Duration in Wave Climate Analysis

Rodney J. Sobey*and Leah S. Orloff†

Abstract

The relationship between sea state intensity, sea state duration and frequency is pursued in the context of wave data from five sites in U.S. waters. A rational methodology is presented for the interpolation and extrapolation of measured trends, based on extreme value series for intensity, given duration. The inevitable short duration data base problem is addressed by routine application of the triple annual maximum methodology. A format is suggested for IDF data preparation and presentation. Examples are given for wave climate data on the Pacific and Atlantic coasts of the United States.

Introduction

The long term history of wave conditions at a particular site exhibits similar variability to other climate variables such as atmospheric pressure, precipitation, stream flow, temperature and wind speed. Wave measurement programs have only recently become commonplace. Apart from the relatively short duration of available data sets, analysis techniques in wave climate have followed, where appropriate, the established practices for other climate variables.

Though by no means universal, a common trend in wave measurement practice is a twenty minute burst sample every hour. This is the practice, for example, of the National Data Buoy Center (NDBC), for the numerous surface buoys in its network throughout U.S. waters. Each burst sample provides a single wave climate intensity estimate, typically the significant wave height H_s.

Wave climate attention has often focused on frequency analyses of the intensity alone. Extreme value analyses, leading to intensity-frequency summaries, are relatively routine, except for extrapolation uncertainties that are a direct consequence of the short duration of existing wave data series.

*Professor, Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, U.S.A.

†Graduate Student, Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, U.S.A.

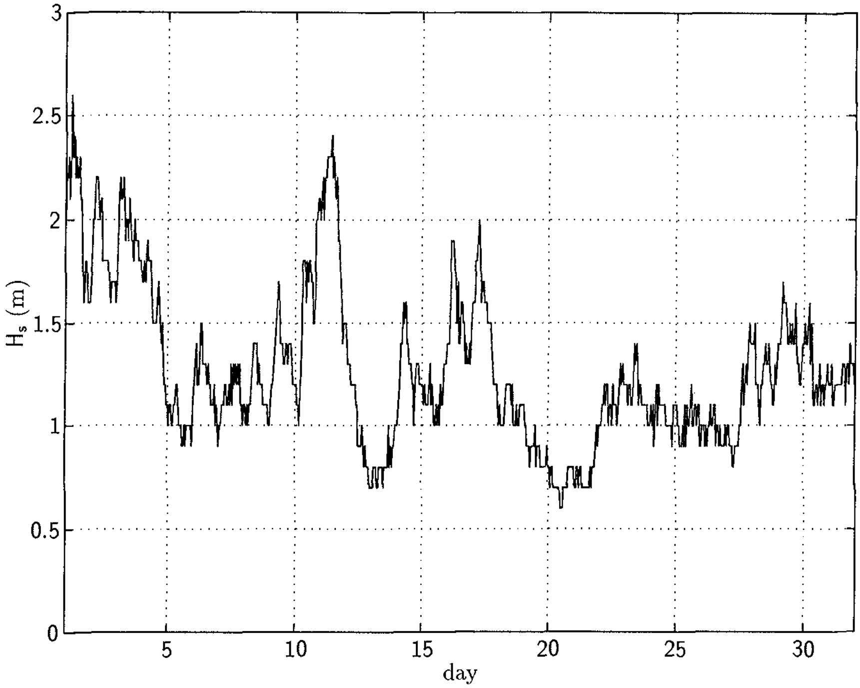


Figure 1: Historical wave climate record at NDBC Buoy 46026 off San Francisco, California for July, 1988.

But there is also a time scale to an historical record of a local wave climate; Figure 1, for example, shows a month of data from the NDBC buoy 46026 off San Francisco, California. The intensity is the significant wave height estimate from hourly burst samples. There is considerable interest in the duration or persistence of the sea state intensity, in applications ranging from the planning of engineering operations at sea to the closure of tidal inlets into coastal wetlands.

In a wave climate, both the intensity or wave height H and the duration D of sea state exceedance of this intensity are random variables. In most wave monitoring programs, burst-sample-averaged intensity is routinely characterized by the significant wave height; H throughout is identified as the significant wave height. Intensity-duration-frequency curves are an interpreted presentation of the joint distribution of intensity and duration, with CDF $F_{H,D}(h,d)$ and PDF $f_{H,D}(h,d)$. h and d are realizations of the random variables H and D respectively.

In surface water hydrology, it is a common practice to summarize precipitation records as IDF (intensity-duration-frequency) curves, and not just as intensity-frequency curves. An IDF-style presentation can be equally useful in wave climate characterization. In this paper, the relationship between sea state intensity, sea state duration and frequency will be pursued. Initial attention will be directed to the trends suggested by measured data. This leads to a rational methodology for the interpolation and extrapolation of measured trends and a suggested format for IDF data preparation and presentation.

Observational Data

Field data suitable for the extraction of coupled intensity-duration information must be able to resolve sea state durations. At a minimum, hourly observations of H seem necessary. Some historical data sets with 3 or 6 hours between burst samples cannot adequately resolve the expected range of sea state duration.

Given that the historical record has adequate time resolution, intensity and duration data may be extracted from the record in three different ways, as

(a). joint intensity-duration observations,

(b). duration, given intensity, observations, and

(c). intensity, given duration, observations.

Each of these approaches has intrinsic value. There is potentially useful information in a focus on both population and extreme value events. In particular, an extreme value analysis of data sets (b) or (c) is a basis of the IDF-style presentations.

The measured wave climate records used in the present study have been taken from the NOAA Marine Environmental Buoy Database (National Oceanic and Atmospheric Administration (1992) and supplements), available on CDROM. Details of the two sites on which attention has been focussed are listed in Table 1. Data gaps of one or two hours in these records were filled by linear interpolation. Exceedance intervals interrupted by longer gaps have been ignored.

Annual maximum series for duration given intensity that are extracted from this data often have no entries in some years for the higher intensities. Unfortunately, this identifies a major weakness of efforts to extrapolate the duration given intensity data to extreme events, like the 100 year event, that are well beyond the overall duration of the data record. Fortunately, the alternative extreme value presentation of this same data, as intensity given duration, largely circumvents this difficulty.

Site	Location	Water depth (m)	Record duration	Climate years	Maximum H_s (m)	Mear H_s (m)
San Francisco, CA NDBC 46026	37.75N, 122.82W	54.3	July 1982 -December 1994	12	7.6	1.7
Portland, ME NDBC 44007	43.53N, 70.14W	18.9	February 1981 -December 1994	13	7.3	0.9

Table 1: Wave Data Base

Intensity, given Duration, Observations

Extreme value analyses is based on direct extraction of sea state intensity at a given duration. A moving window search has been adopted to identify the largest intensity h that is continuously exceeded over the target duration. The search may be applied over any period that is sufficiently long to include several storm hydrographs; a week would generally be sufficient, a month would certainly be. The width of the moving window is the target duration d; values of 1, 2, 3, ... 12, 15, 18, ... 36, 42, 48, ... 72, 84, 90, ... 144 hours have routinely been adopted.

In the expectation of using either annual maximum series (AMS) and triple annual maximum series (TAMS), the moving window search has been applied over all periods of one calendar month in the historical record. For each target duration, this provides a data series of monthly maximum intensities that are maintained in the record for the target duration. These monthly maximum series are then organized into climate years and ranked in descending order of magnitude to define the TAMS series, which includes H_1, the series of annual maximum intensity, H_2, the series of the annual second largest monthly maximum intensity, and H_3, the series of the annual third largest monthly maximum intensity. Extreme value analyses may be based on these data series.

An immediately apparent feature of these observations is the sparsity of the data. The longer records in the entire NOAA data base have twelve to sixteen climate years. Reliable information is concentrated at average recurrence intervals of less than 5 years. Yet extrapolation to 100 year events is routinely expected.

The present study adopts a generic and very direct approach. There is no insistence on a single probability distribution and there is a direct focus on the inevitable interest in the longer recurrence intervals, extending to 100 years. Attention is directed to extreme value series and extreme value distributions.

Intensity, given Duration - Interpolation and Extrapolation of Extreme Value Series

While the existing wave climate literature has mostly focussed on the conditional distribution for duration given intensity, an equally appropriate representation of intensity-duration-frequency curves is available from the conditional distribution for intensity given duration, with CDF $F_H(h \mid d)$ and PDF $f_H(h \mid d)$. The NOAA observational data appears to be more sympathetic to this approach. It also retains the familiarity of intensity-frequency analyses.

Following the adoption of a specific distribution to aide interpolation and extrapolation, the distribution CDF and PDF become $F_H(h \mid d; p_1, p_2, \dots)$ and $f_H(h \mid d; p_1, p_2, \dots)$ respectively. The distribution parameters $p_1(d), p_2(d), \dots$ are dependent on the given intensity d. Once the distribution is established, intensity (h), duration (d) and frequency (the return period or average recurrence interval T_R) are related as

$$1 - \frac{\Delta t}{T_R} = F_H(h \mid d; p_1, p_2, \dots) \qquad (1)$$

in which Δt is the time interval of the observational data series to which $F_H(h \mid d; p_1, p_2, \dots)$ or $f_H(h \mid d; p_1, p_2, \dots)$ has been fitted. For AMS data, Δt is 1 year. An IDF-style presentation might have duration as the abscissa, intensity as the ordinate and frequency as the parameter of a family of curves. The PDF, CDF, mean and variance for a range of candidate two-parameter distributions have been listed in Table 2.

The classical approach would separately address the intensity data at each duration. Given an AMS series for intensity at that duration and the choice of the extreme value distribution (say from Table 2) to aide interpolation and extrapolation, the probability model would be fitted to the AMS data by probability plotting, by the method of moments, by the least squares method or by the method of maximum likelihood. Knowing the parameters of the distribution, the CDF is completely defined together with the preferred interpolation within the range of the data. The preferred extrapolation beyond the range of the data is also defined. Confidence bands on the extrapolation would finally be estimated, often following the Central Limit theorem or the Kite (1975) method of moments.

This classical approach however does not take full advantage of the data. The data sets are very short and AMS series can be supplemented through controlled recognition of marginally less extreme events. Partial duration series (the peak-over-threshold method) or triple annual maximum (TAMS) series have both been used to advantage in this context. The classical approach also does not acknowledge that data at different durations remain samples from the same wave climate. Using all the

Distribution	CDF PDF μ σ^2	Parameters
Extreme Value 1 (Gumbel)	$F_Y(y) = \exp[-\exp(-\frac{y-v}{u})]$ $f_Y(y) = \frac{1}{u}\exp[-\frac{y-v}{u} - \exp(-\frac{y-v}{u})]$ $\mu = \gamma u + v, \quad \gamma = 0.5772\ldots$ $\sigma^2 = \pi^2 u^2/6$	u, v
Extreme Value II	$F_Y(y) = \exp[-(\frac{u}{y})^\alpha]$ $f_Y(y) = \frac{\alpha}{u}(\frac{u}{y})^{\alpha+1}\exp[-(\frac{u}{y})^\alpha]$ $\mu = u\Gamma(1-1/\alpha)$ $\sigma^2 = u^2\Gamma(1-2/\alpha) - \mu^2$	α, u
Extreme Value 111 (Weibull)	$F_Y(y) = 1 - \exp[-(\frac{y}{u})^\alpha]$ $f_Y(y) = \frac{\alpha}{u}(\frac{y}{u})^{\alpha-1}\exp[-(\frac{y}{u})^\alpha]$ $\mu = u\Gamma(1+1/\alpha)$ $\sigma^2 = u^2\Gamma(1+2/\alpha) - \mu^2$	α, u
Log Normal	$F_Y(y) = \frac{1}{2}[1 + \mathrm{erf}(\frac{\ln y - \alpha}{2^{1/2}\beta})]$ $f_Y(y) = \frac{1}{y(2\pi)^{1/2}\beta}\exp[-\frac{(\ln y - \alpha)^2}{2\beta^2}]$ $\mu = \exp(\alpha + \beta^2/2)$ $\sigma^2 = (\exp(\beta^2) - 1)\exp(2\alpha + \beta^2)$	α, β

Table 2: Candidate Extreme Value Distributions for random variable Y

extreme value data together has the potential to provide a superior prediction of the IDF curves. The TAMS approach has been extended to this purpose.

A TAMS Methodology for Wave Climate IDF

Typical record durations from the NOAA data base identify a very familiar problem in wave climate analysis. The record duration, a decade or so, must be significantly extrapolated to provide estimates of the 50 and 100 year events that are expected from frequency analyses. For the NOAA data in Table 1, the AMS series have twelve or thirteen data points, sufficient for a reliable estimate of perhaps a 25 year event. And these are the long established stations in the data base.

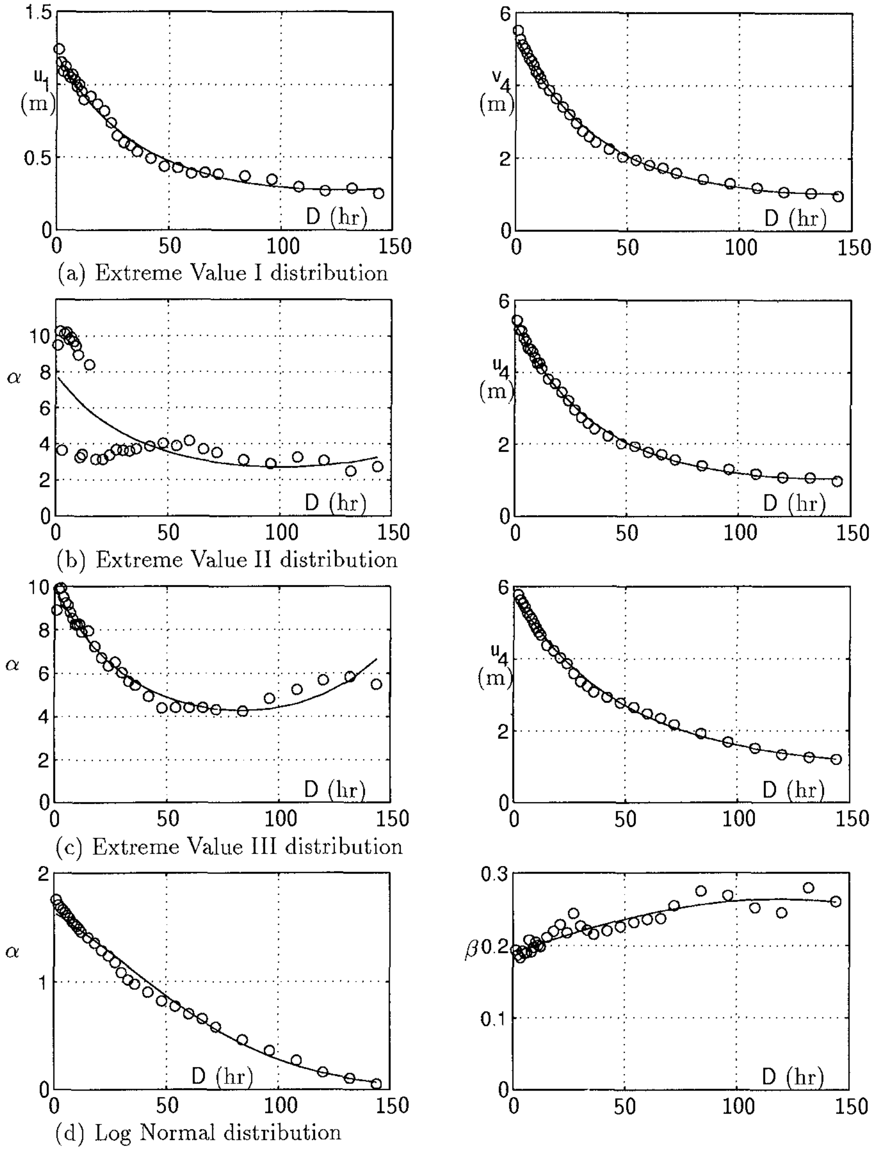
The extrapolation uncertainty can be mitigated by using additional data from the complete duration series, using either a partial duration series or a triple annual maximum series (Sobey and Orloff 1995). The latter approach has been adopted as the basis for the present analysis.

The results of an initial exploratory analysis, using the Sobey and Orloff algorithm on the TAMS data for each duration for the data at Portland, Maine, is shown as the markers in Figure 2. Each of the four Table 2 extreme value distributions has been TAMS-fitted to the data. The TAMS series extracted from the data at the different durations are not unrelated, being extracted from the same historical record. It is not surprising that some of the distribution parameters, such as u and v for Extreme Value I, follow a very consistent trend. Some data scatter is expected, and is an endemic problem of geophysical field data, especially from a relatively short duration record. Nevertheless, the declining probability levels of the longer duration events anticipates smoothly-evolving trends in the distribution parameters. For example, smoothed estimates of both u and v are monotonically decreasing with D in Figure 2a.

Figures 2a through d provide some visual measure of the relative acceptability of each of the four extreme value distributions. The Extreme Value I results (Figures a) look very plausible, the trends being monotonic and the data scatter relatively minor. The Extreme Value II results (Figures b) do not appear acceptable; the u trends is plausible but the jumps in the α response would seem to preclude this distribution as a candidate for data interpolation and extrapolation. The results for Extreme Value III (Figures c) are relatively smooth, except for the α response at large D, where also the response seems no longer monotonic. There is a suggestion that this distribution is also unacceptable. The Log Normal (Figures d) results are relatively encouraging. Both α and β are monotonically evolving, though the data scatter for β is moderately coarse. This result is perhaps acceptable, but seemingly less so than Extreme Value I for this data set.

The trends exhibited by both Table 1 data sets were roughly similar. On a visual basis, the Extreme Value II and III distributions were rejected as suitable candidates for interpolation and extrapolation. The Extreme Value I trends were relatively smooth and monotonic. The Log Normal β parameter trends were less smooth and not always monotonic. On this visual measure, the Extreme Value I distribution is preferred fir these data sets. But selecting the appropriate distribution for data interpolation and extrapolation is a subjective process. It must be guided by the trends of the particular data set and also by measures of analysis acceptability in addition to a purely heuristic interpretation. Data sets from other geographical sites may be very different.

Similar analyses for all Table 1 data sites were completed. Collectively, such figures and tables provide the information upon which a selection of an extreme value



(a) Extreme Value I distribution

(b) Extreme Value II distribution

(c) Extreme Value III distribution

(d) Log Normal distribution

Figure 2: TAMS-predicted distribution parameters for Table 2 Extreme Value distribution from NDBC Buoy 44007 off Portland, Maine. Markers are TAMS predictions at individual durations; solid line is the Equation 2 curve fit to these individual duration predictions.

Site	Distribution	p_i	a_i	b_i	c_i
San Francisco, CA	Extreme Value I	u	-0.0193	-0.0103	3.89e-05
NDBC 46006		v	1.52	-0.0111	3.85e-05
	Log Normal	α	0.459	-0.00855	2.12e-05
		β	-1.66	-0.00143	1.37e-05
Portland, MN	Extreme Value 1	u	0.191	-0.0234	9.29e-05
NDBC 44007		v	1.68	-0.0233	8.15e-05
	Log Normal	α	0.495	-0.00767	-0.000101
		β	-1.66	0.00547	-2.30e-05

Units of a_i, b_i and c_i assume D in hours; u and v in m; α and β are dimensionless.

Table 3: Parameters of Equation 2 curve fits to Extreme Value I and Log Normal distribution parameters.

distribution must be based. As with most extreme value analyses in the natural environment, this choice remains somewhat subjective.

IDF Summaries for US Waters

Given a choice of extreme value distribution, intensity-duration-frequency curves can be constructed.

Interpolation and extrapolation of results such as Figure 2 may be facilitated by smoothing the trends to the empirical curve

$$p_i(d) = \exp\left[a_i + b_i d + c_i d^2\right] \tag{2}$$

where p_i (i = 1,2) are the distribution parameters. There is no fundamental basis for such a relationship, except that it does reasonably follows the $p_i(d)$ trends. Least squares curve fits of Equation 2 to the data in Figure 2 and similar for other sites define the parameters a_i, b_i and c_i. These curve fits for all four of the Table 2 distributions are included as the solid lines on Figure 2 for the Portland, Maine site. Parameters for Extreme Value I and Log Normal at all the Table 1 sites are listed on Table 3. These distributions seem to be the most plausible for these sites, though this may not be the case for other sites.

Given an extreme value distribution and the p_i parameters as a smooth function of duration D, an IDF curve can be constructed for a given average recurrence interval T_R from solutions to Equation 1.

$$1 - \frac{\Delta t}{T_R} = F_H\left(h|d; p_1(d), p_2(d)\right) \tag{3}$$

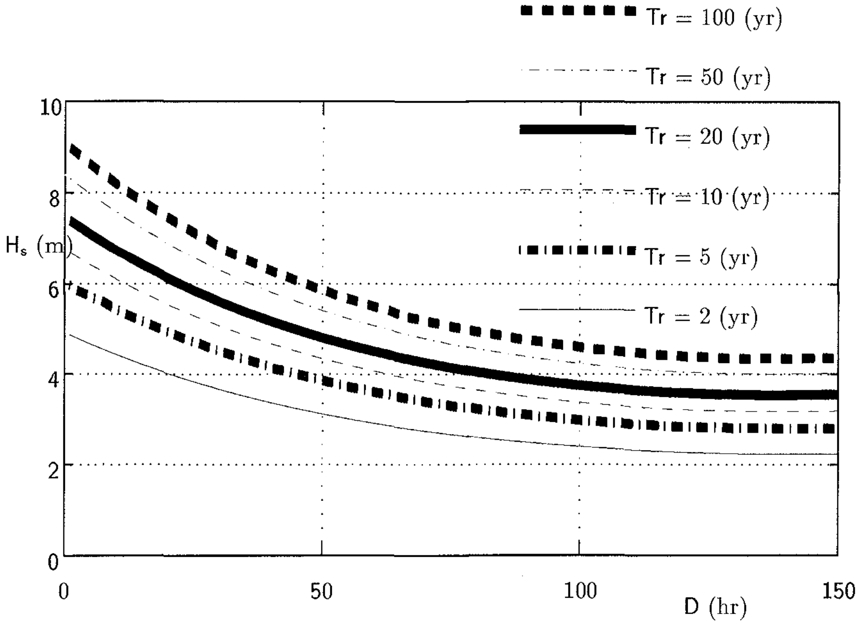


Figure 3: Extreme Value I Prediction for IDF curves at NDBC Buoy 46026 off San Francisco, California.

This is an implicit algebraic equation in $h(d; T_R)$, which can be solved for each duration by standard numerical algorithms such as Newton-Raphson, regula falsi or the secant method. As a result of the smoothing implicit in the Equation 2 curve fits, these solutions will also be smoothly varying. Completing these solutions for a range of appropriate average recurrence intervals, typically 2, 5, 10, 20, 50 and 100 years provides the classical IDF presentation.

IDF predictions from the Extreme Value I and Log Normal distributions for the San Francisco, California and Portland, Maine sites are shown as Figures 3 through 6. These predictions show excellent trend agreement but only moderate magnitude agreement. More can not be expected, these figures once again demonstrating the fragility of extreme value analyses from short duration data.

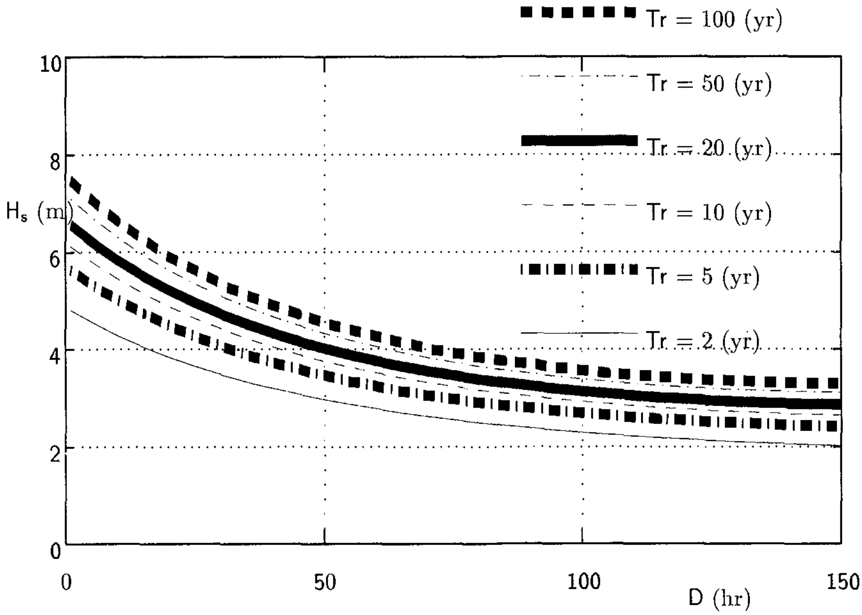


Figure 4: Log Normal Prediction for IDF curves at NDBC Buoy 46026 off San Francisco, California.

Conclusions

The relationship between sea state intensity, sea state duration and frequency is pursued in the context of wave data from the NOAA Marine Environmental Buoy Database at two sites in U.S. waters.

Detailed attention has been given to extreme events, where the popular approach has been a focus on duration, given intensity. Existing studies have mostly utilized wave data from the European and Japanese waters. Analyses have mostly been based on an assumption that the conditional distribution of the population of duration, given intensity, follows the Weibull distribution. This distribution does not seem especially appropriate in U.S. waters.

The present approach is a direct focus on extreme events, through extreme value data extracted from observational records. IDF (intensity-duration-frequency) summaries may originate from the conditional extreme value distribution for duration,

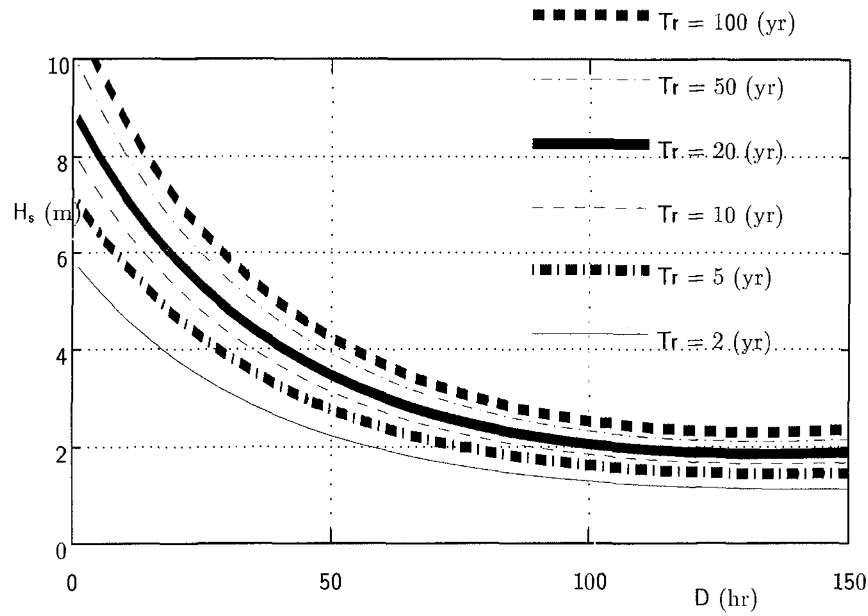


Figure 5: Extreme Value I Prediction for IDF curves at NDBC Buoy 44007 off Portland, Maine.

given intensity, or from the conditional extreme value distribution for intensity, given duration. Extracting the data as duration, given intensity, results in significant data gaps for the more extreme intensities. Intensity, given duration, data appears to be much less problematic.

A rational methodology is presented for the interpolation and extrapolation of measured trends, based on extreme value series for intensity, given duration. The inevitable short duration data base problem is addressed by routine application of the triple annual maximum methodology.

A format is suggested for IDF data preparation and presentation. Examples are given for wave climate data on the Pacific and Atlantic coasts of the United States.

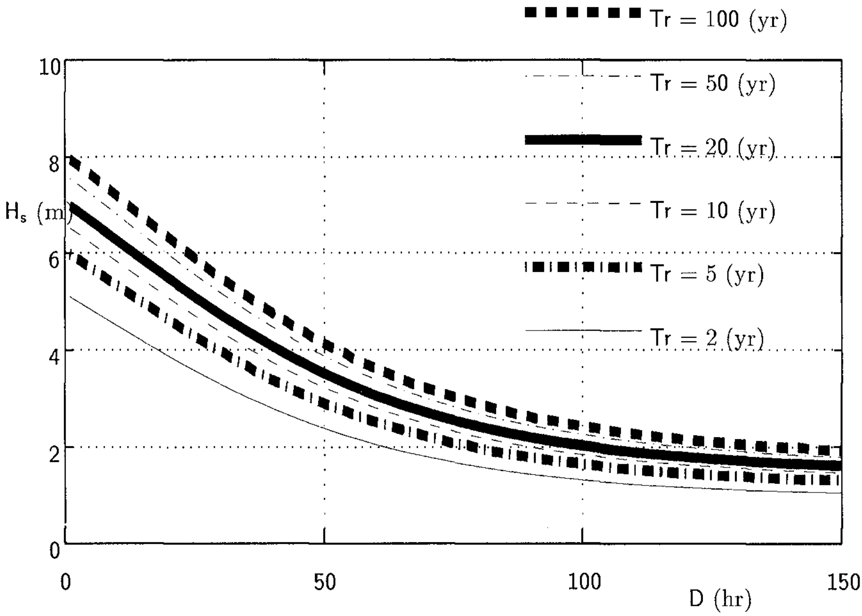


Figure 6: Log Normal Prediction for IDF curves at NDBC Buoy 44007 off Portland, Maine.

Acknowledgements

This paper is funded in part by a grant from the National Sea Grant College Program, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, under grant number NOAA NA66RG0477, project number R/OE-36 through the California Sea Grant College System, and in part by the California State Resources Agency. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its sub-agencies. The U.S. Government is authorized to reproduce and distribute for governmental purposes.

Support through a U.S. Environmental Protection Agency STAR Fellowship to Leah S. Orloff is also acknowledged.

References

Kite, G. W. (1975). Confidence limits for design events. *Water Resources Research 11*, 48–53.

National Oceanic and Atmospheric Administration (1992). NOAA marine environmental database. CD-ROM NODC-32. Updates NODC-38, -39, ...

Sobey, R. J. and L. S. Orloff (1995). Triple annual maximum series in wave climate analysis. *Coastal Engineering 26*, 135–151.