INTERNAL DOCUMENT 91

I.O.S.

EXTREME WAVES:

THE EFFECT OF SEASONAL VARIATION IN WAVE CLIMATE

NATURAL ENVIRONMENT RESEARCH COUNCIL

INSTITUTE OF OCEANOGRAPHIC SCIENCES

INSTITUTE OF OCEANOGRAPHIC SCIENCES

Wormley, Godalming,
Surrey, GU8 5UB.
(042-879-4141)

(Director: Dr. A. S. Laughton)

Bidston Observatory,
Birkenhead,
Merseyside, L43 7RA.
(051-652-2396)
(Assistant Director: Dr. D. E. Cartwright)

Crossway,
Taunton,
Somerset, TA1 2DW.
(0823-86211)
(Assistant Director: M.J. Tucker)

EXTREME WAVES:

THE EFFECT OF SEASONAL VARIATION IN WAVE CLIMATE

by

M J TUCKER

Internal Document
No 91

Institute of Oceanographic Sciences
Crossway
Taunton, Somerset

April 1980

CONTENTS

SUMMARY

It was discovered a year or two ago that predicting extreme waves on a seasonal basis and then combining them into an overall extreme gave a higher answer than putting all the seasons together in the first place. Theoretically, the effect was shown to exist in principle but it could not be quantified. The present paper achieves this theoretical quantification in the general case: it is shown that the effect is maximum when all the extremes come from the same season (or month if monthly values are used), but that its maximum possible value is negligible in all practical cases. The differences found in practice arise from errors in the extrapolation of limited data.

1. INTRODUCTION

There has recently been considerable work in IOS on the subject of the effect of the variation in average wave conditions from month to month throughout the year. For example. Carter and Challenor (Ref 1) prove that this differentiation results in the expected value of the extreme wave in a given period of time being larger than it would be if the same annual average conditions were uniformly distributed throughout the year. However, they were unable to quantify the effect analytically and had recourse to practical demonstrations using 4 actual sets of ocean/meteorological data. The extreme values using month-by-month calculations all came significantly higher than those using the data lumped together, but the differences were within the confidence limits.

While reading the Carter and Challenor paper it occurred to the present author how to demonstrate the true order of magnitude of the effect using a numerical model, and this was done in internal document No 87. While trying to extend this to more subtle cases it became clear how to derive a theoretical value for the effect in the general case, and this is presented in this document.

2. THE BASIC THEORY

2.1 Statement of the problem

The notation will be largely as in Carter and Challenor except that the symbol M will be used for the total number of samples in the period for which the extreme wave is being calculated (for example, if 3 hourly wave records are taken and the highest wave in 50 years is being calculated, $M = 1.46 \times 10^5$). T is the number of years corresponding to M sampling intervals.

The mean monthly wave height varies throughout the year (see, for example Figure 1 which is from Carter and Challenor). Carter and Challenor show that if the months are considered separately and the highest wave occurring during, for example, any November in a T year period is considered, and if these monthly extreme waves are properly combined, then in principle they give a value for the extreme wave in T years which is higher than that derived by lumping all the monthly populations into one.

It is the purpose of this document to evaluate the magnitude of the difference. It is then shown that it is negligible in practical circumstances.

2.2 The maximum wave evaluated considering seasons separately

Following Carter and Challenor, we will notionally divide the year into n sections each of which is homogeneous within itself (these sections could, for example, be calendar months). The variable we are observing is x . Each section has a population π_i where $i = 1\ (1)\ n$. Let the cumulative distribution of π_i be $F_i(x)$ = probability that a random choice of the variable is less than x . Over the period of time of interest (T years), each population contributes m_i samples to the total of M samples.

Then if we take a sample size m_i from π_i , the probability of the maximum $X_{i\,max}$ being less than x is

$$P(X_{i\,max} < x) = \left[F_i(x)\right]^{m_i} \qquad 2.2/1$$

The probability of the overall maximum X_{max} being less than x is

$$P(X_{max} < x) = \left[F_1(x)\right]^{m_1} \cdot \left[F_2(x)\right]^{m_2} \cdot \ldots \left[F_n(x)\right]^{m_n} \qquad 2.2/2$$

This is Carter and Challenor's equation A.1.

We now depart from their argument.

Put $$F_i(x) = 1 - \delta_i \qquad 2.2/3$$

In the region we are concerned with δ_i is small and of order $1/m_i$ (see Section 3.1 below). δ_i is, of course, a function of x.

$$\therefore P(X_{max} < x) = (1-\delta_1)^{m_1}(1-\delta_2)^{m_2} \ldots\ldots (1-\delta_n)^{m_n} \qquad 2.2/4$$

The general term in the product series is

$$(1-\delta)^m = 1 - m\delta + \frac{m(m-1)}{2!}\delta^2 - \ldots\ \delta^m$$

$$= 1 - m\delta + \frac{m^2\delta^2}{2!} - \frac{m^3\delta^3}{3!} + \ldots$$

$$- \frac{m}{2!}\delta^2 + \frac{3m^2}{3!}\delta^3 - \frac{6m^3}{4!}\delta^4 + \ldots \qquad 2.2/5$$

$$= e^{-m\delta} - O(m\delta^2) \qquad 2.2/6$$

$$\therefore P(X_{max} < x) \simeq e^{-m_1\delta_1} e^{-m_2\delta_2} \ldots e^{m_n\delta_n}$$

$$= e^{-(m_1\delta_1 + m_2\delta_2 + \ldots\ m_n\delta_n)}$$

$$= \exp-\left(\sum_{i=1}^{n} m_i\delta_i\right) \qquad 2.2/7$$

=======================================

The error term in 2.2/6 will be considered in more detail in Section 3 below.

2.3 The maximum wave evaluated by lumping all seasons together

If all the sub-populations are put together in the proportions of their m_i values and the samples X of wave height selected at random from the whole population, then the probability of getting a sample from the i^{th} sub-population is m_i/M where $M = \sum_i m_i$

Thus $$P(X<x) = F_A(x) = \sum_i \frac{m_i}{M} F_i(x) \qquad 2.3/1$$

The largest in the sample M then has the cumulative probability

$$P\left(X_{max} < x\right) = \left[\sum_{i=1}^{n} \frac{m_i}{M} F_i(x)\right]^M \qquad 2.3/2$$

This is effectively Carter and Challenor's equation A.2, though for our purpose here it is fruitful to put it in this slightly different form.

Thus $P\left(X_{max} < x\right) = \left[\Sigma \frac{m_i}{M}\left(1 - \delta_i\right)\right]^M$

$$= \left[\Sigma \frac{m_i}{M} - \Sigma \frac{m_i \delta_i}{M}\right]^M$$

$$= \left[1 - \frac{1}{M} \Sigma m_i \delta_i\right]^M$$

$$= 1 - \Sigma m_i \delta_i + \frac{1}{2!}\left(\Sigma m_i \delta_i\right)^2 - \cdots$$

$$+ \frac{1}{M}\left[-\frac{1}{2!}\left(\Sigma m_i \delta_i\right)^2 + \frac{3}{3!}\left(\Sigma m_i \delta_i\right)^3 - \frac{6}{4!}\left(\Sigma m_i \delta_i\right)^4 \cdots\right. \qquad 2.3/3$$

$$= \exp\left(-\Sigma m_i \delta_i\right) - O\left[\frac{1}{M}\left(\Sigma m_i \delta_i\right)^2\right]$$

$$\simeq \exp\left(-\sum_{i=1}^{n} m_i \delta_i\right) \qquad 2.3/4$$

========================

Which is the same as the result from grouped populations (equation 2.2/7).

While it can be seen in a general way that the terms neglected are small, it is important to quantify them in order to make sure that they really can be neglected in all circumstances of practical interest. This is do in the following section. The final equation (2.3 3) is in fact a useful equation in its own right and again one would like to know the limits of its applicability.

3. EVALUATION OF THE DIFFERENCE IN THE HEIGHT OF THE EXTREME DESIGN WAVE DERIVED BY CONSIDERING SEASONS SEPARATELY COMPARED WITH THAT DERIVED BY LUMPING ALL SEASONS TOGETHER IN THE FIRST PLACE

3.1 Definitions of the extreme design wave

It is necessary first to consider these definitions because from them we can get an upper limit for the parameters involved.

The classical method for evaluation of the T-year design wave considers the primary population taking all seasons together and calculates the wave height which is exceeded on average once in T years: that is, in M samples.

$$\text{Probability } (X > X_M) = \frac{1}{M}$$

$$\therefore \quad P(X < X_M) = 1 - \frac{1}{M} \qquad 3.1/1$$

Put $$P(X < x) = F_A(x) = 1 - \delta_A \qquad 3.1/2$$

Thus, when $$x = X_M$$

$$\delta_A = \frac{1}{M} \qquad 3.1/3$$

Assuming that the population is homogeneous (that is, not grouped), the cumulative probability distribution of the highest wave in T years is given by

$$P(X_{MAX} < x) = F_M(x) = [P(X < x)]^M$$

$$= [1 - \delta_A]^M$$

By arguments similar to those used in Section 2.2 this gives

$$P(X_{MAX} < x) \simeq e^{-M\delta_A} \qquad 3.1/4$$

Thus, for the T year return period wave for which $\delta_A = \frac{1}{M}$ (equation 3.1/3)

$$P(X_{MAX} < X_M) = \frac{1}{e} \qquad 3.1/5$$

If $F_M(x)$ is a Fisher-Tippett 1 distribution, then this also corresponds to the mode, that is, the most likely value of the highest wave in a T years period.

The median X_N of the distribution is, of course, given by

$$P(X_{max} < X_N) = 0.5 \qquad 3.1/6$$

Putting this into equation 3.1/4 gives a corresponding value of $M\,\delta_A = 0.7$. To give a feel for the difference, in a particular real case the estimated difference between X_M and X_N is 2.5%.

The "designwave" might be defined in terms of other parameters, for example the average value of the maximum wave in a specified period, but all such likely parameters correspond to values of $P(X_{max} < x)$ which are higher than $1/e$, and therefore in general, whatever definition is used, for the value of δ_A corresponding to the extreme design wave

$$M\,\delta_A \leqslant 1 \qquad 3.1/7$$

Equation 2.3/1 shows how the probability function of the ungrouped parent population is related to those of the component groups. Using the definition of δ_A in 3.1/2 then gives

$$\begin{aligned} 1 - \delta_A &= \sum_i \frac{m_i}{M} F_i(x) \\ &= \sum_i \frac{m_i}{M} (1 - \delta_i) \\ &= 1 - \frac{1}{M} \sum m_i \delta_i \end{aligned}$$

$$\therefore \quad M\,\delta_A = \sum_i m_i \delta_i \qquad 3.1/8$$

Thus $m_i \delta_i \leqslant 1$ for the extreme design wave. 3.1/9

3.2 Evaluation of the error terms in equations 2.2/6 and 2.3/4

Consider again the general term in equation 2.2/4

$$(1-\delta)^m = 1 - m\delta + \frac{m(m-1)}{2!}\delta^2 - \cdots\cdots \delta^m \qquad 3.2/1$$

From the binomial theorem the general term in this series is

$$\frac{m(m-1)(m-2)\ \cdots\cdots\ (m-r+1)}{r!}(-\delta)^r \qquad 3.2/2$$

Expanding the top line of the quotient gives

$$m^r - m^{r-1}\left(\sum_{i=1}^{r-1} i\right) + m^{r-2}\left(\qquad\right) - \cdots$$

The term in m^{r-2} is completely negligible in the present context and our problem is to evaluate the magnitude of the effect of the second term in the final answer.

The sum of an arithmetic series is given by $\sum_{i=1}^{r-1} i = \frac{1}{2} r(r-1)$

Returning to 3.2/2 then gives the general term as approximately

$$\frac{m^r - m^{r-1} \cdot \frac{1}{2} r(r-1)}{r!}(-\delta)^r = \frac{1}{r!} m^r(-\delta)^r - \frac{1}{2} \cdot \frac{1}{(r-2)!} m^{r-1}(-\delta)^r$$

Putting this back in 3.2/1 gives

$$(1-\delta)^m \simeq 1 - m\delta + \frac{1}{2!}m^2\delta^2 - \frac{1}{3!}m^3\delta^3 \cdots + \frac{1}{r!}m^r(-\delta)^r \cdots$$

$$- \frac{1}{2} m\delta^2\left[1 - m\delta + \frac{1}{2!}m^2\delta^2 \cdots + \frac{1}{(r-2)!}m^{r-2}(-\delta)^{r-2} \cdots\right.$$

$$= e^{-m\delta}\left[1 - \frac{1}{2} m\delta^2\right] \qquad 3.2/3$$

Which is a more precise form of 2.2/6.

Equation 2.2/4 now becomes

$$P(X_{max} < x) = e^{-m_1\delta_1}\left(1 - \frac{1}{2}m_1\delta_1^2\right) . e^{-m_2\delta_2}\left(1 - \frac{1}{2}m_2\delta_2^2\right) . e \cdots$$

$$= e^{-\sum_i m_i\delta_i}\left(1 - \frac{1}{2}m_1\delta_1^2\right)\left(1 - \frac{1}{2}m_2\delta_2^2\right)\left(1 - \quad\right) \cdots$$

$$= e^{-\sum_i m_i\delta_i}\left(1 - \frac{1}{2}\sum_i m_i\delta_i^2\right) + O\left[\delta_i^2 \, m_i^2\delta_i^2\right] \qquad 3.2/4$$

It has been shown that $m_i\delta_i \ll 1$ (equation 3.1/9) so that the third term in the bracket is negligible. Going through the similar computation for the ungrouped population in Section 2.3 gives the answer.

$$P'(X_{max} < x) = e^{-\sum_i m_i\delta_i}\left[1 - \frac{1}{2M}\left(\sum_i m_i\delta_i\right)^2\right] \qquad 3.2/5$$

If we can now assume that the year can be divided into n equal parts, each homogeneous in itself, then the problem becomes a little easier. For example, for seasonal distribution $n = 4$ or for monthly distributions $n = 12$.

Then each $m_i = M/n$

Equation 3.1/8 becomes

$$\delta_A = \frac{1}{n} \Sigma \delta_i \ll \frac{1}{M} \qquad 3.2/6$$

And the error term in 3.2/4 becomes

$$-\frac{1}{2} \sum_i m_i \delta_i^2 = -\frac{1}{2} \frac{M}{n} \Sigma \delta_i^2$$

The author has consulted his colleague A G Davies who has shown that the maximum magnitude of this is given when one of the $\delta_i = n \delta_A$ and the rest = 0.

Thus, the maximum effect of grouping occurs when all maxima come from one of the groups. The error term in equation 3.2/3 then becomes

$$-\frac{1}{2} \frac{M}{n} (n \delta_A)^2 = -\frac{1}{2} M n \delta_A^2$$

Compared with $-\frac{1}{2} M \delta_A^2$ for the ungrouped population.

The difference (ungrouped minus grouped) is $\frac{1}{2} M \delta_A^2 (n-1)$ 3.2/7

As explained in 3.1, $M \delta_A \ll 1$ for the conditions of interest.

If E_M is the max difference due to grouping, then

$$E_M \ll \frac{1}{2} \cdot 1 \cdot \frac{n-1}{M} \qquad 3.2/8$$

That is $E_M \ll \frac{n-1}{2M}$

To get a numerical value for this, assume monthly grouping so that $n = 12$ and a 50 year maximum so that $M = 1.46 \times 10^5$.

Then the maximum effect on the probability is 1 part in 2.7×10^4.

This is a proportional change $\frac{\Delta P}{P}$ in the probability. To return to a specified probability requires a change $\Delta x / x$ given by

$$\frac{\Delta x}{x} = -\frac{P}{x}\frac{dx}{dP} \cdot \frac{\Delta P}{P}$$

$$= -\frac{P}{x}\frac{dx}{dP} \cdot E \qquad 3.2/9$$

Values of $\frac{P}{x}\frac{dx}{dP}$ in the extreme value region are typically 0.1, so that the maximum effect on the extreme design wave is approximately 1 part in 3×10^5.

The effect is inversely proportional to M. The smallest possible value of M corresponds to the annual extreme, when it is 2.9×10^3. Clearly the effect is still completely negligible.

4. EXECUTIVE SUMMARY AND DISCUSSION

4.1 Executive summary

The wave climate at a particular point varies with the seasons. The probability distribution of the highest wave in each season in a T-year period can be calculated separately and then combined to give the overall T-year maximum wave probability distribution. Alternatively, the seasons can all be lumped together in the first place before the extremes are calculated. Do estimates of the T-year maximum design wave derived from the two methods differ?

The relationship of the probability distributions of the maximum wave to that of the parent populations is derived for both processes and is found to be the same (equations 2.2/7 and 2.3/4). This is a quite general result and makes no assumptions at all about the parent distributions. However, to reach these equations small terms are neglected.

In Section 3 the derivations are repeated carrying through the small terms and the result for the method taking account of the seasons is given in equation 3.2/4. This again is a perfectly general result.

Assuming that the year is divided into a number n of equal seasons, it is then shown that the proportional difference in the probability in the region of the extreme design wave values is (equation 3.2/8)

$$\bar{E}_n \leqslant \frac{n-1}{2M}$$

where M is the number of independent values of wave height.

This simple result is again completely general. It can be converted into the equivalent difference in design wave value using equation 3.2/9. The differences are shown to be negligible in all practical cases.

4.2 Comparison with a numerical model

Internal document No 87 set up a numerical wave climate model arranged to produce the maximum possible differences, which were then calculated.

Putting the corresponding numbers into equation 3.2/8 gives the proportional difference in probability as

$$E_n \leqslant \frac{9}{2 \times 1.46 \times 10^5}$$

$$= 3.08 \times 10^{-5}$$

Putting the appropriate values into equation 3.2/9 allows this to be converted to proportional errors in the design wave

$$\frac{\Delta x_M}{x_M} \leq 3.17 \times 10^{-6}$$

This compared with the value found in the model of $\frac{38}{26} \times 10^{-6} = 1.50 \times 10^{-6}$

The above inequalities become equations if the 50 year return-period wave is considered whereas the model calculated the median 50 year maximum wave. The agreement is therefore satisfactory.

It is, in fact, possible to go back to the exact equation 3.2/7. The value of δ_A corresponding to the median value is 4.72×10^{-6}. Putting this into the equations gives $\Delta x_M / x_M = 1.51 \times 10^{-6}$. This agreement is within the accuracies of the figures calculated for the numerical model.

4.3 What, then, is the reason for the differences found in practice? These arise from the problems of extrapolating from short data series.

The standard process of extrapolation plots the measured cumulative probability of the wave measurements on axes which will produce a straight line if a certain postulated distribution formula is valid. Several formulae are tried and the one giving the best fit is used. This is then extrapolated to the probability corresponding to the T-year return period.

There is a great deal of judgement involved in this process. Clearly the most relevant data points are those for the highest waves, but the higher the waves the fewer the measurements and the greater the scatter. We do not know enough about the physics involved to predict the shape of the curve.

In practice, the longest data set available (Seven Stones Light Vessel) fits a Fisher-Tippett 1 formula quite well. Plotted month-by-month there is less data but it is difficult to say that each month does not also fit such a formula, and this has been done. However, this is mutually incompatible since one cannot combine monthly distributions following FT-1 formulae into an annual distribition with an FT-1 formula. One or other (or all) must be wrong. Forcing them all to fit produces errors of the sense and magnitude actually found by Carter and Challenor.

The way in which monthly or seasonal distributions should be handled to produce results consistent with the annual distribution will be discussed in another report.

5. REFERENCE

CARTER D J T and CHALLENOR P G. "Estimating return values of environmental parameters" 1980. Private communication

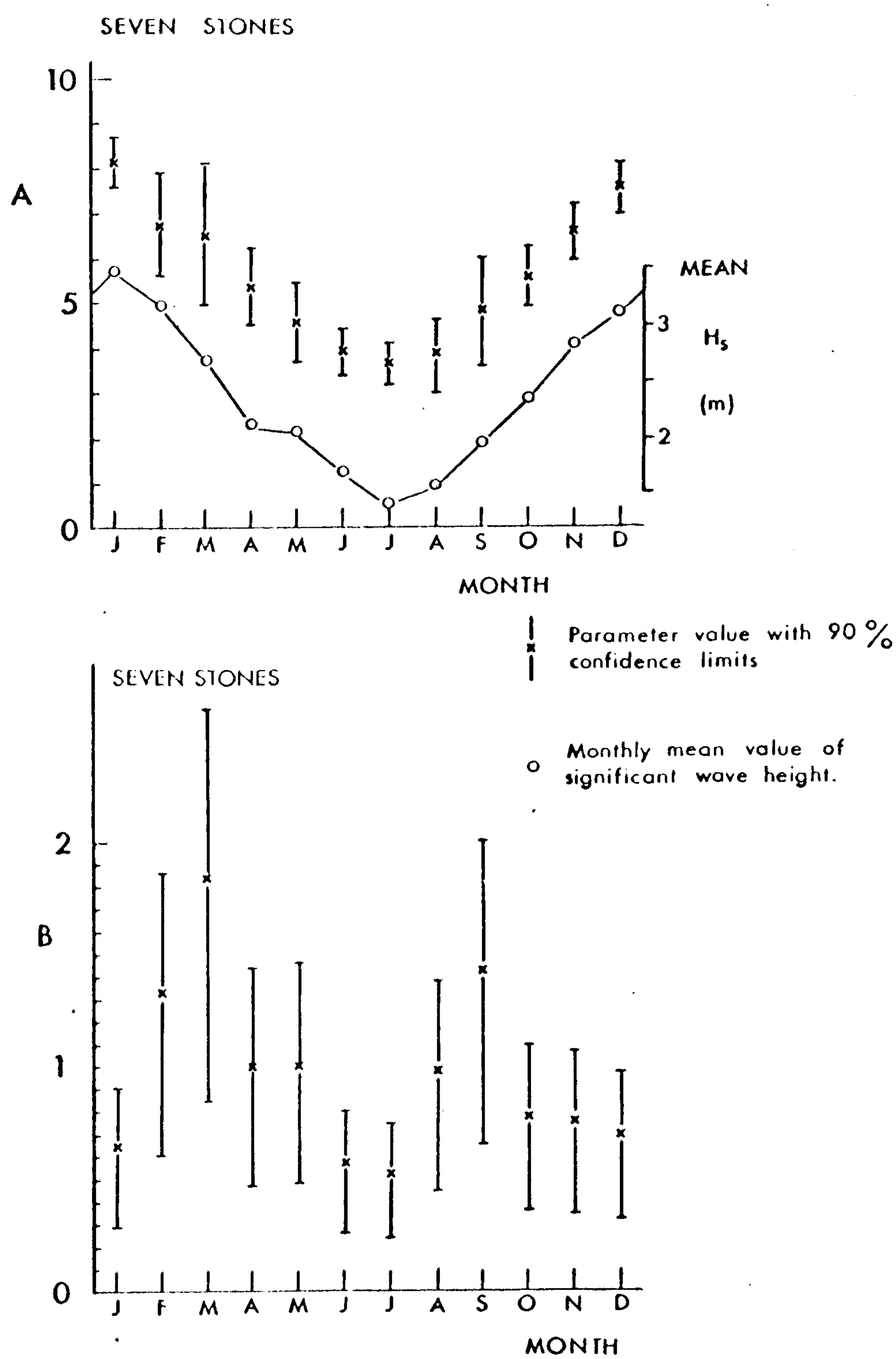

FIG. 1. Month-to-month variations in the values of the Fisher-Tippet Type 1 distribution parameters A & B defined by equation 1.

FIGURE 1