THE EFFECT OF ARCHIVING INTERVAL OF HINDCAST OR MEASURED WAVE INFORMATION ON THE ESTIMATES OF EXTREME WAVE HEIGHTS

E. P. Berek

Metocean, Coastal, and Offshore Technologies, LLC

1. Introduction

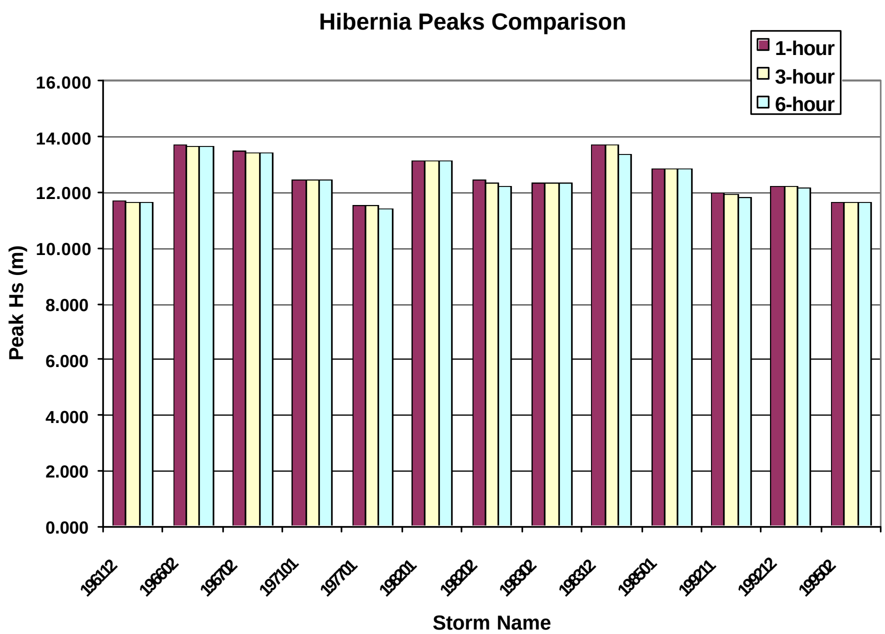
This paper summarizes the results of an investigation into the effect of significant wave height archiving or measurement interval on the determination of extreme value statistics from both hindcast and measured wave data. The archiving interval considered in this study is the interval between archived values of hindcast or between measured significant wave heights; it is not the sampling variability found in the spectral analysis used to determine the archived significant wave heights. The work was motivated primarily by questions that arise in hindcasting relative to the adequacy of the archival interval with respect to the use of the hindcast results in deriving extreme value estimates of significant wave height.

For the hindcast comparison two grid points in the AES-40 hindcast study located near the Hibernia Development on the Grand Banks of Newfoundland were chosen. For the measured data comparisons five data sets with recording periods ranging from 13 to 20 years were examined. In all cases the peaks-over-threshold method of extreme value analysis was used. The Gumbel distribution of extremes was the primary focus, however, the 3-parameter Weibull distribution was also considered in the measured data set analysis.

2. The Hindcast Comparisons

For the study of the effect of archiving interval on extremes derived from hindcast data the premise is that the use of hindcast results taken at longer intervals (3 hours, 6 hours, etc.) will inherently result in a negative bias in extreme value estimates compared to using 1 hour intervals since it is obvious that at least some peak events will be missed in using the longer intervals Although the core AES-40 study archived data at 6-hour intervals, Oceanweather performed a small follow-on study wherein the months containing the most severe Hibernia events were hindcast with the results being saved hourly. Nothing else was changed in the hindcast methodology.

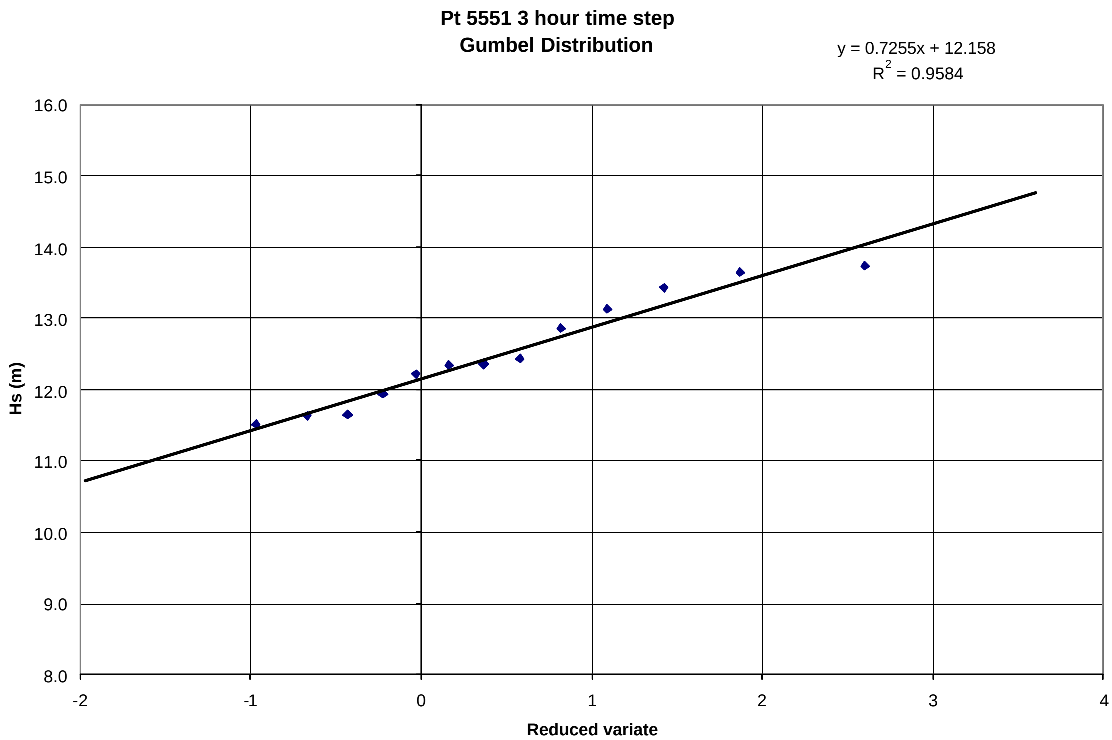
The following plot and table show the 1-hour, 3-hour, and 6-hour peak results for the 13 largest hindcast storms at one of the grid points, point 5551. It is noted that there are not large differences in the peak values as a function of these three archiving intervals with the maximum difference between the one hour and six hour being 0.35 meters, or 2.5%, for the peak in the December 1983 storm.

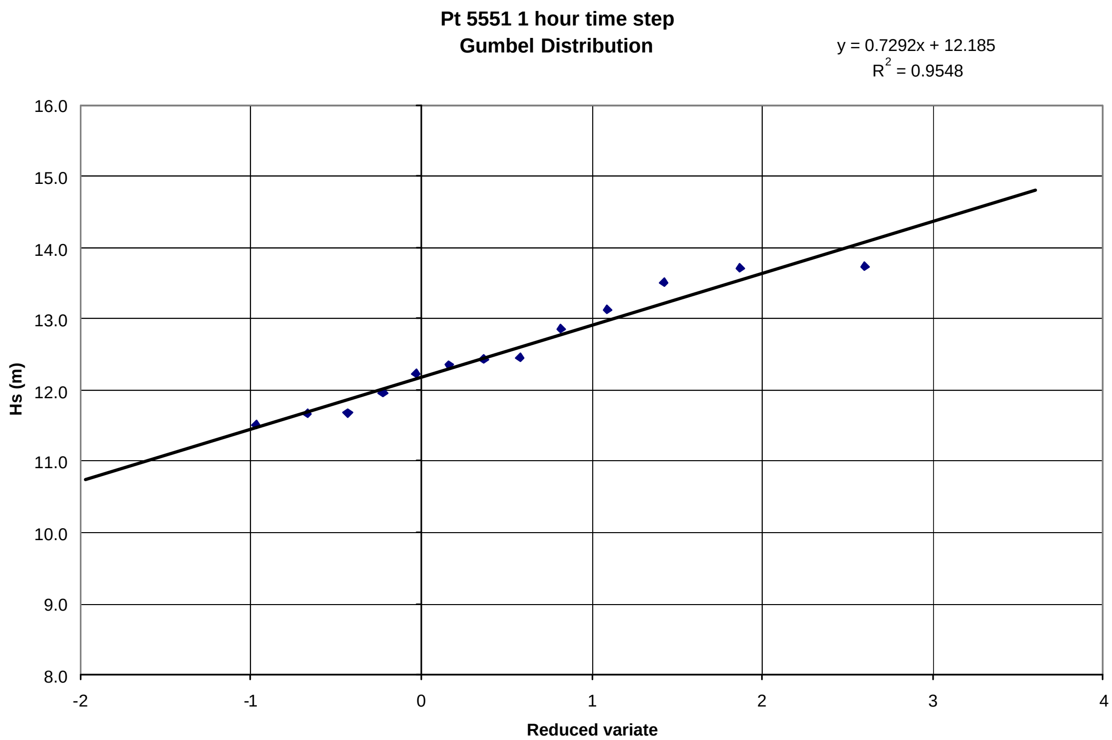


Storm	1-hour	3-hour	6-hour		1hr - 6hr	% diff		1hr - 3hr	% diff
196112	11.685	11.638	11.638		0.047	0.402		0.047	0.402
196602	13.707	13.637	13.637		0.070	0.511		0.070	0.511
196702	13.506	13.427	13.427		0.079	0.585		0.079	0.585
197101	12.426	12.426	12.426		0.000	0.000		0.000	0.000
197701	11.513	11.508	11.427		0.086	0.747		0.005	0.043
198201	13.122	13.122	13.122		0.000	0.000		0.000	0.000
198202	12.448	12.356	12.229		0.219	1.759		0.092	0.739
198302	12.353	12.341	12.320		0.033	0.267		0.012	0.097
198312	13.736	13.736	13.390		0.346	2.519		0.000	0.000
198501	12.854	12.854	12.854		0.000	0.000		0.000	0.000
199211	11.960	11.931	11.842		0.118	0.987		0.029	0.242
199212	12.234	12.217	12.154		0.080	0.654		0.017	0.139
199502	11.665	11.644	11.644		0.021	0.180		0.021	0.180

Table 1. AES-40 Peak Comparisons between 1, 3, and 6-hour Archiving Intervals.

These hindcast results were analyzed to obtain extreme value estimates. The following three figures are the Gumbel distribution plots for the three sampling intervals at point 5551. The lines and equations on the figures are the linear-least-squares fits.





The following information is a tabulation of the extreme value analysis of the hindcast results at point 5551. There are two estimates for each return period; the first, found in the column labeled "Hs – MoM" is the method of moments fit, and the second, labeled "Hs", is a linear-least-squares fit. The confidence limits are on the MoM estimates.

		Point 5551				Point 5551				Point 5551		
		6-hour				**3-hour**				**1-hour**		
N (years)	Hs (m) MoM	Hs(m) LLS	Hs lower (m)	H upper (m)	Hs (m) MoM	Hs(m) LLS	Hs lower (m)	H upper (m)	Hs (m) MoM	Hs(m) LLS	Hs lower (m)	H upper (m)
5	12.14	12.16	11.85	12.51	12.18	12.21	11.89	12.57	12.21	12.23	11.92	12.60
10	12.83	12.72	12.39	13.42	12.89	12.78	12.45	13.50	12.92	12.81	12.48	13.53
25	13.57	13.32	12.85	14.52	13.65	13.40	12.92	14.63	13.69	13.43	12.95	14.67
50	14.09	13.74	13.15	15.32	14.19	13.83	13.23	15.45	14.23	13.87	13.26	15.50
100	14.61	14.16	13.44	16.10	14.72	14.26	13.53	16.25	14.76	14.30	13.56	16.31

Table 2. Extreme Value Estimates Near Hibernia from AES-40 Results Using Different Archiving Intervals.

The results are summarized below:

Return Period (years)	3-hour % Bias (MoM)	6-hour % Bias (MoM)
5	- 0.25	- 0.58
10	- 0.23	- 0.70
25	- 0.29	- 0.88
50	- 0.28	- 0.99
100	- 0.27	- 1.02

The key point is that the maximum bias using the 6 hour time step found in this sample is only 0.15 meters, or 1.02 percent, at the100 year return period level. For the 3 hour time step the maximum bias is only 0.04 meters, or 0.27 percent at the 100-year return period. **These differences are smaller than expected and actually significantly smaller than the differences in the results when using the different extremal analysis techniques, i.e., MoM and LLS**. Obviously the small differences in peak significant wave height lead to the small differences in the extreme value estimates. Possible explanations for the small differences in the peak wave heights include 1.) the storms do not have particularly sharp peaks, and 2.) the limited number of storms, 13, is perhaps too small to see significant differences in the extreme value estimates. **It is imperative to reiterate that the hindcasting interval is the only change in the 1-hour, 3-hour, and 6-hour results – no changes have been made to the wind modelling frequency, nor to the wave modelling procedure.**

3. The Measured Data Sets

The five measured data sets are summarized as follows:

Data Set	Water Depth (m)	Physical Location	Years of record
Magnus	186	North Sea	15
MEDS 016	168	Off St. John’s Newfoundland	16
MEDS 103	40	Off Vancouver Island	15
NDBC 44004	3164	Off Cape May, NJ	20
C46004	3600	Off British Columbia Coast	13

Table 3. The Measured Data Locations.

The Magnus data was collected with an EMI laser water surface elevation meter on the Magnus oil platform operated by BP in the UK sector of the North Sea. It is the UK's most northerly field, located approximately 160 kilometers NE of the Shetland Islands. This location is known to be subject to numerous strong storm events annually and it is relatively far offshore and unprotected from these storms.

MEDS 016, although in relatively moderate water depth, 168 meters, in many ways is a near coastal site. It is located east of St. John's, Newfoundland, and is only about 35 kilometers offshore. Since it is on the eastern side of Newfoundland, this site is not subject to storms from the west.

MEDS 103, is a very shallow coastal site offshore Vancouver Island. Although it is a shallow site this location is exposed to storms from the west, the prevailing storm direction.

NDBC 44004 and C46004 are both data buoys located in the deep ocean, NDBC 44004 being located in the Atlantic Ocean off the East Coast of the United States (Cape May, NJ), and C46004 being located off the West Coast of Canada (Vancouver, BC). The NDBC 44044 data was found to include the so-called "Storm of the Century" and since that storm is much larger than the next highest event, the extreme value results for this data set were greatly affected. In order to not muddle the archival rate effect being investigated in this work the outlier point was eliminated from the data set. (Eliminating ill-fitting data, particularly peak events in a data set, is not being advocated; it is imperative that points like this be considered in any analysis aimed at defining actual design criteria for a site.)

4. Extreme Value Analysis Approach

Since the goal of this investigation was to determine the effect of recording interval on extreme value estimates, it was important to limit the number of other sources of variation in extremes, e.g., probability distributions used, threshold analysis versus cumulative probability analysis, storm threshold, etc. The procedure used herein was to fit the Gumbel distribution, using the method of moments, to all storm peaks exceeding a threshold value of 0.5 or 0.55 times the highest peak significant wave height of the storms in the sample. It was later decided to add the 3-parameter Weibull distribution to the analysis to see if the sampling frequency is more important or less important when the Weibull distribution is used instead of the Gumbel distribution. **These choices are all subject to debate; they were simply made in this study in order to not cloud the archiving interval rate issue and because this approach is used in some circles in deriving extreme value criteria.**

The initial comparisons were made using unsmoothed data, i.e., measured data values. These measurements are typically 18 to 34 -minute samples taken at hourly intervals. To obtain 3-hour values, the measured observations at 0, 03, 06, 09, 12, 15, 18, and 2100 hours were considered. The 6-hour values were obtained from the 0, 06, 12, and 1800 hour observations. A three-point running average filter was later applied to obtain smoothed storm significant wave height values as this is an approach commonly used in validation studies comparing hindcast results to measured data.

5. Results

The results are summarized in the following table 4. For the Gumbel method of moments approach the decrease in the 100-year Hs estimate in going from the 1-hour archiving rate to the 3-hour rate for the measured data sets ranges from 4 to 7.4%. Going to the 6-hour archival rate decreases the 100-year values further, as expected. The range of decrease in the 100-year value using the 6-hour archival rate relative to the 1-hour is from about 5.7% to 17%.

Smoothing the data reduced the effect of the archival interval in all cases, as was expected. The reduction range for the 1-hour to 3-hour archived results is 0.45% to 3%. For the 1-hour to the 6-hour interval the reduction range is from 1.5% to 6.5%.

Using the 3-parameter Weibull distribution yields less consistent results, but the trend is for the archiving interval effect to have a greater effect than what is seen in the Gumbel results. This is an expected result since the Weibull distribution fits the data with three parameters rather than two and is therefore more affected by changes in individual data points. For the unsmoothed data the reduction range for the 1-hour to 3-hour archived results is 3.3% to 32.8%. For the 1-hour to the 6-hour interval the reduction range is from 5.9% to 37.1%.

		Magnus		**MEDS 016 1982-98**		**MEDS 103**		**44004 (w/o SOC)**		**46004**	
		NY = 15	**? = 9.533**	**NY = 16**	**? = 3.6875**	**NY = 15**	**? = 7.267**	**NY = 20**	**? = 3.2**	**NY = 13**	**? = 8.769**
UNSMOOTHED	1 hr data mean, std dev	10.005	2.9090	7.553	1.6910	6.753	0.6570	7.430	0.9935	8.791	2.558
	3 hr data mean, std dev	9.579	2.4065	7.183	1.3616	6.393	0.6587	7.339	0.8810	8.411	2.272
	6 hr data mean, std dev	9.333	2.4708	6.932	1.1031	6.186	0.6762	7.274	0.8521	8.317	2.130
		Hs	?Hs/Hs(1hr)	Hs	?Hs/Hs(1hr)	Hs	?Hs/Hs(1hr)	Hs	?Hs/Hs(1hr)	Hs	?Hs/Hs(1hr)
	100 yr 1 hr Gumbel	18.751		13.439		10.79		12.40		16.95	
	100 yr 3 hr Gumbel	17.506	-6.64%	12.441	-7.43%	10.23	-5.11%	11.91	-3.95%	16.19	-4.48%
	100 yr 6 hr Gumbel	17.327	-7.59%	11.155	-17.00%	10.06	-6.80%	11.69	-5.72%	15.74	-7.14%
	100 yr 1 hr Weibull	18.652		16.530		11.48		12.47		19.74	
	100 yr 3 hr Weibull	15.664	-16.02%	11.105	-32.82%	11.10	-3.31%	11.95	-4.17%	18.56	-5.98%
	100 yr 6 hr Weibull	17.556	-5.88%	10.390	-37.14%	9.93	-13.50%	11.44	-8.26%	16.31	-17.38%
	1 yr 1 hr Gumbel	12.278		8.212		7.55		8.62		10.82	
	1 yr 3 hr Gumbel	11.611	-5.43%	7.735	-5.81%	7.16	-5.17%	8.33	-3.36%	10.44	-3.51%
	1 yr 6 hr Gumbel	11.344	-7.61%	6.768	-17.58%	6.94	-8.08%	8.14	-5.57%	10.14	-6.28%
	1 yr 1 hr Weibull	12.163		8.096		7.69		8.59		10.65	
	1 yr 3 hr Weibull	11.412	-6.17%	7.682	-5.11%	7.27	-5.46%	8.28	-3.60%	10.30	-3.29%
	1 yr 6 hr Weibull	11.355	-6.64%	7.69	-5.01%	6.90	-10.27%	8.03	-6.52%	9.74	-8.54%
SMOOTHED	1 hr data mean, std dev	9.602	2.545	7.181	1.458	6.433	0.6511	7.309	0.815	8.313	2.015
	3 hr data mean, std dev	9.475	2.473	6.960	1.375	6.270	0.6523	7.271	0.803	8.309	2.015
	6 hr data mean, std dev	9.242	2.446	6.773	1.246	6.113	0.6589	7.196	0.789	8.332	2.000
		Hs	?Hs/Hs(1hr)	Hs	?Hs/Hs(1hr)	Hs	?Hs/Hs(1hr)	Hs	?Hs/Hs(1hr)	Hs	?Hs/Hs(1hr)
	100 yr 1 hr Gumbel	17.746		12.659		10.33		11.70		15.65	
	100 yr 3 hr Gumbel	17.488	-1.45%	12.280	-2.99%	10.10	-2.23%	11.59	-0.94%	15.58	-0.45%
	100 yr 6 hr Gumbel	17.233	-2.89%	11.838	-6.49%	9.87	-4.45%	11.44	-2.22%	15.41	-1.53%
	100 yr 1 hr Weibull							12.56		16.51	
	100 yr 3 hr Weibull							12.04	-4.14%	16.12	-2.36%
	100 yr 6 hr Weibull							11.85	-5.65%	15.41	-6.66%
	1 yr 1 hr Gumbel	11.681		7.812		7.20		8.25		10.24	
	1 yr 3 hr Gumbel	11.506	-1.50%	7.573	-3.06%	7.02	-2.50%	8.15	-1.21%	10.15	-0.88%
	1 yr 6 hr Gumbel	11.289	-3.36%	7.356	-5.84%	6.84	-5.00%	8.02	-3.30%	9.95	-2.83%
	1 yr 1 hr Weibull							8.30		10.02	
	1 yr 3 hr Weibull							8.13	-2.05%	9.91	-1.10%
	1 yr 6 hr Weibull							7.99	-3.73%	9.56	-4.59%

Table 4. Summary of Extreme Value Estimates for the Five Measured Data Sets.

Table 4 also includes the mean and standard deviation of the data sets used in the extreme value analysis. As would be expected the mean values always decrease in going from the 1-hour to the 3-hour and 6-hour data sets since less frequent sampling leads to peaks being missed. The changes in standard deviation are not consistent. For the unsmoothed data, at three of the sites (MEDS 016, NDBC 44044, and NDBC 46004) the standard deviations decrease in going from 1-hour to 3-hour, and decrease even further in going to 6-hour; however the MEDS 103 data shows a consistent increase in standard deviation, while the Magnus data shows a decrease in going from 1-hour to 3-hour but a smaller decrease in going from the 1-hour to the 6-hour.

Most of the large differences seen in the results occur in the Weibull analyses. Clearly the Weibull fitting for the unsmoothed Magnus data is problematic since there is no sensible reason for the 3-hour interval 100 year return period value being lower than the 6-hour 100 year return period value. The 100 year Weibull results at MEDS103 also appear suspect since the difference between the 3-hour and 6-hour estimates are so dramatic but the source of this discrepancy is not obvious. For the 46004 buoy the very large difference between the Gumbel and Weibull 1-hour 100 year values (16.95 versus 19.74, an increase of 16%) is also unrealistic.

The results of MEDS016 are arguably the most problematic. The Weibull results again appear to be unreasonably high with the 1-hour 100 year value of 16.53 meters versus the Gumbel value of 13.44, a difference of nearly 23%. The bigger problem, not apparent at the other sites, is the very large drop in the 6-hour 100-year Gumbel estimate (17%). Looking at the sample standard deviation a large decrease is noted in going from the 1-hour sample to the 6-hour sample; 1.69 to 1.10. It is possible, though not certain, that this could be the result of data sampling variability in the actual measurements since the waverider buoy only collects a 20 minute record. If the sampling variability is about 10%, then the hourly values can be as much as 10% too high or 10% too low, but when the archiving interval is greater than one hour only the peaks that are too high are selected for the extreme value analysis, thus biasing the sample.

If the MEDS106 results are not considered representative, then the difference in the 1-hour versus 3-hour and 6-hour archiving interval is likely in the range of 6 to 8% in the 100-year return period value.

5. Words of Caution

All of these results need to be evaluated within the context of the procedure followed in this study. For example, the AES-40 study methodology was not changed in order to generate a more accurate hourly estimate by generating hourly wind fields; all that was changed was the archiving interval. **Obviously changing the wind modelling approach to generate wind fields at intervals less than the 6-hour synoptic interval would change the results significantly.** The use of different extreme value techniques would also change the magnitude of the changes in the extreme value estimates; the changes quoted in this paper will not be exactly the same if different techniques are used (this point is illustrated by the differences between the Gumbel and Weibull results, for example). **The results of this study do not lead to simple factors that can be applied to extreme value estimates that can be used to effectively change 6-hour to 1-hour results, for example**.

6. References

Cox, Andrew T. and Val R. Swail, 2001: A global wave hindcast over the period 1958-1997: Validation and climate assessment. *Journal of Geophys. Research*, **106** (C2), 2313-2329.

Earle, M.D. and L. Baer, 1982: Effects of Uncertainties on Extreme Wave Heights. *Jour. Of Waterway, Port, Coastal, and Ocean Engineering, ASCE*, **108** (5), 216-225.

Forristall, G.Z., J.C. Heideman, I.M. Leggett, B. Roskam and L. Vanderschuren, 1996: Effect of Sampling Variability on Hindcast and Measured Wave Heights," *Jour. Of Waterway, Port, Coastal, and Ocean Engineering, ASCE*, **122**, (5), 216-225.

Gumbel, E.J. 1958: *Statistics of Extremes*. Columbia University Press, New York, 375 pp.

Haring, R.E. and J. C. Heideman, 1980: Gulf of Mexico rare wave return periods. *J. Petrol. Tech*., 35-37.

Haring, R.E., A.R. Osborne, and L.P. Spencer. 1976: Extreme wave parameters based on continental shelf storm wave records. *Proceedings of the 15th Coastal Engineering Conference, Honolulu, July 11-17, 1976*.

Swail, V.R., V.J. Cardone and A.T. Cox, 1998: A Long Term North Atlantic Wave Hindcast. Proc. *5th International Workshop on Wave Hindcasting and Forecasting, Melbourne, FL, January 26-30, 1998*.

Swail, V.R. and A.T. Cox, 2000. On the use of NCEP/NCAR reanalysis surface marine wind fields for a long term North Atlantic wave hindcast. *J. Atmos. Ocean. Technol.*, **17**, 532-545.