i=1}^{N} p_i = \prod_{i=1}^{N} f(y_i)\Delta y \tag{12}$$

Obviously, if the dataset contains lots of measurements that are far from a, then it is a very unlikely dataset: $f(y)$ is small for most points, and so P is small.

Now we turn the question around—we know the y, but not $\tilde{a}$: Given N observations of y, what is the value of a that makes the observed y most likely?

To answer this, we guess a trial value of a, compute the likelihood P according to Eq. (12), and then repeat for all possible values of a. We choose the a that gives the largest P—the one with the *maximum likelihood.* Maximizing P is the same as maximizing $\ln(P)$, so it is more convenient to maximize the sum of the log of individual probabilities:

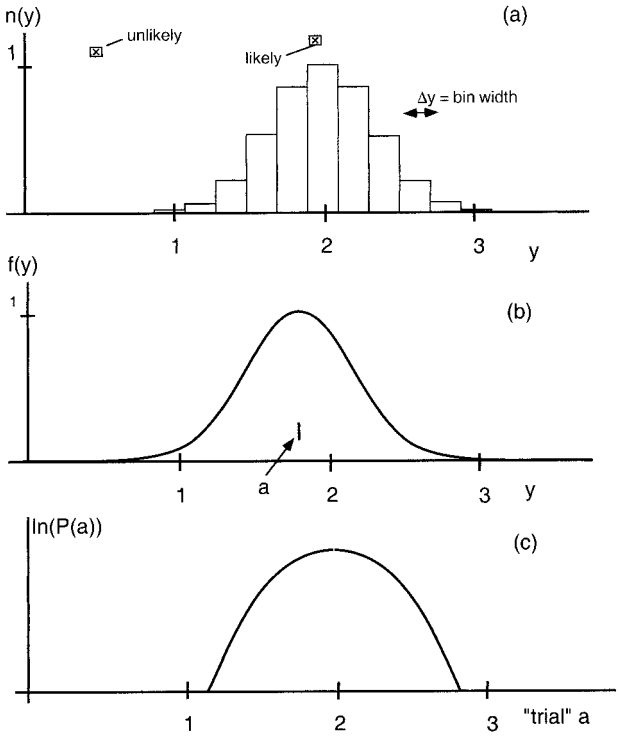

FIG. 4. (a) Histogram of observations of a quantity with true value $\tilde{a} = 2$ and Gaussian error. (b) Hypothesized Gaussian pdf with $a \sim 1.7$. (c) The ln (probability) of our observations as a function of the hypothesized value of a, showing how ln(P) is maximized as a is changed.

$$\ln(P) = \ln\left(\prod_{i=1}^{N} p_i\right) = \sum_{i=1}^{N} \ln(p_i)$$
$$= \sum_{i=1}^{N} \ln[f(y_i)] + N \ln(\Delta y). \quad (13)$$

THE CASE OF GAUSSIAN ERRORS

If the error distribution is Gaussian, then substitution of (10) into (13) yields

$$\ln(P) = \sum_{i=1}^{N} \ln\left\{C\Delta y \exp\left[-\frac{(y_i - a)^2}{2\sigma^2}\right]\right\}$$
$$= \sum_{i=1}^{N} \ln\left\{\exp\left[-\frac{(y_i - a)^2}{2\sigma^2}\right]\right\} + N \ln C + N \ln\Delta y$$
$$= \sum_{i=1}^{N} \left\{-\frac{(y_i - a)^2}{2\sigma^2}\right\} + N \ln C + N \ln\Delta y. \quad (14)$$

Note that σ, N, C, and Δy do not change when a is changed (C is important when the pdf is non-Gaussian, or when estimating σ). So for *Gaussian* error distribution, the a that makes the observed dataset "most likely" is the one that minimizes the sum of squares $S \equiv \Sigma_{i=1}^{N} (y_i - a)^2/2\sigma^2$. If we compute S as a function of the trial a, we will in this case find S is a parabola (Fig. 4c). Minimizing S as a function of a leads to the familiar formula for the arithmetic mean (Priestley 1981):

$$a = \frac{1}{N} \sum_{i=1}^{N} y_i. \quad (15)$$

c. MLE model for spectral fits

The exponential in the Gaussian pdf led in (14) to the sum of squares as the function to minimize. We do not need to assume a Gaussian pdf for the observations, but can work with any other form of pdf. Going back to Eq. (13), we maximize $\ln(P) = \Sigma_{i=1}^{N} [\ln(f(y_i)]$, where f is the expected pdf of y. (Note that the $N\Delta y$ factor is a constant, and so can be ignored.)

Our statistical model assumes the observed spectrum, S_{obs}, is equal to its theoretical value (the Batchelor spectrum plus instrument noise), but is subject to statistical variability (Jenkins and Watts 1968), and is distributed as a χ^2_d pdf, with d degrees of freedom, such that d depends on the spectral technique, window, and averaging methods used (Jenkins and Watts 1968):

$$Y_i \equiv \frac{S_{obs}}{S_{th}} \sim \frac{\chi^2_d}{d}. \quad (16)$$

So the pdf for S_{obs} is

$$f(S_{obs}) = \frac{d}{S_{th}} \chi^2_d\left(\frac{dS_{obs}}{S_{th}}\right), \quad (17)$$

with S_{th} given by (7). The normalizing factor in front makes the integral of f equal 1. From (13), the MLE for k_B is found by maximizing

$$\text{C11} \equiv \ln(P)$$
$$= \sum_{i=1}^{N} \ln\left\{\frac{d}{S_B(k_i; k_B, \chi_\vartheta) + S_n(k_i)} \times \chi^2_d\left[\frac{dS_{obs}(k_i)}{S_B(k_i; k_B, \chi_\vartheta) + S_n(k_i)}\right]\right\} \quad (18)$$

as a function of k_B, with χ_θ constrained by (9). The notation C11 comes from our initial attempts to improve fits using ad hoc cost functions, and is interchangable with MLE. In our experience with fitting a Batchelor spectrum, (18) has a single maximum, so that any simple search technique will suffice. This may not be true for spectra badly contaminated with fine structure, or for other theoretical special shapes, or for much noisier or more poorly resolved signals. If a spectral form with multiple adjustable parameters is fit, then more sophisticated search techniques would need to be used.

There are two special cases of interest. First, when the raw periodogram is used, so that the degrees of freedom for the spectral estimator is $d = 2$, then

$$\chi^2_2(y) = \frac{1}{2} \exp[-(y/2)] \quad (19)$$

(cf. Jenkins and Watts 1968), and C11 reduces to

$$\text{C11}|_{d=2} = -\sum_{i=1}^{N} \frac{S_{obs}}{S_{th}} - \sum_{i=1}^{N} \ln(S_{th}) + N \ln(2) \quad (20)$$

in agreement with Brillinger (1985).

Second, when d becomes asymptotically large (in practice, d must exceed about 60), then the $(1/d)\chi^2_d$ pdf becomes asymptotically normal with mean 1 and variance $2/d$. The MLE estimate C11 becomes a weighted least squares estimate:

$$\text{C11}|_{d\to\infty} \cong -\sum_{i=1}^{N} \left(\frac{S_{obs} - S_{th}}{S_{th}}\right)^2 - \sum_{i=1}^{N} \ln(S_{th}) + N\left(\frac{1}{4} + \sqrt{\frac{d}{4\pi}}\right), \quad (21)$$

which, aside from the second term, is equivalent to C1.

Figure 5c shows the likelihood function exp(C11) as a function of the trial k_B, computed for a SCAMP data segment with a clear signal. The likelihood function is remarkably close to a Gaussian shape, supporting the estimated error bars (shown) that are described in section 4d. Figure 5d shows the SCAMP temperature gradient spectrum, computed with six degrees of freedom (dof's) using a Hanning window, and the MLE-fitted spectrum corresponding to the peak in the likelihood function. We also show the parametric noise model adopted for the fits, and the alternative power law fit (section 4e).

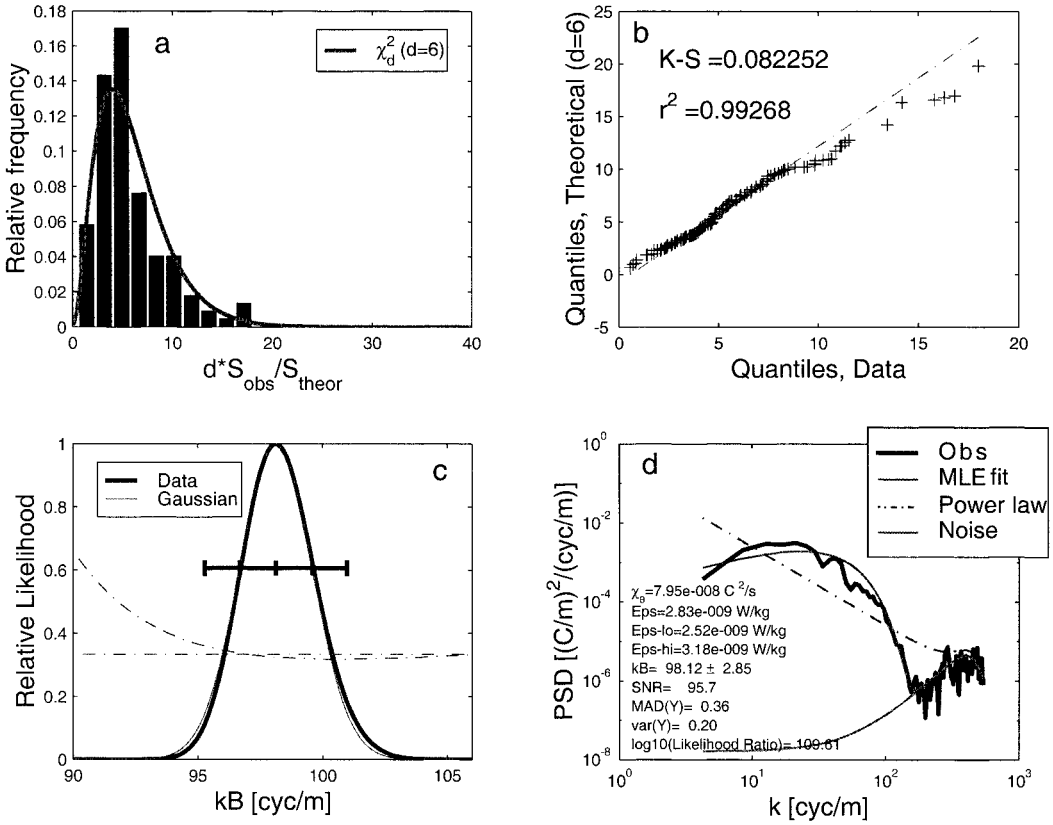

FIG. 5. MLE fit to observed spectrum. (c) Relative likelihood function exp(C11) [Eq. (18)] normalized to a maximum of 1 vs trial k_B, and Gaussian function with same second derivative at maximum. Horizontal error bars indicate one standard deviation according to Eq. (22) by inner vertical bars, plus approx 95% confidence limits by outer vertical bars. Horizontal dash–dot line indicates the theoretical level of the variance [Eq. (23)], and the concave upward dash–dot curve is the total variance (23) as a function of trial k_B, summed over all wavenumbers. (d) Observed spectrum (thick curve) and MLE fit (upper thin curve). Noise model is shown as the lower thin curve. Dash–dot line is least squares fit to a power law plus noise used for likelihood ratio computation (section 4e). MAD(Y) and var(Y) in the legend are computed for the range of wavenumbers for which signal exceeds noise. The signal to noise ratio, SNR, is the ratio of the integrated best-fit Batchelor spectrum to the model noise spectrum. (a) Histogram (bars) of the ratio of observed and theoretical spectra, and theoretical pdf (solid curve) for comparison. (b) Quantile–quantile plot of observed spectral ratio and theoretical pdf. The statistical measures of fit K–S and r^2 are discussed in section 4e.

d. Error bars on fitted parameters

Priestley (1981, section 5.2.2) derives from the Cramer–Rao inequality an expression for the variance of the maximum likelihood estimator (18) for k_B, in terms of the curvature of the log-likelihood function at the maximum:

$$\mathrm{var}(k_B) \geq -\frac{1}{E[\partial^2 \ln(P)/\partial k_B^2]} \approx -\frac{1}{[\partial^2(\mathrm{C11})/\partial k_B^2]}. \quad (22)$$

Thus the curvature of C11 at the peak gives a lower bound estimate for the standard error of the fitted parameter k_B, and corresponding error bounds on ε. This formula works for any parameter estimated via MLE fitting, and can be extended to estimate a covariance matrix for multiple parameter fits (Priestley 1981).

For a Gaussian pdf, (22) can be justified a posteriori by noting (a) from (14) the likelihood function is a Gaussian shape with variance σ^2/N, and (b) the estimate for a is, by (15), the arithmetic mean of N samples drawn from a Gaussian pdf, which is itself a Gaussian with variance σ^2/N.

Thus, for Gaussian observations, the likelihood function and the estimated mean have the same-shaped pdf. This analogy between the likelihood function and the pdf of the estimated parameter allows us to use the curvature of the log-likelihood function to estimate the variance in any MLE-estimated parameter. This analogy is supported by the observation that the computed likelihood function exp(C11) is usually found to be a good match to a Gaussian function with the same curvature (see Fig. 5c).

e. Measures of misfit

There are many situations in which the observed spectrum simply has the wrong shape—one that fails to

match any of the family of theoretical spectra to the expected accuracy. This may be because a data segment contains two or more mixing events with different ε (leading to multiple or indistinct cutoff), because the event is evolving and not in a stationary balance, or for some other reason. In such cases, the spectral fit to k_B should be rejected, although integral variance measures such as χ_θ are probably valid. We present here two possible rejection criteria—objective statistical measures that identify poorly fitting spectra. These should allow automated editing of processed data. We also discuss other goodness of fit measures that did not work very well.

Priestley (1981, section 6.2.6) discusses several ''goodness of fit'' tests designed to answer the following question: Does the spectrum computed from observations conform to the theoretical form hypothesized? However, most of the tests described require that the spectral form be fully specified a priori, without being fit to the observations. Since two parameters, the variance and the Batchelor wavenumber, are constrained by the observations, the tests cannot be used. An exception is Bartlett's homogeneity of the chi-squared test (Priestley 1981, 486–488; Bartlett 1954), which is, according to Priestley, ''asymptotically valid even when [the spectrum] is not fully specified but involves parameters which must be estimated from the data.'' This test involves testing the ratio of the observed and theoretical spectrum to see if it is consistent with a χ^2_d probability distribution. We follow this approach below.

1) PDF TESTS ON SPECTRAL RATIO

When the theoretical and observed spectra match perfectly with the exception of statistical sampling error, the spectral ratio $Y_i \equiv S_{\rm obs}(k_i)/S_{\rm th}(k_i)$ is distributed as d^{-1} times a χ^2_d pdf, which has an expected value of 1 and an expected variance of $2/d$, where d is the equivalent number of degrees of freedom for the spectral estimator used to compute $S_{\rm obs}$. A test of the statistical model (16) is, therefore, to test for this pdf. Figure 5a, shows the observed histogram of Y_i, along with the χ^2_d pdf for $d = 6$ (the number of degrees of freedom appropriate to the spectral window used). The comparison is favorable and certainly makes the point that the spectral errors are far from Gaussian. A more incisive test to see if Y_i, has a χ^2_d pdf is the quantile–quantile (Q–Q) plot in Fig. 5b. A perfect match between theoretical pdf on the y axis and the observed histogram on the x axis would result in a 1:1 straight line. The Q–Q plot is quantified by the r^2 value of the straight-line fit, and/or the Kolmogorov–Smirnov (K–S) test (Priestley 1981, 480–481). The spectral fit is good but not perfect (see the low observed values in the range of 50–90 cpm in Fig. 5d), yet the Q–Q plot is quite close to a straight line, with an r^2 value of 0.97 and a K–S value of 0.08, indicating an acceptable fit. Looking at the histogram or the Q–Q plot requires manual editing, and the r^2 value from the Q–Q plot tended to be high (well above 0.9) even for rather poor fits, and so did not serve as a useful ''rejection statistic.'' We believe that this is because the Y_i statistic tends to be dominated by the large number of spectral estimates from the noise-dominated high wavenumbers. The sensitivity of these measures may be improved by restricting their calculation to wavenumbers for which the signal exceeds the noise, but we did not experiment with that restricted calculation.

Because $Y_i d$ should have a χ^2_d pdf,

$$E[\mathrm{var}(Y_i)] = 2/d. \tag{23}$$

Hence, when the spectrum fits perfectly we expect the variance of Y_i to equal $2/d$ and to increase from that value as the spectrum deviates from the theoretical shape. The ''cost function'' (23), computed for all wavenumbers as a function of k_B, is shown in Fig. 5c, along with the expected value. While the minimum occurs fairly close to the MLE-fit k_B, and is close to (in this case slightly below) the theoretical value, the minimum is very broad compared to the MLE maximum, suggesting that the error in k_B may be larger than that found by MLE. This is found to be true using Monte Carlo tests in section 5.

We also computed the variance of Y_i over the range of wavenumbers for which the fitted spectrum (S_B) exceeds the noise (S_n); the value is shown in the legend of Fig. 5d. This seems to be a potentially useful indicator of spectral misfit (analogous to residual sum of squares); a rejection criterion would be a number significantly larger than that expected. We see in Fig. 6a of the next section that very few variance values meet the perfect fit criterion (23), and we suggest a possible rejection level of $17(2/d)$, or about 5.7. However, experience in using this measure on real data found it to not be as robust as a similar measure, the mean absolute deviation (MAD), also computed over the wavenumber range where signal to noise exceeds 1:

$$\mathrm{MAD2} \equiv \frac{1}{n}\sum_{k_i=k_1}^{k_n}\left|\frac{S_{\rm obs}}{S_{\rm th}} - \left\langle\frac{S_{\rm obs}}{S_{\rm th}}\right\rangle\right|. \tag{24}$$

In (24), k_n indicates the wavenumber where signal matches noise. Experience with both (23) and (24) on real SCAMP data leads us to prefer (24) as being more robust, with MAD values greater than 1.2 (for $d = 6$) indicating unacceptable fits.

One advantage held by var(Y) is the theoretical result that for a perfect fit var(Y) should be $2/d$. We determined the equivalent ''perfect fit'' value for MAD(Y) by computing the expected value of $|\chi^2_d/d - \langle\chi^2_d/d\rangle|$ in a Monte Carlo fashion. We found that an upper bound for MAD2(Y) in the case of perfect fit is $(2/d)^{1/2}$, the square root of E[var(Y)]. We take twice this value as our rejection level.

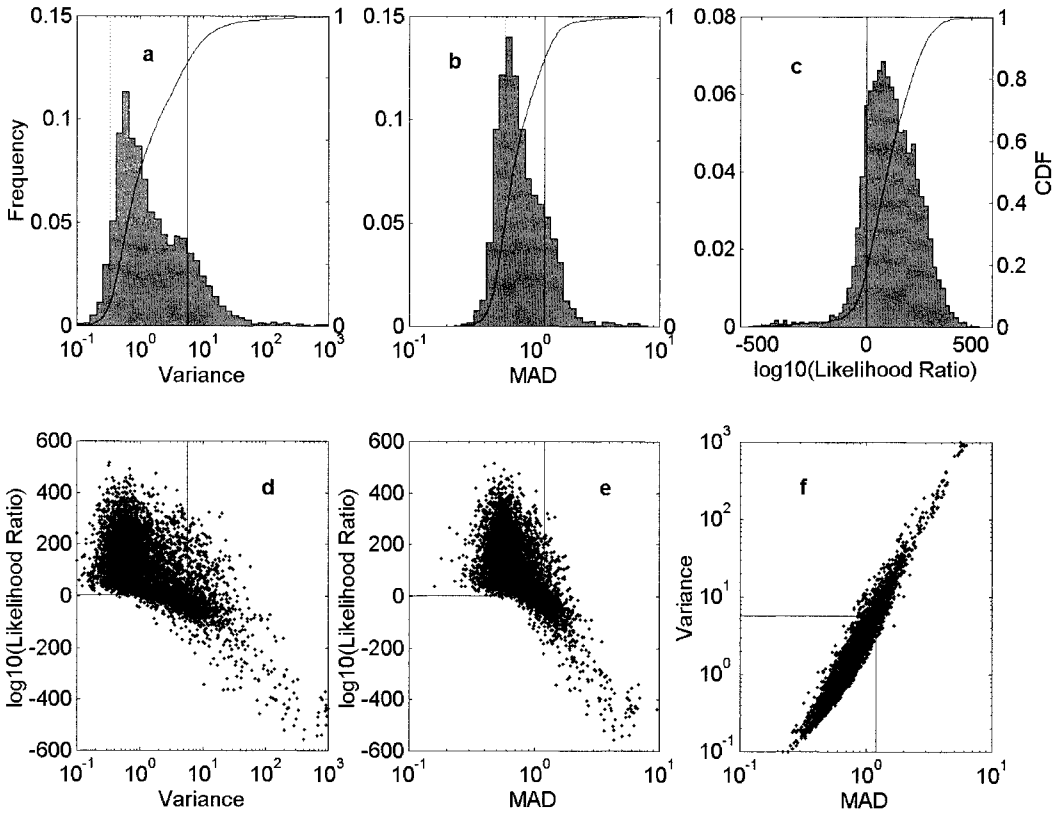

FIG. 6. (a) Histogram and cumulative distribution function of var(Y) from a dataset with 5498 segments and $a = 6$. Dotted vertical line shows theoretical perfect fit variance ($2/d = ⅓$), and vertical thin line shows our suggested rejection criterion of approximately 5.7. (b) As for (a) but for MAD(Y). Here the vertical dotted line is $(\overline{2/d})^{1/2}$, an approximate upper bound for perfect fits verified by Monte Carlo methods. The thin vertical line indicates our suggested rejection limit of $2(\overline{2/d})^{1/2} = 1.2$. (c) As for (a) but for $\log_{10}$(likelihood ratio). The thin vertical line indicates our suggested rejection limit of 2. Note that values *larger* than 2 should be accepted. (d) Scatterplot of $\log_{10}$(likelihood ratio) vs var(Y). Solid lines indicate suggested acceptance values for each measure in (d)–(f). (e) Scatterplot of $\log_{10}$(likelihood ratio) vs MAD(Y). (f) Scatterplot of var(Y) vs MAD(Y).

2) SPECTRAL SMOOTHING—THE SELECTION OF D

Although formula (18) is valid for general dof's, there appears to be little advantage to doing a lot of spectral smoothing. The situation is analogous to the task of fitting a straight line to, say, 20 points. One could group them in clumps of five and fit a straight line to the four averaged points. The statistical fitting error is not reduced, but some bias is introduced, due to lumping of different data points into single averaged values. For the Batchelor spectrum, with a sharp cutoff, spectral smoothing will smooth the cutoff and, therefore, may introduce bias into the fit. The advantage to smoothing is that more spectral smoothing will reduce the scatter of the observed spectrum and hence reduce the expected value of var(Y_i) and MAD(Y), making misfit spectra easier to detect.

3) ALTERNATIVE SPECTRAL MODEL AND THE LIKELIHOOD RATIO

The thin dash–dot line in Fig. 5d is a MLE fit of S_{obs} to a power law $S_{th} = Ak^{-b} + S_n$, with A constrained to preserve the spectral variance. The idea is to compare the Batchelor spectrum fit (which has a sharp cutoff), to a simpler one with no cutoff—a simple power law. If the fit to the Batchelor spectrum does not provide a significantly better fit than the power law fit, then that segment should be rejected as not having a clear cutoff. The two spectral fits are compared by computing the log-likelihood (18) for the power law fit, subtracting it from the log-likelihood for the Batchelor fit and translating the difference to base 10 for easier interpretation. The likelihood ratio, shown in the legend of Fig. 5, is $\log_{10}(P_{Batch}/P_{power})$. Segments with the Batchelor fit being significantly more likely are accepted. Although the likelihood values from two different spectral models (each with one free parameter) cannot be compared with formal statistical validity, both likelihoods are computed with the same statistical model [Eq. (18)], using the same noise model, and a theoretical spectrum with one free parameter. The likelihood ratio appears to be a useful indicator of poor spectral fits. It appears from experience that ratios greater than 10^2 (i.e., $\log_{10}$ values larger than 2) can be confidently accepted.

4) COMPARISON OF REJECTION CRITERIA

Of the rejection criteria that we tried, only var(Y), MAD(Y) (both computed only over the wavenumber range where signal exceeds noise), and the likelihood ratio seemed to reliably indicate unacceptable spectral fits. We investigated and compared the behavior of each on real SCAMP datasets, looking to see if one or the other measure seemed more reliable or robust than the others.

Figure 6 shows in the top row histograms and cumulative distribution functions for each of the measures, computed from a SCAMP dataset. It is comforting to see that the rejection criteria we have settled on are not severe; only a relatively small fraction of the fits would be rejected (about 22% in this dataset). The correspondence between the three measures is shown in the second row. It appears that var(Y) and MAD(Y) are well correlated (Fig. 6f), although var(Y) is porportional to the cube of MAD(Y), rather than the expected second power. This is likely due to var(Y) being more sensitive to outliers and, hence, perhaps less robusts q. The other measure, $\log_{10}$(likelihood ratio), is better correlated with MAD(Y) than var(Y), although the scatterplots in Fig. 6d and 6e are similar. We conclude that the most successful pair of rejection measures is MAD(Y) and $\log_{10}$(likelihood ratio).

5) SUMMARY OF REJECTION CRITERIA

For SCAMP Batchelor fits we suggest that a given data segment be rejected if any of the following criteria are true:

1) integrated signal–noise ratio is less than 1.3,
2) MAD(Y) computed over the wavenumber range where signal exceeds noise is larger than $2(2/d)^{1/2}$, and
3) $\log_{10}$(likelihood ratio) is less than 2.

Criterion 1 avoids the difficult ''low signal'' region delineated by Monte Carlo tests in section 5. For criterion 2 we prefer MAD(Y) over var(Y) as being more robust to outliers. Criterion 3 appears to be the most useful in terms of automated rejection of bad fits. In the dataset of Fig. 6, these criteria result in the rejection of 22.2% of the data segments.

5. Monte Carlo tests

For each of an array of k_B and χ_θ values covering a range suitable for SCAMP fitting, we generated 300 random samples of Batchelor spectra plus noise conforming to the statistical model (17), with $d = 6$ degrees of freedom. These represent datasets for which ''good'' fits should be possible, and for which we knew the answers. These were then analyzed by our fitting routines, using the MLE algorithm and a variety of least squares and cost function algorithms, and the bias (estimated minus true value) and standard deviation, both normalized by the true value of k_B, were computed and expressed in percent. The purpose was to 1) test the MLE method for bias, and to see if the predicted standard deviation of the estimated k_B agreed with the estimated fit error from Eq. (22), and 2) to compare the MLE fit bias and standard deviation with those from the other fit algorithms.

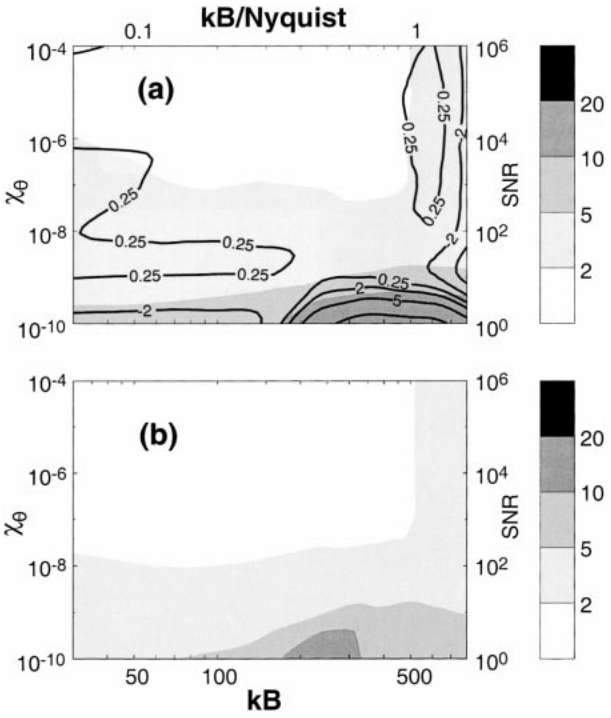

FIG. 7. (a) Standard deviation in percent of the MLE estimated k_B as a function of the ''true'' k_B and χ_θ, shown in grayscale with white contours. The MLE bias in percent is shown as heavy black contours. (b) Estimated standard deviation of the MLE-estimated k_B as a function of the true k_B and χ_θ, computed from the average of 300 realizations of Eq. (22) (i.e., from the average curvature of the 300 log-likelihood functions). Both panels also show k_B scaled by the Nyquist wavenumber and χ_θ (expressed as SNR) on the top and right-hand axes, respectively.

In Fig. 7a, the grayscale contours show the computed standard deviation of the k_B estimates. These should be compared with the standard error of k_B estimated from formula (22) in Fig. 7b. The agreement is excellent except in the lower right-hand corner of Fig. 7b, where the fit becomes noisy due to the coincidence of the noise and Batchelor peaks.

Figure 7a also shows the estimated bias in percent of the MLE estimate of k_B (heavy black contours). It can be seen that the bias is, in practical terms, zero, with the exception of the lower right corner of Fig. 7a, where all fitting methods had a difficult time.

The following least squares and cost function estimates for k_B were made using the Monte Carlo datasets. In each case, the improvements described in section 4a

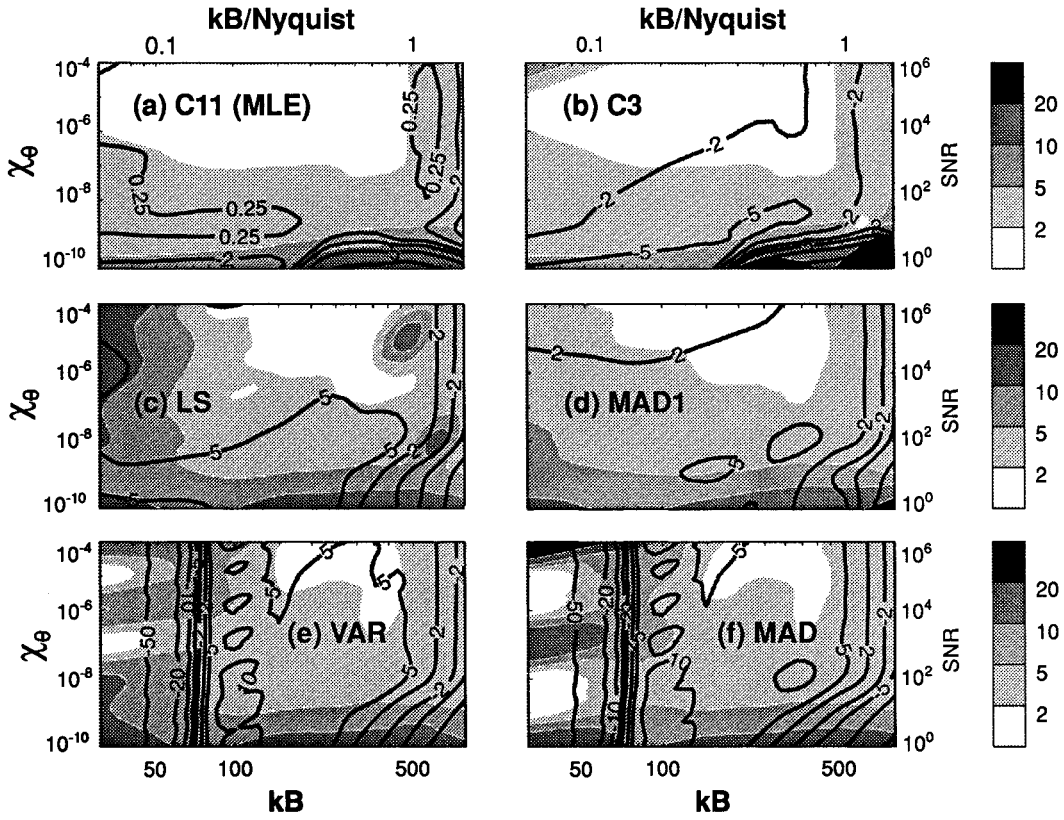

FIG. 8. (a) Standard deviation, in percent, of the MLE-estimated k_B as a function of the true k_B and χ_θ, shown in grayscale with white contours. The MLE bias in percent is shown as heavy black contours. (b) As for (a) but for log-space least squares fitting, C3. (c) As for (a) but for weighted least squares. (d) As for (a) but for MAD1 minimization. (e) As for (a) but for var(Y) minimization. (f) As for (a) but for MAD minimization. All panels show k_B scaled by the sampling Nyquist wavenumber and χ_θ (expressed as signal–noise ratio) on the top and right-hand axes, respectively.

were implemented: the theoretical spectrum included instrument noise, and χ_θ was constrained by the observed variance less noise, leaving a single-parameter fit. The quantities below, summed over all wavenumbers, were maximized as a function of k_B:

1) C11 (MLE fit),
2) C3 [Eq. (6)] least squares minimization in log space,
3) LS—weighted least squares minimization on Batchelor plus noise spectra [as for C1, Eq. (5), with S_B replaced by S_{th}];
4) $$\text{MAD1} \equiv \frac{1}{n}\sum_{k_i=k_1}^{k_n}\left|\frac{S_{\text{obs}}}{S_{\text{th}}} - 1\right|;$$

   mean value of $|Y - 1|$ minimized;
5) var(Y) minimized [Eq. (23)]; and
6) MAD—MAD($Y - \langle Y\rangle$) minimized [Eq. (24)]. This is a robust version of var(Y).

The performance of these estimators is compared to that of the MLE in Fig. 8, which the following can be seen.

- C11 (MLE) has the smallest bias (less than ¼%), and the smallest standard deviation overall (Fig. 8a).
- C3 has, in most of the k_B–χ_θ space, similar standard deviation to the MLE estimate. However, in the high k_B–low χ_θ region it has much larger standard deviations, up to about 100%. In the low k_B–high χ_θ corner the standard deviation is also high. In addition, C3 has a fairly uniform bias of about −2% (Fig. 8b).
- LS suffers from a 5% bias and rather large standard deviation at the low and high ends of the k_B scale (Fig. 8c).
- MAD1 minimization is only slightly inferior in performance to MLE in standard deviation, but is biased about 2% high (Fig. 8d).
- var(Y) minimization has unacceptably high standard deviation and bias (Fig. 8e).
- MAD minimization is very noisy and biased, similar to var(Y) minimization (Fig. 8f).

The conclusion is that while some estimators work nearly as well as MLE (MAD1 in particular), none performs quite as well or is as unbiased as MLE, and the

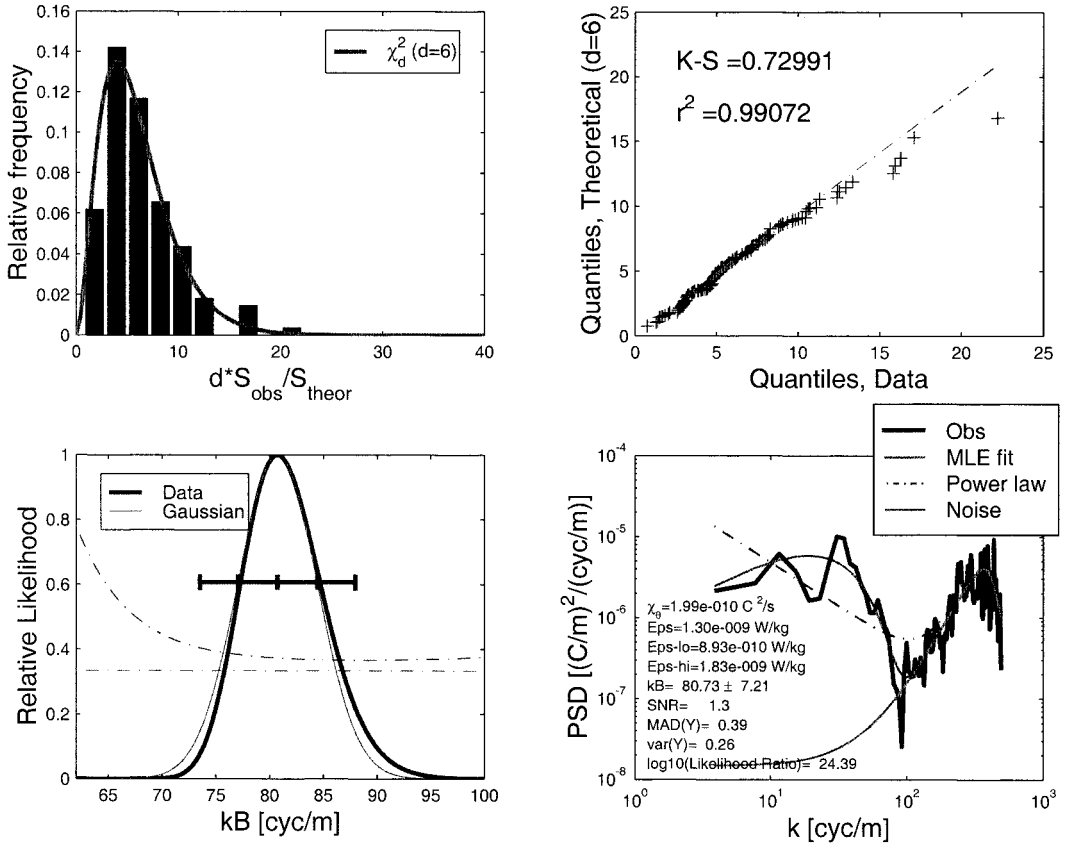

FIG. 9. Example of real data fit with very low signal level. All indications point to a good fit. See Fig. 5 caption for detailed information on each panel.

MLE error estimate (22) is supported by the Monte Carlo tests. MAD1 deserves further study, as it seems to work nearly as well as MLE, and it is robustness properties could make it less sensitive to ''poorly shaped'' spectra. (Summing absolute deviations is a common way to create a robust estimator similar to least squares.)

6. Further examples of MLE fits

Figure 9 shows a MLE fit to a data segment with a low signal level (S–N = 1.3), but with a clear Batchelor spectrum. The fitted spectrum plus noise is a very good match to the observed spectrum, the likelihood function is a reasonable match to a Gaussian, and the estimated error in Batchelor wavenumber (about 9%) is larger than in Fig. 5. The spectral ratio (Y) is a good match to the theoretical pdf, and the Q–Q plot and it is associated statistics so indicate. The rejection indicators both indicate that the fit is acceptable: the Batchelor fit is 10^{24} times more probable than a power law fit, and the MAD is 0.38, which is small. The variance of Y is also small and close to the theoretical value.

Figure 10 shows a fit to a SCAMP data segment that is unacceptable and is flagged as such by the likelihood ratio. The likelihood ratio computation states that a Batchelor spectral fit is $10^{-0.95}$ times more likely (i.e., about 10 times less likely) than a simple power law. This matches what the eye says—the spectrum is the ''wrong shape.'' The other rejection indicators fail to flag this as a bad fit. The deviation between the observed and theoretical spectra is somewhat larger than expected from time series theory. The variance of the spectral ratio (0.6) is significantly larger than the theoretical value of 0.33, and even more so (1.88) when computed over the range of wavenumbers where signal exceeds noise, but not larger than our suggested rejection limit of 5.7. The MAD is also moderately large, 0.78, but also smaller than the suggested rejection limit of 1.2. Yet the histogram and the Q–Q plot in the upper quadrants of the plot seem to say that the fit is acceptable. This is because these plots are dominated by the large number of data points in the wavenumber range for which noise exceeds signal.

7. Discussion

A method to parametrically fit theoretical spectra to observations was developed and demonstrated with fits

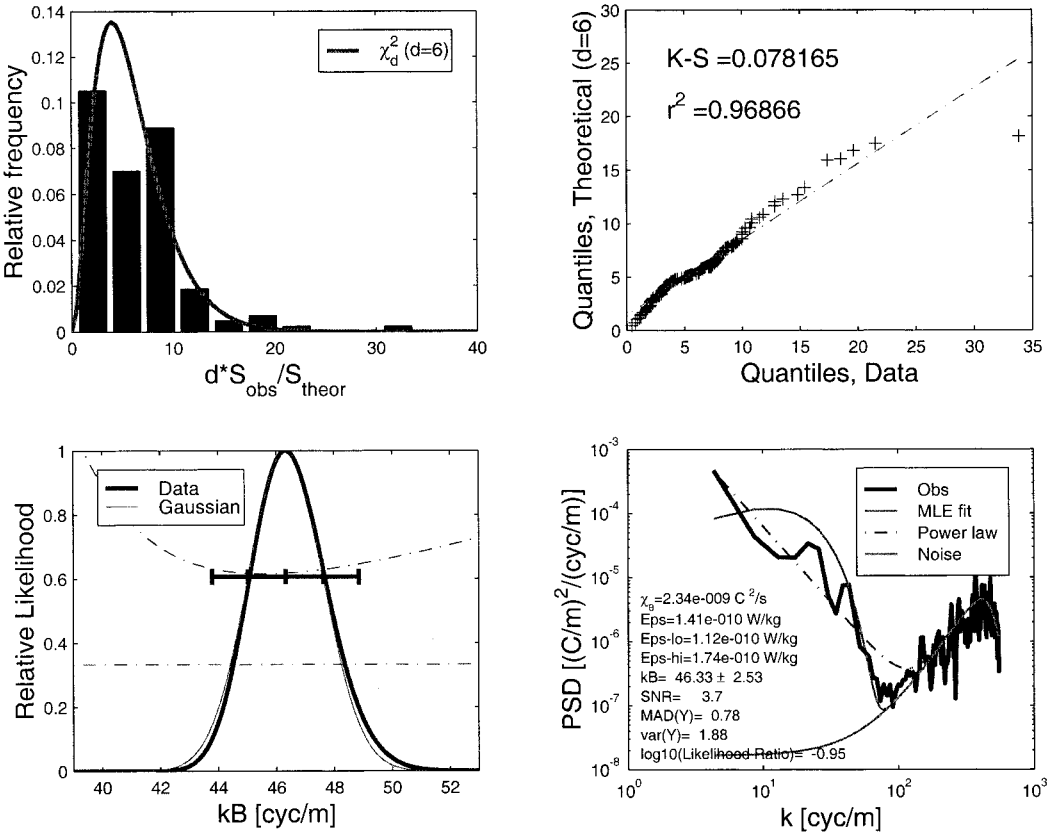

FIG. 10. Example of a ''bad'' fit—one for which the spectral shape is not an acceptable match to the Batchelor form for any value of k_B. See Fig. 5 caption for detailed information on each panel.

to the Batchelor spectrum. The useful techniques include the following key points.

1) Explicit incorporation of an instrumental noise spectrum in the fitted model. This helps the fit routines to automatically ignore regions dominated by instrument noise.
2) Use of available constraints to reduce the dimension of the fitted parameter space.
3) Direct application of the maximum likelihood technique to yield efficient, unbiased spectral fits. This is important because the statistical error in spectral estimation is not Gaussian, and so least squares is not the most appropriate technique.
4) The curvature of the log-likelihood function near its peak gives an estimate of the error in the fitted parameters.
5) We develop and discuss several measures of spectral misfit, useful for rejecting data segments for which the best fit spectrum is unacceptable. These allow automated processing of large numbers of spectra, for which manual editing is tedious.

The MLE technique was compared with other least square and cost function techniques using Monte Carlo–generated datasets that satisfy the expected statistical distribution. The standard deviation and the bias of fitted Batchelor wavenumber are estimated by repeated fits for each of several values of true Batchelor wavenumber and signal level. The MLE technique is found to be the best one, with bias typically less than ¼%, the smallest standard deviation overall, and estimated fit error in agreement with the standard deviation.

Acknowledgments. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), and Canadian GLOBEC. It is a pleasure to acknowledge collaborative discussions with and generous advice from Bruce Smith, Neil Oakey, and Dan Kelley. Mike Head kindly provided the SCAMP noise routine and the algorithms behind much of the SCAMP processing system. Adam Webber, supported by NSERC and the Nova Scotia links program, performed many of the early computations during the development of the fitting algorithms. A. Anis was partly supported by a grant from the German–Israel Foundation.

APPENDIX

MLE Spectral Fit Pseudocode Outline

MATLAB versions of the m files used in this paper for MLE spectral fitting, and also for computing Batchelor spectra, may be requested directly from the authors.

1) Compute noise spectrum versus frequency (call instrument noise subroutine); change to wavenumbers using drop speed of instrument.
2) Compute χ_θ using observed spectral variance less noise variance [Eq. (9)].
3) Loop for trial k_B:
4) compute trial Batchelor spectrum, Eq. (4) (call Batch_Spec subroutine);
5) add computed noise spectrum from line 1;
6) calculate log-likelihood function C11 at each wavenumber [Eq. (18)] (call Chi2pdf transformation function); sum C11 over all wavenumbers;
7) compute var(Y) = var(S_{obs}/S_{th}) summed over all wavenumbers [Eq. (23)].
8) End of loop.
9) Compute location of maximum in C11. This is the best-fit k_B.
10) Compute curvature in C11 function at maximum.
11) Use Eq. (22) to compute standard error in k_B.
12) Plot exp(C11) as likelihood function. Plot Gaussian curve [Eq. (10)] with same standard deviation and peak. Plot error bars corresponding to line 11. Plot line 7 [var(Y) vs trial k_B and theoretical minimum value $2/d$].
13) Compute ε and upper/lower bounds on ε from k_B results in lines 9 and 11 [Eq. (3)].
14) Compute integrated S–N ratio from lines 1 and 2.
15) Compute $Y = S_{obs}/S_{th}$ at each k, for k_B of best fit.
16) Compute var(Y) [Eq. (23)] over wavenumbers for which S_B exceeds S_n.
17) Compute MAD($Y - 1$) [Eq. (24)] over wavenumbers for which S_B exceeds S_n.
18) Loop for power law fit:
19) repeat steps 5–7 using power law instead of Batchelor spectrum; constrain the variance of power law plus noise spectrum to match the observed variance.
20) End of power law loop.
21) Compute location of C11 maximum to find best power law.
22) Take difference of log(likelihood) from lines 9 and 21. Convert to $\log_{10}$ to get likelihood ratio.
23) Plot best-fit spectrum, noise spectrum, and observed spectrum. Superpose best-fit power law spectrum. Annotate this plot with results from lines 2, 13, 9, 11, 14, 16, and 17.
24) Plot histogram of d^*Y, and χ^2_d pdf for comparison.
25) Plot Q–Q of d^*Y compared with χ^2_d cdf. Compute r^2 value and K–S test value.

Subroutines

Instrument noise routine, computes S_n as a function of frequency.

Batch Spec subroutine, computes Batchelor spectrum as function of k given χ_θ and k_B.

Chi2pdf subroutine, computes the χ^2_d pdf. Since this is simply an exponential function and a power, with a normalization constant that involves a gamma function of d, it may save time to compute the gamma function once and store the result, leaving only the power and exponential functions to be evaluated.

REFERENCES

Bartlett, M. S., 1954: Problèmes de l'analyse spectrale des sèries temporelles stationnaires. *Publ. Inst. Stat. Univ. Paris,* **3,** 119–134.

Batchelor, G. K., 1959: Small-scale variation of convected quantities like temperature in turbulent fluid. Part 1. General discussion and the case of small conductivity. *J. Fluid Mech.,* **5,** 113–133.

Brillinger, D. R., 1985: Fourier inference: Some methods for the analysis of array and non-Gaussian series data. *Water Resour. Bull.,* **21,** 743–756.

Dillon, T. M., and D. R. Caldwell, 1980: Batchelor spectrum and dissipation in the upper ocean. *J. Geophys. Res.,* **85** (C4), 1910–1916.

Ivey, G. N., and J. Imberger, 1991: On the nature of turbulence in a stratified fluid. Part I: The energetics of mixing. *J. Phys. Oceanogr.,* **21,** 650–680.

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

Kocsis, O., H. Prandke, A. Stips, A. Simon, and A. Wüest, 1998: Comparison of dissipation of turbulent kinetic energy determined from shear and temperature microstructure. *J. Mar. Syst.,* **21,** 67–84.

Luketina, D. A., and J. Imberger, 2001: Determining turbulent kinetic energy dissipation from Batchelor curve fitting. *J. Atmos. Oceanic Technol.,* in press.

Moum, J. N., 1996: Efficiency of mixing in the main thermocline. *J. Geophys. Res.,* **101,** 12 057–12 069.

Oakey, N. S., 1982: Determination of the rate of dissipation of turbulent energy from simultaneous temperature and velocity shear microstructure measurements. *J. Phys. Oceanogr.,* **12,** 256–271.

Osborn, T. R., 1980: Estimates of the local rate of vertical diffusion from dissipation measurements. *J. Phys. Oceanogr.,* **10,** 83–89.

——, and C. S. Cox, 1972: Oceanic fine structure. *Geophys. Fluid Dyn.,* **3,** 321–345.

Priestley, M. B., 1981: *Spectral Analysis and Time Series.* Academic Press, 890 pp.

Ruddick, B. R., D. Walsh, and N. Oakey, 1997: Variations in apparent mixing efficiency in the North Atlantic central water. *J. Phys. Oceanogr.,* **27,** 2589–2605.

St. Laurent, L., and R. W. Schmitt, 1999: The contribution of salt fingers to vertical mixing in the North Atlantic Tracer Release Experiment. *J. Phys. Oceanogr.,* **29,** 1404–1424.