Bias of Effective Degrees of Freedom of a Spectrum

By Steve Elgar[1]

Introduction

Statistical variability must be accounted for when estimating ocean parameters from observations. For example, estimates of power spectral levels fluctuate about their true values according to a chi-square distribution. Consequently, confidence limits for the estimated levels can be obtained based on the number of degrees of freedom used for each spectral estimate. Increasing the number of degrees of freedom, by either merging neighboring spectral estimates (i.e., frequency merging) or ensemble averaging the individual estimates from several realizations of the process, or both, leads to increasingly stable estimates. Integral properties of the spectrum (e.g., significant wave height) also fluctuate about their true values, and thus, meaningful data analyses must include the appropriate confidence limits. These confidence limits also depend on the available number of degrees of freedom. However, as shown by Donelan and Pierson (1983), Medina, et al. (1985), and Young (1986), the effective number of degrees of freedom for the total spectrum is not simply the sum of the number of degrees of freedom for each spectral estimate, but is a weighted sum that is spectral shape dependent. A narrow spectrum has fewer effective degrees of freedom than a broad spectrum.

Using the expression developed in Medina, et al. (1985), for effective degrees of freedom (Eq. 1, to be given subsequently), the theoretical number of effective degrees of freedom for integral properties of the spectrum can be calculated given an analytical spectral shape. Donelan and Pierson (1983), and Young (1986) present equations to calculate the effective number of degrees of freedom from data. However, as will be shown, the effective number of degrees of freedom estimated from data is biased by as much as a factor of two. These analytical results are verified by numerical simulations and field data.

Bias of Effective Number of Degrees of Freedom

The effective number of degrees of freedom for a spectrum α is (Medina, et al. 1985):

[1]Asst. Prof., Electrical and Computer Engrg. Dept., Washington State Univ., Pullman, WA 99164-2752.

Note.—Discussion open until June 1, 1987. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on April 30, 1986. This paper is part of the *Journal of Waterway, Port, Coastal and Ocean Engineering*, Vol. 113, No. 1, January, 1987. ©ASCE, ISSN 0733-950X/87/0001-0077/$01.00. Paper No. 21178.

$$\alpha = \frac{2\left[\int_0^\infty S(f)df\right]^2}{\Delta f \int_0^\infty S^2(f)df} \quad (1)$$

where $S(f)$ is the power spectral density at frequency f; and Δf is the frequency resolution of the data (Δf = 1/record length). The term in brackets in the numerator of Eq. 1 is the variance of the record m_0. Thus

$$\alpha = \frac{2m_0^2}{\Delta f \int_0^\infty S^2(f)df} \quad (2)$$

In terms of discrete spectral values, the expression for α becomes (Donelan and Pierson 1983; Young 1986):

$$\alpha = \frac{l\left[\sum_{n=1}^{N} S(f_n)\right]^2}{\sum_{n=1}^{N} S^2(f_n)} \quad (3)$$

where l is the number of degrees of freedom used to estimate each of the N discrete spectral levels $S(f_n)$, where $f_n = n\Delta f$. Thus, if there is no frequency merging, nor any ensemble averaging, the $S(f_n)$ terms are periodogram estimates, and $l = 2$.

The estimated effective number of degrees of freedom $\tilde{\alpha}$ obtained from data (e.g., using measured values of S in Eq. 3) is a biased estimate of the true value of α. This is demonstrated as follows. Let the Fourier representation of the sea surface, $\eta(t)$, be given by

$$\eta(t) = \sum_{n=1}^{N} a_n \cos 2\pi f_n t + b_n \sin 2\pi f_n t \quad (4)$$

where a_n, b_n are independent Gaussian-distributed random variables with zero mean and variance $S(f_n)\Delta f$. Here $S(f_n)$ is the true spectral density at frequency $f_n = n\Delta f$. From a finite length time series only the estimates $\tilde{a}_n$, $\tilde{b}_n$ of the true values of a_n, b_n can be obtained. The random variable

$$\tilde{Z}_n^2 \equiv \frac{(\tilde{a}_n^2 + \tilde{b}_n^2)}{[S(f_n)\Delta f]} \quad (5)$$

is chi-square distributed, with 2 degrees of freedom. From the probability density function of a chi-square-distributed random variable x with ν degrees of freedom, it is readily shown that

$$E[x] = \nu \quad (6a)$$

$$E[x^2] = \nu^2 + 2\nu \quad (6b)$$

where $E[\ ]$ is the expected value, or average, operator. Since the estimated value of $S(f_n)$ is $(\tilde{a}_n^2 + \tilde{b}_n^2)/2\Delta f$, the expected value of $\tilde{\alpha}$ is, using Eq. 2:

$$E[\tilde{\alpha}] = E\left[\frac{2m_0^2}{\Delta f \sum_{n=1}^{N} \frac{(\tilde{a}_n^2 + \tilde{b}_n^2)^2 \Delta f}{4\Delta f^2}}\right] = \frac{2m_0^2}{\sum_{n=1}^{N} E[(\tilde{Z}_n^2)^2] \frac{S^2(f_n)\Delta f^2}{4}}$$

$$= \frac{2m_0^2}{\frac{(\nu^2 + 2\nu)\Delta f}{4} \sum_{n=1}^{N} S^2(f_n)\Delta f} = \frac{4}{\nu^2 + 2\nu}\alpha = \frac{\alpha}{2} \qquad (7)$$

because $\nu = 2$.

If several realizations of the process, j, say, are available for ensemble averaging, or if neighboring frequency components are merged together, ν will increase. In this case, the average $(\bar{T}_n^2)^2$

$$(\bar{T}_n^2)^2 = \frac{1}{j^2} \sum_{k=1}^{j} \frac{(\tilde{a}_{nk}^2 + \tilde{b}_{nk}^2)^2}{[S^2(f_n)\Delta f^2]} \qquad (8)$$

may be used in place of $(\tilde{Z}_n^2)^2$, yielding

$$E[\tilde{\alpha}] = \frac{2m_0^2}{\Delta f \sum_{n=1}^{N} E\left[\frac{1}{j^2} \sum_{k=1}^{j} \frac{(\tilde{a}_{nk}^2 + \tilde{b}_{nk}^2)^2 \Delta f}{4\Delta f^2}\right]} = \frac{2m_0^2}{\frac{(\nu^2 + 2\nu)}{4j^2} \Delta f \sum_{n=1}^{N} S^2(f_n)\Delta f}$$

$$= \frac{4j^2}{\nu^2 + 2\nu}\alpha = \frac{1}{1 + \frac{2}{\nu}}\alpha \qquad (9)$$

because $\nu = 2j$.

Thus, the value of α estimated from a finite length time series is biased. The mean value of $\tilde{\alpha}$ obtained by averaging the individual $\tilde{\alpha}$ from unsmoothed spectra will be one-half the α value of the true spectrum. On the other hand, ensemble averaging or frequency merging the unsmoothed spectra, and calculating α from the ensemble will produce a better estimate of α in accordance with Eq. 9. In fact, Eq. 9 can be used to remove the bias when estimating α from data.

Practical Application

The bias of $\tilde{\alpha}$ was verified for standard JONSWAP spectral forms (Hasselman, et al. 1973) by means of numerical simulations and measured shallow water wave data. Realizations of JONSWAP spectra were produced by numerically generating Gaussian-distributed, zero-mean, unit variance random variables which were multipled by the appropriate JONSWAP spectral level (Andrew & Borgman 1981). This procedure yields the a_n, b_n of Eq. 4. The individual spectra represented wave records of 2,048 points sampled at 2 Hz. For each of four spectral widths, 128 realizations of the JONSWAP spectrum were simulated. Mean values of α were estimated by averaging individual $\tilde{\alpha}$ values obtained from the collection of unsmoothed spectra. The degrees of freedom associated with

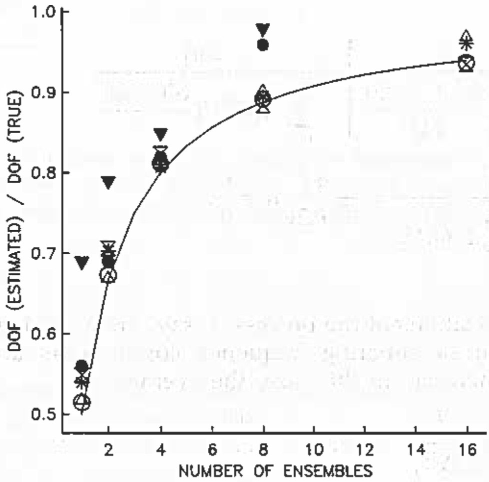

FIG. 1.—Ratio of Estimated Effective Number of Degrees of Freedom to the True Number of Effective Degrees of Freedom versus the Number of Ensembles (*j*) from Which Estimates Are Made

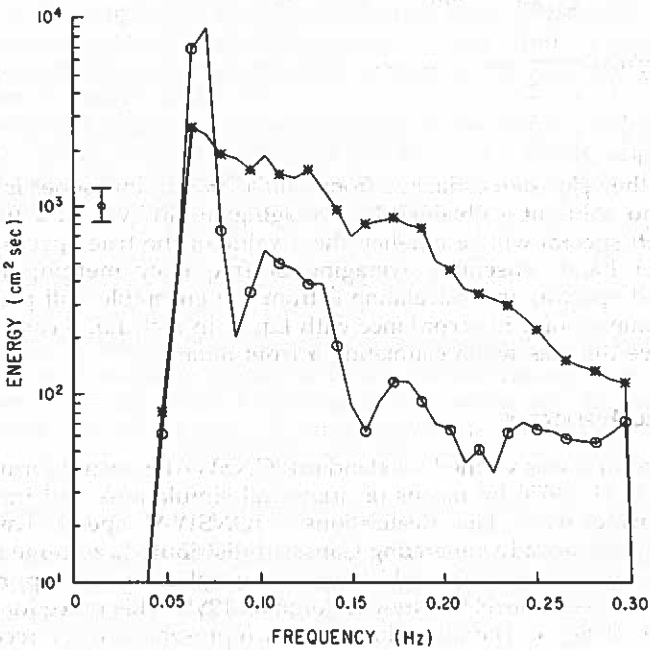

FIG. 2.—Power Spectral Density of Sea Surface Elevation in Water 9 m Deep: Narrow Band (Circle); Broad Band (Asterisk)

each spectral estimate was next increased by ensemble averaging the individual unsmoothed JONSWAP realizations. Thus, the 128 raw spectra were used to produce 128 α estimates with $\nu = 2$, 64 α estimates with $\nu = 4$, 32 α estimates with $\nu = 8$, etc. The α values for each ν were averaged together. The ratio of estimated (i.e., average) α to the true value of α (obtained from analytic JONSWAP spectra) is shown in Fig. 1, along with the theoretical ratio from Eq. 9. In Fig. 1, the open symbols are from numerical simulations of a JONSWAP spectrum with $f_p = 0.2$ Hz, $\sigma_a = 0.07$, $\sigma_b = 0.09$, and γ values of 1.0 (circles), 3.3 (triangles), 4.2 (asterisks), and 7.0 (hourglasses). The closed symbols are the ratios from the spectra shown in Fig. 2; the closed circle is the broad spectrum, the closed triangle is the narrow spectrum.

The estimated effective number of degrees of freedom as a function of the number of ensembles was also calculated for the two shallow water (depth = 9 m) spectra shown in Fig. 2. In this figure, circles represent the narrow-band data, and asterisks the broad-band data. The individual spectral estimates displayed here have 128 degrees of freedom, and the 90% confidence limits are indicated by the bars. For these ocean data the total record lengths of 32,768 points (sampled at 2 Hz) were broken into 32 segments 1,024 points long. As shown in Fig. 1, ensemble averaging the raw spectra reduces the bias in α estimates in close agreement with Eq. 9. Since only relatively few ensembles are available for estimating the mean value of α, the scatter is expected. The laboratory and field data of Donelan and Pierson (1983: table 4) are also consistent with Eq. 9.

Conclusions

The effective number of degrees of freedom for integral properties of a spectrum are given by Eq. 1, and are spectral shape dependent. The mean value of estimates of α from data is biased, as shown by Eq. 9, which is verified by numerical simulations and field data (Fig. 1). Ensemble averaging or frequency band averaging spectra before calculating α reduces the bias. However, oversmoothing the spectrum by frequency merging will artificially increase the effective number of degrees of freedom. Indeed, in the limit of maximum frequency smoothing the spectrum is reduced to a single value, and $\alpha = \nu N$. Thus, spectra that are undersmoothed yield an α which is biased low, while oversmoothed spectra give an α which is biased high.

Acknowledgments

R. T. Guza made useful (and unbiased) suggestions. Support was provided by the College of Engineering, University of Idaho, and the Office of Naval Research.

Appendix.—References

Andrew, M. E., and Borgman, L. E., (1981). "Procedures for studying wave grouping in wave records from California Coastal Data Collection Program." U.S. Army Corps of Engrs., San Francisco, Calif.

Donelan, M., and Pierson, W. J. (1983). "The sampling variability of estimates

of spectra of wind-generated gravity waves." *J. Geophys. Res.*, 88, 4,381–4,392.
Hasselman, K., et al. (1973). "Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP)." *Erg. Deutsch. Hydrograph. Zeitschr.* Reihe A(8), 12, 1–95.
Medina, J. R., Aguilar, J., and Diez, J. J., (1985). "Distortions associated with random sea simulators." *J. Waterway Port Coast. Ocean Engrg.*, ASCE, 111(4), 603–628.
Young, I. R. (1986). "Probability distribution of spectral integrals." *J. Waterway Port Coast. Ocean Engrg.*, ASCE, 112(2), pp. 338–341.