

CHAPTER 29

WAVE HEIGHT AND PERIOD DISTRIBUTIONS FROM LONG-TERM WAVE MEASUREMENT

Chung-Chu Teng¹ and Ian M. Palao²

Abstract

Both the univariate and joint distributions of the significant wave height (H_s) and the zero-crossing wave period (T_z) are studied using 16 years of hourly wave data measured from five National Data Buoy Center (NDBC) buoy stations. For the univariate distribution, the log-normal distribution fits both the H_s and the T_z well. The modified log-normal distribution proposed by Fang and Hogben (1982) does improve the fit at the high H_s end and at the peak for some stations. For H_s , the Weibull distribution, with parameters computed from the maximum likelihood (ML) method, fits the upper tail of the cumulative distribution; however, it underpredicts both the probability peak and the probability density at the high end. For the joint distribution, the marginal Weibull/conditional log-normal distribution best describes the measured data of the steeper sea states and has the best overall fit.

Introduction

The theoretical distribution and characteristics of individual wave heights and periods have been extensively studied and verified using measured data (see Tucker 1991). Unlike individual waves, distributions of H_s and T_z , which are important for ocean and coastal engineering, need to be determined empirically from real measurements. Thus, both the univariate and joint distributions of the two wave parameters have not been extensively studied in the past due to a lack of reliable, long-term wave measurements. However, the early studies conducted by Ochi (1978) and Fang and Hogben (1982) provide a very useful basis for studying distributions of H_s and T_z . Some recent studies (e.g., Ochi (1992); Mathisen and Bitner-Gregersen (1990); and Athanassoulis, et al. (1994)) present many ideas and useful results on this topic.

In this study, 16 years of hourly wave data measured at five NDBC buoy stations are used to study both the univariate and the joint distributions of H_s and T_z .

¹ Principal Engineer/Project Manager, National Data Buoy Center, Stennis Space Center, MS 39529-6000, U.S.A.

² Senior Scientist, Computer Sciences Corporation, Stennis Space Center, MS 39529-6000, U.S.A.

Measured Wave Data

Data used in the present study were obtained from five NDBC buoy stations, designated as stations 46001, 46002, 46003, 46005, and 46006. All five buoy stations, as shown in Figure 1, are located in the northeastern Pacific Ocean. Water depths (in meters) at these buoy stations are listed in parentheses below the station numbers in the figure. These buoy stations were chosen for this study mainly because they are located in very deep water, so that wave height and period are not affected by shallow water effects, and they have longer periods of available wave data. These buoy stations measure both meteorological and wave data and report hourly data in nearly real time through the Geostationary Operational Environmental Satellite (GOES).

Wave data are processed onboard the buoys by transforming the time-series data into wave frequency spectra. The processed wave data, together with meteorological data, are transmitted to shore via GOES. Then, the wave information received are further processed and analyzed (e.g., conducting transform function, noise correction, and data quality control). H_s and T_z are derived from the wave spectrum: $H_s = 4(m_0)^{1/2}$ and $T_z = 2\pi(m_0/m_2)^{1/2}$, where m_0 and m_2 are the zero-th and the second spectral moments, respectively.

Sixteen years (from 1980 through 1995) of hourly wave data measured at these buoy stations are used in this study. During these 16 years, periods of missing data were present in the wave records for all of the stations. These data gaps were caused by a number of different factors: sensor failure, payload failure, periodic loss of data during transmission, or the buoy not being on station due to refurbishment or mooring failure. Table 1 shows the monthly distribution of the available data points during the 16 years for all five stations. Data for these five stations cover 76% (46006) to 90% (46001) of the all possible hourly data. From the distribution, it is clear that the data gaps or missing data did not concentrate on any

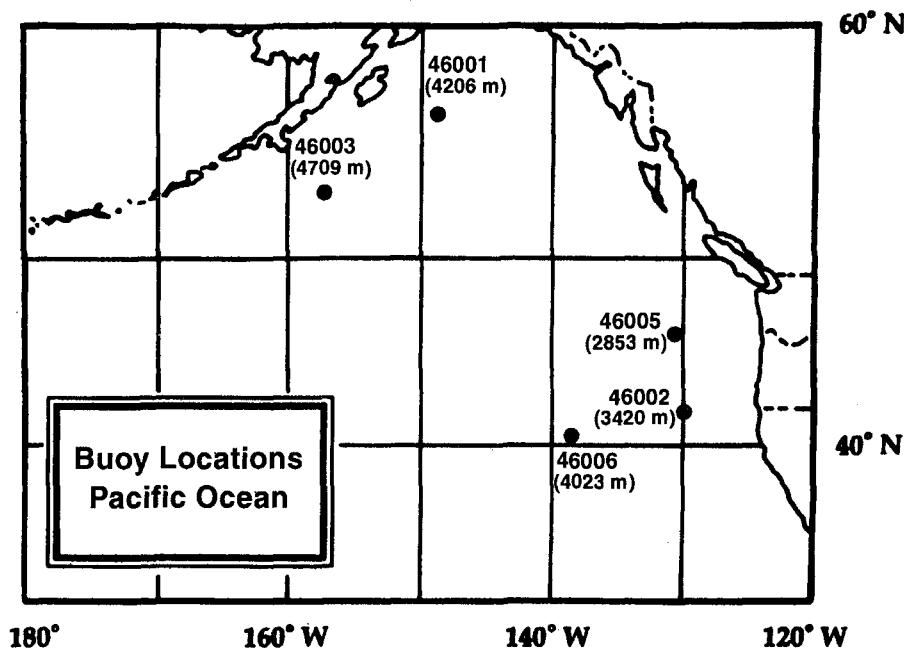


Figure 1. Locations of the five NDBC buoy stations used in this study

Table 1. Monthly distributions and total numbers of available data

	46001	46002	46003	46005	46006
January	10,917	9,368	9,322	10,141	9,023
February	9,483	8,703	7,053	9,343	7,931
March	9,886	9,448	8,348	10,179	8,271
April	8,897	8,602	8,765	9,546	7,526
May	10,351	8,104	8,134	9,347	7,880
June	10,567	9,459	8,616	9,031	7,975
July	10,836	9,878	9,289	10,145	8,752
August	11,326	9,975	9,029	10,915	9,649
September	11,069	9,930	10,447	9,133	9,838
October	10,994	10,473	9,570	9,347	10,242
November	10,610	9,583	9,016	8,471	9,792
December	10,876	9,433	10,242	9,216	9,650
Total	125,692	112,956	107,831	114,814	106,529

particular month or season. In addition, the whole data set was carefully examined, and no data gap that could significantly affect long-term wave statistics was found. Thus, it can be assumed that the missing data or data gaps in this data set will not affect the long-term distributions of H_s and T_z .

Data Analysis

Univariate Distributions

For univariate (or marginal) distributions of H_s and T_z , two commonly used probability distribution functions are studied: the log-normal and the three-parameter Weibull distributions. Parameters in these two distributions are computed by three different methods: the least square (LS) method, the ML method, and the method of moments. More details of these methods are presented in Teng, et al. (1994) and Palao, et al. (1994).

(1) Log-Normal Distribution

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

where μ and σ are parameters of the distribution.

The modified log-normal distribution, which includes a skew factor to make a better fit at high values (Fang and Hogben, 1982), is also examined in this study.

(2) *Three-Parameter Weibull Distribution*

$$p(x) = \frac{\beta(x-\gamma)^{\beta-1}}{\alpha^\beta} e^{\left[-\left(\frac{x-\gamma}{\alpha}\right)^\beta\right]}$$

where α , β , and γ are the shape, scale, and location parameters, respectively.

Joint Distributions

For joint distributions of H_s and T_z , three frequently used distribution functions are examined: bivariate log-normal distribution, bivariate two-parameter Weibull distribution, and marginal three-parameter Weibull/conditional log-normal distribution. More details of these distributions can be found in Mathisen and Bitner-Gregersen (1990). In the following equations, x represents T_z , and y represents H_s .

(1) *Bivariate Log-Normal Distribution*

$$p(x,y) = \frac{0.5}{\sqrt{1-\alpha_{xy}^2} \pi \sigma_x \sigma_y xy} e^{-\frac{0.5}{1-\alpha_{xy}^2} \left[\frac{(\ln(x)-\mu_x)^2}{\sigma_x^2} - \frac{2\alpha_{xy}(\ln(x)-\mu_x)(\ln(y)-\mu_y)}{\sigma_x \sigma_y} + \frac{(\ln(y)-\mu_y)^2}{\sigma_y^2} \right]}$$

where μ_x , μ_y , σ_x , σ_y , and α_{xy} are parameters of the distribution.

Computationally, μ_x and μ_y are the expected values of $\ln(x)$ and $\ln(y)$; σ_x and σ_y are the standard deviations of $\ln(x)$ and $\ln(y)$; and α_{xy} is the correlation coefficient between $\ln(x)$ and $\ln(y)$. The modified bivariate log-normal distribution, proposed by Fang and Hogben (1982), is also examined in this study for reference purposes. The modified distribution is identical to the bivariate log-normal distribution except for the addition of a multiplicative term describing the skewness of $\ln(y)$.

(2) *Bivariate Weibull Distribution*

$$p(x,y) = \left[\frac{\zeta y^{\zeta-1}}{\eta^\zeta} e^{-\left(\frac{y}{\eta}\right)^\zeta} \right] \left[\frac{\beta x^{\beta-1}}{\alpha^\beta} e^{-\left(\frac{x}{\alpha}\right)^\beta} \right]$$

where α , β , η , and ζ are parameters of the distribution.

This distribution uses two two-parameter Weibull distributions (i.e., one for H_s and one for T_z) as a joint model. The parameters are calculated using the LS method. The marginal distribution is used for T_z while the estimation of the distribution of H_s is conditional on T_z through the use of the η parameter.

(3) *Marginal Weibull/Conditional Log-Normal Distribution*

$$p(x,y) = \left[\frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} \right] \left[\frac{\beta(y-\gamma)^{\beta-1}}{\pi\alpha^\beta} e^{-\left(\frac{y-\gamma}{\alpha}\right)^\beta} \right]$$

where α , β , γ , σ , and μ are parameters.

This distribution is calculated by using a three-parameter Weibull for H_s along with a log-normal for T_z . The parameters of this distribution are identical to the corresponding univariate distributions. The LS method is used to calculate the three Weibull parameters and the two log-normal parameters. Note the log-normal parameters are calculated for every 0.5-m wave height interval.

Results and Discussion

Figure 2 shows the comparisons between histograms of the measured H_s , and the univariate distributions for the five buoy stations. From visual judgment, the log-normal distribution, in general, fits the data better than the Weibull distribution. For the log-normal distribution, results obtained from the three different methods of parameter estimation are almost identical. The log-normal distribution slightly overpredicts the peak for the two northern stations (46001 and 46003) and slightly misses the peak location for two of the three southern stations (46005 and 46006). For the Weibull distribution, different parameter estimate methods produce different parameters and have different effects on the distribution shapes. The LS method always underpredicts the high wave range and overpredicts the location of the peak. The ML method provides much better prediction of the peak location and at high H_s , but significantly underpredicts the height of the peak and overpredicts at the lower range. The Weibull fits, based on the method of moments, are much poorer than the other two methods and are not presented in the figures.

Two goodness-of-fit statistics, the Kolmogorov-Smirnov statistic (K-S) and Chi-square statistic (χ^2), are used to evaluate the degree of fit between the data and the fitted distributions. Table 2 summarizes the values of these two statistics for the two distributions for each of the five buoy stations. Note the values of χ^2 are computed based on a wave height interval of 0.1 m. For the Weibull distribution, the goodness-of-fit values for the ML method, which are listed in the table, are always smaller (i.e., the fits are better) than those from the other two methods. Based on the two goodness-of-fit statistics, the log-normal distribution fits the data better than the Weibull distribution, regardless of the parameter estimate method used.

To further examine the fit of the univariate distributions of H_s , comparisons between the empirical cumulative distribution from the measured data and the theoretical cumulative distributions are made. Figure 3(a) shows an example of this comparison for station 46001. The log-normal distribution fits the data well, except for the upper tail (i.e., cumulative probability greater than 0.9) where it slightly underpredicts the empirical cumulative distribution. This trend was also reported by many previous studies (e.g., Ochi (1978); Fang and Hogben (1982)). However, among the five stations, stations 46001 and 46003 clearly show

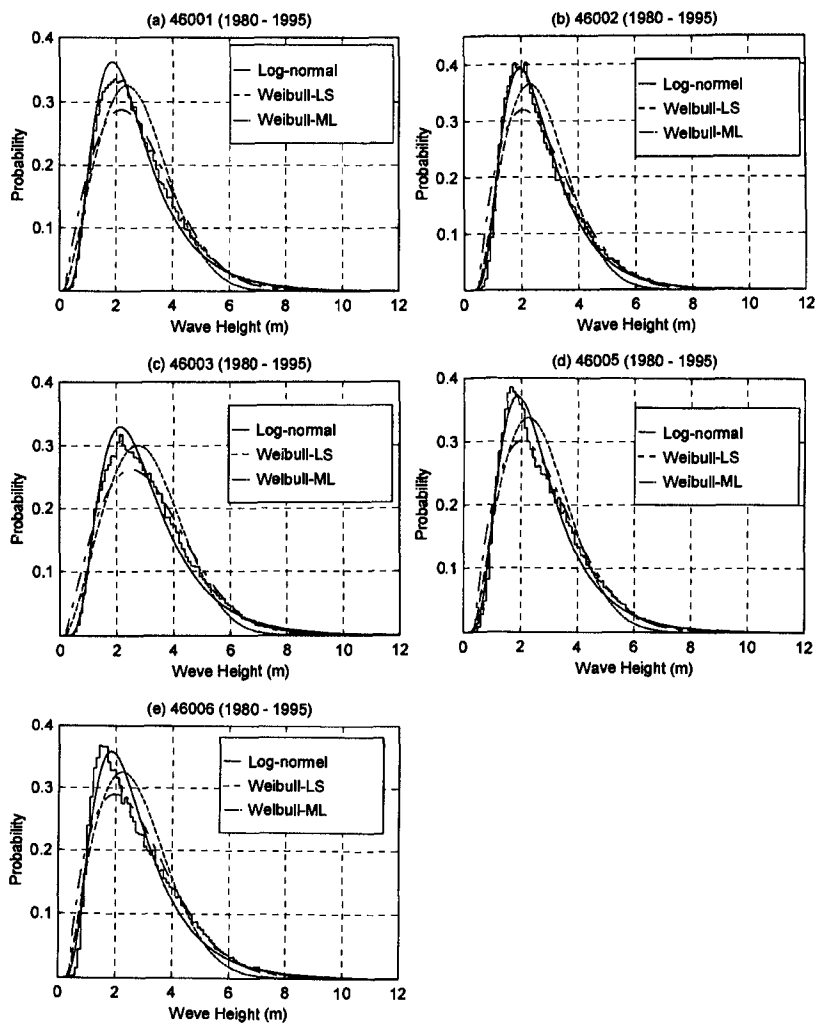


Figure 2. Comparisons of H_s between histograms and the univariate distributions for all five buoy stations

Table 2. Values of χ^2 and K-S for distributions of the zero-crossing wave period

	Log-Normal		Weibull	
	K-S	$\chi^2 (\times 10^{-6})$	K-S	$\chi^2 (\times 10^{-6})$
46001	0.01560	1.01749	0.03438	1.01889
46002	0.03862	0.91254	0.03086	0.91537
46003	0.01440	0.87029	0.03172	0.87135
46005	0.03021	0.92694	0.03337	0.92846
46006	0.02717	0.86185	0.02981	0.86325

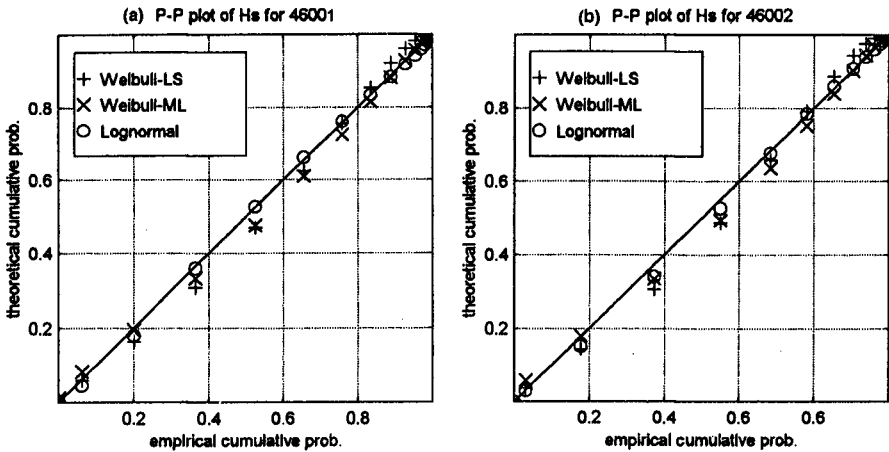


Figure 3. Comparisons of cumulative distribution for H_s between measured data and univariate distributions for (a) station 46001 and (b) station 46002

the underprediction. For other stations, there is very little or no underprediction from the log-normal distribution at the upper tail, as shown in Figure 3(b) for station 46002. For the Weibull distribution, the LS method overpredicts both the upper and lower tails for all five stations, while the ML method predicts well at the upper tail but underpredicts at the lower tail.

Figure 4 shows, for H_s at 46001, the histograms and univariate distributions including the Weibull, the log-normal, and the modified log-normal (Fang and Hogben (1982)) distributions. It is clear that the modified log-normal distribution significantly improves the fitting at high values and at the peak. The improvement is significant for stations 46001 and 46003. For other stations, there is very little or no improvement. Note the Weibull distribution, with parameters computed from the ML method, underpredicts the distribution at high values despite the better fit of the cumulative distribution at the high end. This trend is valid for all five stations.

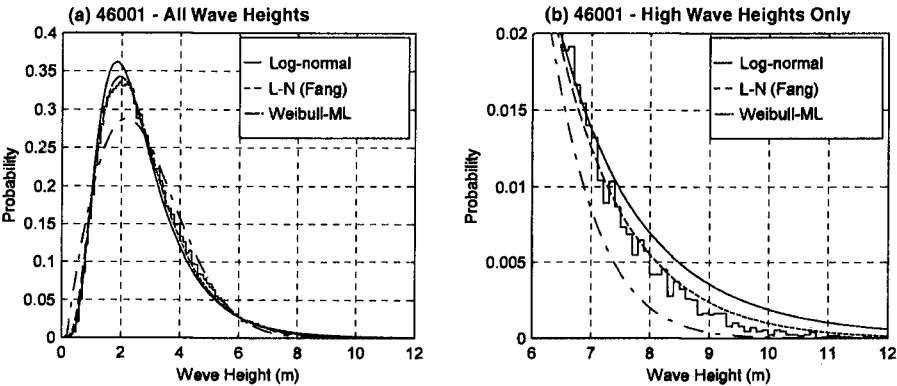


Figure 4. Comparisons of H_s between histograms and the univariate distributions for station 46001

For T_z , comparisons between the histograms and the univariate distributions are presented in Figure 5. Similar to H_s , the parameters of the log-normal distribution for T_z computed from the three methods of parameter estimation are almost identical. In addition, variations of these parameters from station to station are very small. The parameter μ varies from 1.88 to 1.96, and the parameter σ varies from 0.187 to 0.202. The log-normal distribution fits very well for the two northern stations (i.e., 46001 and 46003). For the three southern stations (i.e., 46002, 46005, and 46006), the histograms show broader peaks, and neither of the two distributions predicts the peak very well. Similar to their effects on H_s , the Weibull distribution based on the ML method predicts the upper range best and underpredicts the peak, while the LS method underpredicts the high range and does not predict the peak well. Again, the method of moments is much worse than the other two methods, and its results are not presented.

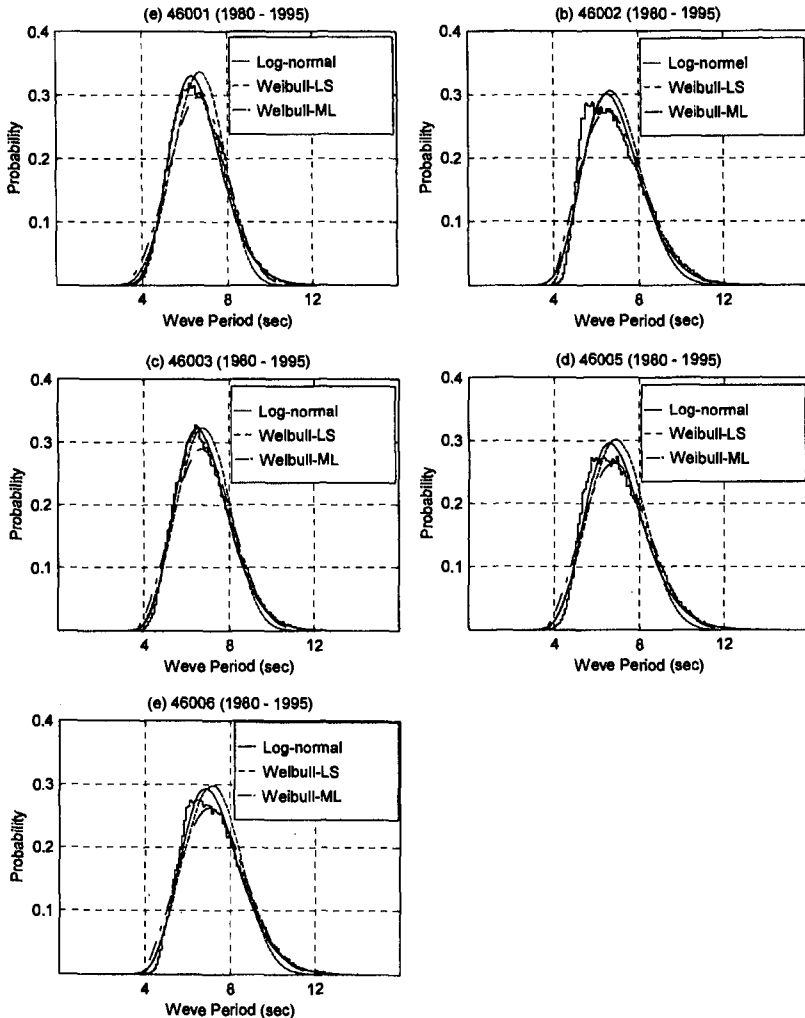


Figure 5. Comparisons of T_z between histograms and the univariate distributions for all five buoy stations

Table 3 presents values of K-S and χ^2 for T_z . Based on the two statistics, the log-normal distribution fits better than the Weibull distribution (using the ML method). Figure 6 presents comparisons between the empirical and fitted cumulative distributions of T_z for stations 46001 and 46002. For the log-normal distribution, unlike that for H_s , there is no underprediction at the upper tail of the cumulative distribution of T_z for all five stations. Performances of the Weibull distribution in modeling T_z using the LS and ML methods are similar to those for H_s .

The contour plots of the joint probability between H_s and T_z for station 46001 are presented in Figure 7. It is clear that both the bivariate log-normal and the bivariate Weibull distributions overpredict the joint probability for the steeper sea states, as indicated by the straight line. The marginal Weibull/conditional log-normal distribution fits this steeper sea range well. This trend was also shown by Mathisen and Bitner-Gregersen (1990) using wave data collected from the Norwegian Sea. The bivariate Weibull distribution also performs poorly at high wave periods. Except for steep seas, the bivariate log-normal distribution generally fits the data well. The modified bivariate log-normal distribution, proposed by Fang and Hogben (1982), was also studied. Its results are very similar to those from the original distribution and, thus, are not presented in this paper. Although the marginal Weibull/conditional log-normal distribution generally fits the data, it seems that it does not predict the very high probability location very well. The trends observed for station 46001 are also applicable to all other stations, but these results are not shown in this paper due to space limitations.

Table 3. Values of χ^2 and K-S for distributions of the significant wave height

	Log-Normal		Weibull	
	K-S	$\chi^2 (\times 10^{-6})$	K-S	$\chi^2 (\times 10^{-6})$
46001	0.02270	1.01465	0.04155	1.01859
46002	0.02113	0.91402	0.04693	0.91566
46003	0.02720	0.86907	0.03510	0.88058
46005	0.02652	0.92650	0.04303	0.92890
46006	0.03183	0.86031	0.04202	0.86336

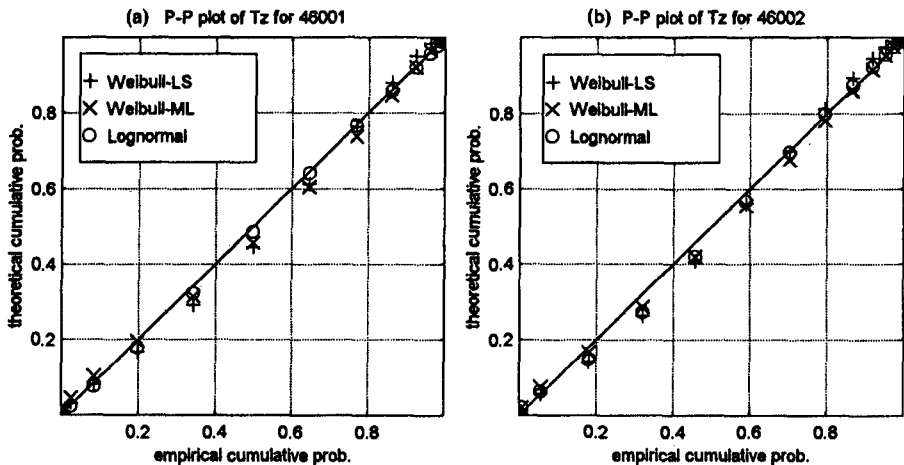


Figure 6. Comparisons of cumulative distribution for T_z between measured data and univariate distributions for (a) station 46001 and (b) station 46002

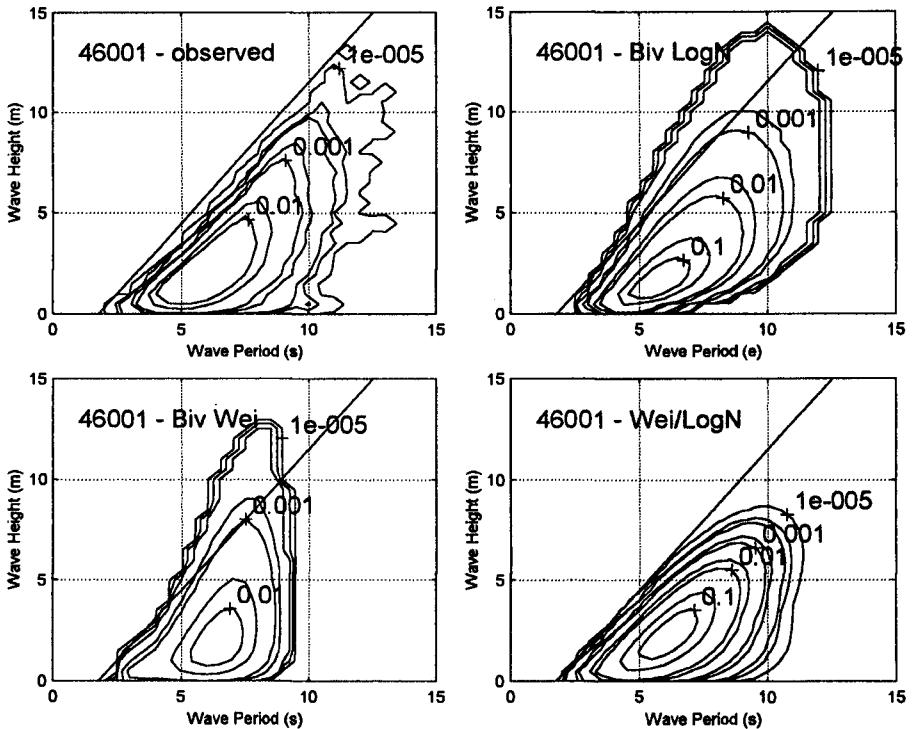


Figure 7. Contour plots of joint probability between H_s and T_z for (a) measured data, (b) bivariate log-normal distribution, (c) bivariate Weibull distribution, and (d) marginal Weibull and conditional log-normal distribution

The theoretical joint probability distributions are compared to the distribution of the observed data based on the correlation coefficients. For H_s , a correlation coefficient is computed for each 0.5-m wave height bin. For each wave height bin, the corresponding values of observed frequencies and frequencies calculated from the joint distributions across the 0.5-s T_z bins form a paired data set. A correlation analysis is then performed on the two paired data (i.e., observed frequencies versus calculated frequencies). The correlation coefficients for every 0.5-s T_z bin are computed similarly. Note the correlation coefficient at both the upper and lower ends may not be reliable due to limited data observations. Negative coefficient values show a polarity between the deviations of observed frequencies and the deviations of calculated frequencies. This shows a very poor fit between the measured data and the joint distribution. Figure 8 shows the correlation coefficients for all the H_s and T_z bins for all five stations. For H_s , it is clear that the marginal Weibull/conditional log-normal distribution fits better than other distributions. The bivariate Weibull distribution's performance is very poor and, therefore, unacceptable. All joint distributions have high correlation coefficients for the T_z fit. Note the bivariate Weibull distribution does not have correlation values for periods higher than 10 seconds because there are no calculated frequencies from which to compute the coefficients.

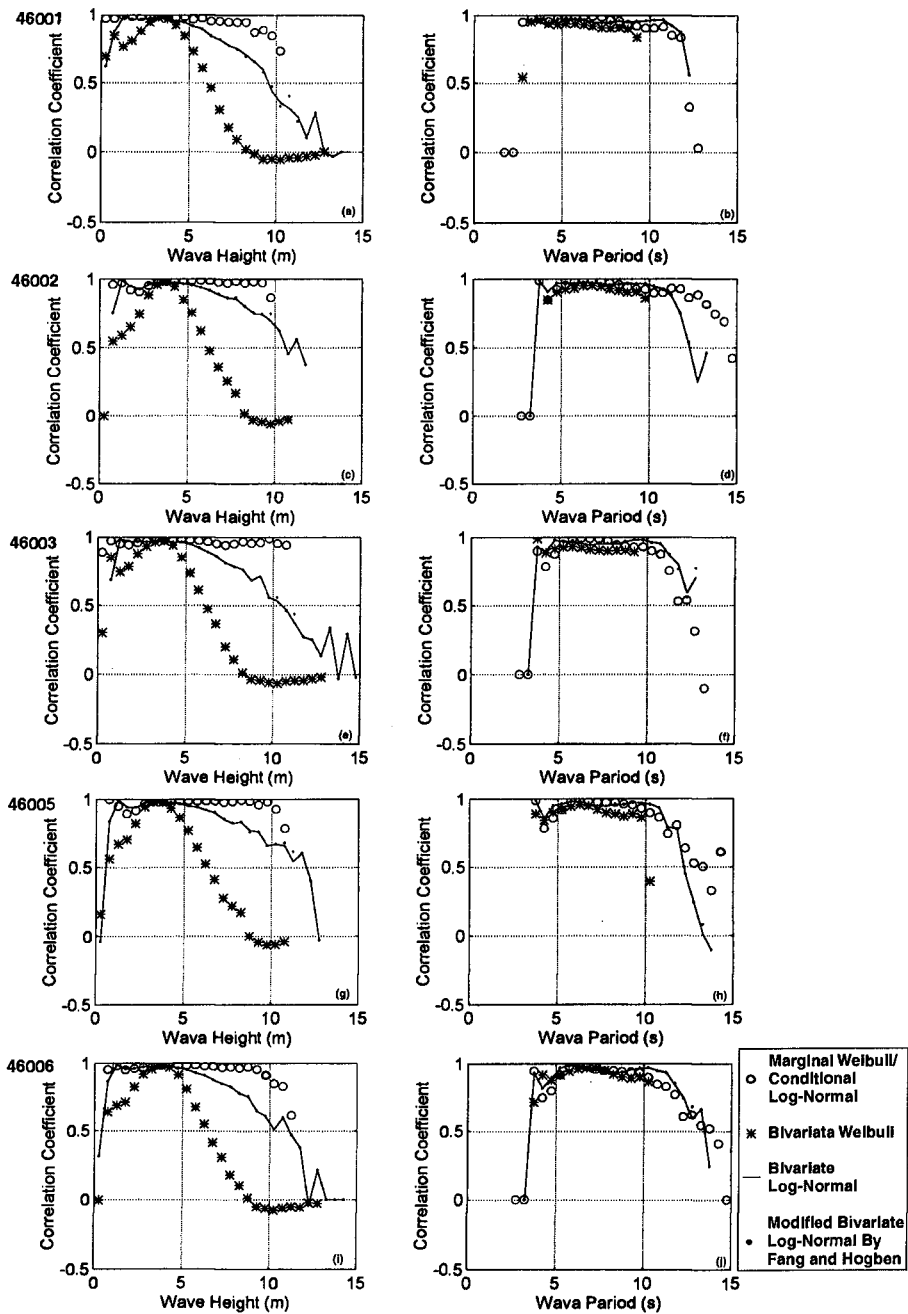


Figure 8. Coefficients of correlation between measured data and joint distributions for various H_s and T_z values for all five buoy stations

Conclusions

The results show that the log-normal distribution best fits both H_s and T_z for the five buoy stations. The modified log-normal distribution, proposed by Fang and Hogben (1982), significantly improves the fitting at the high H_s end and at the peak for some stations. For the three-parameter Weibull distribution, different methods of parameter estimation produce different parameters and distribution shapes. In general, the ML method is better than both the LS method and the method of moments. Although the Weibull distribution with parameters computed from the ML method fits the upper tail of the cumulative distribution for the H_s , it underpredicts both the probability peak and the probability density values at the high values.

Although the fit between various joint distributions and the measured data depends on the wave height and period range, the marginal Weibull/conditional log-normal distribution has the best overall fit based on the visual inspection and the correlation coefficients. Also, this joint distribution better estimates the data of the steeper sea states. Both the bivariate log-normal and the bivariate Weibull distributions overpredict the steeper sea states. The bivariate Weibull distribution also poorly fits the high wave period data.

References

- ATHANASSOULIS, G.A., SKARSOULIS, E.K., and BELIBASSAKIS, K.A., "Bivariate Distributions With Given Marginals With an Application to Wave Climate Description," *Applied Ocean Research*, Vol. 16, pp. 1-17, 1994.
- FANG, Z.S., and HOGBEN, N., "Analysis and Prediction of Long-Term Probability Distributions of Wave Heights and Periods," Technical Report, National Maritime Institute, London, 1982.
- MATHISEN, J., and BITNER-GREGERSEN, E., "Joint Distributions for Significant Wave Height and Wave Zero-Up-Crossing Period," *Applied Ocean Research*, Vol. 12, No. 2, pp. 93-103, 1990.
- OCHI, M.K., "Wave Statistics for the Design of Ships and Ocean Structures," *Transactions, Society of Naval Architects and Marine Engineers*, Vol. 86, pp. 47-76, 1978.
- OCHI, M.K., "New Approach for Estimating the Severest Sea State From Statistical Data," *Proceedings of International Conference on Coastal Engineering*, pp. 512-525, 1992.
- PALAO, I.M., TENG, C.C., TIMPE, G.L., and BROWN, D.A., "An Extremal Analysis of Significant Wave Height Measured From NDBC Data Buoys," Technical Report 1804-07.01, National Data Buoy Center, May 1994.
- TENG, C.C., TIMPE, G., and PALAO, I., "The Development of Design Waves and Wave Spectra for Use in Ocean Structure Design," *Transactions, Society of Naval Architects and Marine Engineers*, Volume 102, pp. 475-499, 1994.
- TUCKER, M.J., *Waves in Ocean Engineering*, Ellis Horwood, 431 pp., 1991.