Teaching time-series analysis. II. Wave height and water surface elevation probability distributions

Dennis J. Whitford, Jennifer K. Waters and Mario E. C. Vieira

Citation: American Journal of Physics **69**, 497 (2001); doi: 10.1119/1.1331301
View online: https://doi.org/10.1119/1.1331301
View Table of Contents: https://aapt.scitation.org/toc/ajp/69/4
Published by the American Association of Physics Teachers

Teaching time-series analysis. II. Wave height and water surface elevation probability distributions

Dennis J. Whitford[a)]
Department of Oceanography, U.S. Naval Academy, 572M Holloway Road, Annapolis, Maryland 21402

Jennifer K. Waters
Department of Naval Architecture and Ocean and Marine Engineering, U.S. Naval Academy, 590 Holloway Road, Annapolis, Maryland 21402

Mario E. C. Vieira
Department of Oceanography, U.S. Naval Academy, 572 Holloway Road, Annapolis, Maryland 21402

(Received 18 March 1999; accepted 18 September 2000)

This paper describes the second of a two-part series of pedagogical exercises to introduce students to methods of time-series analysis. While these exercises are focused on the analysis of wind generated surface gravity waves, they are cross-disciplinary in nature and can be applied to other fields dealing with random signal analysis. Two computer laboratory exercises are presented which enable students to understand many of the facets of random signal analysis with less difficulty and more understanding than standard classroom instruction alone. The first pedagogical exercise, described in the previous article, uses mathematical software on which the students execute the manual arithmetic operations of a finite Fourier analysis on a complex wave record. The results are then compared to those obtained by a fast Fourier transform. This article, the second of this two-part pedagogical series, addresses analysis of a complex sea using observed and theoretical wave height and water surface elevation probability distributions and wave spectra. These results are compared to a fast Fourier transform analysis, thus providing a link back to the first exercise.

[DOI: 10.1119/1.1331301]

I. INTRODUCTION

Analysis of water surface waves is a fundamental learning objective for undergraduate and graduate students majoring in ocean engineering and physical oceanography. Students usually find the transition from studying monochromatic waves, i.e., single-frequency waves, to studying a more realistic complex and random sea composed of many different frequencies to be difficult. Typical oceanic waves, commonly referred to as ''random waves,''[1] are often represented as the result of the interaction of multiple single-frequency waves with differing wavelengths, frequencies, wave heights, and directions of propagation. Another common method of representing random waves is by statistical means, in which the statistics and probabilities of water surface elevations encountered by a stationary observer are analyzed. The goal of this article, the second of a two-part series,[2] is to provide a pedagogical exercise which illustrates the statistical and probabilistic descriptions of random waves.

Water surface elevations[3] (η) of random waves have a theoretical probability density function (PDF) which is Gaussian:[4]

$$p(\eta)=\frac{1}{\sigma_\eta\sqrt{2\pi}}\exp\left(\frac{-\eta^2}{2\sigma_\eta^2}\right), \tag{1}$$

where σ_η is the standard deviation of the water surface elevation data. Note that σ_η is linked to the first exercise[2] since the variance (σ_η^2) equals the area under the energy density $[S(f)]$ spectral curve.[5] This relationship will be further addressed in Sec. V.

However, the analysis of water surface elevation distributions and statistics is not the only statistical interest of scientists and engineers dealing with water waves. What is referred to as wave height, as shown in Figs. 1(a) and 1(b) for regular and random waves, respectively, is often of more concern than water elevations since wave heights, or in more general terms, the relative proximate extrema of water surface elevations, may indicate more important characteristics of the wave field than water elevation alone. Wave heights are found by dividing the water elevation time series into individual waves, specifically, a new wave begins when the water level crosses the still water level in a particular direction, either in the upcrossing or downcrossing direction. The maximum water surface elevation minus the minimum water surface elevation between each of these individual wave delineations defines the wave height for that portion of, or wave within, the time series.[6] The time elapsed between upcrossings (or, alternatively, downcrossings) is the individual wave period. Since a random wave signal can have many individual waves defined within its water elevation time series, and therefore many individual associated wave heights and periods, statistics on these measurements can be performed in an effort to analyze and characterize the underlying signal. The distribution of wave heights is very different from wave elevations; and as classically demonstrated by Longuet–Higgins,[7] these wave heights (H) from a random wave signal follow a Rayleigh probability distribution with the resultant PDF:

$$p(H)=2\,\frac{H}{H_{\mathrm{rms}}^2}\exp\left[-\left(\frac{H^2}{H_{\mathrm{rms}}^2}\right)\right], \tag{2}$$

where the root-mean-square wave height (H_{rms}) is defined as:

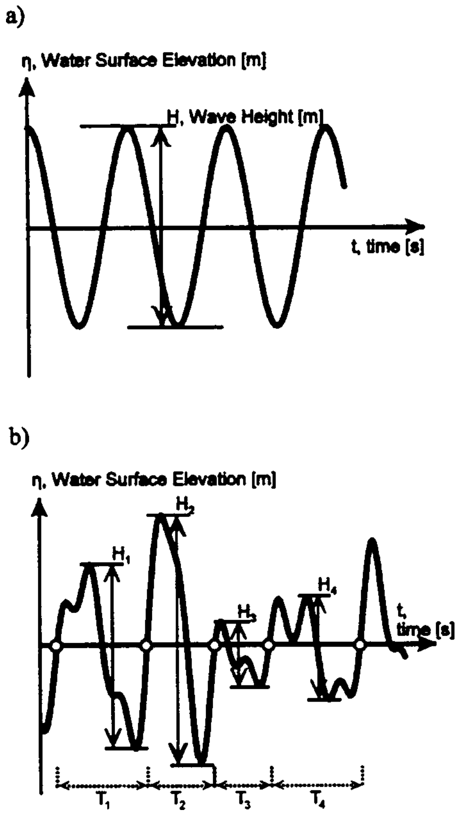

Fig. 1. Wave height definition for (a) regular and (b) irregular or random wave time series. Note that (b) also illustrates the zero-upcrossing wave height analysis method for water surface elevation where T is wave period and H is wave height.

$$H_{\rm rms}=\sqrt{\frac{1}{M}\sum_{j=1}^{M} H_j^2}, \tag{3}$$

where M is the number of waves under consideration, and H_j is the jth wave of the group M. As with most science laboratories, this exercise would follow a homework reading assignment and thorough classroom discussion of this background material.

Students usually find this exercise more difficult and time consuming to complete than the first. However, the benefits gained from this second analysis of the data are worth it. The students tend to learn not just a method of ocean wave analysis, but they also learn and appreciate the process in which many scientific and engineering phenomena are analyzed, and more importantly, the interrelation between different methods of analysis.

II. RAW DATA

Figure 2 is the analog representation of the random water surface elevation signal which the students will analyze. Equivalent exercise data can be taken from a water surface elevation recorder using wave tank laboratory data or from actual field data.[8] We provide the data of Fig. 2 in a digitized form with a duration of 320 s and a sampling period of 0.0625 s. This record was obtained from a wave tank which was programmed to generate an irregular wave signal, simulating a typical random sea.[9]

Any spreadsheet software will suffice for this exercise. The entire exercise uses approximately 2 Mbytes of storage

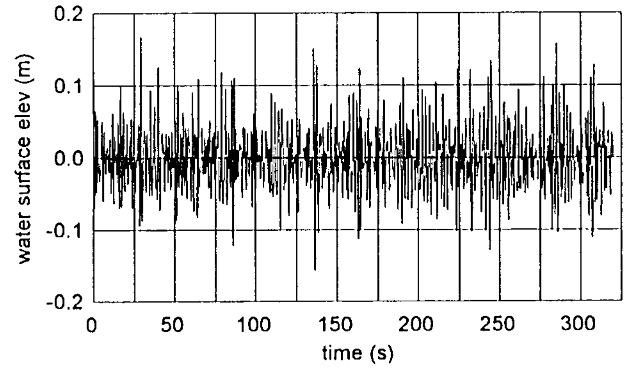

Fig. 2. Analog representation of the random water surface elevation which is to be analyzed.

and thus is too large to fit on a single 3.5 in. disk of 1.44 Mbytes storage capacity. Therefore we recommend that this spreadsheet exercise be stored on a computer hard drive or 100 Mbytes Zip™ drive. If 3.5 in. disks must be used, two disks will be required. The exercise will contain five spreadsheet pages labeled:

- Raw Data: the provided raw water surface elevation data ≃300 kbytes
- Water Surface Elevation Analysis≃200 kbytes
- Wave Height Analysis≃700 kbytes
- Fast Fourier Analysis≃700 kbytes and
- Summary≃20 kbytes.

These 5120 discrete values of time and water surface elevation are placed vertically in the first and second columns of the Raw Data notebook page. Table I is an abbreviated illustration of the Raw Data notebook page.

The students now demean the data so that tidal phase and atmospheric forcing of the water surface are removed from the signal. We take this step even though our signal's duration is short, so that the students will consider demeaning for later cases when the signal is much longer in duration. These demeaned data of 5120 η values are placed in the third column and the mean, maximum, minimum, and standard de-

Table I. Abbreviated illustration of the Raw Data notebook page where η is water surface elevation.

	η (m)	Demeaned η (m)
Mean	−0.001	1.4E−19
Max		0.168
Min		−0.157
σ_η		0.042
H_s		0.166

Time (s)	η (m)	Demeaned η (m)	Bin left edge (m)	Bin center (m)
0.0000	0.021	0.022	−0.160	−0.150
0.0625	0.021	0.022	−0.140	−0.130
0.1250	0.017	0.018	−0.120	−0.110
⋯	⋯	⋯	−0.100	−0.090
319.7500	0.028	0.029	⋯	⋯
319.8125	0.021	0.023	0.160	0.170
319.8750	0.011	0.012	0.180	0.190
319.9375	0.000	0.001		

viation (σ_η) of the demeaned η values are determined using built-in spreadsheet functions. Although the procedure of demeaning and further detrending real oceanic data—such as from tidal effects—is discussed in class, no further detrending of laboratory data is necessary here. In the laboratory, tidal effects and other typical sources of low frequency fluctuation and data drift are either negligible or nonexistent.

Short-term wave statistics describe the probabilities of occurrence of wave heights and periods that occur within one observation or measurement of random waves from a stationary point. Such an observation consists of a wave record over a short time period; this time period here can range from a few minutes to perhaps 15–30 min or more, but ultimately depends on the characteristics of the signal analyzed. The most important consideration concerning sampling period is to ensure that a reasonable data set with a sufficient number of constituent waves is acquired. From these observations, a single characteristic wave height and period may be determined which represents the random waves. It is important to understand, however, that wave heights and periods have some statistical variability about these characteristic values.

Further data analysis considerations are also discussed with students before, during, and after the laboratory experiment. For example, the use of data filters' effect on data integrity, resolution, and analysis results is described and illustrated. In addition, it is mentioned that windowing data and ensemble averaging for frequency analysis may improve spectra results. While the value of these data analysis procedures and tools is discussed with the students, this pedagogical exercise does not incorporate all of them explicitly. The authors believe that the incremental improvement in the results does not justify the additional complexities introduced into the numerical analysis. Perhaps the application of these considerations could be incorporated into a graduate course; but for undergraduates first experiencing the application of these methods of data analysis, we find it most effective to concentrate on the basic concepts and discuss the more advanced tools and methods qualitatively.[10]

III. WATER SURFACE ELEVATION ANALYSIS

The demeaned η data are now transferred from the Raw Data notebook page to a new notebook page entitled Water Surface Analysis. While the majority of this portion of the analysis deals with manipulation of the water surface data, the students are first instructed to calculate the most widely used statistical wave height parameter, Significant Wave Height (H_s), from[11]

$$H_s = 4\sigma_\eta, \tag{4}$$

where σ_η is the standard deviation of the demeaned η data. This definition is quite important since it provides a relationship between our two analyses of water surface elevation and wave heights and is discussed further in Sec. V. An alternate and independent definition of significant wave height ($H_{1/3}$) is the average of the highest 1/3 waves in the group being analyzed. For most open ocean waves in deep water, $H_{1/3}$ has the same statistical expectation value as H_s determined from Eq. (4).[12] In a practical sense, significant wave height is the wave height value approximately observed and recorded by a human when making a visual observation of waves.[13] In other words, the human eye has a natural filtering that tends to exclude the smallest waves and focuses instead on the larger waves in the sea state. H_s and $H_{1/3}$ will be used later in this exercise.

Table II. Abbreviated illustration of the Water Surface Elevation Analysis notebook page which depicts histogram results for observed data and Gaussian theory.

Bin left edge (m)	Bin center (m)	No. of occurrences	Observed pdf (m^{-1})	Gaussian pdf (m^{-1})
−0.160	−0.150	5	0.0488	0.0142
−0.140	−0.130	7	0.0684	0.0719
−0.120	−0.110	31	0.3027	0.2886
...	...	...	...	...
0.020	0.030	822	8.0273	7.3986
...	...	...	...	...
0.180	0.190	0	0.0000	0.0003
	total=	5120	50.0	49.9974

We now have the students select data bins into which they will place their η values. From their previously determined maximum and minimum η values, the students determine that a bin range of -0.160 to 0.190 m will include all their data. We specify that the data bins will be 0.02 m apart and will range from the lowest to highest demeaned η value. Bin centers and bins are shown on the right-hand side of Table I. An example interpretation of the bin centers and bins in Table I would be that all waves from -0.140 m up to, but not including, -0.120 m would be included in the bin whose center is -0.130 m.

We then introduce the students to probability density functions by having them build a histogram. We copy the specified Bin Centers and Bins from our Raw Data notebook page directly into the Water Surface Elevation Analysis notebook page illustrated in Table II. We now wish to determine a frequency distribution of η. We accomplish this by creating a ''frequency'' column which represents the number of η value occurrences within each bin. Fortunately, this laborious process is automated using a histogram subroutine found in most spreadsheet software where the input block is our demeaned η values, bin block is the center points of our bins (e.g., $-0.150, -0.130, \ldots, 0.190$), and output block is one cell above where we wish to place the calculated frequencies of occurrence for each bin.[14] We can now interpret the data in a meaningful way. For example, in Table II, there were 7 η values out of 5120 that were in the range $-0.140 \text{ m} \leq \eta < -0.120$ m.

Students must now determine the experimental, or observed, water surface elevation PDF, $p(\eta)$, for their data. We lead them in this direction by having them first note that the percentage of η values in bin i is:

$$\frac{N_i}{N_T}, \tag{5}$$

where N_i, is the number of η values in bin i and N_T is the total number of η values in all bins. Although this step is obvious for most students, the next step is not. They must now determine the probability density for bin i, $p(\eta_i)$, using

$$p(\eta_i) = \frac{N_i}{N_T \Delta\eta}, \tag{6}$$

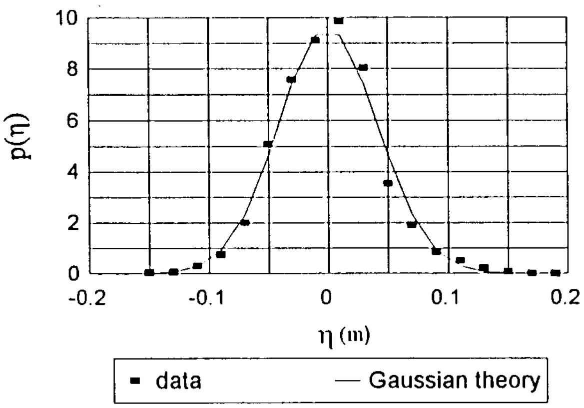

Fig. 3. Water surface elevation (η) probability density functions determined from data (discrete points) and Gaussian theory (continuous line).

where $\Delta\eta$ is the bin size and the units of $p(\eta)$ are m^{-1}.[15] If the students have difficulty with this concept, we remind them of the basic definition of the PDF:[16]

$$\int_{-\infty}^{\infty} p(\eta)\, d\eta \equiv 1. \tag{7}$$

Therefore, for the finite function of our example, we can combine Eqs. (6) and (7) to write

$$\sum_{-0.160}^{0.190} \frac{N_i}{N_T \Delta\eta}\,\Delta\eta = \sum \frac{N_i}{N_T} \equiv 1. \tag{8}$$

The percentage of occurrence is:

$$\frac{N_i}{N_T} = \frac{822}{5120} = 0.16 = 16\%, \tag{9}$$

whereas in Table II, for example, the observed PDF for the bin with $\eta = 0.030\,\mathrm{m} \pm 0.01\,\mathrm{m}$ is:

$$p(\eta_{+0.030}) = \frac{822}{5120 \times 0.02} = 8.0273\ \mathrm{m}^{-1}. \tag{10}$$

We now have the students calculate the theoretical Gaussian PDF using Eq. (1), noting that the only data-dependent value required is σ_η, which we obtained when developing the Raw Data notebook page. The students then plot a graph of the probability density of $\eta[p(\eta)]$ vs η itself for both the data (as discrete points) and for the theoretical Gaussian probability distribution (as a continuous curve). Students' understanding of the applicability of the theoretical Gaussian PDF is reinforced by its subsequent agreement with their data values as illustrated in Fig. 3.

IV. WAVE HEIGHT ANALYSIS

The demeaned η data are again transferred from the Raw Data notebook page to a new notebook page entitled Wave Height Analysis. This page is illustrated in Table III. The students are now directed to conduct a zero-upcrossing analysis[17] of the water surface elevation time history record. Basically, this is an analysis where the student defines an individual wave period (T_i) as the time between successive upcrossings of η across the mean water line. The wave height is then the difference between the extreme maximum and extreme minimum η value within that particular period. Figure 1(b) illustrates the zero-upcrossing method. Depending on the spreadsheet or computer programming experience of your students, the instructor should decide whether the students should program the concept or if the spreadsheet formulas for this step will be provided.

Table III. Illustration of Wave Height Analysis notebook page for the first 31 values of water surface elevation (η). Maximum crest is the largest η value in the crest between previous and following upcrossings whereas minimum trough is the smallest η value in the trough between upcrossings. The last column is the sorted wave heights (H) in descending order determined from the entire record of 5120 data values. Note that there are many more η values than H values which do not show because the H values are truncated for illustration.

η (m)	Crest (m)	Maximum crest (m)	Trough (m)	Minimum trough (m)	Wave height (m)	Sorted wave ht (m)
0.022	0.022					0.309
0.022	0.022					0.264
0.018	0.018					0.255
0.011	0.011	0.022		−0.069	0.092	0.241
−0.000			−0.000			0.234
−0.014			−0.014			0.231
−0.028			−0.028			0.226
−0.040			−0.040			0.225
−0.052			−0.052			0.223
−0.062			−0.062			0.199
−0.067			−0.067			0.195
−0.069			−0.069			0.193
−0.064			−0.064			0.192
−0.056			−0.056			0.191
−0.047			−0.047			0.183
−0.032			−0.032			0.181
−0.015			−0.015			0.174
0.005	0.005					0.174
0.025	0.025					0.172
0.043	0.043					0.169
0.055	0.055					0.169
0.064	0.064					0.168
0.066	0.066					0.166
0.054	0.054					0.166
0.038	0.038					0.166
0.020	0.020					0.166
0.007	0.007	0.066		−0.047	0.113	0.163
−0.008			−0.008			0.161
−0.020			−0.020			0.159
−0.029			−0.029			0.156
−0.038			−0.038			0.153

Separating the η values in Table III into columns of wave crests and wave troughs, plus placing the crest values in blue and the trough values in red, allows the students to better visualize the waves in their digital data records. It is now a simple matter for the students to determine the maximum crest and minimum trough value within a particular period using the spreadsheet software's maximum/minimum function or by visual examination. The wave height is then determined by subtracting the minimum trough from the maximum crest. The resultant wave heights are then converted to ''values'' to remove their relational dependence on specific spreadsheet cell values (otherwise, when moved and/or sorted, the spreadsheet software would be looking for a minimum or maximum of irrelevant numbers), and then sorted from highest to lowest using a spreadsheet sort routine.

We now wish to lead the students to developing a PDF for wave heights. We proceed by following the same approach as we did for water surface elevations. We want to group the

Table IV. Abbreviated illustration of the wave height analysis histogram results for probability density functions (pdfs) and probability of exceedence based on observed data and Rayleigh theory. This histogram would be found on the Wave Height Analysis notebook page.

Bin left edge (m)	Bin center (m)	No. of waves	Observed pdf (m^{-1})	Rayleigh pdf (m^{-1})	Observed prob. of exceed.	Rayleigh prob. of exceed.
0.000	0.010	4	1.105	1.418	1.000	1.000
0.020	0.030	11	3.039	4.018	0.978	0.972
0.040	0.050	21	5.801	5.974	0.917	0.892
0.060	0.070	30	8.287	7.046	0.801	0.773
0.080	0.090	31	8.564	7.208	0.635	0.633
⋯	⋯	⋯	⋯	⋯	⋯	⋯
0.300	0.310	1	0.276	0.046	0.055	0.002
0.320	0.330	0	0.000	0.019	0.000	0.001
	Total=	181	50.000	50.107		

just determined wave heights into ''bins'' of height $H_i \pm(\Delta H)/2$, using a bin width (ΔH) of 0.020 m, and then count the number of waves in each bin. The students can build this histogram again, using a spreadsheet histogram tool where the input block is our sorted wave heights, bin block is the center points of our bins (e.g., 0.010, 0.030, . . . , 0.330), and output block is one cell above where we wish to place the calculated frequency of occurrence for each bin. The result is illustrated in Table IV. The observed probability density function is determined for each bin by dividing the number of waves in the bin (H_i) by the total number of waves in the record (H_T) and the bin width. This PDF is mathematically expressed as:

$$p(H_i)=\frac{H_i}{H_T\Delta H}. \tag{11}$$

The total number of waves is determined by the zero-upcrossing method. The theoretical Rayleigh probability density function is obtained using Eq. (2). The students then are required to graph the observed PDF values with the Rayleigh PDF values and comment on any differences. Students' understanding of the applicability of the theoretical Rayleigh PDF is reinforced by its subsequent agreement with their data values as illustrated in Fig. 4.

The cumulative probability distribution function, $P(H)$, is the percentage of waves having a height equal to or less than H, and can be written as:

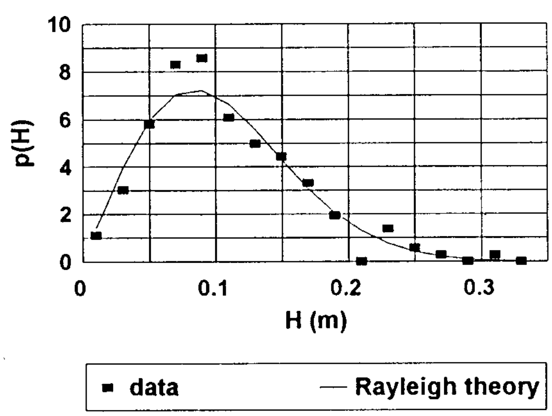

Fig. 4. Wave height (H) probability density functions determined from data (discrete points) and Rayleigh theory (continuous line).

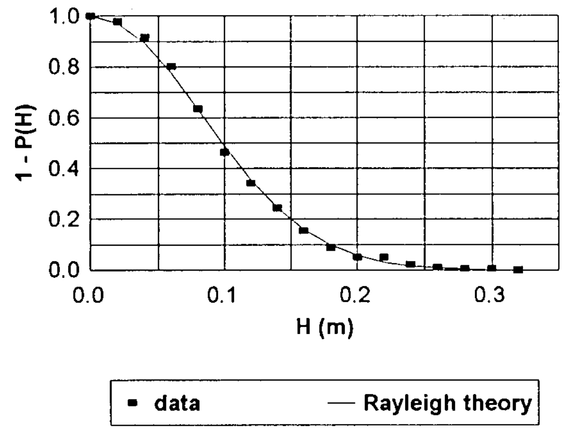

Fig. 5. Wave height (H) probability of exceedence determined from data (discrete points) and Rayleigh theory (continuous line).

$$P(H)=\int_0^H p(H)dH=1-\exp\left[-\left(\frac{H}{H_{\mathrm{rms}}}\right)^2\right]. \tag{12}$$

However, in designing structures to be placed in the water, ocean engineers are concerned with how often large waves will be encountered. Therefore, the percentage of waves having a height greater than a given height is of significant interest. This term is called the probability of exceedence and is defined as:

$$1-P(H)=\exp\left[-\left(\frac{H}{H_{\mathrm{rms}}}\right)^2\right]. \tag{13}$$

The students are now required to compare and comment on the observed probability of exceedence with the Rayleigh probability of exceedence values as listed in Table IV and illustrated in Fig. 5.

We now ask some questions of the students to ensure that they understand the material:

(1) What is the observed and theoretical probability density that wave heights will be between 0.020 and 0.040 m? [*Answer*. 3.039; 4.018 m^{-1}.]
(2) What is the observed and theoretical probability of occurrence that wave heights will be between 0.020 and 0.040 m? [*Answer*. (3.039)(0.020)=6.1%; (4.018)(0.020) =8.0%.]
(3) What is the observed and theoretical cumulative probability of occurrence that wave heights will be less than 0.040 m? [*Answer*. (1.015+3.039)(0.020)=8.3%; (1.418+4.018)(0.020)=10.9%.]

The students complete this section by using their wave height data and Rayleigh theory to determine common wave height parameters. If we assign a rank number, m, to the sorted wave heights where $m=1$ is the largest wave height and $m=M$ is the smallest wave height, then we may define common wave height statistical parameters as:
average wave height:

$$H_{\mathrm{avg}}=\frac{1}{M}\sum_{i=1}^{M} H_i, \tag{14}$$

significant wave height:

Table V. Rayleigh theory interrelationship between common wave height parameters. For example, $H_{avg} = 0.886\, H_{rms}$.

	σ_η	H_{avg}	H_{rms}	$H_{1/3}$	$H_{1/10}$
σ_η	1	0.399	0.3536	0.25	0.1964
H_{avg}	2.506	1	0.886	0.625	0.4926
H_{rms}	2.828	1.1284	1	0.706	0.566
H_s	4	1.598	1.416	1	0.787
$H_{1/10}$	5.091	2.03	1.8	1.273	1

$$H_{1/3} = \frac{1}{M/3} \sum_{m=1}^{M/3} H_m, \tag{15}$$

one-tenth highest wave height:

$$H_{1/10} = \frac{1}{M/10} \sum_{m=1}^{M/10} H_m. \tag{16}$$

These wave height parameters are used in conjunction with Eqs. (3) and (4) later in this exercise.

A significant feature of the Rayleigh distribution is that all characteristic wave height parameters are interrelated (see Table V). Therefore, once a single wave height parameter (e.g., H_{rms}) is determined from observations, Rayleigh theory can provide all the other wave height parameters.[18] Rayleigh theory can also provide an estimate of the maximum expected wave height, which is a function of H_{rms} and data set length, using

$$H_{max} = \sqrt{\ln M} H_{rms}. \tag{17}$$

Average wave period may be determined from

$$T_{avg} = \frac{1}{M} \sum_{i=1}^{M} T_i. \tag{18}$$

Alternatively, T_{avg} may be determined by dividing the record length in seconds by the number of waves in the record.

V. FAST FOURIER TRANSFORM

The essence of the Fourier transform of a waveform is to decompose or separate the waveform into a theoretically infinite sum of sinusoids of a series of continuous frequencies. If these sinusoids sum to the original waveform, then we have determined the Fourier transform of the waveform. In this exercise, we use a discrete Fourier transform to obtain a finite sum of waveforms of discrete frequencies to closely represent the observed waveform.

To relate the wave analysis efforts of the first exercise[2] to that which we have accomplished in this exercise, we now turn our attention to the fast Fourier transform (FFT). Using the demeaned η data from our Wave Height Analysis notebook page, the students conduct a FFT analysis on a notebook page entitled Fast Fourier Analysis. Since some spreadsheet FFT subroutines are limited to analyzing a maximum of 1024 data points, we utilize an ergodicity[19] assumption for our data set. This allows us to perform ensemble averaging by conducting a FFT on the first ensemble of 1024 data points, the second ensemble of 1024 data points, followed by the third, fourth, and fifth ensembles of 1024 data points. An ensemble average energy density, $S(f)$, is then determined from these five $S(f)$'s.[20]

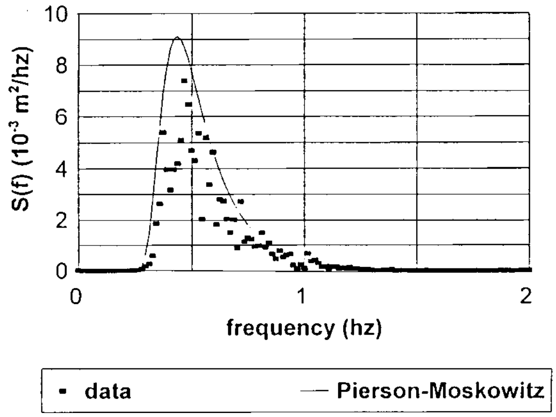

Fig. 6. Wave spectra determined from a fast Fourier transform of data (discrete points) and the empirically derived Pierson–Moskowitz formulation (continuous line).

The students are then asked to graph and compare the discrete $S(f)$ values obtained from the ensemble method against the common, empirical Pierson–Moskowitz (PM) spectrum of ocean waves determined by[21,22]

$$S(f) = \frac{\alpha_{PM} g^2}{(2\pi)^4 f^5} \exp\left[-1.25\left(\frac{f_m}{f}\right)^4\right], \tag{19}$$

where $\alpha_{PM} = 0.0081$, g=gravitational acceleration, f=frequency (Hz), and $f_m = 0.77/T_{avg}$,[23] where T_{avg} is the average wave period of the record. Figure 6 illustrates this comparison. The students conclude this section by calculating H_s from an expanded Eq. (4) where[24]

$$H_s = 4\sigma_\eta = 4(\sigma_\eta^2)^{1/2}$$
$$= 4(\text{area under the } S(f) \text{ curve})^{1/2}. \tag{20}$$

Using these procedures, more advanced wave analysis exercises can be developed using three other one-dimensional spectra such as the Bretschneider,[25] JONSWAP,[26] and TMA[27] spectra. These four spectra are presented because they represent a variety of conditions (see the references) and are extensively used in the field.

VI. SUMMARY OF RANDOM WAVE ANALYSIS METHODS

At this time, students are then asked to interpret their results and provide comments, including answers to the following additional questions:

(4) Comparisons:

(a) Compare and discuss how Gaussian theory replicated the distribution of the water surface elevation data.

(b) How did the results from the water surface elevation analysis, zero-upcrossing analysis, Rayleigh theory, and FFT analysis differ for the common wave parameters? What was the range of answers for the significant wave height?

(5) How can you improve the correlation of the results from the different methods? (*Answer*. More data, specifically, greater *N*.)

(6) Would you expect to get the same results from your wave height analysis if you used ''downcrossings'' instead of ''upcrossings''? (*Answer*. For an infinitely long data set, the answers should be the same, but for a finite data set such as

Table VI. Illustration of the Summary notebook page which provides a comparison of common wave parameters as determined by four random wave analysis methods.

Analysis method	σ_η (m)	H_{avg} (m)	H_{rms} (m)	H_s (m)	$H_{1/3}$ (m)	$H_{1/10}$ (m)	H_{max} (m)	T_{avg} (s)
Water surface	0.042			0.166				
Zero upcrossing		0.104	0.118		0.163	0.229	0.229	1.768
Rayleigh theory		0.105	0.118		0.167	0.213	0.270	
Fast Fourier					0.166			

the one used here, there will be a noticeable difference in the results.)

(7) If you were to set up an apparatus to measure a time series in a particular situation (e.g., real ocean waves), what do you need to know beforehand in order to choose reasonable values of N and sampling interval (Δt)? [*Answer*. You would need to know what general frequencies you were interested in. Analyzing ocean surface gravity waves would require different values of N and Δt than analyzing longer period waves such as tides. Knowing the upper and lower bounds to your frequencies of interest would allow an appropriate selection of the fundamental frequency and Nyquist frequency, and hence Δt and N, respectively.[28]]

To conclude the exercise, the students are asked to construct a table on the Summary notebook page which compares common wave height parameters (H_s, H_{avg}, etc.) determined from their four analysis methods (i.e., water surface elevation analysis, zero-upcrossing analysis, Rayleigh theory, and FFT analysis). Table VI illustrates this result for our data set. The students are then directed to copy Tables II and IV, along with Figs. 3–6, onto the Summary notebook page. The students are then tasked to articulate as compactly as possible on the Summary page:

(8) What are the three most important concepts you have learned from this exercise?

This page is then delivered to the instructor for evaluation along with the students' interpretation of their results and answers to the above questions.

VII. SUMMARY

There are typically few opportunities throughout science and engineering curricula in which students analyze a physical phenomenon using multiple and considerably different analysis methods. In the procedure discussed in this pedagogical paper, random water waves are analyzed using three different methods: (1) a statistical analysis of water surface elevation, (2) a statistical analysis of wave heights (after performing a zero-upcrossing analysis on the data to extract wave heights from the water surface elevation data), and (3) a spectral analysis employing use of a fast Fourier transform. The results from the water surface elevation data are compared to Gaussian probability theory. The results from the wave height statistics are compared to those obtained from Rayleigh theory, using the experimentally obtained root-mean-square wave height (H_{rms}).

It is the interpretation and correlation of the results from these widely varying methods that imparts to the students an appreciation for both time-domain and frequency-domain analyses of random signals. While the procedures discussed here are easily applicable to physical phenomena other than water waves, it is the familiarity that the students have with water waves that makes such a numerically intense analysis exercise more tolerable and ultimately more meaningful.

[a]Electronic mail: whitford@nadn.navy.mil

[1]Note that random waves as observed in the ocean are not typical ''white noise'' signals with a ''flat line'' frequency distribution. They possess distinguishable, narrow-band (observable surface gravity wind waves fall between 0.03 and 1.0 Hz) frequency distribution features which are Gaussian.

[2]The first part is contained in D. J. Whitford, M. E. C. Vieira, and J. K. Waters, ''Teaching Time-Series Analysis. I. Finite Fourier Analysis of Ocean Waves,'' Am. J. Phys. **69**, 490–496 (2001).

[3]Since this exercise's procedures are equally applicable to lakes, bays, and rivers, we have adopted the term ''water surface elevations'' for this article rather than the more traditional term of ''sea surface elevations.''

[4]M. J. Tucker, *Waves in Ocean Engineering: Measurement, Analysis, Interpretation* (Ellis Horwood, New York, 1991), 41 p. Further information on Gaussian, or normal, distributions can be found in any introductory probability text such as M. R. Spiegel, *Schaum's Outline of Theory and Problems of Probability and Statistics* (McGraw–Hill, New York, 1975), pp. 109–111.

[5]For additional information, see R. M. Sorensen, *Basic Coastal Engineering*, 1st ed. (Wiley, New York, 1978), p. 121.

[6]Note that elevation (η) is a signed zero-to-value amplitude, that is, it measures from the still water level or baseline, and can be positive or negative. In contrast, wave height (H), as used in this paper, is a peak-to-peak measure of water surface elevation, and is a positive value (i.e., never less than zero).

[7]M. S. Longuet-Higgins, ''On the statistical distribution of the heights of sea waves,'' J. Mar. Res. **11** (5), 245–266 (1952): Alternatively, see R. G. Dean and R. A. Dalrymple, *Water Wave Mechanics for Engineers and Scientists* (World Scientific, Singapore, 1991), pp. 187–193, or B. Kinsman, *Wind Waves* (Dover, New York, 1984), pp. 342–347.

[8]Sample water surface elevation data from the U.S. Naval Academy's wave tank can be found and downloaded at http://www.usna.edu/NAOE/courses/en475.htm under ''Additional Course Materials.'' Other web sites, such as the National Oceanic and Atmospheric Administration's National Data Buoy Center (http://seaboard.ndbc.noaa.gov/) have actual historical and real-time oceanic data available for downloading in a wide variety of formats.

[9]Note that although the signal generated is not truly random (i.e., the signal is known and can be repeated), the wave field it creates simulates an irregular or ''random'' sea. Thus it is referred to as a ''random wave signal'' and the generated waves are called ''random waves.'' Students are cautioned to appreciate the difference between what is truly random and what is a programmed simulation of a random process. For additional information, see J. Zseleczky, ''Wave Generating Software at the U.S. Naval Academy,'' 24th American Towing Tank Conference, Texas A&M University, College Station, Texas, November 1995 (unpublished).

[10]Further information on more advanced topics in time series analysis (e.g., filtering and windowing) can be found in C. Chatfield, *The Analysis of Time Series—An Introduction* (Chapman and Hall, New York, 1984), 3rd ed. or G. Jenkins and D. Watts, *Spectral Analysis and Its Applications* (Holden–Day, Oakland, CA, 1968), among others.

[11]Reference 5, p. 163.

[12]For steep deep water waves, as well as waves in intermediate and shallow water depths, significant wave height determined by the wave-by-wave method will be increasingly larger than significant wave height determined by the standard deviation of the demeaned η data. See E. F. Thompson and

C. L. Vincent, ''Significant Wave Height for Shallow Water Design,'' Journal, Waterway Port Coastal and Ocean Engineering Division, American Society of Civil Engineers, September, 828–842 (1985).

[13]Reference 5, p. 153.

[14]Your spreadsheet software will dictate the tabular format of the histogram output. Thus your resulting format may not be identical to Tables I and II.

[15]This renormalization by $\Delta\eta$ is similar, but different, from that carried out in Ref. 2 using Δf.

[16]M. R. Spiegel, *Schaum's Outline of Theory and Problems of Probability and Statistics* (McGraw–Hill, New York, 1975), pp. 40–41.

[17]W. J. Pierson, ''An Interpretation of the Observable Properties of Sea Waves in Terms of the Energy Spectrum of the Gaussian Record,'' Trans. Am. Geophys. Union **35**, 747–757 (1954).

[18]Several authors have compared the Rayleigh distribution with measured wave heights and found that the distribution yields acceptable values for most storms. See M. D. Earle, ''Extreme Wave Conditions During Hurricane Camille,'' J. Geophys. Res. **80** (3), 377–379 (1975) or S. K. Chakrabarti and R. P. Cooley, ''Statistical Distribution of Periods and Heights of Ocean Waves,'' *ibid.* **82** (9), 1363–1368 (1971). For more explanatory information on the Rayleigh distribution beyond Longuet–Higgins' (1952) original work, one may peruse R. G. Dean and R. A. Dalrymple, *Water Wave Mechanics for Engineers and Scientists* (World Scientific, Singapore, 1991), pp. 187–193 and R. M. Sorensen, *Basic Coastal Engineering* (Wiley, New York, 1997), 2nd ed., pp. 155–159.

[19]In time series analysis, ''ergodicity'' means that '' . . . for most stationary processes which are likely to be met in practice, the sample moments of an observed record length T converge to the corresponding population moments as $T\rightarrow\infty$. In other words, time averages for a single realization converge to ensemble averages.'' (Chatfield, 1985, p. 65). For further information on ergodicity, see C. Chatfield, *The Analysis of Time Series—An Introduction* (Chapman and Hall, New York, 1984), 3rd ed, p. 65, or G. Jenkins and D. Watts, *Spectral Analysis and Its Applications* (Holden–Day, Oakland, CA, 1968), p. 179, among others.

[20]Although other math software programs such as MATLAB are capable of performing a FFT on this entire data set, it is still useful to use a spreadsheet program here, despite a possible 1024-point limitation. In addition to introducing the concept of ensemble averaging to the student, performing this task tends to bring about a general discussion of the limitations that exist with all software. Virtually all data analysis programs and computer hardware reach some limitation in real data analysis. Discovering methods to work within these limitations is many times part of conducting research.

[21]This spectrum was based on wave records acquired in the North Atlantic Ocean for 20–40 knots. It applied to deep water, fully arisen seas, i.e., seas in equilibrium with the wind speed which had blown over unlimited fetch and for unlimited duration. See W. J. Pierson and L. Moskowitz, ''A Proposed Spectral Form for Fully Developed Wind Seas Based on Similarity Theory of S. A. Kitaigorodskii,'' J. Geophys. Res. **69**, 5181–5190 (1964).

[22]There have been many spectra developed which address ocean waves. Most of them, including Eq. (19), follow an accepted format of $S(f)=(A/f^5)\exp[-B/f^4]$, where A and B adjust the shape and scale of the spectra. By combining terms, Eq. (19) could be more simply written as: $S(f)=(0.0005/f^5)\exp[-1.25(f_m/f)^4]$.

[23]M. K. Ochi, *Ocean Waves, the Stochastic Approach* (Cambridge U.P., Cambridge, UK, 1991), p. 39.

[24]Reference 5, pp. 119–121.

[25]This spectrum was based on Eq. (2), was empirically derived, and assumed no correlation between individual wave periods and heights. See C. L. Bretschneider, ''Wave Variability and Wave Spectra for Wind-Generated Gravity Waves,'' Technical Memorandum No. 118, U.S. Army Beach Erosion Board, Washington, DC, 1959.

[26]This spectrum was a modification of the PM spectrum to account for fetch-limited seas. See K. Hasselmann, T. P. Barnett, E. Bouws, H. Carlson, D. E. Cartwright, K. Enke, J. A. Ewing, H. Gienapp, D. E. Hasselmann, P. Kruseman, A. Meerburg, P. Muller, D. J. Olbers, K. Richter, W. Snell, and H. Walden, ''Measurements of Wind-Wave Growth and Swell Decay During the Joint North Sea Wave Project (JONSWAP),'' Report to the German Hydrographic Institute, Hamburg, Germany, 1973.

[27]This spectrum accounted for waves that were moving into shallower water. See E. Bouws, H. Gunther, W. Rosenthal, and C. L. Vincent, ''Similarity of the Wind Wave Spectrum in Finite Depth Water. I. Spectral Form,'' J. Geophys. Res. **90**, 975–986 (1985).

[28]Fundamental frequency ($f_1=1/[N\Delta t]$) and Nyquist frequency ($f_c=1/[2\Delta t]$) are discussed in detail in the first exercise described in Ref. 2.

THERMO PROBLEMS

What distinguishes a textbook from a technical monograph is problems in the former but not in the latter. And what distinguishes thermodynamics textbooks is problems even more boring than the text that precedes them. A typical problem is something like the following: What is the final temperature of two kilograms of nitrogen initially at room temperature when its pressure is doubled adiabatically? This is the kind of drill question rarely if ever asked outside the pages of textbooks, a question to which no one wants to know the answer, what students refer to contemptuously as ''plug and chug.'' Yet questions in textbooks should be genuine, by which we mean ones that inquisitive people ask in the course of their attempts to understand the world around them. An example of a genuine question is one we were asked by a writer for a motorcycle magazine. He called to ask about wind chill. Our response was duly published in the magazine. A question about thermodynamics asked by tattooed and bearded men wearing nose rings and riding chrome-plated Harley–Davidson motorcycles is by our reckoning a good question.

Craig F. Bohren and Bruce A. Albrecht, *Atmospheric Thermodynamics* (Oxford University Press, New York, 1998), p. xix.