file

INTERNAL DOCUMENT 87

I.O.S.

THE 50 YEAR EXTREME WAVE

Demonstration that the variation of average wave conditions month by month has no significant effect

[This document should not be cited in a published bibliography, and is supplied for the use of the recipient only].

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)

THE 50 YEAR EXTREME WAVE

Demonstration that the variation of average wave conditions month by month has no significant effect

by

M J TUCKER

Internal Document
No 87

Institute of Oceanographic Sciences
Crossway
Taunton, Somerset

March 1980

CONTENTS

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/met 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.

The present writer has felt for some time that it was difficult to believe that the differences could be significant, but his arguments were rather woolly and he has been searching for a way of demonstrating this straightforwardly. Reading the paper by Carter and Challenor suggested such a way, and this demonstration is given below.

2. THE PRINCIPLE OF THE DEMONSTRATION

A set of 5 years of Hmax (3hr) values calculated from measured values of Hs at the Seven Stones Light Vessel is taken, and a probability distribution curve is fitted to the whole of it lumped together. This analytical expression is then used to calculate the 50 year extreme wave on two different bases.

(1) Using all the data in the conventional way.
(2) Assuming that the highest 10% of values all occur randomly in a given 36.5 day period of the year, and that the lowest 90% of the data occur randomly in the rest of the year. This is an extreme case of grouping, but is not an enormous simplification of what actually happens (see figure 1A reproduced from Carter and Challenor, loc cit). It will presumably somewhat exaggerate any effect from the slightly less extreme grouping which actually takes place.

Note that Carter and Challenor used the actual primary data to do their calculations, which presumably results in sampling error differences between the equivalent results, and the present writer's thesis is that this produces the differences found in their examples which are therefore spurious. The present method avoids this random sampling error completely in so far as differences between the grouped and ungrouped results are concerned.

Note also that the author has used Hmax (3hr) for convenience since this data set is conveniently presented for the present purposes by Fortnum and Tann (Ref 2). For practical purposes, Hmax (3hr) $\simeq$ constant x Hs for a given data set.

3.0 THEORY

The Seven Stones data fit a Fisher-Tippett I distribution reasonably well (Fortnum and Tann loc cit) giving the formula for the cumulative probability

$$P(x < x) = \exp\left[-\exp(-ax + b)\right] \qquad 3.0/1$$

where x = Hmax (3hr)

a = 0.54 m^{-1}

b = 1.86

(As a matter of passing interest, 10% of the values lie above 7.61 m.)

The median value of the extreme distribution will be used because it greatly simplifies the theory. For large samples this bears effectively constant ratios to the mode and the mean. Symbols will be the same as those in Carter and Challenor so far as is practicable.

3.1 All values selected randomly from the whole population (ungrouped)

The chance that all of n^* selections will be below x is

$$P(X_{max} < x) = \left\{\exp\left[-\exp(-ax + b)\right]\right\}^{n^*}$$

$$= \exp\left[-n^* \exp(-ax + b)\right]$$

The median X_m of this distribution is given when $P(X_{max} < X_m) = 0.5$

or $$\log_e 0.5 = -n^* \exp(-a X_m + b)$$

$$\therefore \quad \exp(-a X_m + b) = \frac{0.69314718}{n^*}$$

$$\therefore \quad X_m = \frac{\log_e 1.4426950\, n^* + b}{a} \qquad 3.1/1$$

3.2 Population grouped

The distribution is divided at $P(X<x) = 1 - \frac{m}{n^*}$. m samples are then taken from the upper part and $n^* - m$ are taken from the lower part. It is clear that the extreme value must be in the first part of the sample which is therefore the only part which need be considered here. Its cumulative probability distribution is

$$P^1(X<x) = 10\left\{\exp\left[-\exp(-ax+b)\right]-0.9\right\} \text{ for } x > X_1$$
$$= 0 \quad \text{for } x < X_1$$

where X_1 is the lower limit of x for this part of the distribution.

The chance that all m selections will be below x is

$$P(X'_{max} < x) = \left(10\left\{\exp\left[-\exp(-ax+b)\right]-0.9\right\}\right)^m$$

(This can be seen from Carter and Challenor's equation A 1)

The median X'_m of this distribution is given when $P(X_{max} < x) = 0.5$

or $$(0.5)^{1/m} = 10\left\{\exp\left[-\exp(-ax+b)\right]-0.9\right\} \qquad 3.2/1$$

4.0 NUMERICAL EXAMPLES

4.1 Small sample

The author's qualitative reasoning (Appendix 1) indicates that the differences in extremes are only significant for small samples (small values of n^*) and decrease rapidly as the sample size is increased. To demonstrate that the above theory gives a real difference (that is, that we have not in some way chosen a type of grouping which fortuitously produces no difference) the calculations will first be done for a very small sample: $n^* = 10$ and $m = 1$.

From equation 3.1/1 putting in values from equation 3.0/1

$$X_m = \frac{\log_e 14.4269 + 1.86}{0.54} \text{ metres}$$

$$= 8.387 \text{ metres} \qquad 4.1/1$$

From equation 3.2/1

$$0.5 = 10\left\{\exp\left[-\exp(-aX'_m+b)\right]-0.9\right\}$$

or $$\exp(-a x'_m + b) = -\log_e 0.95 = .0512933$$

$$a x'_m - b = -\log_e .0512933$$

$$= 2.97020$$

Putting in values of a and b from equation 3.0/1 gives

$$X'_m = 8.949 \text{ metres} \qquad 4.1/2$$

Thus, a significant difference is, in fact, produced and the value from the grouped distribution is higher, as expected.

4.2 The 50 year maximum value

For the present purpose it will be assumed that the 3 hour samples of Hmax (3hr) are independent, but this assumption is not critical.

Thus $n^* \simeq 8 \times 365 \times 50 = 1.46 \times 10^5$

$m \simeq 1.46 \times 10^4$

4.21 Population ungrouped

From equations 3.0/1 and 3.1/1

$$X_m = \frac{\log_e (1.4426950 \times 1.46 \times 10^5) + 1.86}{0.54}$$

$$= 26.144213 \qquad 4.21/1$$

4.22 Population grouped

From equation 3.0/1 and 3.2/1

$$(0.5)^{\frac{1}{1.46 \times 10^4}} = 10\left\{\exp\left[-\exp(-0.54 X'_m + 1.86)\right] - 0.9\right\}$$

$$\therefore \exp\left[-\exp(-0.54 X'_m + 1.86)\right] = 1 - 4.74770 \times 10^{-5}$$

$$\therefore \exp(-0.54 X'_m + 1.86) = 4.74758 \times 10^{-5}$$

$$= \frac{1.86 + 12.257896}{0.54} \text{ metres}$$

$$= 26.144252 \text{ metres} \qquad 4.21/2$$

(Fourteen significant figures were actually carried through the calculation so that the two final figures are correct to the accuracy given.)

It is interesting to note that the value of Hmax (3hr) corresponding a return period of 50 years obtained by Fortnum and Tann (loc cit) from this data set varied over a range of about 2 m depending on the method used, but the most comparable value was 25.1 m.

5. DISCUSSION

The difference between the 50 year maximum waves calculated from the grouped and ungrouped data is negligible.

The author is aware that one or two assumptions in the above calculations are questionable. The most obvious is that the 3 hourly values of Hmax are independent. However, even reducing the number of effectively independent events by a factor of 10, say, would not significantly affect the conclusion.

The only other questionable assumption which the author can see is the form assumed for the grouping. It is just possible that the sharp division used is in some way less effective in affecting the maximum than the smoother grouping in the real world. The author cannot see why this should be so, and unless and until such an effect can be proved, the author feels he has produced a reasonably convincing demonstration that the grouping effect is negligible when determing the expected maximum wave height for practical purposes.

6. REFERENCES

1. CARTER D J T and CHALLENOR P G. "Estimating return values of environmental parameters" 1980. Private communication
2. FORTNUM B C H and TANN . "Waves at Seven Stones Light Vessel (50° 04' N, 06° 04' W) January 1968 to June 1974" IOS Report No 39, 1977.

APPENDIX 1

Qualitative argument to show why one would expect the grouping effect to be negligible when dealing with large samples.

One has first to try to gain a picture of how the effect comes about, and for this it is useful to consider the type of grouping assumed in the main paper: that is, that the top 10% of the population all occur in a fixed $35\frac{1}{2}$ day period of the year. In effect we have 2 urns:

A contains the lowest 90% of the population
B contains the highest 10% of the population
X_1 is the value of the variable dividing the two populations

If we make 10m selections, precisely m come from urn B.

The difference becomes clearest if we put m = 1. Then our sample consists of 9 random selections from urn A and 1 from urn B. This constrains the maximum to be always greater than X_1, and its average value is the average of the population in urn B.

Suppose we now mix all the population together and take 10 selections at random. There is now a 35% chance that the maximum will be below X_1, and a roughly similar chance that 2 or more will be taken which are greater than X_1. In the latter case, only the biggest of these will of course be the maximum of that sample. Thus when we take many samples 10 at a time and list their maximum values, there will be some below X_1, and some of the individual selections whose value is above X_1 will be exceeded by others in the same sample and not be listed. It is clear that the nett effect relative to the first case is to replace the unlisted values exceeding X_1 by the values below X_1, thus reducing the mean value.

Thus, the effect depends on the probability of the maximum value coming from urn A in the ungrouped case, and this rapidly decreases as the sample size is increased. The computation of this probability is of course, simple in the example used. The chance of selecting a value below X_1 is 0.9. The chance of all of n* values being below X_1 is $(0.9)^{n^*}$. This gives the following probabilities:

Sample size	Probability of max X_1
10	.349
100	2.67×10^{-5}
1000	10^{-45}

Thus, for large samples the maximum effectively always comes from that part of the population which is above X_1, and as the sample gets larger, the probability of finding such values clearly depends less and less on whether the population is or is not grouped.

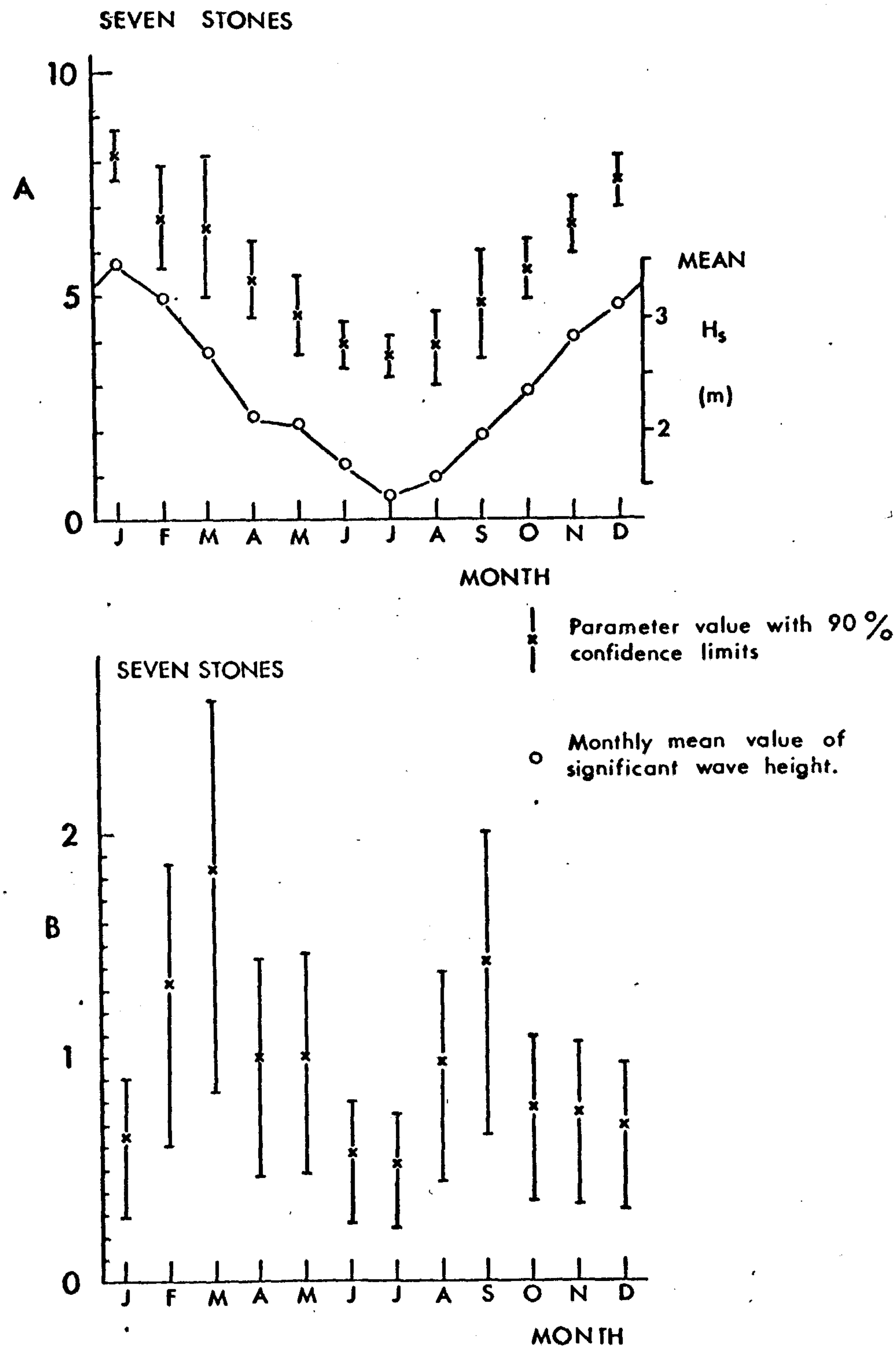

FIG. 1. Month-to-month variations in the values of the Fisher–Tippet Type I distribution parameters A & B defined by equation 1.