I.O.S.

RETURN WAVE HEIGHTS OFF SOUTH UIST ESTIMATED FROM SEVEN YEARS OF DATA

BY

B.R. STANTON

REPORT NO. 164

1984

NATURAL ENVIRONMENT RESEARCH COUNCIL

INSTITUTE OF OCEANOGRAPHIC SCIENCES

INSTITUTE OF OCEANOGRAPHIC SCIENCES

Wormley, Godalming,
Surrey, GU8 5UB.
(0428 - 79 - 4141)

(Director: Dr. A.S. Laughton FRS)

Bidston Observatory,
Birkenhead,
Merseyside, L43 7RA.
(051 - 653 - 8633)

(Assistant Director: Dr. D.E. Cartwright)

Crossway,
Taunton,
Somerset, TA1 2DW.
(0823 - 86211)

(Assistant Director: M.J. Tucker)

When citing this document in a bibliography the reference should be given as

STANTON, B.R. **1984** Return wave heights off South Uist estimated from seven years of data. *Institute of Oceanographic Sciences, Report*, No. 164, [56pp.]

INSTITUTE OF OCEANOGRAPHIC SCIENCES

WORMLEY

Return wave heights off South Uist estimated from seven years of data

by

*B.R. Stanton

I.O.S. Report No. 164

1984

* *Present address:*
New Zealand Oceanographic Institute
P.O. Box 12-346
Wellington North
New Zealand

This report is part of a continuing study at IOS of questions relating to the statistics of ocean waves including their extremes, and as such its conclusions do not represent an agreed IOS view.

CONTENTS

1. INTRODUCTION

Waves have been recorded off the island of South Uist in the Outer Hebrides since March 1976 in a continuing programme of wave climate studies. This report is principally concerned with the data collected at the Offshore site over the seven year period up to February 1983. This site is 15 km west of South Uist in a water depth of 42m. Some use has been made of the data collected for shorter periods at the other South Uist sites, in particular the Deepwater site some 16 km further offshore i.e. West from the Offshore buoy. Details of the South Uist sites are given in Table 1. All positions and depths are approximate as minor variations occur when buoys are replaced on service visits.

All wave measurements have been made using Datawell Waverider buoys. Details of individual buoy deployment and calibrations are given in Fortnum (1981) and a series of internal reports (R. Gleason and J.A. Crabb - personal communication).

The Waverider buoy measures the vertical acceleration of the water surface which is then integrated twice to give the elevation relative to the mean water surface. The signal is telemetered to a shore station and digitised every half second. Each data record gives a sample length of about 17 minutes and successive records start at three hourly intervals. A fast fourier transform is performed on 2048 data points to produce a wave spectum every three hours. Details of the method used including response corrections and data validation are given in Fortnum, Humphery and Pitt (1979). The significant wave height, H_s, and the zero-up-crossing wave period, T_z are calculated from the sample spectral moment, m_n, using the relations

$$H_s = 4\,(m_o)^{1/2}$$

$$T_z = (m_o/m_2)^{1/2}$$

where $m_n = {}_o\int^{\infty} f^n E(f)\,df$

and E(f) is the spectral density at frequency f.

A recent comprehensive review of the statistical parameters and techniques used in estimating return wave heights has been given by Carter and Challenor (1981a), so these will only be summarised here. The significant wave height, H_s, being proportional to the standard deviation of the sea surface elevation, is a measure of the roughness and is generally taken to be representative of the three hour period during which the sample record was taken. Assuming a statistically stationary sea surface during a period p, the most likely maximum zero-up-crossing wave height $H_{max,p}$ is given approximately by

$$H_{max,p} = H_s \left(\frac{\ln Nz}{2}\right)^{1/2}$$

where N_z is the number of zero-up-crossing waves in the period p i.e. $N_z = P/T_z$. Implicit in this equation is the assumption of a reasonably narrow banded spectrum so that the individual up crossing waves follow a Rayleigh probability distribution. Design or return wave heights are generally given in terms of H_s or $H_{max,3hrs}$ which requires some knowledge of the joint H_s, T_z probability distribution although the above equation is only a slowly varying function of T_z.

The N year return value of wave height (or any other environmental parameter) is the value which is exceeded on average once in N years. Hence if M values of wave height are obtained each year the N year return value, H_N, has a probability given by

$$Prob(H<H_N) = P(H_N) = 1-1/MN$$

Consequently the 50 year return wave height for 3 hourly samples has a probability given by

$$P(H_{50}) = 1-1/(8x365.25x50) = 0.99999316$$

2. WAVE DATA FROM THE OFFSHORE SITE

2.1 Data Return

The overall data return from the Offshore site was 69% for the seven year period March 1976 to February 1983, i.e. 14099 valid wave spectra were available. The monthly data returns (Table 2) show the distribution of the data throughout the seven year period. Data returns were high during the first two years and this data has been analysed by Fortnum (1981). Following this the returns were initially lower because of a radio interference problem along with the more usual intermittent buoy or mooring malfunctions. Data returns improved again with the allocation of a new radio frequency. Occasionally major logistic problems in servicing this remote site have led to long data gaps, the worst being the last three months of 1980.

2.2. Wave Height Statistics

The mean and variance of significant wave height in monthly groups (fig. 1) show the seasonal variation in wave climate at the Offshore Site. Wave heights were greatest during the winter with a maximum during November and a secondary peak in March. The seasonal trend in variance of H_S was highly correlated with the seasonal trend in the mean and consequently the largest waves could be expected to occur around November and March. The largest measured value of H_S was 10.61m recorded on 18 November 1979. The minimum value of H_S was 0.05m and the overall mean value was 2.35m.

The monthly data returns (fig. 1) show that the data set is biassed towards the summer months when wave heights are lowest. In the following statistical analysis a correction for this sampling bias has been applied to determine the effect of such a bias on the estimation of return wave heights.

2.3 Distribution of H_S and T_Z

The joint H_S, T_Z distribution derived from all the data is shown in a scatter plot (fig. 2) with class intervals of 0.5m in H_S and 0.5 sec in T_Z; a second modified H_S, T_Z distribution was computed to correct for the bias in sampling towards the summer months. Monthly H_S, T_Z distributions were found and then the frequencies scaled downwards so that each month had the same number of samples as the least sampled month (November). The monthly distributions were then recombined to form the modified joint distribution shown in figure 3. From the All Data and Modified joint H_S, T_Z distributions the marginal distributions of H_S and T_Z were found (fig. 4). In these figures the effect of modifying the frequency distibution can be seen to give greater importance to the larger winter waves as would be expected. The marginal distribution of H_s and T_z were moved slightly towards the higher values (fig. 4) while on the joint distribution (fig. 3) the modal peak in the distribution is reduced. On the joint distribution plots the average wave steepness (defined as $H_s/2\pi T_z{}^2$) was less than 1:12.

2.4 Distribution of Data Gaps

The modified frequency distribution derived above corrects only for the bias in sampling towards the summer (lower waves) in the data received. Such a correction would only produce a true frequency distribution if the data gaps were randomly distributed throughout the sample with respect to wave roughness. Any tendency for the buoy to fail in high wave conditions would result in the higher waves not being adequately sampled. As a coarse check on this point the H_s values immediately preceding a data gap (independent of the length of the data gap) were taken and the frequency distribution of these values were calculated. This so called Gaps distribution was taken to approximate the sea conditions at the beginning of a data gap (since H_s changes only slowly between each 3 hour recording). If the gaps occur randomly then they would form a randomly drawn sample from the parent distribution and so would be expected to have a similar frequency distribution to the parent population. In the comparison of the Gaps H_s distribution and the Modified H_s Distribution (fig. 5) it can be seen that the Gap distribution is biassed towards the higher waves. As the missing data cannot be recovered it can but be noted that this bias exists and its effect on the derived return values is not known.

3. ESTIMATION OF RETURN WAVE HEIGHTS FROM LONG TERM STATISTICS

In this commonly used method, a set of H_s or $H_{max,3\ hrs}$ data for a year or more is fitted to a probability distribution and then extrapolation into the tail of the distribution is used to derive the desired return value statistic. Choice of probability distribution is arbitrary being based only on how well the data fits the distribution which in turn gives confidence to the extrapolation used. The Weibull, Lognormal, Fisher-Tippett Type I and Type III distributions have all been used in the past. Goodness of fit is assessed by how well the data gives a straight line fit on appropriately constructed probability paper.

In this report the H_s data in 0.5m classes were used to derive cumulative probabilities using the Weibull plotting positions (see e.g. Carter and Challenor, 1981a). The mid point of the m^{th} class was plotted at the probability

$$S_m/(S_n + 1)$$

where S_m = Cumulative number of occurrences in the first m classes and n is the total number of classes. Both the All data and Modified data sets were used throughout in an effort to assess the effect of sampling bias on the derived return values. Best straight line fits on the probability plots were determined by least squares over all or part of the data, as stated in each case.

3.1 Weibull Distribution

The Weibull cumulative probability density function is given by

$$\text{Prob}(X<x) = F(x) = 1-\exp[-\{(x-A)/B\}^C]$$

The two parameter Weibull distribution (with the location parameter A, set to zero) is shown in fig. 6 for both the All data and Modified data set. Both data sets fit reasonably well and the least squares line fitted to the upper 16 points gave parameters B = 2.132, 2.249; C = 1.401, 1.432 for the All data and Modified data sets respectively. Extrapolation to a return period of 50 years gave H_s values of 12.5 and 12.7 from which it can be seen that the sampling bias had only a small effect on the estimated return value.

The associated values of T_z for these return values of H_s were found, by extrapolation of the joint H_s, T_z distributions (fig. 2, 3), to be around 13.5 sec and hence the 50 year return values of $H_{max,3hrs}$ were 22.9m and 23.2m for the All data and Modified Data distributions.

Fitting a three parameter Weibull distribution gave a good fit over the whole height range with A=0.3m (fig. 7). The straight lines fitted to all the data gave 50 year return significant wave heights of 12.4m and 12.6m respectively for the two distributions. In this case B = 1.973, 2.097; C = 1.365, 1.402 respectively. Corresponding 50 year return values of $H_{max,3hrs}$ were 22.7m and 23.0 m.

Fortnum (1981) found that the first two years of South Uist Offshore data gave a good fit to a three parameter Weibull distribution with parameters A = 0.15, B = 2.29, C = 1.55. The resultant 50 year return value of H_s was 11.5m, around 1m lower than the present estimates with the larger data set.

3.2 Lognormal Distribution

The variate x is lognormally distributed if $z = l_n(x-A)$ is normally distributed with mean, μ, and variance, σ^2. The two parameter lognormal plot (with A set to zero) of the South Uist data (fig. 8) did not give a convincing fit to this distribution although the upper tail approached a straight line. The least squares line fitted to the upper 10 points gave 50 year return values for H_s of 13.5m and 13.9m for the All data and Modified Data sets respectively. These values are around one metre higher than those obtained for the Weibull distribution but the data seemed to fit a Weibull more closely than a lognormal distribution. Using the three parameter lognormal distribution gave no real improvement in overall fit for realistic values of A.

3.3 Fisher-Tippett Type I Distribution

The Fisher-Tippett type I (FT-I) or Gumbel distribution given by

$$F(x)=\exp\{-\exp[-(x-A)/B]\}$$

is an unbounded, two parameter probability distribution.

The data gave a good fit to the FT-I distribution (fig. 9). The straight line fitted to the upper 21 points gave parameter values of A = 1.580, 1.668, B = 0.9932, 1.0144 for the All data and Modified data sets respectively. Corresponding 50 year return values of H_s were 13.4m and 13.7m. The 50 year return values of $H_{max,3\ hrs}$, were 24.5m and 25.0m using a T_z value of 13.7 sec. These values are all higher than those obtained with the Weibull distribution.

3.4 Fisher-Tippett Type III Distribution

The FT-III distribution is a three parameter distribution bounded at A and given by

$$F(x) = \exp\left[-\left(\frac{A-x}{B}\right)^{C}\right] \qquad x<A$$
$$= 1 \qquad x>A$$

If the data fits an FT-1 distribution, as the South Uist data appears to do, it cannot fit an FT-III distribution except for large values of A (when the FT-III distribution tends to the FT-I distribution). However the FT-III plot with A chosen to be 15m (fig 10) is given largely for comparison with a similar plot for the first two years of data in Fortnum (1981). Clearly the present data sets fit a FT-I distribution more closely than an FT-III whereas Fortnum (1981) found the reverse to be the case. His analysis gave FT-III parameter values of A = 15.0, B = 13.76 and C = 9.76 with a resulting 50 year return value for H_s of 10.9m.

3.5 Discussion

The Offshore H_s data gave a good fit to an FT-I distribution (a 2 parameter distribution) and a reasonable fit to a 2 parameter Weibull distribution. Three parameter distributions give greater scope for distribution fitting and in the case of the Weibull distribution an excellent fit was achieved. The 50 year return values of H_s for the FT-1 and Weibull distribution were in the range 12.4m to 13.7m with the larger values being obtained with the

FT-I distribution.

The present return values of H_s were higher than the 11.5m obtained by Fortnum (1981) using the first 2 years of data. Some subtle differences exist between the observed frequency distributions of the two data sets. While both sets could be fitted to the Weibull distribution an intriguing aspect was the good FT-III fit obtained by Fortnum (1981) compared to the good FT-I fit with the present data. These differences imply some interannual changes in wave conditions which would be of concern when using these methods with only one year of data as has frequently been done. In the present case the largest waves were recorded in 1979 and 1982 rather than in the first two years analysed by Fortnum (1981).

In the present analysis the return values obtained using the Modified data set were only around 0.5m higher than those obtained using the All data set. It seems that the sampling bias towards the summer months in the data set had only a small effect on the estimated return values. However this conclusion is subject to the missing data not being severely biassed and, as discussed in section 2.4, there is some evidence that proportionally more high wave data were missed.

4. RETURN VALUES OF ZERO-UP-CROSSING WAVE HEIGHT

The probability distribution of H_s or $H_{max,3\ hours}$ is not known and the previously described method is empirical in that the form of the distribution chosen is based only on how well it fits the data in a given situation. Consequently the extrapolation into the tail, to the low probability return event, is somewhat conjectural. A further reservation arises from the assumption that the largest wave will come from the three hour period with the largest H_s value whereas there is a finite probability of larger waves occurring in storms with lower values of H_s. Some improvement with respect to these two reservations can be achieved by deriving a distribution of individual wave heights. Using the method proposed by Battjes (1970) the marginal distribution of H_z, the zero-up-crossing wave height, is given by

$$P(H_z<h) = \int Prob(H_z<h|H_s).p(H_s)dH_s$$

where $p(H_s)$ is the probability density function of H_s. In practice the integral is replaced by a summation and $Prob(H_z<h \mid H_s)$ is taken to be the Rayleigh distribution. The density function $p(H_s)$ is estimated from the observed joint distribution of H_s and T_z (fig 2,3). If N_{ij} is the number of occurrencies of H_{si} and T_{zj} in the scatter plot then the number of waves associated with a particular H_{si}, T_{zj} pair is

$$N_{ij}\ 3hr/T_{zj}$$

hence $p(H_s) = \dfrac{\sum_j N_{ij}/T_{zj}}{\sum_i \sum_j N_{ij}/T_{zj}}$

and therefore

$$P(H_z<h) = \frac{\sum_i \sum_j \{1-\exp[-2(h/H_{si})^2]\} N_{ij}/T_{zj}}{\sum_i \sum_j N_{ij}/T_{zj}}$$

Following Battjes (1970) proposal, values of $P(H_z<h)$ were computed at 1m increments up to twice the maximum value of H_s observed which is approximately the most probable maximum individual wave height observed. A Weibull distribution was fitted to these values and extrapolated to the 50 year return value. The 50 year return probability for individual waves is given by

$$P_{50} = 1-1/N_{50}$$

where N_{50} is the total number of waves in 50 years. In the present case with the scatter plots (figs 2,3) based on 3 hourly observations and N_{ij} expressed in parts per thousand this is given by

$$N_{50} = \frac{50 \times 8 \times 365.25}{1000} \sum_i \sum_j N_{ij}/T_{ij}$$

The computed distribution of zero-up-crossing wave heights for both the All data and Modified Data sets were fitted to Weibull distributions. For the two parameter distribution (fig. 11) the least squares line fitted to the upper 19 points gave parameter values of B = 1.079, 1.142 and C = 0.947, 0.961 for the two data sets. The corresponding values of N_{50} were 2.671×10^8, 2.625×10^8 leading to 50 year return zero-up-crossing wave heights of 24.7m and 25.0m for the All data and Modified data sets respectively.

The distribution fit was improved using the three parameter

Weibull distribution with A = 0.4 (fig. 12). The straight line fitted to all data points gave the parameter values B = 0.938, 1.001 and C = 0.909, 0.924. The corresponding 50 year return values of H_Z were 24.9m and 25.2m for the All data and Modified data sets respectively.

The return heights found by this method were around 2m higher than those found by fitting the H_S data to a Weibull distribution but only marginally greater than those derived by fitting H_S data to an FT-I distribution.

Battjes (1970) noted that the shape parameter, C, was close to one indicating that the distribution of zero-up-crossing waves was nearly exponential. He analysed data from seven stations and found values of C (using a two parameter Weibull distribution) between 0.93 and 1.06 with the exception of 0.85 for Morecambe Bay data. Carter and Draper (1979) and Draper and Carter (1981) have reanalysed some of this data including the Morecambe Bay data and by fitting a three parameter Weibull distribution have obtained zero-up-crossing distributions that are close to exponential particularly in the upper tail. In the present data the values of C between 0.91 and 0.96 were close to one but still significantly different from one. However this may be accounted for by the missing high wave data which would be expected to give C values lower than the true value for Weibull distributed data. In the present case the three parameter distribution was less exponential (i.e. lower C) than the two parameter Weibull distribution.

5. ANALYSIS OF EXTREMES

Where data for several years exist the analysis of the extreme values can be used with the advantage that the distribution of the population from which the maxima arise does not need to be known in detail. Extreme value theory shows that the distribution of maxima from independent and identically distributed observations approach one of only three asymptotic forms as the sample size increases provided that a limiting distribution exists at all. The three extreme value distributions are the Fisher-Tippett Type I, Type II and Type III distributions. These distributions are also known by a variety of other names. The distributions differ in character with Type I being unbounded and Types II and III being bounded below or above respectively. The three distributions are related by simple transformations and can be shown to be the different forms of a General Extreme Value (GEV) Distribution (Jenkinson, 1955). The choice of which distribution to use is based on some knowledge of the underlying distribution or can be estimated from the GEV parameters. The Weibull and other distribution generally used in wave analysis are in the domain of attraction of the FT-I extreme value distributions. Carter and Challenor (1978) have derived return wave heights by fitting annual and monthly maxima to the FT-I distribution using seven years of data from the Sevenstones Light Vessel and Famita Rescue Ship. The theoretical basis of this method and its application has been reviewed by Carter and Challenor (1981a).

5.1 Monthly and Annual Maxima of H_s

The first requirement for an analysis of extremes was a set of independent monthly and annual H_s maxima from the Offshore data set. Care was required to ensure that higher values had not been missed during a gap in the data recording. From the time series plot of three hourly H_s values, gaps were identified and available covariate information used to assess the probability that a higher maximum had been missed. Covariate information was available from wave data or weather records at adjacent sites.

The H_s values obtained at the various South Uist sites were highly correlated and consequently it was possible in some cases to estimate missing maxima at the Offshore site using data from an alternative site. The correlation coefficients and linear regression coefficients for Offshore H_s values regressed on the values from other sites were calculated (J.A. Crabb - personal communication) using available coincident observations, and are given in Table 3.

Where it was necessary to use the regression model data the Deepwater buoy, if available, was used in preference to minimise any effect of water depth on the higher waves. The depth effects were negligible at least up to the highest waves observed as can be seen from the joint distribution of H_s at the Deepwater and Offshore buoys (fig. 13). The linear regression relationship holds up to the higher values with no apparent fall off. Any depth effects in using the Inshore regression model would be less serious since they would underestimate H_s at the Offshore site. In the final result, 10 monthly maxima were derived from Deepwater buoy data and two from the Inshore data.

On one occasion (February, 1982) when the monthly maximum storm was recorded at all three sites, the regression maxima from the Deepwater and Inshore sites were considerably higher than that observed at the Offshore site. In this one case the Deepwater data was used to overwrite the observation from the Offshore buoy.

Where a gap in wave data occurred at all sites the only recourse was to available weather records. The principal sources were the Daily Weather Reports published originally by the Meteorological Office and lately by the London Weather Centre. Six hourly values of wind speed and direction at Benbecula airport (57° 28'N; 7° 22'W) along with the surface pressure maps for the North Atlantic region, were used. Ewing (1980) has shown that wind waves at the Offshore site and winds at Benbecula are highly correlated. The wave generating potential of weather systems were assessed following the guidelines developed for the North Sea Wave Model (NORSWAM) study by Harding and Binding (1978). On this basis the size and intensity of any storm missed relative to the storm which generated the maximum recorded waves was assessed.

Where short data gaps of, say, one to five records length occurred close to a maxima in H_s additional problems arose. In some cases the underlying trend in H_s could be determined but where maxima occurred as an isolated record this could not be done. In these cases the monthly maximum must be considered a lower bound on the true maximum.

The monthly and annual maxima for H_s used in the extreme value analysis are given in Table 4 along with a two digit quality flag. The quality flag gives the data source (flag 1) and a reliability assessment based on the covariate wave and weather data (flag 2), as follows.

FLAG 1

1 H_s from Offshore waverider buoy

2 H_s from linear regression with Deepwater buoy

3 H_s from linear regression with Inshore buoy

9 No data or insufficient data

FLAG 2

0 Highest quality data i.e. Virtually complete month of data from the Offshore buoy

1 H_s(max) probably observed i.e. Data gaps unimportant as covariate data shows lesser storm(s) in gaps

2 H_s(max) reasonable value probably to within 1m. Similar storm occurred during data gaps and/or small gaps adjacent to H_s(max) but trend discernable.

3 H_s(max) probably missed so value is a lower bound estimate of the true maximum i.e. Large storms occurred during gaps and/or large gaps adjacent to H_s(max) such that no trend throughout the maximum storm was discernable.

9 No data or insufficient data.

5.2. Fitting H_s Maxima to an FT-1 Distribution

The FT-1 extreme value distribution is given by

$$F(x) = \exp\{-\exp[-(x-A)/B]\}$$

Annual maxima of environmental parameters such as wave height and wind speed have been fitted to this distribution and it is assumed the data meets the criteria of being independent and identically distributed with a large sample size. The distribution parameters, A,B are best determined using Maximum Likelihood Estimators (MLE). Studies with simulated wave data (Carter and Challenor, 1981b, 1983) have shown this method is preferable since it has the smallest mean square error. The small bias in these estimators can be corrected for if necessary. The other commonly used methods are the method of moments and graphical methods using least squares fitted straight lines. All three methods are used here for comparison purposes.

Carter and Challenor (1981b) have proposed that fitting annual maxima for wave data to an FT-1 distribution underestimates the return values. This arises because wave data shows a strong seasonal change (see e.g. fig 1) and hence is not identically distributed throughout the year. They propose that seasonal variations be incorporated by fitting FT-1 distributions to the monthly maxima and compounding these monthly distributions, F_m, to give an annual distribution given by

$$\mathrm{Prob}(X<x) = \prod_{m=1}^{12} F_m(x)$$

Dividing the year into months was done as a convenient subdivision. Turner (1983) has reanalysed the Carter and Challenor (1978) data using different subdivisions and this suggests that monthly divisions are close to optimal.

The monthly and annual H_s maxima (Table 4) were fitted to FT-I distributions using Maximum Likelihood Estimators and the distribution parameters along with the 50 year return values are given in Table 5. No bias correction has been applied but Carter and Challenor (1978) found that for 6 or 7 years data the bias correction would increase the return values by 4%. The 90% confidence limits are derived from Challenor (1979). The 50 year return values were calculated using a probability of

$P = 1 - 1/50 = 0.98$

Taking the product of the twelve separate monthly FT-I distributions and solving numerically gave a 50 year return significant wave height of 15.35m compared to the 11.67m value obtained by fitting the annual maxima to an FT-I distribution. This difference is larger than has been found when applying this method to other wave data sets (Carter and Challenor, 1981b).

The distribution parameters and return values using method of moments estimators and graphical methods are given in Table 6. The graphical method used a least squares fit to the maxima plotted on FT-I paper with the Gringorten plotting positions. Details of these methods are given in Carter and Challenor (1981a).

5.3 Censored Maximum Likelihood Estimates

A few of the monthly and annual maxima for H_s (Table 4) must be considered to be lower bounds on the true maxima because of the nature of the data gaps, as explained in section 5.1. These values, identified by a second quality flag of 3, (Table 4) can be considered to be censored at the observed level and the maximum likelihood estimates revised accordingly (P.G. Challenor - personal communication). If F(x) and f(x) are the cumulative and density functions respectively for an FT-I (or any other extreme value distribution) the usual likelihood function for n observed extreme values, x_i, is given by

$$L = \prod_{i=1}^{n} f(x_i)$$

However if an observation, x_j, is considered to have been censored at the level $x_j = C_j$ then the probability of the true value exceeding this level is $1-F(C_j)$. For m censored values the likelihood function becomes

$$L = \prod_{j=1}^{m} (1-F(C_j)) \prod_{i=m}^{n} f(x_i)$$

The log likelihood function is then maximised in the usual way and the resulting equations solved numerically (Carter and Challenor, 1981a).

This method was applied to the monthly and annual maxima of H_S (Table 4) with modified distributions being obtained for the months of January, February, March, July, November, December as well as the annual maxima (Table 7). The 50 year return value of H_S from the annual distribution was 13.5 while combining the monthly distribution gave a value of 18.6m. These values are higher than those obtained earlier, assuming no data censoring, as would be expected.

5.4 Sinusoidal Seasonal Trend Model

When monthly maxima of H_S are fitted to FT-1 distributions the resulting location parameters (A) show a seasonal trend similar to that of the mean values of H_S. The maximum likelihood estimates of A derived in section 5.2 are shown in fig. 14 and the seasonal trend can be compared with that in fig. 1. Challenor (1982) has utilized this trend to derive a new annual distribution where the A parameters are assumed to follow a sine curve; given by

$$P(X<x) = \prod_{i=1}^{12} \exp(-\exp(-(\frac{x-\gamma-\theta \, \mathrm{Cos}(i\pi/6+\phi)}{B_i})))$$

The three parameters γ, θ, ϕ along with the 12 B parameters are determined simultaneously by maximising the likelihood function of this distribution. Provided the seasonal variation is adequately modelled by a sine curve this method should improve the statistical inferences since the effect of outliers are reduced and only 15 parameters are estimated compared with the 24 parameters estimated in the original method for monthly maxima.

This method was applied to the South Uist data and the resulting parameters are given in Table 7 and plotted in figure 11. The 50 year return value of H_s from this annual distribution was 16.7m.

A further refinement was tried based on the fact that the B parameters appear to be randomly scattered with respect to month (fig. 14). Assuming that the scale parameter take the same value for all months results in only four parameters having to be estimated. Such a model was applied to the South Uist data and gave 17.4m for the 50 year return value of H_s.

The relative merits of the three seasonal models were assessed using likelihood ratio tests. These tests showed that the 24 parameter model gave a significantly better fit to the data at the 99.9% confidence level than the 15 parameter model using the sinusoidal model for the A parameters. In turn the 15 parameter model was a better fit at the 90% level than the 4 parameter model where B is assumed to be constant.

5.5 Peaks Over Threshold Methods

An alternative approach to extreme value analysis is to examine the peak exceedances over a high threshold. Peaks Over Threshold (POT) methods incorporating seasonal variations have recently been applied to wave data (Turner, 1982; Smith, in press) but these require a reasonably unbroken data set. The methods were examined using a pseudo four year time series of H_s values obtained by combining the most complete records from the South Uist data set. The simple model of Turner (1982) where peak exceedances of the seasonally adjusted logrithmic wave heights are assumed to be exponentially distributed gave reasonable results. However the extension of this model (Smith, in press) which incorporates a Generalised Pareto Distribution for the exceedances gave unrealistic results. As there were some shortcomings in the synthetic data set used it was impossible to assess the validity of these methods which must still be considered experimental. However these methods have a good theoretical basis and warrant further development and testing.

5.6 Discussion

The asymptotic extreme value methods require independant and identically distributed observations while wave data has a clear seasonal trend. Hence methods which allow for the seasonal variation are preferable to those which do not. The seasonal model of Carter and Challenor (1981b), where fitted monthly FT-1 distributions are combined to give an annual distribution, gave the best results with the present data set. The estimated 50 year return value of H_s was 15.4m and taking an associated T_z of 14.5s gives a return $H_{max,3hours}$ of 28.0m. While confidence limits for this distribution are difficult to evaluate they will be of similar order to those found in fitting FT-1 distributions to monthly or annual maxima (Table 5). The effect of some extreme values being missed because of data gaps was assessed using censored maximum likelihood estimators which suggest that the return H_s value could be underestimated by as much as 3m.

6. GENERAL DISCUSSION AND CONCLUSIONS

The 50 year return wave height parameters for the South Uist offshore site as found in the present study and by Fortnum (1981) are given in Table 8. The most striking feature of these data are the wide range of estimates which is to be expected when short time series are used to predict low probability events. Of the methods which fit all the data to a frequency distribution the best fits were obtained with the FT-1 distribution, the 3 parameter Weibull distribution and the individual wave model. The first two gave 50 year return values of $H_{(max,3\ hours)}$ of 25.0m and 23.0m while the 50 year return value of the zero up crossing wave was 25.2m from the individual wave model. Of the methods based on analysis of measured monthly or annual extremes, that using the combined monthly maxima FT-1 distributions was preferred and gave a return $H_{(max,3\ hours)}$ of 28.0m. However all these values are probably underestimated because the biggest waves were inadequately sampled. Analysis of the gaps in the data set showed that even after the frequency distribution was corrected for the over all higher data returns in summer, the gaps in the record were still biassed towards times of high wave activity. For the extreme value method covariate wave and meteorological data were used to assess the reliability of recorded extreme values and this showed that some extremes were missed, which suggests that the method of combining distributions of monthly maxima using censored data techniques should give the best estimate of return value, indicating a 50 year return value of $H_{max,3hrs}$ of 33.8m. However, all the methods used in this report for estimating return wave heights assume that the physical constraints on the predicted large waves are similar to those on the observed smaller wave field; but this would not be correct at the offshore site for waves of 33.8m which is 0.8 of the water depth of 42m. Therefore the wave height of 33.8m is probably more appropriate to deeper water in the vicinity.

Comparison of the present results with those obtained by Fortnum (1981) using the first two years of data (Table 8) show that the return value estimates have increased by around 2m where comparable methods have been used. Furthermore where Fortnum (1981) found a good fit to an FT-III distribution the present data gave a better fit to an FT-1 distribution. Such changes suggest interannual variability exists which is of concern as all presently used methods of estimating return wave heights assume that interannual variability is negligible.

7. ACKNOWLEDGEMENTS

Wave data were collected and processed at IOS Taunton and for this report the help of J.A. Crabb and R. Gleason is acknowledged. Much advice was received from members of the wave group at IOS Wormley and in particular I thank P.G. Challenor and D.J.T. Carter for many useful discussions. This report was written while the author was a visiting scientist from the New Zealand Oceanographic Institute.

8. REFERENCES

Battjes J.A. 1970 Long term wave height distribution at seven stations around the British Isles. NIO Internal Report No. A44.

Carter D.J.T. & Challenor P.G. 1978 Return wave heights at Seven Stones and Famita estimated from monthly maxima. IOS Report No. 66.

Carter D.J.T. & Challenor P.G. 1981a Estimating return values of wave height IOS Report No. 116.

Carter D.J.T. & Challenor P.G. 1981b Estimating return values of environmental parameters. Quarterly Journal of the Royal Meteorological Society 107(451): 259-266.

Carter D.J.T. & Challenor P.G. 1983 Methods of fitting the Fisher-Tippett type 1 extreme value distribution. Ocean Engineering 10(3): 191-199.

Carter D.J.T. & Draper L. 1979 Waves at Ocean Weather Station Alpha. IOS Report No. 69.

Challenor P.G. 1979 Confidence limits for extreme value statistics. IOS Report No. 82.

Challenor P.G. 1982 A new distribution for annual extremes of environmental variables. Quarterly Journal of the Royal Meteorological Society 108: 975-980.

Draper L. & Carter D.J.T. 1981 Waves at Morecambe Bay Light Vessel during 1957. IOS Report No. 113.

Ewing J.A. 1980 Observations of wind-waves and swell at an exposed coastal location. Estuarine and Coastal Marine Science 10: 543-554.

Fortnum B.C.H. 1981 Waves recorded off South Uist in the Hebrides. IOS Report No. 115.

Fortnum B.C.H., Humphery J.D. & Pitt E.G. 1979 Contoured wave data off South Uist. IOS Report No. 71.

Harding J. & Binding A.A. 1978 The specification of wind and pressure fields over the North Sea and some areas of the North Atlantic during 42 gales from the period 1966 to 1976. IOS Report No. 55.

Jenkinson A.F. 1955 The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society 81: 158-171.

Smith R.L. in press. Threshold methods for sample extremes. Proceedings of NATO Advanced Study Institute "Statistical Extremes and Applications" Vimeiro, Portugal, 31 August - 14 September 1983.

Turner M.J. 1982 Estimation of the fifty year return wave height at Seven Stones. M.Sc. and D.I.C. Report, Department of Mathematics, Imperial College.

TABLES

TABLE 1 - Particulars of the South Uist Wave Recording Sites.

SITE	DEPTH m	POSITION	RECORDING PERIOD
Offshore	42	057°18'N 007°38'W	March 1976 to Present
Deepwater	98	057°18'N 007°54'W	August 1980 to Present
Inshore I		057°20'N 007°27'W	June 1978 to August 1979
Inshore II	23	057°20'N 007°29'W	August 1979 to January 1981
Inshore III	25	057°17'N 007°29'W	March 1981 to June 1982

TABLE 2 - Monthly Percentage Data Return from the South Uist offshore waverider, March 1976 to February 1983

	J	F	M	A	M	J	J	A	S	O	N	D
1976			85	97	99	98	96	100	94	98	97	99
1977	90	98	86	89	97	81	86	90	95	89	84	94
1978	79	67	43	0	0	14	92	96	83	33	0	67
1979	46	61	0	75	71	91	57	58	25	54	44	40
1980	67	90	64	94	97	94	92	78	90	0	0	0
1981	21	8	0	5	96	98	99	90	98	95	88	97
1982	99	94	94	78	85	71	95	98	98	90	56	0
1983	0	31										

TABLE 3 - Linear regression offshore H_s values on H_s from other South Uist sites.

SITE	REGRESSION SLOPE	LINE INTERCEPT	CORRELATION COEF.	No. PAIRS
Deepwater	0.826	0.078	0.970	4012
Inshore I	1.575	0.016	0.951	1051
Inshore II	1.073	0.188	0.970	1993
Inshore III	1.110	0.188	0.958	2810

TABLE 4 - Monthly and annual maxima for H_s at the offshore site along with associated data quality flag (FF). Year = March to February.

	1976		1977		1978		1979		1980		1981		1982		1983	
	H_s	FF	H_s	FF	H_s	FF	H_s	FF	H_s	FF	H_s	FF	H_s	FF	H_s	FF
Jan			6.05	12	7.27	11	4.77	13	4.48	12	8.09	13	5.89	10	10.53	23
Feb			5.39	12	6.07	12	4.88	12	5.13	11	6.91	23	9.46	21	8.96	22
Mar	7.64	10	6.45	11	8.94	13	8.99	33	5.42	12	5.70	22	8.27	12		
April	5.85	10	6.33	12	-	99	5.31	12	4.43	11	3.79	21	4.90	12		
May	4.56	10	5.09	10	-	99	4.61	12	3.21	10	3.52	12	7.63	12		
June	5.42	10	4.95	11	-	99	5.51	11	7.29	10	4.47	10	2.23	12		
July	3.13	10	3.95	11	3.88	10	3.42	13	3.60	11	3.38	10	3.97	11		
Aug	3.42	10	5.39	10	3.58	10	3.73	12	4.05	11	3.80	11	6.51	12		
Sept	3.83	10	9.06	10	6.97	12	-	99	5.72	11	3.94	10	6.00	10		
Oct	7.49	10	7.20	11	7.47	12	7.40	32	8.31	22	7.70	21	8.26	12		
Nov	8.29	10	8.98	12	-	99	10.61	13	5.32	21	7.31	12	10.56	13		
Dec	4.83	10	7.99	10	4.79	11	8.64	13	5.78	23	5.08	10	8.21	23		
Year	8.29	10	9.06	10	8.94	13	10.61	13	8.31	22	9.46	21	10.56	13		

TABLE 5 - Fifty year return H_s and FT-I parameters (A,B) from monthly and annual maxima using Maximum Likelihood Estimates.

	No. Data	A	B	H_s (50 yr.) (m)	90% Confidence Limits (m)
Jan	7	5.8296	1.4766	11.59	9.5 - 18.8
Feb	7	5.8913	1.2645	10.83	9.1 - 17.0
Mar	7	6.6465	1.2531	11.54	9.8 - 17.7
April	6	4.6757	0.7853	7.74	6.6 - 12.4
May	6	4.1337	1.0236	8.13	6.6 - 14.2
June	6	4.1950	1.5614	10.29	7.9 - 19.5
July	7	3.4658	0.2838	4.57	4.2 - 6.0
Aug	7	3.8958	0.6679	6.50	5.6 - 9.8
Sept	6	5.0666	1.4595	10.76	8.6 - 19.4
Oct	7	7.5045	0.3021	8.68	8.3 - 10.2
Nov	6	7.5534	1.8547	14.79	12.0 - 25.7
Dec	7	5.7012	1.2739	10.67	8.9 - 16.9
Year	7	8.8962	0.7108	11.67	10.7 - 15.2

TABLE 6 - Fifty year return H_S and FT-I parameters from monthly and annual maxima using Moment and Least Squares Estimators.

	METHOD OF MOMENTS			LEAST SQUARES		
	A	B	H_s (50)	A	B	H_s (50)
Jan	5.7771	1.6434	12.2	5.7750	1.7984	12.8
Feb	5.8506	1.4468	10.7	5.8728	1.5376	11.9
Mar	6.6716	1.1654	11.2	6.7100	1.1999	11.4
April	4.6832	0.7251	7.5	4.6869	0.7944	7.8
May	4.0628	1.2253	8.8	4.0809	1.3197	9.2
June	4.2351	1.2875	9.3	4.2756	1.3457	9.5
July	3.4719	0.2541	4.5	3.4809	0.2605	4.5
Aug	3.8352	0.8993	7.3	3.8576	0.9396	7.5
Sept	5.0359	1.5317	11.0	5.0421	1.6812	11.6
Oct	7.4954	0.3371	8.8	7.5027	0.3543	8.9
Nov	7.6005	1.5787	13.8	7.6549	1.6407	14.1
Dec	5.6955	1.3493	11.0	5.7012	1.2739	10.7
Year	8.8871	0.7475	11.8	8.9020	0.7880	12.0

TABLE 7 - Fifty year return H_S and FT-I parameters from monthly and annual maxima using Censored Maximum Likelihood Estimates and Sinusoidal Season trend model.

	CENSORED M.L.E.			SINUSOIDAL MODEL		
	A	B	H_s (50)	A^+	B	H_s (50)
Jan*	6.5302	2.1059	14.7	6.7843	1.9500	14.4
Feb*	6.0194	1.4608	11.7	6.3160	1.4390	11.9
Mar*	6.8469	1.6240	13.2	5.5362	1.4981	11.4
April	4.6757	0.7853	7.7	4.6539	0.7803	7.7
May	4.1337	1.0236	8.1	3.9054	1.0009	7.8
June	4.1950	1.5614	10.3	3.4914	1.5443	9.5
July*	3.5139	0.3150	4.7	3.5227	0.3018	4.7
Aug	3.8958	0.6679	6.5	3.9910	0.6991	6.7
Sept	5.0666	1.4595	10.8	4.7708	1.4213	10.3
Oct	7.5045	0.3021	8.7	5.6531	1.5144	11.6
Nov*	7.9331	2.4792	17.6	6.4016	1.9372	14.0
Dec*	5.2296	2.0464	14.2	6.8156	1.8971	14.2
Year*	9.171	1.1167	13.5	-	-	

* Distribution changed using the censored data method.

+ From $A_i = 5.1535 - 1.7056 \cos(\frac{i\pi}{6} + 2.9153)$

TABLE 8 - Summary of 50 year return wave height estimates (metres) at the South Uist Offshore Site

	H_s	$H_{max,3\ hours}$	H_z
2 YEARS DATA - Fortnum (1981)			
Weibull (3 parameter)	11.5	21.3	
FT-III	10.9	20.4	
Individual Wave Model	-		22.5
7 YEARS DATA - Present Report			
Weibull (3 parameter)	12.6	23.0	
FT-1	13.7	25.0	
Individual Wave Model	-		25.2
Annual Extremes, FT-1	11.7	21.5	
Monthly Extremes, FT-1	15.4	28.0	
Sinusoidal Seasonal Model	16.7	30.4	
Censored Annual Extremes	13.8	25.1	
Censored Monthly Extremes	18.6	33.8	

FIGURES

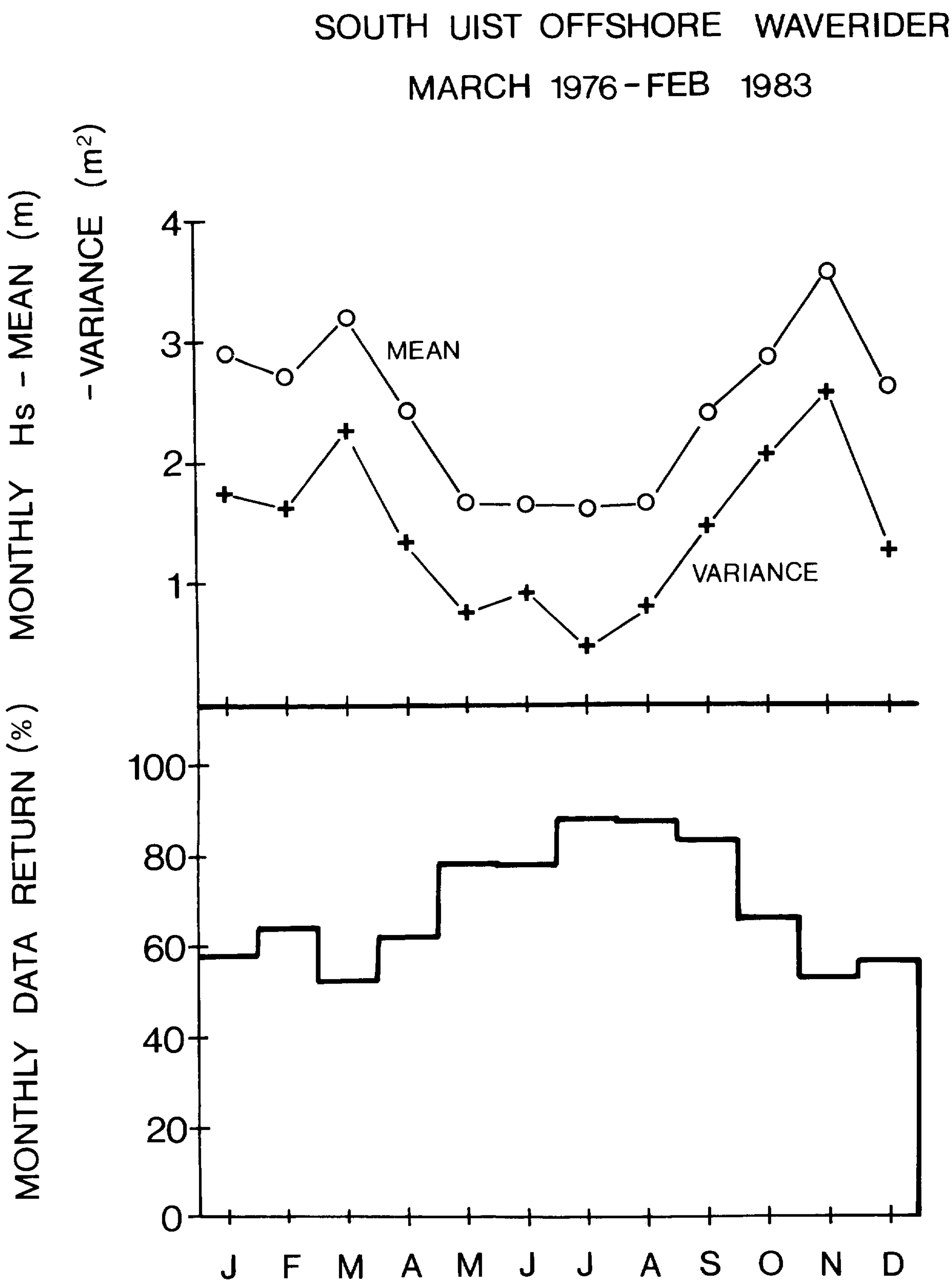

Fig. 1 Mean and variance of H_s along with percentage data return for each month from the Offshore wave data set.

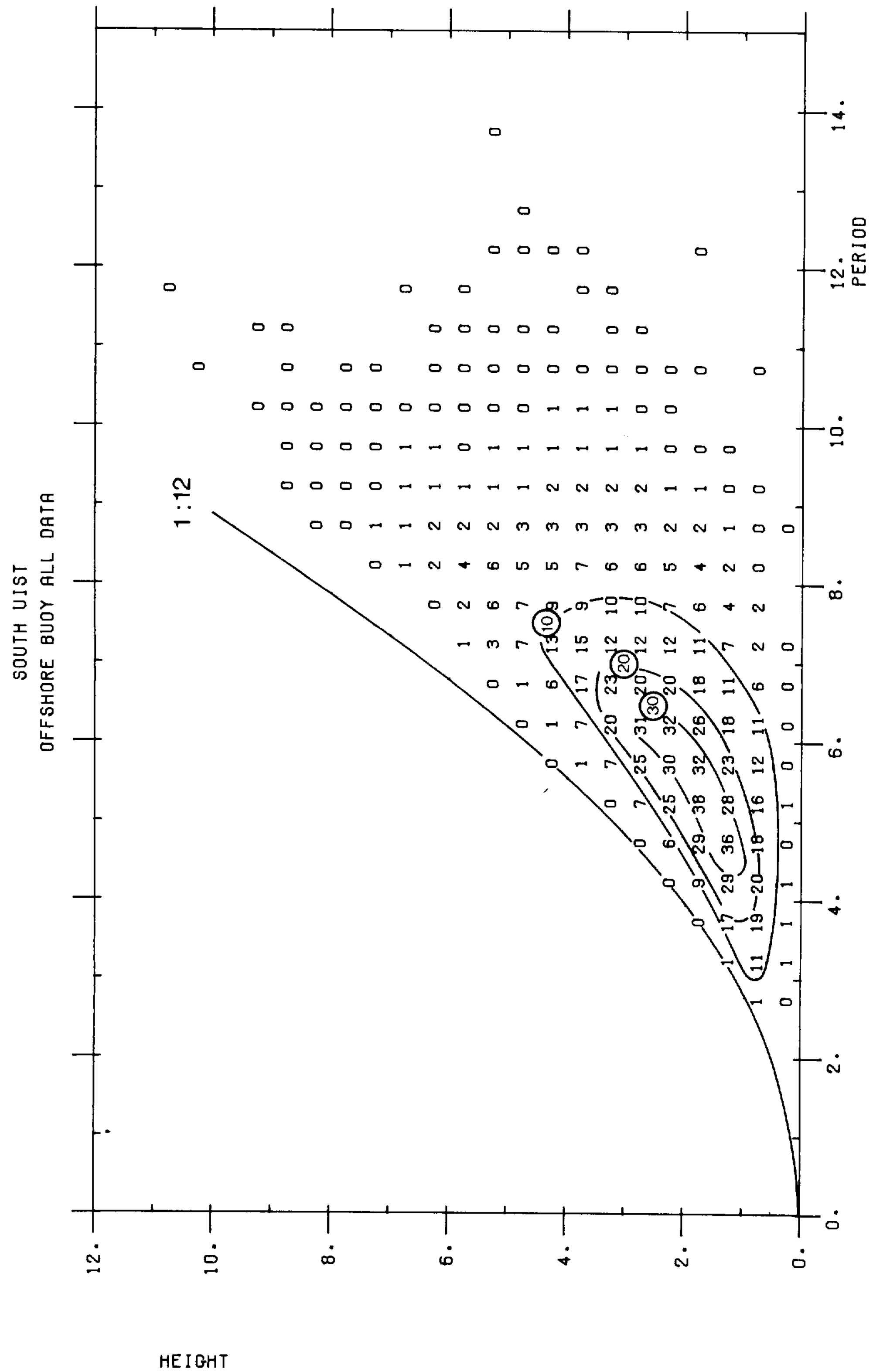

Fig. 2 Joint distribution (parts per thousand) of wave height H_s(m) and period T_z(s) - All data set (0:<0.5°/oo)

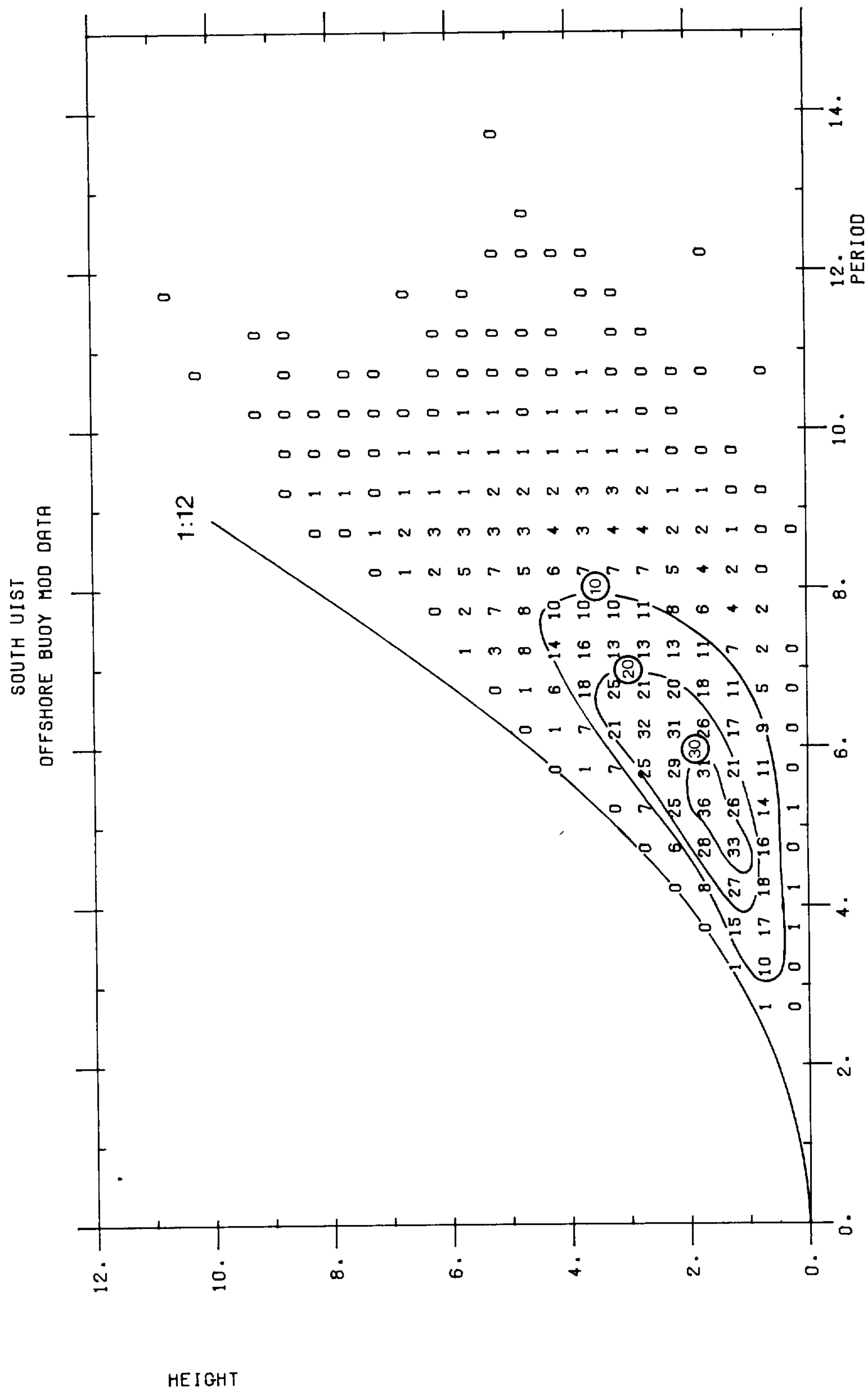

Fig. 3 Joint distribution (parts per thousand) of wave height H_s(m) and period T_z(s) - Modified data set (0:<0.5°/oo)

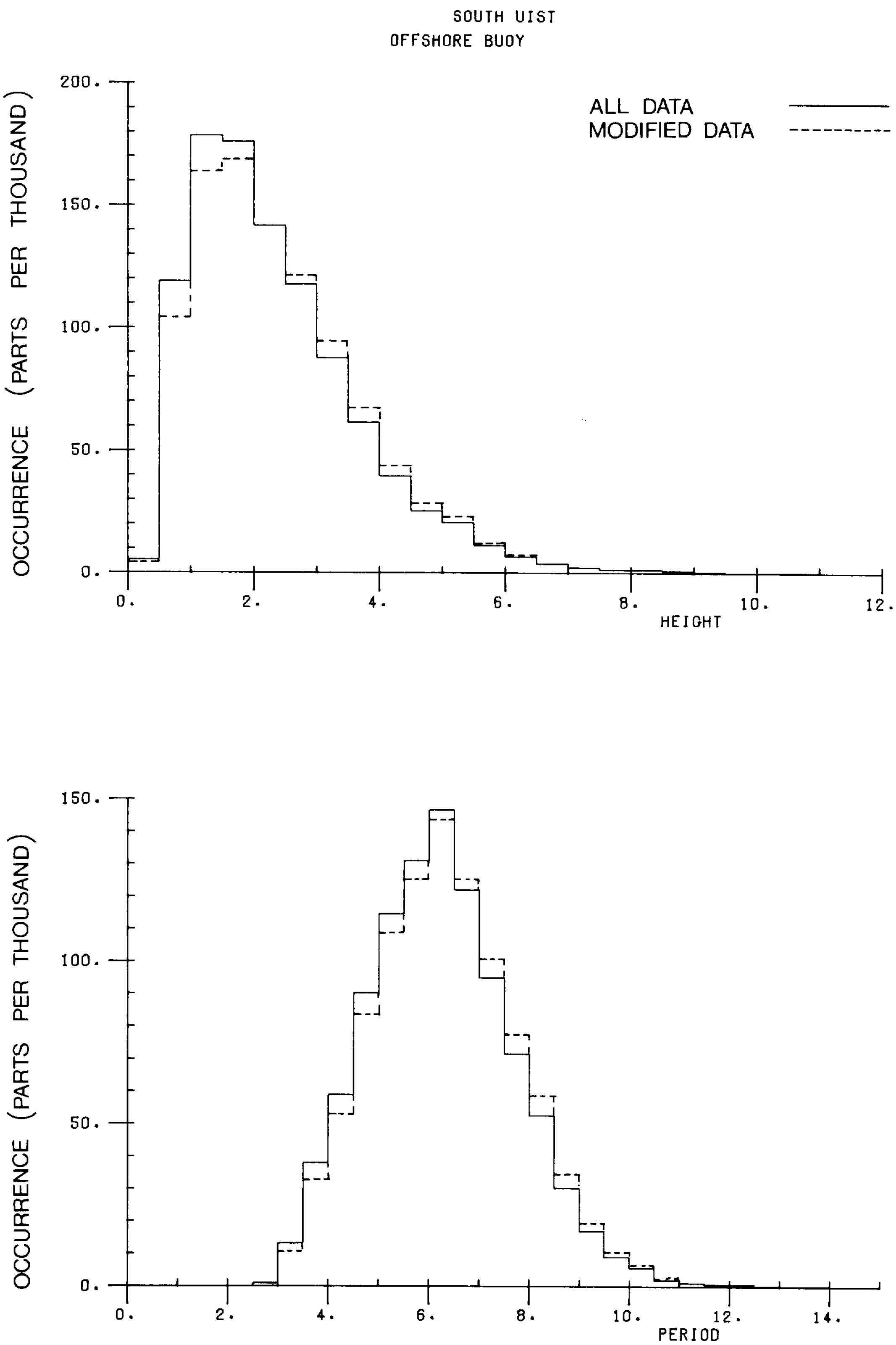

Fig. 4 Marginal distributions of wave height H_s(m) and period T_z(s). All data set ———. Modified data set ......

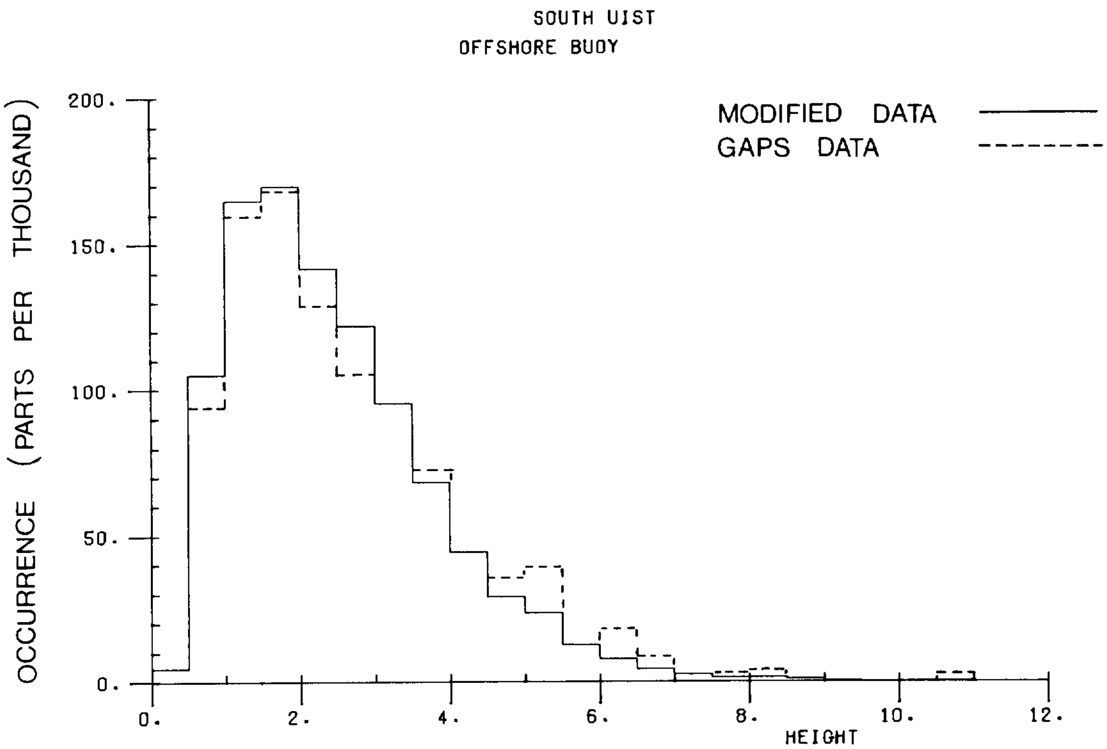

Fig. 5 Marginal distribution of wave height H_s(m). Modified data set ————. Gaps data set .....

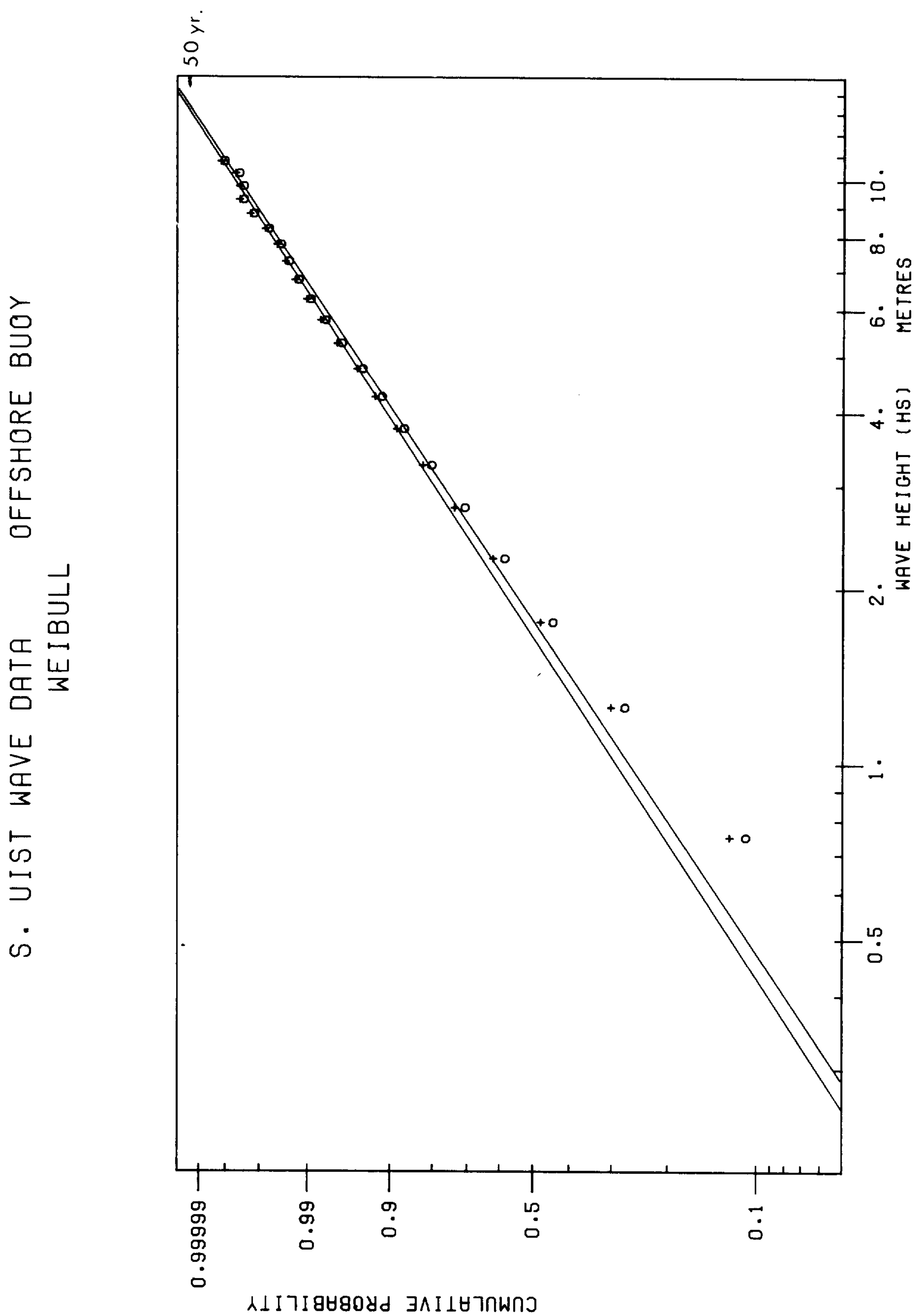

Fig. 6 Cumulative distribution of H_s: 2 parameter Weibull distribution

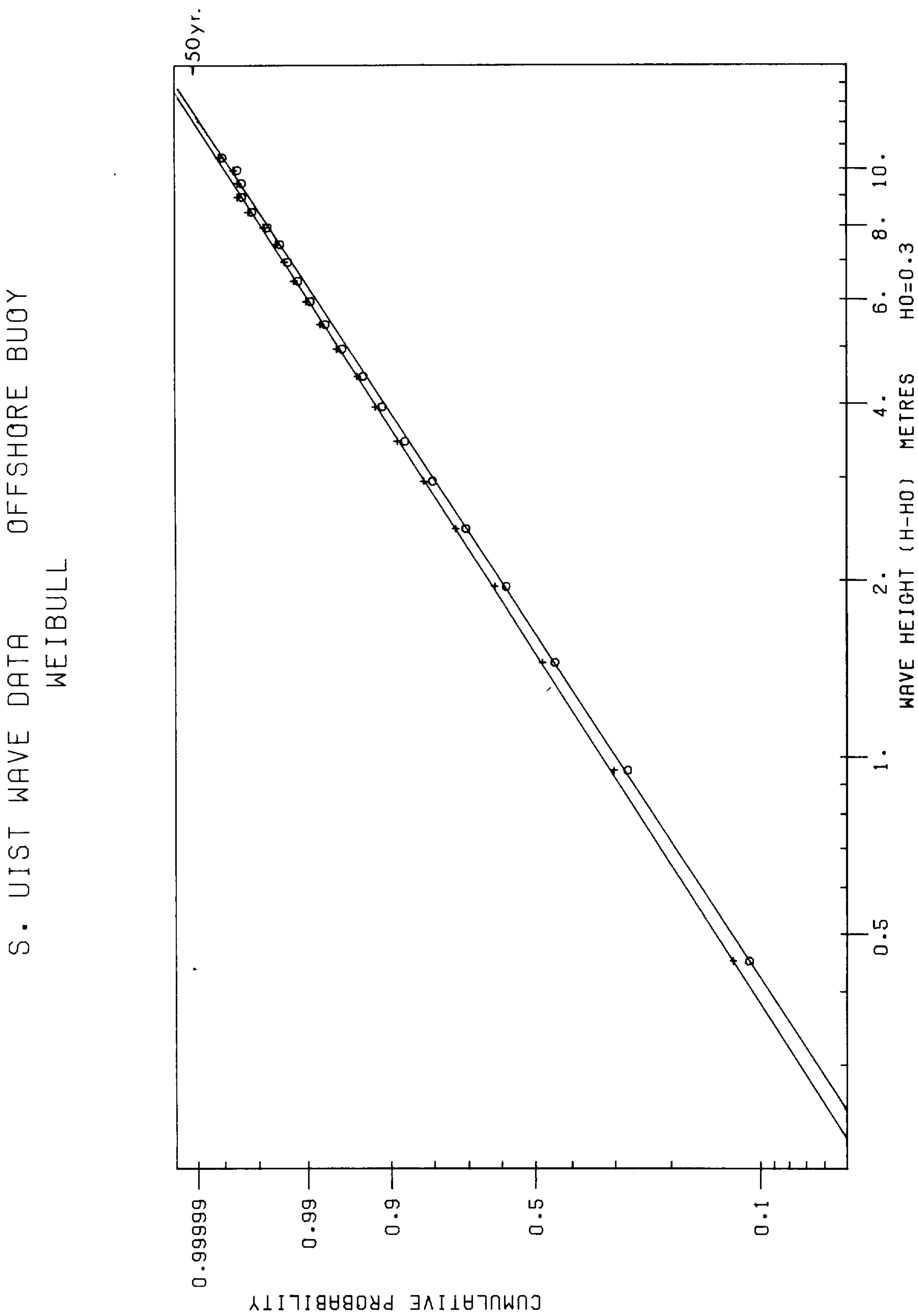

Fig. 7 Cumulative distribution of H_s: 3 parameter Weibull distribution. Location parameter HO=0.3m.

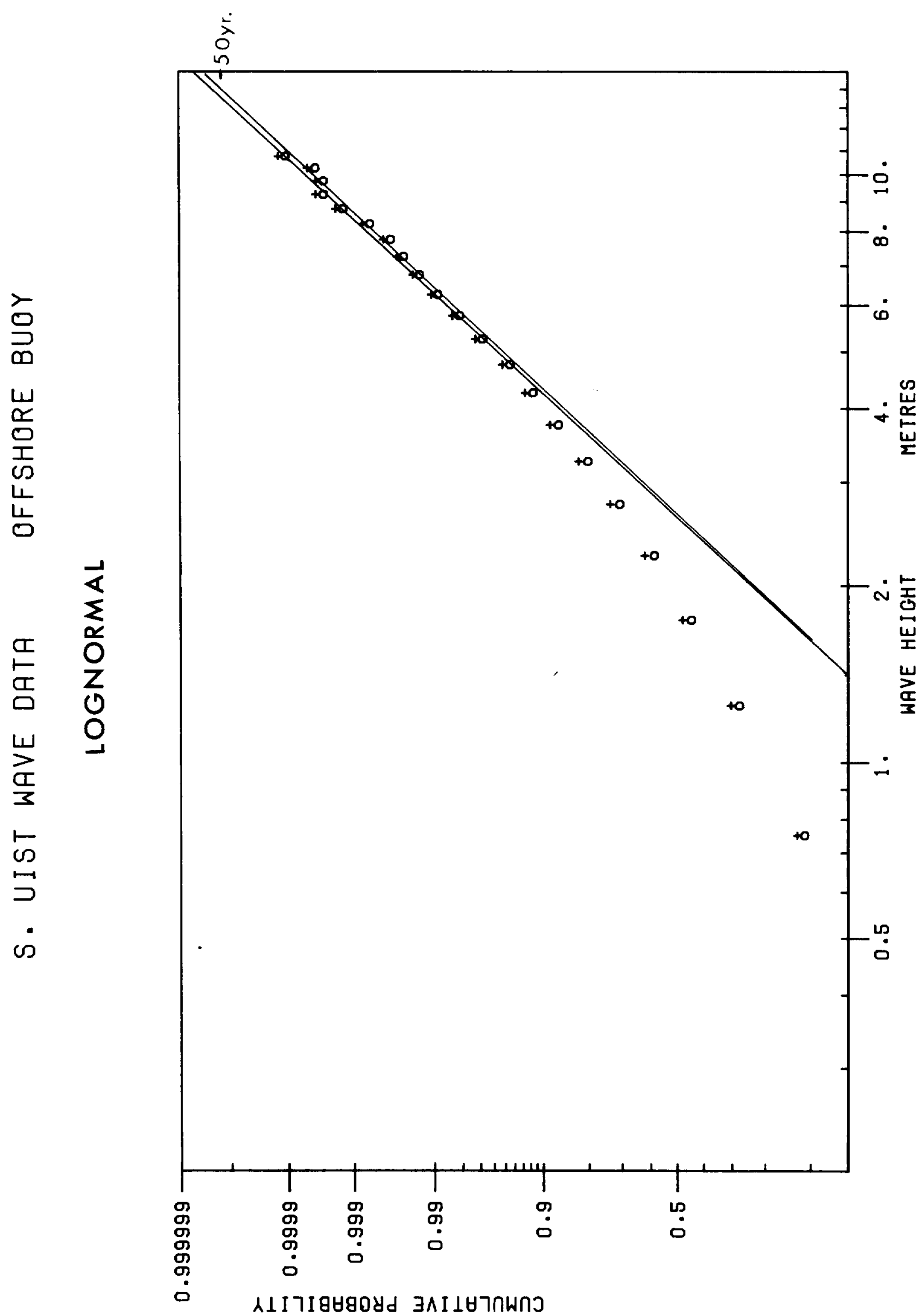

Fig. 8 Cumulative distribution of H_s: 2 parameter lognormal distribution

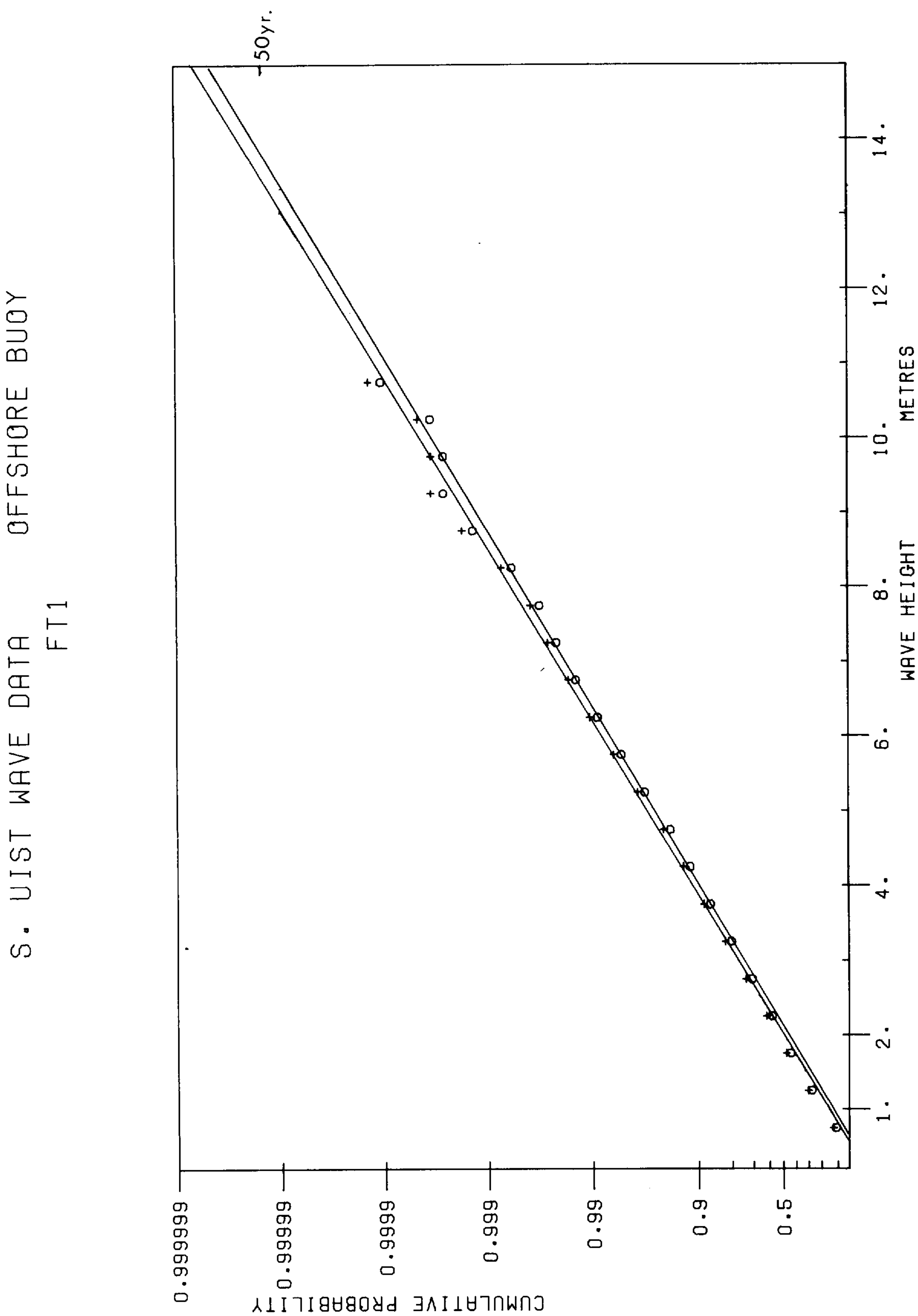

Fig. 9 Cumulative distribution of H_s: Fisher Tippett Type I distribution

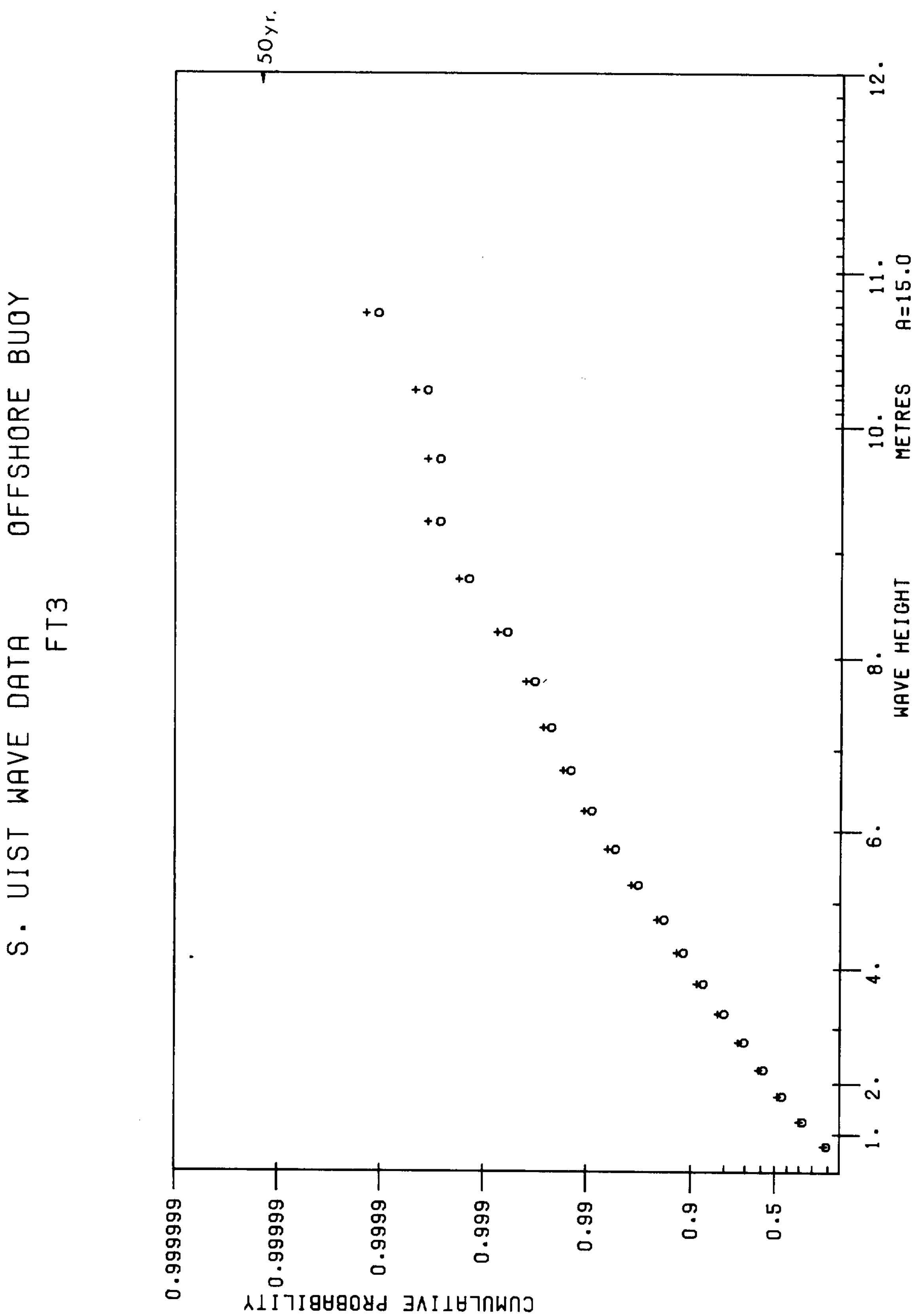

Fig. 10 Cumulative distribution of H_s: Fisher Tippett Type III distribution. Location parameter A=15.0m.

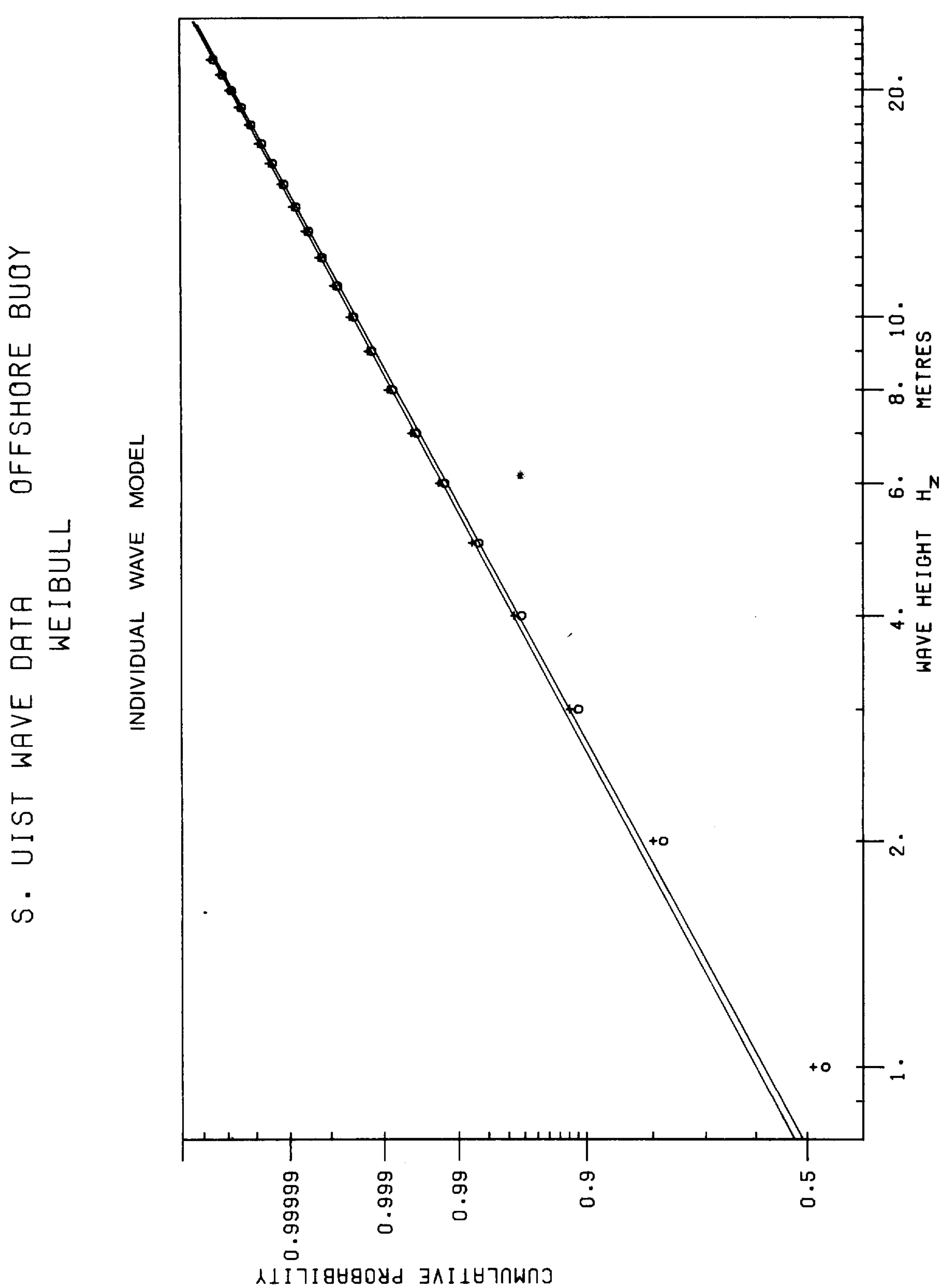

Fig. 11 Cumulative distribution of H_Z: 2 parameter Weibull distribution

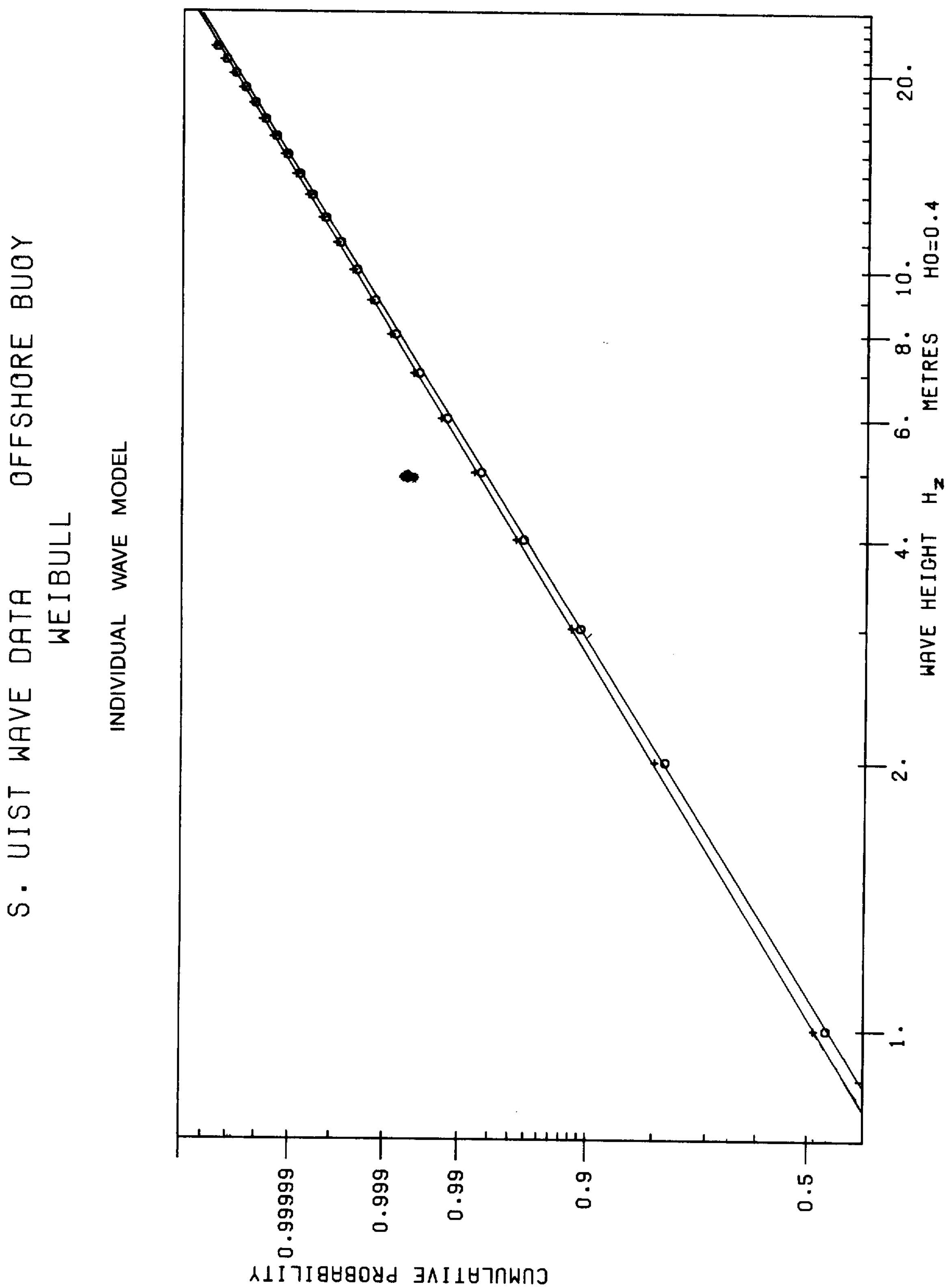

Fig. 12 Cumulative distribution of H_z: 3 parameter Weibull distribution. Location parameter HO=0.4m.

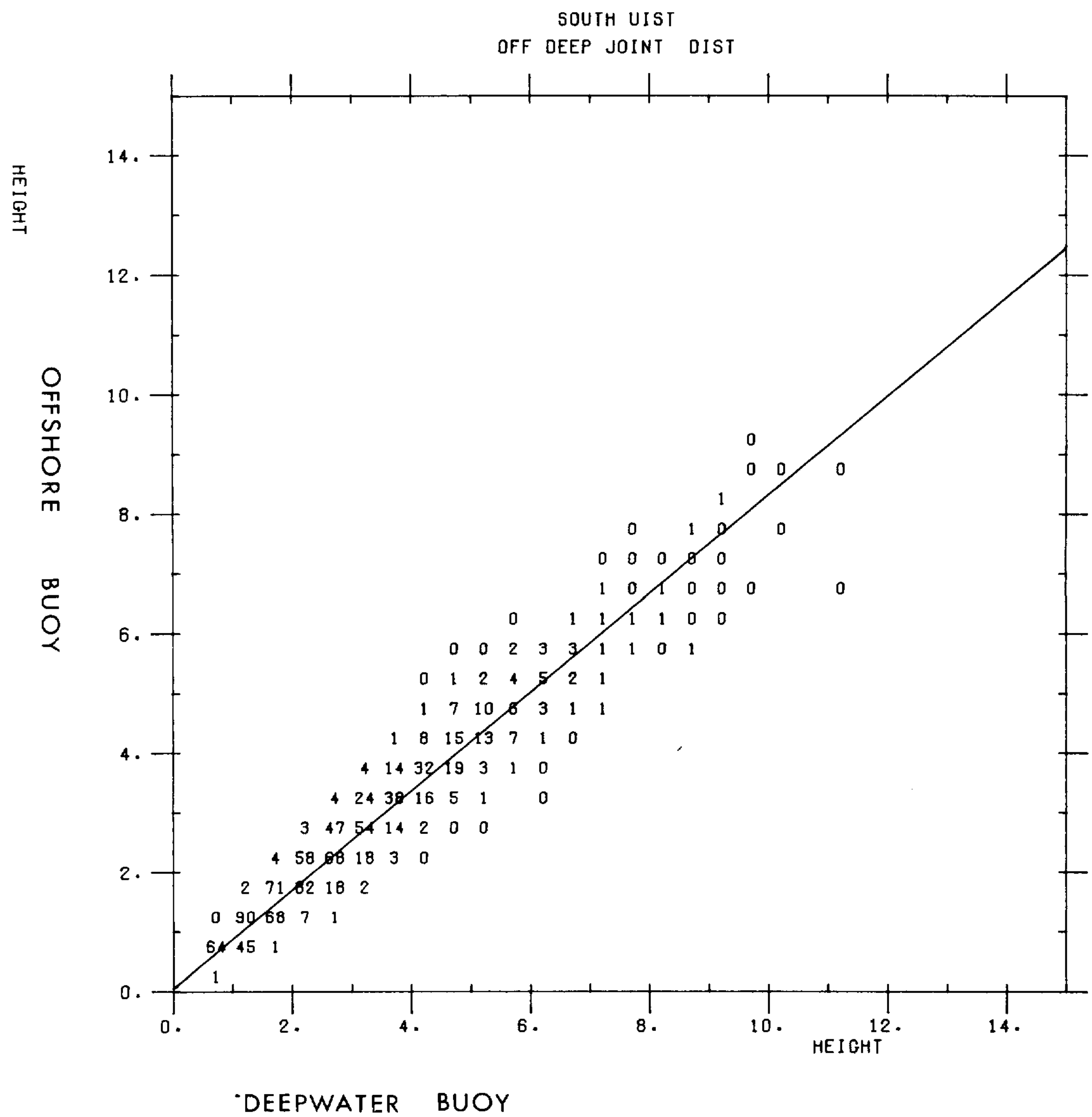

Fig. 13 Joint distribution of H_s (Offshore Buoy) and H_s (Deepwater Buoy)

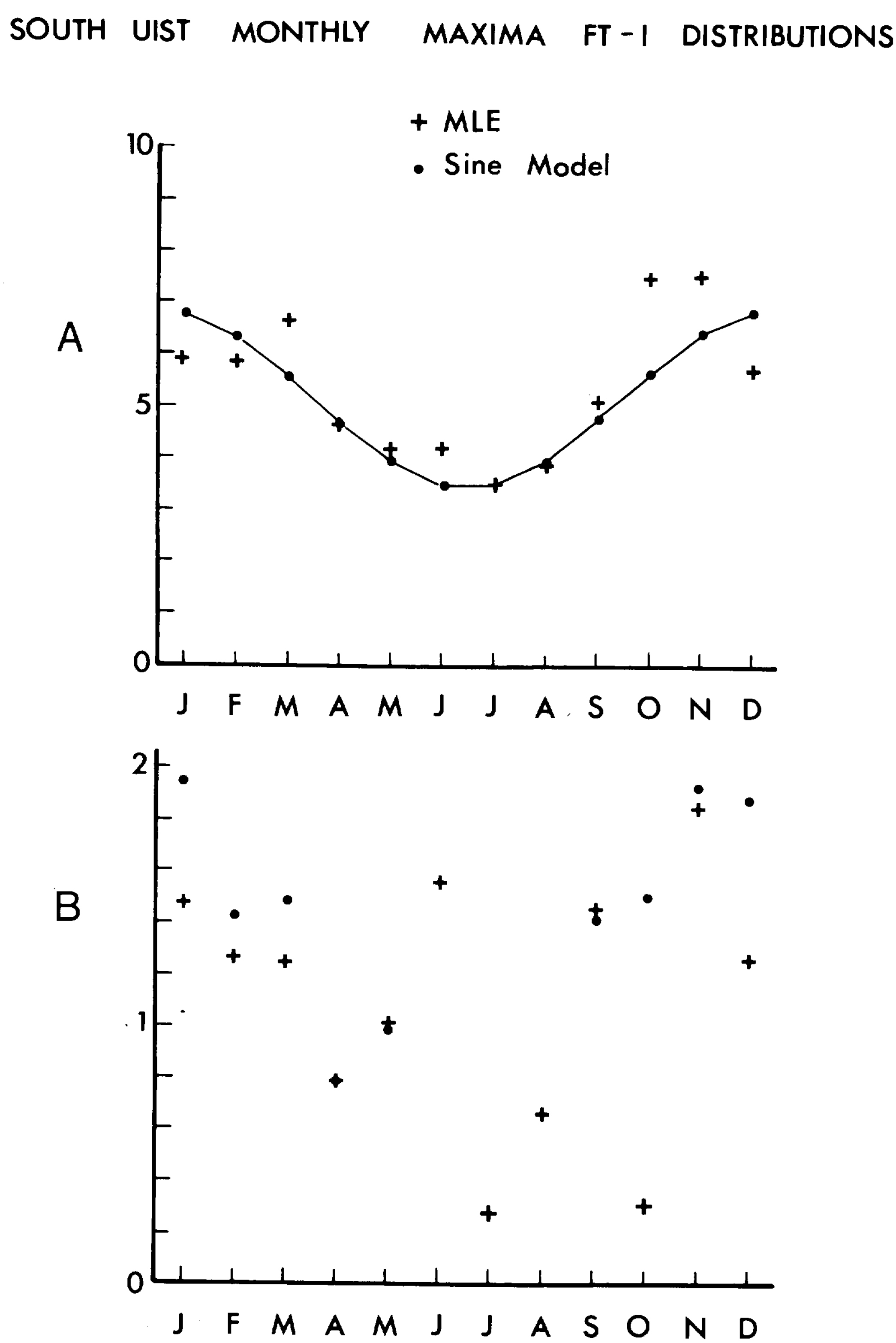

Fig. 14 Location and scale parameter of monthly FT-1 extreme value distributions from maximum likelihood estimates and the sinusoidal seasonal trend model.