The Variance of Mean Sea-Ice Thickness: Effect of Long-Range Dependence

D. B. Percival and D. A. Rothrock

Applied Physics Laboratory, University of Washington, Seattle, Washington,

USA

A. S. Thorndike

Department of Physics, University of Puget Sound, Tacoma, Washington,

USA

T. Gneiting

Department of Statistics, University of Washington, Seattle, Washington,

USA

D. B. Percival, Applied Physics Laboratory, University of Washington, Box 355640, Seattle, WA 98195–5640, USA. (dbp@apl.washington.edu)

D. A. Rothrock, Applied Physics Laboratory, University of Washington, Box 355640, Seattle, WA 98195–5640, USA. (rothrock@apl.washington.edu)

A. S. Thorndike, Department of Physics, University of Puget Sound, 1500 N. Warner, Tacoma, WA 98416, USA. (thorndike@ups.edu)

T. Gneiting, Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195–4322, USA. (tilmann@stat.washington.edu)

Abstract. Measured sea-ice draft exhibits variations on all scales. We regard a draft profile over small scales, tens of kilometers, as being drawn from a random process whose probability density function is the same along the profile. We focus on the estimation of the mean draft $\overline{H}$ of the process. This elementary statistic is typically estimated from a profile segment of length L and has some uncertainty, or sampling error, that is quantified by its variance σ_L^2. We answer the question: how efficiently can the variance of $\overline{H}$ be reduced by the use of more data, that is, by increasing L? Three properties of the data indicate the need for a non-standard statistical model: the variance σ_L^2 of $\overline{H}$ falls off more slowly that L^{-1}; the autocorrelation sequence does not fall rapidly to zero; and the spectrum does not flatten off with decreasing wavenumber. These indicate that ice draft exhibits, as a fundamental geometric property, 'long-range dependence.' One good model for this dependence is a fractionally differenced process, whose variance σ_L^2 is proportional to $L^{-1+2\delta}$. From submarine ice draft data in the Arctic Ocean, we find $\delta = 0.27$. In addition, we find that the variance at $L = 50$ km is independent of $\overline{H}$ to a good approximation. Mean draft estimated from a 50-km sample has a sample standard deviation of 0.40 m; for 200 km, it is 0.29 m. Tabulated values provide the sample standard deviation σ_L for various values of L for samples both in a straight line and in a rosette or spoke pattern.

1. Introduction

Sea-ice draft, such as can be measured by a submarine transiting the Arctic Ocean, exhibits variations on all horizontal scales. We take the point of view that small scale variations up to several hundred kilometers, are manifestations of a random process; i.e., we regard the draft profile over such distances as being drawn from a random process whose probability density function is the same along the profile [*Thorndike et al.*, 1975]. Complete geophysical information is contained not in the particular profile, but in the statistical ensemble from which our profile is one representative. An elementary and crucial statistic is the mean draft $\overline{H}$. This statistic is of particular geophysical interest since, as an estimator of the first moment of the probability density function, it is directly related to the volume and mass of sea ice per unit area. Under this framework, variations on larger scales of hundreds to thousands of kilometers represent variations in the underlying probability density function. The small scale variations in draft are random and unpredictable. The large scale variations in any statistic such as $\overline{H}$ are related to predictable phenomena in the atmosphere and ocean and are themselves predictable.

This point of view allows us to address questions of the accuracy of an observation of $\overline{H}$. By accuracy, here we mean representativeness, not instrumental accuracy, which is an entirely separate issue [*Rothrock and Wensnahan*, 2006]. Our attempt to measure $\overline{H}$ is based typically on a profile segment of length L and inevitably has some uncertainty, or sampling error. In general using more data in the estimation reduces this sampling error. The practical issue we address here is how efficiently the uncertainty in $\overline{H}$ can be reduced by the use of more data, that is, by increasing L. This issue is central to all attempts

to use observations to characterise the state—or changes in the state—of sea-ice draft or thickness. To address this issue, we need to assume some sort of model for the spatial variations of ice draft.

We argue in this paper that the nature of sea-ice drafts dictates the use of nonstandard statistical models. The simplest standard model would assume that the measurements are uncorrelated. As we demonstrate empirically in section 2, this assumption is problematic and leads to estimates of the variance in $\overline{H}$ that are much smaller than are realistic. A common way to account for correlation in climate research is to use a first-order autoregressive process as a model [*von Storch and Zwiers*, 1999]. We demonstrate empirically in section 2 that this more refined standard model is still not a good match for the actual measurements. The problem with the autoregressive model is that it assumes that the autocorrelation between sea-ice drafts at two distinct locations decays exponentially as the distance between the locations increases. Again we show empirically this assumption does not match reality. We then argue in section 3 that the measurements are better described by a model exhibiting 'long-range' (or 'long-memory') dependence. The key feature of this nonstandard model is that the autocorrelation decays as a power law, a much slower rate. It should be emphasized that long-range dependence is not just a statistical characterization that affects only sampling errors, but is a fundamental geometric property of the ice pack, akin to roughness and fractal properties, lead and floe size distributions, and ridge and keel properties.

The data used in this paper are acoustic measurements of sea-ice draft, taken by upward looking sonars mounted on U.S. Navy submarines. In particular, we use data from two SCientfic ICe EXpedition (SCICEX) cruises within the Arctic Ocean: one in September

and October of 1996 and one in September of 1997. The data are archived at the National Snow and Ice Data Center in Boulder, Colorado, USA, along with data from numerous other cruises that spanned three decades. Sea-ice draft, which is the submerged portion of sea ice, is about 89% of the ice thickness and is a good proxy for thickness.

2. Interpreting Variability in Mean Drafts via Standard Statistical Methods

We begin by considering a set of ten draft profiles collected in a rosette pattern in October 1996 (thin curves in the right-hand part of Figure 1). Although data are available at one meter spacing, we take each profile to consist of a certain number of one kilometer averages $\overline{H}_1$. In all, there are 1814 such averages, so the average length of the ten profiles is 181.4 km. The sample mean of these averages is 2.36 m, and their sample variance is $\hat{\sigma}_1^2 = 0.390$ m^2, which gives us an empirical measure of the sampling variability in any given $\overline{H}_1$. We then take the one kilometer averages for each of the ten profiles and average them pairwise to form two kilometer averages $\overline{H}_2$. There are 904 such averages (there would have been 907 averages if all profiles had had an even number of data points). The sample mean of these two kilometer averages is again 2.36 m; however, their sample variance is $\hat{\sigma}_2^2 = 0.248$ m^2, which is notably smaller than $\hat{\sigma}_1^2$. This demonstrates empirically that a mean draft over one kilometer has greater variability than a two kilometer mean draft, which is no surprise.

We can continue in this manner to build up a picture of how the variability in draft changes as we increase the distance over which measurements are averaged. We do so by taking the two kilometer averages and averaging them pairwise to form four kilometer averages $\overline{H}_4$, yielding $\hat{\sigma}_4^2 = 0.154$ m^2. Subsequent pairwise averaging yields values for $\hat{\sigma}_8^2$,

$\hat{\sigma}^2_{16}$ and $\hat{\sigma}^2_{32}$. Figure 2a shows a plot of $\hat{\sigma}^2_L$ versus averaging lengths $L = 1, 2, 4, 8, 16$ and 32 (circles). As expected, the variability decreases as L increases.

We can also demonstrate a similar monotonic decrease in variability using a longer single draft profile of one kilometer averages collected in 1997 (thick curve in the right-hand part of Figure 1). This profile passes through the 1996 rosette region. Here there is a noticeable overall linear trend in the draft profile, which we have eliminated via a least squares fit. The residuals from the fit are shown in Figure 3. If we apply the same scheme as before to these residuals, we obtain the values for $\hat{\sigma}^2_L$ shown in Figure 2b (circles). Although the variability is systematically lower here, the pattern of decay is similar to what we observed before.

Note that, on a log/log plot, the patterns of decay in the variances look fairly linear in both Figures 2a and b. In an attempt to explain this observed rate of decay, suppose we entertain the simplest possible standard statistical model for our measurements, namely, that the one kilometer averages can be considered to be independent realizations from some distribution with an unknown variance σ^2_1. Standard statistical theory then says that the population variance σ^2_L in $\overline{H}_L$ should be related to the variance σ^2_1 in $\overline{H}_1$ via the equation

$$\sigma^2_L = \sigma^2_1 \times L^{-1}; \tag{1}$$

i.e., the rate of decay is L^{-1}, which, on a log/log plot, becomes a line with a slope of -1. In Figures 2a and b, we have plotted as thin lines $\hat{\sigma}^2_1/L$ versus L for the rosette and SCICEX 97 profiles. Clearly the empirically observed rate of decay is much slower than under the independence assumption.

A more realistic model must take into account the fact that draft measurements are not independent but rather are correlated. A standard way of handling a correlated time series in climatology is to assume that its correlation properties are dictated by a first-order autoregressive model. Under this model, the correlation between $\overline{H}_1$ measurements that are adjacent to each other in the SCICEX 97 profile is ϕ, where typically $0 \leq \phi < 1$ (when $\phi = 0$, the measurements are uncorrelated, which is implied by the assumption of independence). We note two important implications of the autoregressive model. First, the correlation between $\overline{H}_1$ measurements that are separated by d km is ϕ^d. Second, the variances of $\overline{H}_L$ and $\overline{H}_1$ are related by

$$\sigma_L^2 = \sigma_1^2 \times \frac{1+\phi}{1-\phi}\left(1 - \frac{2\phi(1-\phi^L)}{L(1-\phi^2)}\right) \times L^{-1} \tag{2}$$

(see Appendix A for details). In Figure 2b the dotted curve shows σ_L^2 versus L in equation (2) with σ_1^2 replaced by $\hat{\sigma}_1^2$. The autoregressive model fails to predict the observed rate of decay as L increases. Note that equation (2) reduces to

$$\sigma_L^2 \approx \sigma_1^2 \times \frac{1+\phi}{1-\phi} \times L^{-1} \text{ as } L \to \infty. \tag{3}$$

The difference between the above approximation and equation (1) is just a multiplicative factor $(1+\phi)/(1-\phi)$. This says that, as L gets large, the rate of decay for the autoregressive model is proportional to L^{-1}, just as under the independence assumption, which explains why the thin curve in Figure 2b is parallel to the thick line for large L. (For higher order autoregressive models, σ_L^2 is also proportional to L^{-1} for large L, so the observed rate of decay cannot be explained using an asymptotic approximation by just increasing the autoregressive order.)

Why does the autoregressive model fail here? One telltale sign is the autocorrelation sequence. Figure 4a shows the sample autocorrelation sequence for the SCICEX 97 profile (circles), i.e., the observed correlation between values of the profile that are separated by d km. For $d = 1$ km, we obtain an observed correlation between adjacent profile measurements of $\hat{\phi} = 0.36$. The autoregressive model predicts the correlation between profile measurements separated by $d > 1$ km to be $\hat{\phi}^d$, indicated by the thick curve. The thin curves above and below the thick curve indicate limits within which we can reasonably expect to find the sample autocorrelation sequence (at any given d, these are 95% confidence limits under the autoregressive model). Eight out of the first ten values from this sequence are above the upper 95% confidence limit. Clearly the autoregressive model systematically underestimates the observed autocorrelation sequence and postulates a sequence that damps down to zero much more quickly than is actually observed.

A second telltale is the spectrum (also known as the power spectral density function). Figure 5 shows an estimate of the spectrum (thin jagged curve), along with a 95% confidence interval for the true spectrum based upon this estimate at the lowest displayed wavenumber (0.003 cycles/km, which corresponds to a period of 341 km; see Appendix B for details about this estimate). The dotted curve shows the spectrum implied by the autoregressive model. This spectrum falls outside of the displayed 95% confidence interval, and the same statement holds at other low wavenumbers. While the autoregressive model is capable of capturing the high wavenumber portion of the estimated spectrum, it significantly underestimates the observed spectrum at the lowest wavenumbers, which correspond to the longest lags or scales.

3. Interpreting Variability in Mean Drafts via Long-Range Dependence

In the previous section we noted empirically that the variability in mean drafts cannot be adequately explained either by a model that assumes independence or by one that allows correlation as dictated by a first-order autoregressive model. The three symptoms noted are that the variance in the mean draft falls off more slowly than L^{-1}, the autocorrelation sequence does not fall to zero sufficiently rapidly, and the spectrum does not flatten off with decreasing wavenumber. These symptoms indicate long-range dependence. The hydrologist Hurst is credited with sparking interest in long-range dependence amongst geophysicists in a study of long-term storage capacity of reservoirs in the Nile River Basin [*Hurst*, 1951]. Models with long-range dependence have been applied to such diverse topics as assessing power output from wind-driven turbines [*Haslett and Raftery*, 1989], investigating long-term trends in global temperatures [*Bloomfield*, 1992; *Smith*, 1993], interpreting variability in North Pacific sea level pressure time series [*Percival et al.*, 2001] and studying hydroclimatological time series [*Cohn and Lins*, 2005]. Good general references are *Beran* [1994], *Doukhan et al.* [2003] and *Rangarajan and Ding* [2003].

A simple model that exhibits long-range dependence is a fractionally differenced process [*Granger and Joyeux*, 1980; *Hosking*, 1981]. The precise definition of this process is deferred to Appendix A, but suffice it to say that one can construct a good approximation to a fractionally differenced process by averaging together a collection of first-order autoregressive processes with different ϕ's [*Mandelbrot*, 1971; *Granger* 1980; *Beran*, 1994; see also equation (26)]. Any given first-order autoregressive process has a 'decorrelation length' of $(1 + \phi)/(1 - \phi)$ lags, which can be interpreted as the approximate range of dependence for such a process [*von Storch and Zwiers*, 1999, section 17.1]. The aver-

aged process does not have a single well-defined decorrelation length, but rather exhibits nonnegligible dependencies between portions of the process that are far apart.

Just as the correlation properties of an autoregressive process are determined by a single parameter ϕ, those of a fractionally differenced process are also characterized by a single parameter, denoted as δ; i.e., the autoregressive and fractionally differenced models are equally simple in the sense that both involve the same number of parameters. The long-range parameter δ is of interest to us in that it determines the rate of decay of variance with averaging length:

$$\sigma_L^2 \approx \sigma_1^2 \times \frac{\Gamma(1-\delta)}{(2\delta+1)\Gamma(1+\delta)} \times L^{-1+2\delta} \text{ as } L \to \infty \tag{4}$$

[*Percival*, 1985]. In the limiting case $\delta = 0$, the process is the same as a white noise process, so the rate of decay reverts to the more familiar L^{-1} given in equations (1) to (3). If $0 < \delta < 1/2$, then the fractionally differenced process is stationary and exhibits long-range dependence. In the extreme case $\delta = 1/2$, the process becomes nonstationary and is very similar to so-called $1/f$ or flicker noise [*Solo*, 1992]. Figure 6 shows four realizations of stationary fractionally differenced processes with similar short range fluctuations but differing degrees of long-range dependence, along with an example of white noise ($\delta = 0$). Note that, although the eye can perceive differences among the realizations in this figure, one would be hard pressed to judge how strong is the long-range dependence in any particular realization, that is, to assign even a rough value of δ to a wiggly curve. Visually, long-range dependence is a fairly subtle property.

Let us now see how the fractionally differenced model corrects the problems that the independence and autoregressive models experienced in representing the decrease in variance of the sample mean and in matching the observed autocorrelation sequence and spectrum.

Using a maximum likelihood procedure, we estimate the parameter δ for the SCICEX 97 profile as $\hat{\delta} = 0.27$ (see Appendix B for details). Using this estimate and the observed sample variance $\hat{\sigma}_1^2$ for the one kilometer averages, we can compute the expected pattern of $\hat{\sigma}_L^2$ versus L under the fractionally differenced model for both the rosette profiles and the SCICEX 97 profile. These are shown in Figures 2a and b as the thick curves, which are a better overall match to the observed decrease (circles) than either the independence or autoregressive models. Similarly, the thick curve in Figure 4b shows the expected value of the sample autocorrelation sequence for the SCICEX 97 profile under the fractionally differenced model, along with upper and lower 95% confidence limits (thin curves). In contrast to the autoregressive model, we now see that the sample autocorrelation sequence (circles) is consistently within the confidence limits for the fitted model (only the value at lag $d = 76$ km falls outside the limits).

A comparison of the thick curves in Figures 4a and b shows the main qualitative difference between the autocorrelation sequences for the autoregressive and fractionally differenced models. The autoregressive model predicts essentially zero correlations for observations that are about $d \geq 7$ km apart, whereas the fractionally differenced model postulates small—but persistent and slowly decaying—correlations beyond 7 km. Visually the expected pattern of the autocorrelation sequence under the fractionally differenced model is a much better fit than that of the autoregressive model. Additionally, the thick curve in Figure 5 shows the spectrum for the SCICEX 97 profile under the fractionally differenced model and is clearly a better overall match to the observed spectrum than the dotted curve for the autoregressive model.

Our conclusion from this study is that the fractionally differenced model adequately explains the variance of the sample mean of sea-ice thickness over one kilometer and greater averages and is to be preferred over a model assuming independence or an autoregressive formulation.

4. Practical Implications

We now have a method based upon Equation (4) of determining the variance for a sample from a profile averaged over length L:

$$\sigma_L^2 \approx \sigma_1^2 \times C_\delta \times L^{-1+2\delta}, \tag{5}$$

where our best estimate of δ for the SCICEX 97 profile is 0.27, σ_1^2 is the variance at a sample length of 1 km and $C_\delta = 0.9$ when $\delta = 0.27$. This equation requires that we know σ_1^2. Alternatively, if we know $\sigma_{L_0}^2$ for some $L_0 > 1$ instead, we can express the above as

$$\sigma_L^2 \approx \sigma_{L_0}^2 \times \left(\frac{L}{L_0}\right)^{-1+2\delta}, \tag{6}$$

which is a good approximation if the averaging lengths L and L_0 are both at least as long as 5 km. In an earlier exploration of the statistical sampling problem, *Wadhams* [1997] used $L_0 = 50$ km and presumed a white noise model, which corresponds to setting $\delta = 0$ rather than $\delta = 0.27$ as we advocate. He also supposed that σ_{50} would be proportional to mean draft $\overline{H}_{50}$ and, based on results from near the North Pole with fairly high values for $\overline{H}_{50}$ in the range 3.671 to 4.008 m (from his Table 3), suggested that $\sigma_{50} \approx 0.1275 \times \overline{H}_{50}$. This approach allows indirect determination of σ_{50}^2 via $\overline{H}_{50}^2$ for use in (6).

To explore the Wadhams approach, we used draft data from NSIDC to examine the relationship between σ_{50} and $\overline{H}_{50}$. We grouped together ten nearby samples with $L_0 \approx$ 50 km. The ten samples in each group are from the same cruise and roughly from the

same 500-km-wide patch of ice. (The boxes containing each group of ten have a mean hypotenuse of 477 km and a maximum of 688 km.) There were 140 such groups. The standard deviation for each group is plotted against its corresponding mean in Figure 7, where σ_{50} appears to have a slight dependence on $\overline{H}_{50}$. We used ordinary least squares to fit the models

$$\sigma_{50} = a \text{ (a constant model)}, \tag{7}$$

$$\sigma_{50} = b\overline{H}_{50} \text{ (the Wadhams model) and} \tag{8}$$

$$\sigma_{50} = a + b\overline{H}_{50} \text{ (a linear model).} \tag{9}$$

For the constant model we obtained an estimate of $\hat{a} = 0.403 \pm 0.023$. For the Wadhams model, we found $\hat{b} = 0.126 \pm 0.007$, which is virtually identical to the result reported by Wadhams even though our range of $\overline{H}_{50}$ is much broader. For the linear model (equation (9)), we obtained $\hat{a} = 0.206 \pm 0.073$ and $\hat{b} = 0.064 \pm 0.022$, so both parameters would be deemed significantly different from zero at a level of significance of 0.05. Because the observed standard deviations σ_{50} have a markedly non-Gaussian distribution (see Figure 7b), we also computed a robust MM–estimate for the linear model [*Yohai et al.*, 1991; *Venables and Ripley*, 2002]. This yielded estimates of $\hat{a} = 0.276 \pm 0.049$ and $\hat{b} = 0.022 \pm 0.015$, which calls into question the significance of the slope—but not the intercept—term. The four fitted models are shown in Figure 7a, where the dashed, dotted, thin solid and thick solid lines are for the constant model, Wadhams model, and the linear model fit by ordinary and robust least squares. For all four fitted models, the residual standard deviations are virtually the same (0.266, 0.266, 0.260 and 0.270), indicating that none of the models with a slope term is markedly better at predicting σ_{50} than just using the mean of the 140 observed values of σ_{50}.

Under the simplifying assumption that σ_{50} does not depend on $\overline{H}_{50}$ and is represented by its mean value of 0.4 (the dashed lines in both plots of Figure 7), the top row of Table 1 lists the standard deviation σ_L of mean ice thickness for sampling length L ranging from 5 to 500 km. For this table we assume a value for $\sigma_{50} = 0.40$ m estimated from the NSIDC draft data, and use the value $\delta = 0.27$ from just the SCICEX 1997 profile shown in Figure 3. For comparison the middle and bottom rows show standard deviations based upon the white noise and autoregressive models. While the latter two models both predict a decrease in σ_L of approximately an order of magnitude as L goes from 5 km up to 500 km, the fractionally differenced model predicts a substantially smaller decrease.

Table 2 shows standard deviations of mean ice thickness for data collected in an ideal rosette pattern with n profiles, again under the fractionally differenced model. The case $n = 1$ is a degenerate rosette consisting of a single draft profile of length L, so the top rows of Tables 1 and 2 are identical. When $n > 1$, the angle θ between adjacent spokes in the rosette is $180^\circ/n$. The case $n = 10$ corresponds to an idealized version of the rosette shown in Figure 1. A rosette of 10 profiles, each of length L, has approximately the same number of measurements as collected along a single profile of length $10L$. A comparison of the numbers in bold font in the top and bottom rows in Table 2 indicates that there is a greater reduction in standard deviation by sampling along a single profile ($n = 1$) of length $10L$ rather than along ten profiles ($n = 10$) of length L in a rosette. Additionally the table indicates that there is virtually no decrease in standard deviation when going from $n = 4$ to $n = 10$ legs, even though the latter involves about 2.5 times more samples. This result is in stark contrast to what is predicted by both the white noise and autoregressive models.

5. Conclusions

This paper addresses the question: how accurately can one estimate the mean draft from a finite sample provided by a profile of ice draft measurements of length L? Or, put differently, what is the sampling error in ice draft observations? The assumption has been that the underlying random process producing draft profiles is the same over any length L, ignoring the fact that the polar oceans are of finite size and, at very large lengths, are subject to regional variations that are incompatible with the notion of stationarity. Ice draft is not well represented by a white noise process or by an autoregressive process, for which the variance of the sampling error falls off as L^{-1} for large L. It is better represented as a process with long-range dependence, for which the variance of the sampling error falls off as $L^{-1+2\delta}$ (Figure 2). A stochastic process with long-range dependence that serves as a useful model of ice draft is a fractionally differenced process. From observations acquired by U.S. Navy submarines (Figure 3), the long-range parameter δ is estimated to be 0.27. Symptoms of this long-range dependence are that the autocorrelation sequence falls slowly to zero as the lag increases (Figure 4), and that the spectrum never levels off as the wavenumber gets small (Figure 5).

Relying on analysis of all available ice draft observations from U.S. submarines, we advocate using the approximation that σ_{L_0} has a value of about 0.40 m for $L_0 = 50$ km, independent of the mean draft (Figure 7). With these two parameters, δ and σ_{L_0}, one can construct tables that give the values of the standard deviation of the sampling error for various sample lengths L in a single straight-line profile (Table 1) and for various combinations of sample length L and number n of rosette legs (Table 2). These tables illustrate how difficult it is to 'beat down' sampling errors by taking larger and larger

samples. Quadrupling the amount of data from $L = 50$ km to $L = 200$ km would reduce the standard deviation of a white noise process by 50%, but, for a fractionally differenced process that mimics sea ice, it is reduced from 0.40 only to 0.29 m (Table 1). For a rosette it is difficult to reduce the sampling error below 0.20 m no matter how much more data one acquires. These sampling errors must be kept in mind when using data to draw conclusions about the temporal or spatial variations in ice draft [e.g., *Wadhams and Horne*, 1980; *McLaren et al.*, 1994; *Wadhams*, 1990; *Rothrock et al.*, 1999; *Tucker et al.*, 2001]

We need to point out three caveats about our results. First, we have focused on a single SCICEX draft profile, for which we estimated the long-range parameter to be $\hat{\delta} = 0.27$. A 95% confidence interval for the assumed true δ is $[0.21, 0.33]$. Analyses of other SCICEX draft profiles give estimates $\hat{\delta}$ that are compatible with the one presented here, but with some exceptions. There does not seem to be any obvious pattern to the exceptions, but how much δ varies spatially and temporially is an open question. Second, the theory of long-range dependence has been most thoroughly worked out under an assumption of Gaussianity, which is reasonable to make here because we are dealing with one kilometer averages. For lesser amounts of averaging, the Gaussianity assumption is not reasonable. A model for the original one meter data would need to take into account non-Gaussianity. Third, the error discussed in this paper is the error in estimating the mean value of a process using a finite observational sample. A problem still to be addressed – but also of great interest – is estimating the error in the 'observation mean' or 'realization mean' as distinct from the 'process mean'.

Appendix A: Three Models for SCICEX Draft Profiles

In sections 2 and 3, we consider three Gaussian models for the detrended SCICEX draft profile of 1 km averages, namely, white noise, a first-order autoregressive process and a fractionally differenced process. Here we describe these three models in more detail using a unified formalism, which allows us to see how a fractionally differenced process compares qualitatively to the other two processes. In particular we formulate the autocorrelation sequences and spectra for all three processes and review formulae for evaluating how the variance of the sample mean depends upon the length L of the average.

Let $\overline{H}_{1,n}, n = 0, \ldots, N-1$, represent the SCICEX profile, where here $N = 803$. Let μ be a real-valued constant, and let $\{\epsilon_n\}$ be a Gaussian white noise process with mean zero and variance σ_ϵ^2, where n ranges over all integers. All three models for $\overline{H}_{1,n}$ can be defined in terms of μ and ϵ_n. The first model simply states that $\overline{H}_{1,n} = \mu + \epsilon_n$; the autoregressive model dictates that

$$\overline{H}_{1,n} = \mu + \phi(\overline{H}_{1,n-1} - \mu) + \epsilon_n = \mu + \sum_{j=0}^{\infty} \phi^j \epsilon_{n-j}, \tag{10}$$

where $|\phi| < 1$; and the fractionally differenced model says that

$$\overline{H}_{1,n} = \mu + \sum_{j=0}^{\infty} \frac{\Gamma(j+\delta)}{\Gamma(j+1)\Gamma(\delta)} \epsilon_{n-j}, \text{ where } \frac{\Gamma(j+\delta)}{\Gamma(j+1)\Gamma(\delta)} \approx \frac{j^{\delta-1}}{\Gamma(\delta)} \text{ for large } j, \tag{11}$$

and $|\delta| < 1/2$ [*Granger and Joyeux*, 1980; *Hosking*, 1981; *Beran*, 1994]. The process mean for each model is μ, and the process variances σ_1^2 are, respectively, σ_ϵ^2, $\sigma_\epsilon^2/(1-\phi^2)$, and $\sigma_\epsilon^2\,\Gamma(1-2\delta)/\Gamma^2(1-\delta)$.

All three models can be reexpressed as

$$\overline{H}_{1,n} = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{n-j}, \tag{12}$$

with a suitable definition for the weights ψ_j, which are nonnegative and satisfy $1 = \psi_0 \geq \psi_1 \geq \psi_2 \geq \cdots$. For any n and $j > 0$, the white noise deviate ϵ_n has some influence on $\overline{H}_{1,n+j}$ as long as ψ_j is positive. The weight ψ_j is thus a measure of how much influence ϵ_n has on $\overline{H}_{1,n+j}$. If ψ_j damps down to zero rapidly as j increases, the influence of ϵ_n on $\overline{H}_{1,n+j}$ also decreases rapidly. This is certainly the case for white noise, for which $\psi_j = 0$ for all $j > 0$. The ratio ψ_j/ψ_{j-1} of adjacent weights for an autoregressive process is ϕ, which implies a rapid decrease in ψ_j when ϕ is set equal to its estimated value ($= 0.36$) for the SCICEX 97 profile (see Appendix B). This is demonstrated in Figure 8, where the thin curve shows ψ_j versus j. Since ψ_j is already below 0.01 when $j = 5$, the influence of ϵ_n on $\overline{H}_{1,n+5}$ and beyond is quite limited. By contrast, the ratio of adjacent weights for a fractionally differenced process is $(j+\delta-1)/j$, which becomes closer and close to unity as j gets large. The thick curve in Figure 8 shows ψ_j versus j with δ set to its estimated value ($= 0.27$) for the SCICEX 97 profile. The ψ_j weights decay at a much slower rate for the fractionally differenced process and first dip below 0.01 when $j = 103$. The influence of ϵ_n thus extends much further than in the autoregressive case, which is why a fractionally differenced process is said to have long-range dependence. A comparison of the weights for large j is another way of seeing the difference between the models. In the autoregressive case, ψ_j decays at an exponential rate ϕ^j, whereas it decays as a power law $j^{\delta-1}$ for a fractionally differenced process.

The autocorrelation sequence $\rho_d \equiv E\{(\overline{H}_{1,n} - \mu)(\overline{H}_{1,n+d} - \mu)\}/\sigma_1^2$ for a white noise process is $\rho_0 = 1$ and $\rho_d = 0$ for any nonzero integer d. The corresponding sequences for the autoregressive and fractionally differenced processes are given by

$$\rho_d = \phi^{|d|} \text{ and } \rho_d = \frac{\Gamma(|d|+\delta)\Gamma(1-\delta)}{\Gamma(|d|+1-\delta)\Gamma(\delta)} \approx \frac{\Gamma(1-\delta)}{\Gamma(\delta)}|d|^{2\delta-1}, \tag{13}$$

where the approximation is valid for large $|d|$. For $d \geq 1$, these sequences satisfy the respective recursions $\rho_d = \phi\rho_{d-1}$ and $\rho_d = \frac{d+\delta-1}{d-\delta}\rho_{d-1}$. Note that ρ_d/ρ_{d-1} is constant for the autoregressive process, whereas, for the fractionally differenced process, this ratio becomes closer and closer to unity as d increases. These recursions indicate that the rate of decay of the autocorrelation sequence is much slower for the fractionally differenced process. Whereas ρ_d decays at an exponential rate $\phi^{|d|}$ for an autoregressive process, it decays as a power law $|d|^{2\delta-1}$ for a fractionally differenced process, which is a primary characteristic of long-range dependence.

The spectrum for a white noise process is given by $S(f) = \sigma_\epsilon^2$, $|f| \leq 1/2$, whereas those for the autoregressive and fractionally differenced processes are given by

$$S(f) = \frac{\sigma_\epsilon^2}{1 - 2\phi\cos(2\pi f) + \phi^2} \text{ and } S(f) = \frac{\sigma_\epsilon^2}{[4\sin^2(\pi f)]^\delta}. \tag{14}$$

Examples of these latter two spectra are shown by the dotted and thick curves in Figure 5, again using values of ϕ and δ appropriate for the SCICEX 1997 profile. Long-range dependence is indicated in the spectrum by the fact that $S(f)$ increases unboundedly as the wavenumber f decreases to zero, an indication of the dominating importance of low wavenumbers in the construction of the fractionally differenced process. By contrast, the spectrum for the autoregressive process flattens out at low wavenumbers (large lags).

Given $\overline{H}_{1,n}, n = 0, \ldots, N-1$, from one of the three processes, we are interested in the statistical properties of length L averages of the form

$$\overline{H}_{L,m} = \frac{1}{L}\sum_{l=0}^{L-1} \overline{H}_{1,mL+l}, \quad m = 0, 1, \ldots, M-1, \tag{15}$$

where $M = \lfloor N/L \rfloor$ (i.e., the largest integer less than or equal to N/L). By stationarity, the variance of $\overline{H}_{L,m}$ is independent of m and is given by

$$\sigma_L^2 = \frac{\sigma_1^2}{L} \sum_{d=-(L-1)}^{L-1} \left(1 - \frac{|d|}{L}\right) \rho_d. \tag{16}$$

[*Brockwell and Davis*, 1991, p. 219; *Fuller*, 1996, p. 310]. If H_l is a white noise or autoregressive process, the above reduces to equations (1) or (2). If H_l is a fractionally differenced process, we can compute σ_L^2 exactly by substituting ρ_d from (13) into the above or approximately using equation (4). A natural estimator for σ_L^2 is the sample estimator

$$\hat{\sigma}_L^2 = \frac{1}{M} \sum_{m=0}^{M-1} \left(\overline{H}_{L,m} - \overline{H}_{LM,0}\right)^2, \text{ where } \overline{H}_{LM,0} = \frac{1}{M} \sum_{m=0}^{M-1} \overline{H}_{L,m}. \tag{17}$$

Because

$$E\{\hat{\sigma}_L^2\} = \sigma_L^2 - \sigma_{LM}^2, \tag{18}$$

where σ_{LM}^2 is defined as in equation (16) with L replaced by the product LM, it follows that $\hat{\sigma}_L^2$ is a biased estimator of σ_L^2 and that the bias is always toward zero [*David*, 1985]. In Figure 2b, we have plotted $E\{\hat{\sigma}_L^2\}$ versus L for the three models under consideration, again using values for ϕ and δ appropriate for the SCICEX profile. Here the difference between $E\{\hat{\sigma}_L^2\}$ and σ_L^2 is very small for the white noise and autoregressive processes because σ_{LM}^2 is negligible; however, this term is nonnegligible for the fractionally differenced process, which is why it is important to compare the observed values in the figure to the expected value of the sample variance $E\{\hat{\sigma}_L^2\}$ rather than the theoretical process variance σ_L^2 (or the large L approximation to it in equation (4)).

Appendix B: Data Preparation and Estimation Procedures

Data from the SCICEX cruises were recorded at a 1 m resolution, but there were a significant number of gaps in the recorded data. Here we explain how the gaps were filled

by means of a stochastic simulation scheme so that these gap-filled data could be used to compute a sample autocorrelation sequence and a sample spectrum.

We used nominal 1 km averages as our starting point for analyzing the SCICEX 97 profile. We based the nth such average on all available 1 m measurements taken between n and $n+1$ km from the beginning of the profile, where n ranges between 0 and 802 (the total length of the profile was 803 km). If there were fewer than 500 data points available within a given 1 km window, we declared the corresponding 1 km average to be missing. The resulting gappy profile of 632 averages has a notable linear trend, which was removed by subtracting off a line whose slope and intercept were estimated using ordinary least squares. We take this detrending operation to yield a set of gappy 1 km averages that can be regarded as a realization of $\overline{H}_{1,n_j}, j = 0, \ldots, 631$, where $\overline{H}_{1,n}$ is as described in Appendix A with $\mu = 0$ (here n_j is the index of the jth available 1 km average; thus $n_0 = 0$ and $n_{631} = 802$). In effect we are treating detrending as a preprocessing step that yields a realization of a zero mean process. A refined, but considerably more complicated, approach would take into account this preprocessing, but would not substantively alter the conclusions we have drawn using the simpler approach.

Fitting the white noise model to the detrended gappy profile $\overline{H}_{1,n_j}$ requires only the estimation of the process variance via

$$\hat{\sigma}_1^2 = \frac{1}{632} \sum_{j=0}^{631} \overline{H}_{1,n_j}^2. \tag{19}$$

To fit the autoregressive and fractionally differenced models, we use the exact maximum likelihood method as adapted to handle missing observations [*Jones*, 1980; *Palma and Chan*, 1997]. These yield parameter estimates of $\hat{\phi} = 0.360$ and $\hat{\sigma}_\epsilon^2 = 0.2033$ for the

autoregressive model and $\hat{\phi} = 0.268$ and $\hat{\sigma}_\epsilon^2 = 0.1984$ for the fractionally differenced model.

Both fitted models can be used to parametrically estimate the autocorrelation sequence and the spectrum for the SCICEX 97 profile (Figure 5 shows these parametric spectrum estimates). It is of interest to compare these parametric estimates to the sample autocorrelation sequence and sample spectrum commonly used in time series analysis, but these descriptors require series that are gap free. To form the sample autocorrelation sequence and spectrum displayed in Figures 4 and 5, we used a stochastic interpolation scheme to convert the gappy $\overline{H}_{1,n_j}$ to a gap-free series, as follows. Suppose that $n_j - n_{j-1} > 1$ for a given j; i.e., there are $N_j = n_j - n_{j-1} - 1$ missing observations between $\overline{H}_{1,n_{j-1}}$ and $\overline{H}_{1,n_j}$. Let $\mathbf{H}_a = [\overline{H}_{1,n_{j-1}}, \overline{H}_{1,n_j}]^T$, where $\mathbf{H}_a$ is a two-dimensional column vector, and 'T' denotes the transpose operation. Let $\mathbf{H}_b = [\overline{H}_{1,n_{j-1}+1}, \ldots, \overline{H}_{1,n_j-1}]^T$ represent the missing profile values. Let $s_d = \rho_d \sigma_1^2$ represent the autocovariance sequence as specified by either the fitted autoregressive or fractionally differenced model. Under a Gaussian assumption, the distribution of $\mathbf{H}_b$ conditional on $\mathbf{H}_a$ is multivariate normal with a mean vector and covariance matrix given by, respectively, $\mu_{b|a} \equiv \Sigma_{ab}^T \Sigma_{aa}^{-1} \mathbf{H}_a$ and $\Sigma_{b|a} \equiv \Sigma_{bb} - \Sigma_{ab}^T \Sigma_{aa}^{-1} \Sigma_{ab}$, where Σ_{aa} is the 2×2 covariance matrix for $\mathbf{H}_a$ and has diagonal and off-diagonal elements given by, respectively, s_0 and $s_{n_j - n_{j-1}}$; Σ_{aa}^{-1} denotes the inverse of Σ_{aa}; Σ_{bb} is the $N_j \times N_j$ covariance matrix for $\mathbf{H}_b$, and its (l,m)th element is s_{l-m}; and Σ_{ab} is a $2 \times N_j$ matrix containing the covariances between the elements of $\mathbf{H}_a$ and $\mathbf{H}_b$ (the first row of Σ_{ab} has $s_1, s_2, \ldots, s_{N_j}$, while the second row has $s_{N_j}, s_{N_j-1}, \ldots, s_1$) [*Anderson and Moore*, 1979]. Let $\Sigma_{b|a} = LL^T$ represent the Cholesky factorization of the covariance matrix. Given a vector $\mathbf{Z}$ of N_j uncorrelated deviates from a normal distribution with zero mean and unit

variance, then $\mu_{b|a} + L\mathbf{Z}$ is a realization from the conditional distribution of $\mathbf{H}_b$ given $\mathbf{H}_a$ and can be used as a surrogate for the missing values $\mathbf{H}_b$. (This procedure could be refined by defining $\mathbf{H}_a$ to comprise all available data, and $\mathbf{H}_b$ all missing data, but conditioning on two 'boundary' values only is computationally simpler.)

This stochastic interpolation technique can be used to fill in all of the gaps in the SCICEX 97 profile with values that have sampling variations consistent with the rest of the profile. One particular interpolation is shown in Figure 3. Let $\overline{H}_{1,n}$ represent the gap-filled series, and let $\overline{H}_{N,0}$ be its sample average. The sample autocorrelation at lag d is given by

$$\hat{\rho}_d = \frac{\sum_{n=0}^{N-|d|-1}(\overline{H}_{1,n} - \overline{H}_{N,0})(\overline{H}_{1,n+|d|} - \overline{H}_{N,0})}{\sum_{n=0}^{N-1}(\overline{H}_{1,n} - \overline{H}_{N,0})^2} \tag{20}$$

We repeated the above procedure 5000 times using the fractionally differenced model and another 5000 times using the autoregressive model, giving us a total of $10,000$ sample autocorrelation sequence. The sample autocorrelation sequence shown by the circles in both plots of Figure 4 is the average of all $10,000$ individual sequences. While there is a small systematic difference between the average sample autocorrelation sequences generated by the two models individually, the variation in the individual sequences is quite small (generally less than the diameter of the circles), particularly when compared to the uncertainty due to the finite sample size N. To assess this latter source of uncertainty, we generated $25,000$ realizations of length N from both an autoregressive process and a fractionally differenced process (with values of ϕ and δ again dictated by estimates from the SCICEX 97 profile) using an appropriate exact simulation procedure [*Kay*, 1981; *Davies and Harte*, 1987; *Wood and Chan*, 1994; *Dietrich and Newsam*, 1997; *Gneiting*, 2000; *Craigmile*, 2003]. We computed $\hat{\rho}_d$ for each realization and then formed the sample

average of the $\hat{\rho}_d$'s to estimate $E\{\hat{\rho}_d\}$, which are displayed as thick curves in Figure 4. The long-range dependence in the fractionally differenced process shows up in the slow decay of $E\{\hat{\rho}_d\}$ toward zero as d increases, in contrast to the rapid decay exhibited by the autoregressive process. While $E\{\hat{\rho}_d\} \approx \rho_d$ in the autoregressive case, the same is not true for the fractionally differenced process, for which $\hat{\rho}_d$ is biased toward zero [*Newbold and Agiakloglou*, 1993]. We also estimated the distribution of $\hat{\rho}_d$ for both processes. The lower and upper thin curves in Figure 4 show the 2.5% and 97.5% percentage points from the estimated distributions.

In a similar manner we can use the stochastic interpolations in order to estimate the spectrum of the SCICEX 97 profile. The simplest estimator of the spectrum is the periodogram, which unfortunately has high variability. To reduce the variability somewhat, we used a multitaper estimator based upon five orthogonal sine tapers [*Thomson*, 1982; *Percival and Walden*, 1993; *Riedel and Sidorenko*, 1995]. The sample spectrum that is shown in Figure 5 is the average of the multitaper spectra formed for the $10,000$ individual stochastic interpolations. The center of the small circle in the figure indicates the value of the sample spectrum at the wavenumber $f = 0.003$ cycles/km (corresponding to a period of $1/f = 341$ km). A 95% confidence interval for the true spectrum at f is indicated by the vertical line bisecting the circle and is based upon a chi-square distribution with ten degrees of freedom. Because the spectrum is plotted on a log scale, confidence intervals at other frequencies have an identical apparent height and can be formed by mentally moving the circle and its associated vertical line so that its center captures a given value of the sample spectrum. Again the additional variability due to uncertainty in the gaps is small compared to uncertainty represented by the 95% confidence intervals.

Appendix C: Treatment of Rosette Data

The rosette data from October 1996 were recorded along ten co-located profiles (see Figure 1). While each individual profile can be treated in a manner similar to that of the SCICEX 1997 profile, their co-location must be taken into account if we want to understand how variability across profiles is influenced by averaging within profiles. Here we describe how this can be done.

As in case of the SCICEX 1997 profile, we formed 1 km averages $\overline{H}_{1,n}$ based upon all available 1 m data along each of the ten rosette profiles. The profiles have a mean length is 181.4 km, which is considerably shorter than the SCICEX 1997 profile (803 km). Because of this, there is no need to detrend via a linear regression as in the case of the SCICEX 1997 profile; however, the sample mean of each profile was used to center the profile prior to using the same maximum likelihood methods described in Appendix B to compute estimates of ϕ and δ. The individual estimates for the ten rosette profiles were systematically smaller than the ones for the SCICEX 1997 profile, possibly attributable to biases due to the small profile lengths. Pending additional research on how to form estimates of ϕ and δ using all ten profiles jointly, for simplicity we have just adopted the estimates for the SCICEX 1997 profile in what follows.

The sampling pattern for the rosette profiles requires use of a spatial process if we want to model how sampling variability is affected by different averaging schemes. Accordingly, we now denote the 1 km averages by $\overline{H}_{1,\mathbf{x}_n}$, where $\mathbf{x}_n$ is a two-dimensional vector indicating the location of the center of all the 1 m measurements involved in forming the average. Here n ranges from 0 to 1813 since the total number of 1 km averages in the ten profiles is $N = 1814$. We assume that $\overline{H}_{1,\mathbf{x}_n}$ can be modeled as a spatially stationary and isotropic

random process with variance σ_1^2, which requires that

$$\operatorname{cov}\{\overline{H}_{1,\mathbf{x}_n}, \overline{H}_{1,\mathbf{x}_n+\mathbf{d}}\} \equiv \sigma_1^2 \rho_{|\mathbf{d}|} \tag{21}$$

depends on just the magnitude $|\mathbf{d}| = \sqrt{(d_0^2+d_1^2)}$ of the two-dimensional vector $\mathbf{d} = [d_0, d_1]^T$ and not on $\mathbf{x}_n$ [*Chilès and Delfiner*, 1999, Chapter 2]. Note that $|\mathbf{d}|$ can assume any nonnegative real value, whereas d in equation (13) is integer valued. For simplicity (and in keeping with the isotropy assumption), we are ignoring the directionality in our basic data; i.e., each $\overline{H}_{1,\mathbf{x}_n}$ is formed using 1 m samples along a line in a certain direction.

We entertain three models for $\overline{H}_{1,\mathbf{x}_n}$ that are obvious analogs to those for the SCICEX 1997 profile. The first model assumes uncorrelatedness between distinct $\overline{H}_{1,\mathbf{x}_n}$'s:

$$\rho_{|\mathbf{d}|} = \begin{cases} 1, & \text{if } |\mathbf{d}| = 0; \text{ and} \\ 0, & \text{for all } |\mathbf{d}| = |\mathbf{x}_n - \mathbf{x}_m|, \end{cases} \tag{22}$$

where $\mathbf{x}_n$ and $\mathbf{x}_m$ are any two distinct locations. The other two models assume either

$$\rho_{|\mathbf{d}|}^{(ar)} = \phi^{|\mathbf{d}|} \text{ or } \rho_{|\mathbf{d}|}^{(fd)} = \frac{\Gamma(|\mathbf{d}|+\delta)\Gamma(1-\delta)}{\Gamma(|\mathbf{d}|+1-\delta)\Gamma(\delta)}, \tag{23}$$

where, in general, $0 < \phi < 1$ and $0 < \delta < \frac{1}{2}$, but we set ϕ and δ to agree with the estimates from the SCICEX 1997 profile ($\hat{\phi} = 0.36$ and $\hat{\delta} = 0.27$). The $\rho_{|\mathbf{d}|}^{(ar)}$ model is such that, if we were to extract a profile by taking equally spaced samples along a line, we would obtain the autoregressive model for the SCICEX 1997 profile; similarly, the $\rho_{|\mathbf{d}|}^{(fd)}$ model yields the fractionally differenced model. While it is well known that $r_{|\mathbf{d}|}^{(ar)}$ is a valid correlation function for a stationary and istropic two-dimensional random process [*Chilès and Delfiner*, 1999, p. 84] the use of $\rho_{|\mathbf{d}|}^{(fd)}$ as such has evidently not been proposed before in the literature. To show that $\rho_{|\mathbf{d}|}^{(fd)}$ is a valid correlation function, note that

$$\rho_{|\mathbf{d}|}^{(fd)} = \frac{\Gamma(1-\delta)}{\Gamma(\delta)\,\Gamma(1-2\delta)} \, \frac{\Gamma(|\mathbf{d}|+\delta)\,\Gamma(1-2\delta)}{\Gamma(|\mathbf{d}|+1-\delta)} \tag{24}$$

$$= \frac{\Gamma(1-\delta)}{\Gamma(\delta)\,\Gamma(1-2\delta)} \int_0^1 u^{|\mathbf{d}|+\delta-1}(1-u)^{-2\delta} du \tag{25}$$

$$= \frac{\Gamma(1-\delta)}{\Gamma(\delta)\,\Gamma(1-2\delta)} \int_0^\infty e^{-s|\mathbf{d}|}\, e^{-s\delta}(1-e^{-s})^{-2\delta} ds \tag{26}$$

by the definition of the Beta integral and a straightforward change of coordinates. Hence $\rho_{|\mathbf{d}|}^{(fd)}$ can be written as a probability mixture of $\rho_{|\mathbf{d}|}^{(ar)} = \phi^{|\mathbf{d}|} = e^{-s|\mathbf{d}|}$, where $s = -\log\phi$ ranges from 0 to ∞. Since probability mixtures of valid correlation functions are correlation functions [*Stein, 1999*, Section 2.3], the proof is complete. (The very same argument shows that $\rho_{|\mathbf{d}|}^{(fd)}$ is a valid correlation function for a stationary and isotropic random process in Euclidean spaces of any dimension.)

Let $\overline{\mathbf{H}}_1$ be a vector whose nth element is $\overline{H}_{1,\mathbf{x}_n}$, and let $\mathbf{1}_N$ be a vector of the same size N, all of whose elements are one. The analog of the estimator in equation (17) for σ_1^2 is $\hat{\sigma}_1^2 = \overline{\mathbf{H}}_1^T\overline{\mathbf{H}}_1/N - (\mathbf{1}_N^T\overline{\mathbf{H}}/N)^2$, and we have

$$E\{\hat{\sigma}_1^2\} = \frac{\mathrm{trace}(\Sigma)}{N} - \frac{\mathbf{1}_N^T\Sigma\mathbf{1}_N}{N^2}, \tag{27}$$

where Σ is an $N \times N$ covariance matrix whose (i,j)th component is $\mathrm{cov}\,\{\overline{H}_{1,\mathbf{x}_i}, \overline{H}_{1,\mathbf{x}_j}\}$. The expression for $E\{\hat{\sigma}_L^2\}$ for general L is analogous to the above, with the elements of Σ being adjusted to take into account the averaging operation. Evaluation of $E\{\hat{\sigma}_L^2\}$ under the three models yields the three curves shown in Figure 2a.

Acknowledgments. This work was generously supported by the National Science Foundation under Grants OPP–9910331, ARC–0453825 and DMS–0134264.

References

Anderson, B. D. O., and J. B. Moore (1979), *Optimal Filtering*, 357 pp., Prentice–Hall, Englewood Cliffs, New Jersey.

Beran, J. (1994), *Statistics for Long-Memory Processes*, 315 pp., Chapman & Hall, New York.

Bloomfield, P. (1992), Trends in global temperature, *Climate Change, 21*(1), 1–16.

Brockwell, P. J., and R. A. Davis (1991), *Time Series: Theory and Method* (Second Edition), 577 pp., Springer, New York.

Chilès, J.–P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, 695 pp., John Wiley & Sons, New York.

Cohn, T. A., and H. F. Lins (2005), Nature's style: Naturally trendy, *Geophys. Res. Lett., 32*(23), L23402.

Craigmile, P. F. (2003), Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes, *J. Time Ser. Anal., 24*(5), 505–511.

David, H. A. (1985), Bias of S^2 under dependence, *Amer. Statist., 39*(3), 201.

Davies, R. B., and D. S. Harte (1987), Tests for Hurst effect, *Biometrika, 74*(1), 95–101.

Dietrich, C. R., and G .N. Newsam (1997), Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix, *SIAM J. Sci. Comput., 18*(4), 1088–1107.

Doukhan, P, G. Oppenheim, and M. S. Taqqu, editors (2003), *Theory and Applications of Long-Range Dependence*, 719 pp., Birkhäuser, Boston.

Fuller, W. A. (1996), *Introduction to Statistical Time Series* (Second Edition), 698 pp., John Wiley & Sons, Inc., New York.

Gneiting, T. (2000), Power-law correlations, related models for long-range dependence, and their simulation, *J. Appl. Probab., 37*(4), 1104–1109.

Granger, C. W. J. (1980), Long memory relationships and the aggregration of dynamic models, *J. Econometrics, 14*(2), 227–238.

Granger, C. W. J., and R. Joyeux (1980), An introduction to long-memory time series models and fractional differencing, *J. Time Ser. Anal., 1*(1), 15–29.

Haslett, J., and A. E. Raftery (1989), Space-time modelling with long-memory dependence: assessing Ireland's wind power resource, *Appl. Stat., 38*(1), 1–50.

Hosking, J. R. M. (1981), Fractional differencing, *Biometrika, 68*(1), 165–176.

Hurst, H. E. (1951), Long-term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers, 116*, 770–799.

Jones, R. H. (1980), Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics, 22*(3), 389–395.

Kay, S. M. (1981), Efficient generation of colored noise, *Proc. IEEE, 69*(4), 480–481.

Mandelbrot, B. (1971), A fast fractional Gaussian noise generator, *Water Resources Research, 7*(3), 543–553.

McLaren, A.S., R.H. Bourke, J.E. Walsh, and R.L. Weaver (1994), Variability in sea-ice thickness over the North Pole from 1958 to 1992, in *Polar Oceans and Their Role in Shaping the Global Environment*, edited by O. M. Johannessen, R. D. Muench, and J. E. Overland, pp. 363–371, American Geophysical Union, Washington, D. C.

Newbold, P., and C. Agiakloglou (1993), Bias in the sample autocorrelations of fractional noise, *Biometrika, 83*(3), 698–702.

Palma, W., and N. H. Chan (1997), Estimation and forecasting of long-memory processes with missing values, *J. Forecast., 16*(6), 395–410.

Percival, D. B. (1985), On the sample mean and variance of a long memory process, Technical Report No. 69, 22 pp., Department of Statistics, University of Washington, Seattle.

Percival, D. B., J. E. Overland, and H. O. Mofjeld (2001), Intepretation of North Pacific variability as a short- and long-memory process, *J. Clim., 14*(24), 4545–4559.

Percival, D. B., and A. T. Walden (1993), *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Technique*, 583 pp., Cambridge Univ. Press, Cambridge, United Kingdom.

Rangarajan, G, and M. Ding, editors (1999), *Processes with Long-Range Dependence: Theory and Applications*, 392 pp., Springer, Berlin.

Riedel, K. S., and A. Sidorenko (1995), Minimum bias multiple taper spectral estimation, *IEEE Trans. Signal Process., 43*(1), 188–195.

Rothrock, D. A., Y. Yu, and G. A. Maykut (1999), Thinning of the Arctic sea-ice cover, *Geophys. Res. Lett., 26*, 3469–3472.

Rothrock, D. A, and M. R. Wensnahan (2006), The accuracy of sea-ice drafts measured from U. S. Navy submarines, submitted to *J. Atmos. and Oceanic Technology*.

Smith, R. L. (1993), Long-range dependence and global warming, in *Statistics for the Environment*, edited by V. Barnett and F. Turkman, pp. 141-161, John Wiley, Hoboken, N. J.

Solo, V. (1992), Intrinsic random functions and the paradox of l/f noise, *SIAM J. Appl. Math., 52*(1), 270–291.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, 247 pp., Springer, New York.

Thomson, D. J. (1982), Spectrum estimation and harmonic analysis, *Proc. IEEE, 70*(9), 1055–1096.

Thorndike, A. S., D. A. Rothrock, G. A. Maykut, and R. Colony (1975), The thickness distribution of sea ice, *Journal of Geophysical Research, 80*(C33), 4501–4513.

Tucker, W. B., J. W. Weatherly, D. T. Eppler, D. Farmer, and D. Bentley (2001), Evidence for the rapid thinning of sea ice in the western Arctic Ocean at the end of the 1980s, *Geophys. Res. Lett., 28*, 2851–2854.

Venables, W. N., and B. D. Ripley (2002), *Modern Applied Statistics with S* (Fourth Edition), 495 pp., Springer, New York.

von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, 484 pp., Cambridge Univ. Press, Cambridge, United Kingdom.

Wadhams, P. (1990), Evidence for thinning of the Arctic ice cover north of Greenland, *Nature, 345*, 795–797.

Wadhams, P. (1997), Ice thickness in the Arctic Ocean: the statistical reliability of experimental data, *J. Geophys. Res., 102*(C13), 27,951–27,959.

Wadhams, P., and R. J. Horne (1980), An analysis of ice profiles obtained by submarine sonar in the Beaufort Sea, *J. Glac., 25*(93), 401–424.

Wood, A. T. A., and G. Chan (1994), Simulation of stationary Gaussian processes in $[0,1]^d$, *J. Comput. Graph. Statist., 3*(4), 409–432.

Yohai, V., W. A. Stahel, and R. H. Zamar (1991), A procedure for robust estimation and inference in linear regression, in *Directions in Robust Statistics and Diagnostics, Part II*, edited by W. A. Stahel and S. W. Weisberg, pp. 365–374, Springer–Verlag, New York.

Figure Captions

Figure 1: Ten rosette profiles from SCICEX October 1996 (short thin curves in right-hand plot) and a single profile from SCICEX 1997 (long thick curve), shown to scale in the Arctic ocean (left) and enlarged (right). The coordinate system is Cartesian overlaid on a Lambert azimuthal equivalent projection; its origin is at the north pole, and the x axis is positive along 35°E.

Figure 2: Plots of $\hat{\sigma}_L^2$ versus L (circles) for (a) rosette draft profiles and (b) the 1997 SCICEX profile. In (a) and (b), the thin lines depict the expected pattern of $\hat{\sigma}_L^2$ versus L under an assumption of independence. The dotted curves give the expected pattern under the assumption of that the data are correlated, with a correlation structure in keeping with a first-order autoregressive process. The thick curves gives the corresponding patterns for a fractionally differenced process, discussed in section 3.

Figure 3: Residual draft profile from SCICEX 1997. As explained in Appendix B, a model-based stochastic interpolation scheme was used to fill in missing values in this residual profile at locations indicated by the dashes near the bottom of the plot.

Figure 4: Sample autocorrelation sequence for residual draft profile from SCICEX 1997 (circles in both plots). Plot (a) shows the mean value (thick curve) and 2.5% and 97.5% percentage points (thin curves) from the distribution of the sample autocorrelation sequence for an autoregressive process. The autocorrelation sequence for this process depends on just the unit lag autocorrelation ϕ, which was estimated from the SCICEX profile via the maximum likelihood method. In (b) the curves are for a fractionally differenced process, with its associated parameter δ estimated again by maximum likelihood.

Figure 5: Sample spectrum for residual draft profile from SCICEX 1997 (thin wiggly line). The theoretical spectra are also shown for the fitted fractionally differenced (thick curve) and

autoregressive (dotted) processes as presented in Appendix A. A 95% confidence interval to the true spectrum based upon the sample spectrum at the lowest displayed wavenumber is also plotted. Whereas the fractionally differenced spectrum falls within this confidence interval, the autoregressive spectrum does not.

Figure 6: Plots of white noise ($\delta = 0$) and four fractionally differenced series with long-range dependence ($\delta = 0.1$, 0.2, 0.3 and 0.4). Each series was formed using an 'exact' simulation technique that transforms 1024 random numbers from a standard normal distribution into a correlated time series of length 512 [*Davies and Harte*, 1987; *Wood and Chan*, 1994; *Dietrich and Newsam*, 1997; *Gneiting*, 2000; *Craigmile*, 2003]. To better illustrate how δ influences a time series, we used the same random numbers to create each series.

Figure 7: Sample standard deviations versus sample means for 140 groups of $\overline{H}_{50}$ (circles in plot (a)). The dotted line is the least squares fit for the Wadhams model $\sigma_{50} = b\overline{H}_{50}$; the horizontal dashed line, for the constant model $\sigma_{50} = a$; the thin solid line, for the linear model $\sigma_{50} = a + b\overline{H}_{50}$; and the thick solid line, for the linear model also, but now fit using a robust least squares procedure. Plot (b) shows an estimated probability density function for the sample standard deviations, with their sample mean indicated by the dashed line (this is the same as the least squares fit to a in the constant model).

Figure 8: Weights ψ_j used in Appendix A to create autoregressive (thin curve) and fractionally differenced (thick) processes from a weighted average of white noise. The weights for the autoregressive process are below 0.01 for $j \geq 5$, while those for the fractionally differenced process are above this threshold for $j \leq 102$.

Table 1. Standard Deviation of Mean Sea-Ice Thickness for Various Sampling Lengths[a]

L (km)	5	10	15	20	50	100	150	200	500
σ_L (m), fractionally differenced	0.68	0.58	0.53	0.49	0.40	0.34	0.31	0.29	0.24
σ_L (m), white noise	1.26	0.89	0.73	0.63	0.40	0.28	0.23	0.20	0.13
σ_L (m), autoregressive	1.17	0.86	0.72	0.62	0.40	0.28	0.23	0.20	0.13

[a] With $L_0 = 50$ km and $\sigma_{L_0} = 0.40$, the values in the top row are based on equation (6) with $\delta = 0.27$; in the middle row, on equation (1); and in the bottom row, on equation (2) with $\phi = 0.36$.

Table 2. Standard Deviation of Mean Sea-Ice Thickness for Various Sampling Lengths and Rosette Patterns under the Fractionally Differenced Model[a]

L (km)	5	10	15	20	50	100	150	200	500
σ_L (m), $n = 1$	0.68	0.58	0.53	0.49	**0.40**	**0.34**	**0.31**	**0.29**	**0.24**
σ_L (m), $n = 2$ ($\theta = 90°$)	0.62	0.53	0.48	0.46	0.37	0.31	0.29	0.27	0.22
$n = 3$ ($\theta = 60°$)	0.60	0.52	0.47	0.44	0.36	0.31	0.28	0.26	0.21
$n = 4$ ($\theta = 45°$)	0.59	0.51	0.46	0.44	0.36	0.30	0.28	0.26	0.21
$n = 5$ ($\theta = 36°$)	0.59	0.51	0.46	0.44	0.35	0.30	0.28	0.26	0.21
$n = 10$ ($\theta = 18°$)	**0.58**	**0.51**	**0.46**	**0.43**	**0.35**	0.30	0.27	0.26	0.21

[a] The first row of tabulated values is repeated from Table 1. Using the same settings as in that table, the remaining rows are given by $(\mathbf{1}_N^T \Sigma \mathbf{1}_N)^{1/2}/N$, where N is the number of distinct measurements in a rosette pattern, and $\mathbf{1}_N$ and Σ are as given in equation (27) of Appendix C.

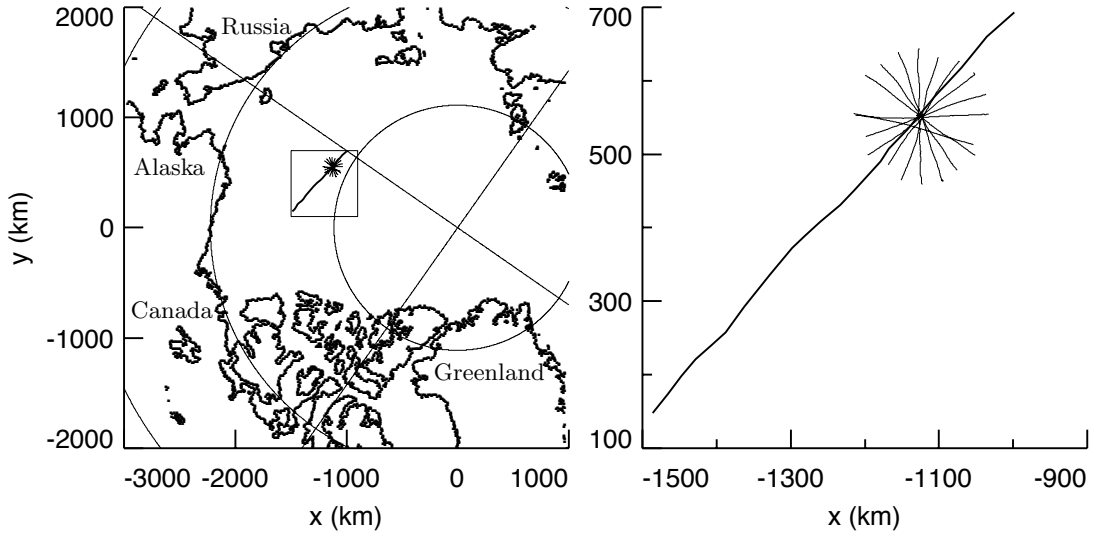

Figure 1. Ten rosette profiles from SCICEX October 1996 (short thin curves in right-hand plot) and a single profile from SCICEX 1997 (long thick curve), shown to scale in the Arctic ocean (left) and enlarged (right). The coordinate system is Cartesian overlaid on a Lambert azimuthal equivalent projection; its origin is at the north pole, and the x axis is positive along 35°E.

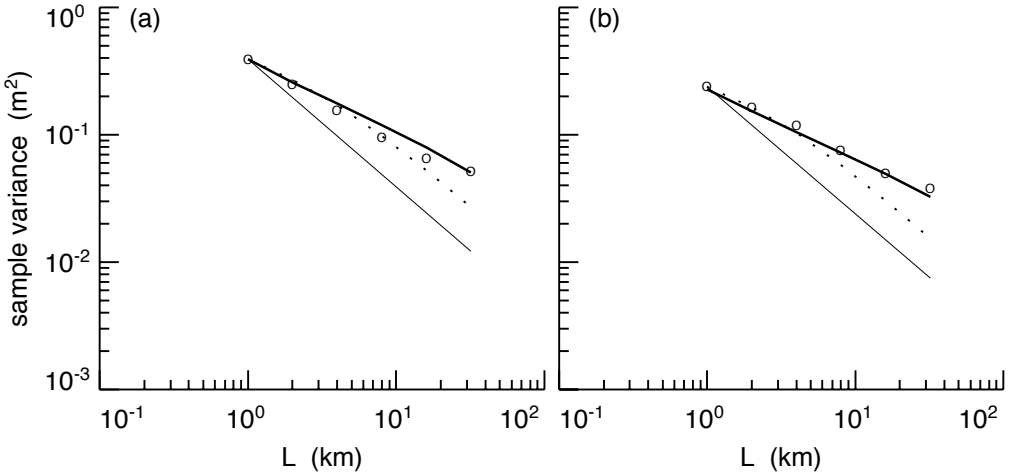

Figure 2. Plots of $\hat{\sigma}_L^2$ versus L (circles) for (a) rosette draft profiles and (b) the 1997 SCICEX profile. In (a) and (b), the thin lines depict the expected pattern of $\hat{\sigma}_L^2$ versus L under an assumption of independence. The dotted curves give the expected pattern under the assumption of that the data are correlated, with a correlation structure in keeping with a first-order autoregressive process. The thick curves gives the corresponding patterns for a fractionally differenced process, discussed in section 3.

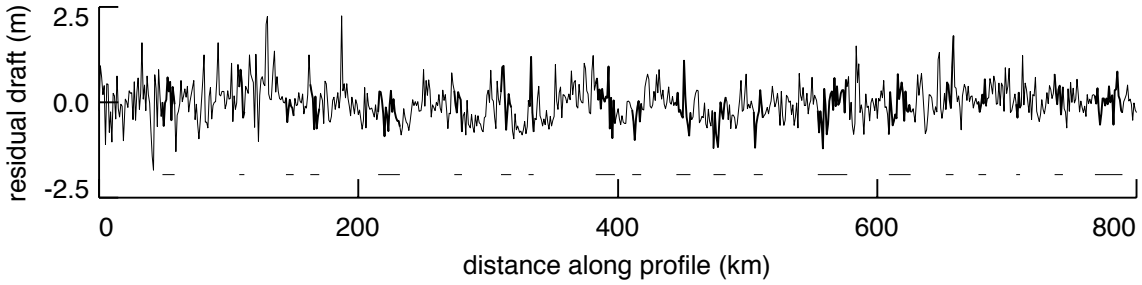

Figure 3. Residual draft profile from SCICEX 1997. As explained in Appendix B, a model-based stochastic interpolation scheme was used to fill in missing values in this residual profile at locations indicated by the dashes near the bottom of the plot.

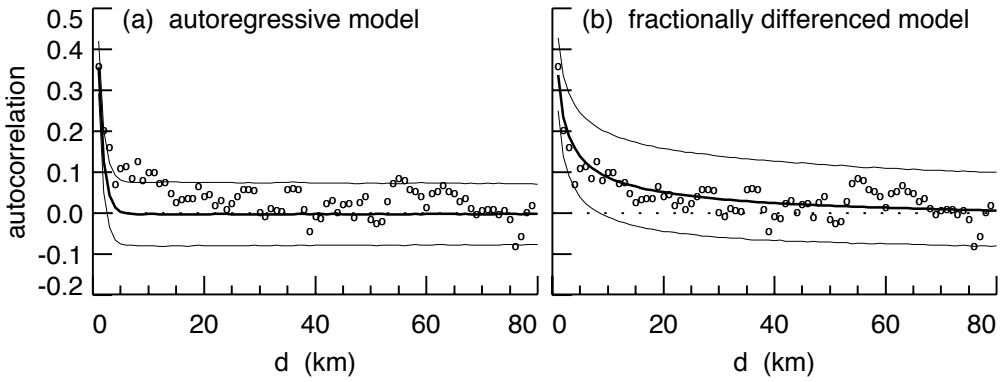

Figure 4. Sample autocorrelation sequence for residual draft profile from SCICEX 1997 (circles in both plots). Plot (a) shows the mean value (thick curve) and 2.5% and 97.5% percentage points (thin curves) from the distribution of the sample autocorrelation sequence for an autoregressive process. The autocorrelation sequence for this process depends on just the unit lag autocorrelation ϕ, which was estimated from the SCICEX profile via the maximum likelihood method. In (b) the curves are for a fractionally differenced process, with its associated parameter δ estimated again by maximum likelihood.

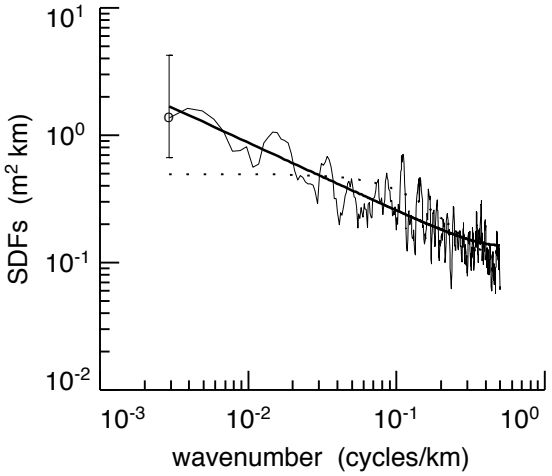

Figure 5. Sample spectrum for residual draft profile from SCICEX 1997 (thin wiggly line). The theoretical spectra are also shown for the fitted fractionally differenced (thick curve) and autoregressive (dotted) processes as presented in Appendix A. A 95% confidence interval to the true spectrum based upon the sample spectrum at the lowest displayed wavenumber is also plotted. Whereas the fractionally differenced spectrum falls within this confidence interval, the autoregressive spectrum does not.

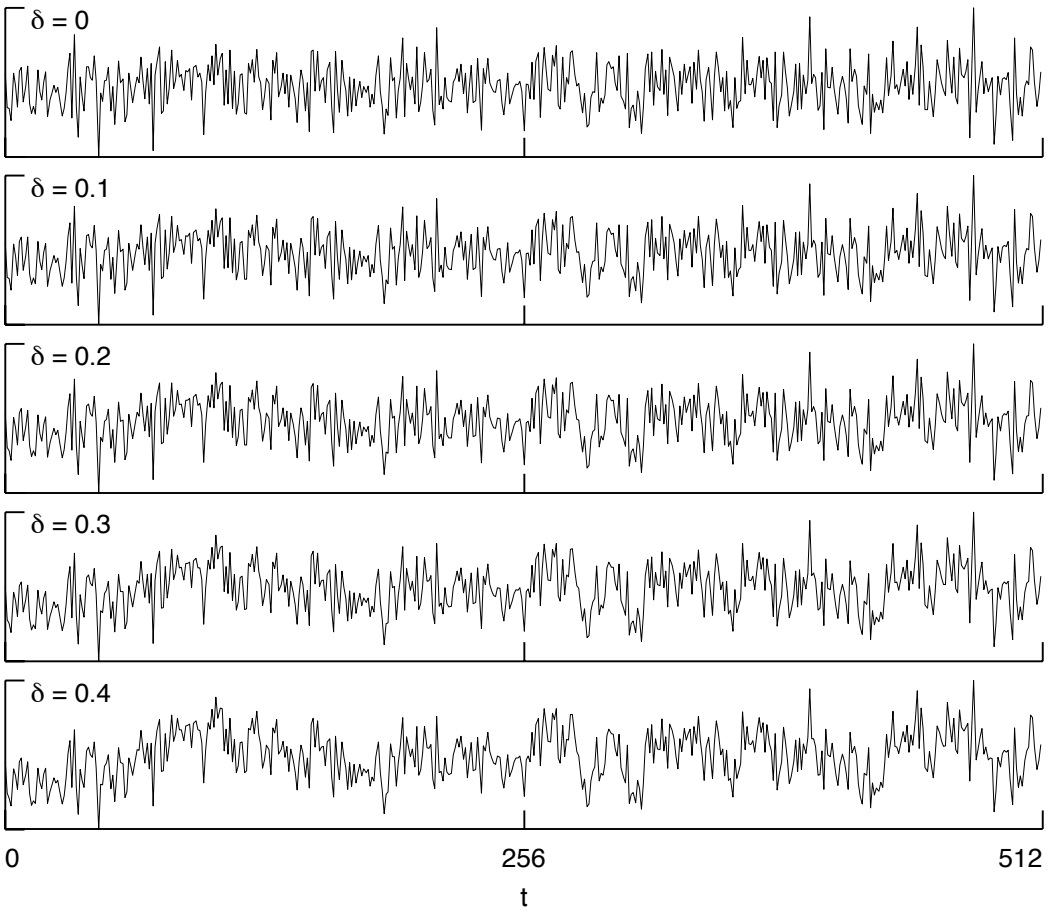

Figure 6. Plots of white noise ($\delta = 0$) and four fractionally differenced series with long-range dependence ($\delta = 0.1$, 0.2, 0.3 and 0.4). Each series was formed using an 'exact' simulation technique that transforms 1024 random numbers from a standard normal distribution into a correlated time series of length 512 [*Davies and Harte*, 1987; *Wood and Chan*, 1994; *Dietrich and Newsam*, 1997; *Gneiting*, 2000; *Craigmile*, 2003]. To better illustrate how δ influences a time series, we used the same random numbers to create each series.

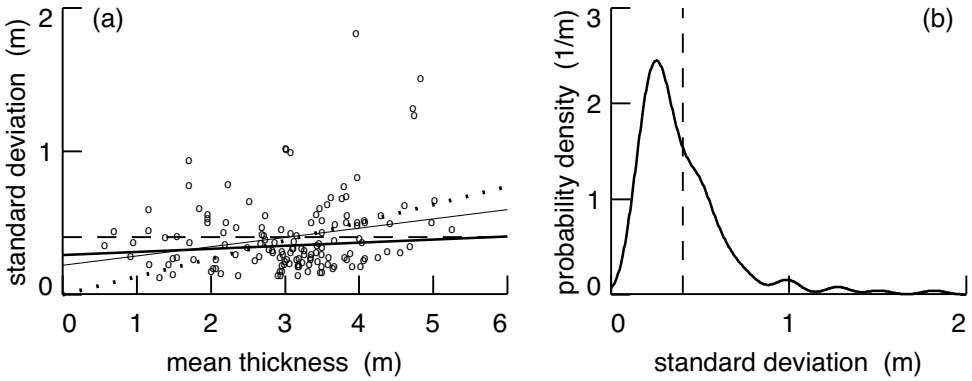

Figure 7. Sample standard deviations versus sample means for 140 groups of $\overline{H}_{50}$ (circles in plot (a)). The dotted line is the least squares fit for the Wadhams model $\sigma_{50} = b\overline{H}_{50}$; the horizontal dashed line, for the constant model $\sigma_{50} = a$; the thin solid line, for the linear model $\sigma_{50} = a + b\overline{H}_{50}$; and the thick solid line, for the linear model also, but now fit using a robust least squares procedure. Plot (b) shows an estimated probability density function for the sample standard deviations, with their sample mean indicated by the dashed line (this is the same as the least squares fit to a in the constant model).

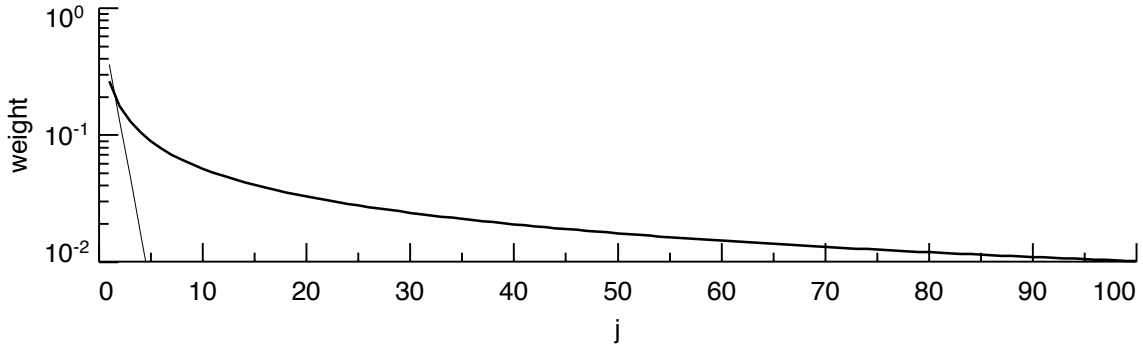

Figure 8. Weights ψ_j used in Appendix A to create autoregressive (thin curve) and fractionally differenced (thick) processes from a weighted average of white noise. The weights for the autoregressive process are below 0.01 for $j \geq 5$, while those for the fractionally differenced process are above this threshold for $j \leq 102$.