CHAPTER 48

STATISTICAL PROPERTIES OF THE MAXIMUM RUN OF IRREGULAR SEA WAVES

by Akira Kimura

ABSTRACT

The probability distribution of the maximum run of irregular wave height is introduced theoretically. Probability distributions for the 2nd maximum, 3rd maximum and further maximum runs are also introduced. Their statistical properties, including the means and their confidence regions, are applied to the verification of experiments with irregular waves in the realization of a "severe sea state" in the test.

1. INTRODUCTION

The importance of bringing the statistical nature of irregular sea waves into the design concepts of coastal structures has been widely recognized. Hydraulic tests with irregular waves in a wave tank have become one of the standard techniques to investigate dynamic properties of coastal and ocean structures. Simulation and generation techniques of irregular sea waves have been developed and applied for this purpose.

The group nature of irregular sea waves, on the other hand, has been clarified recently. Since there is a correlation between consecutive wave heights (and period), high waves (or low waves) tend to arrive in groups. This means that high waves tend to arrive intermittently in groups and that there are long durations between them, in which low waves consecutively arrive. Irregular wave simulation techniques are mainly directed to simulate the averaged properties of irregular sea waves, i.e. wave spectrum, probability distribution for wave height and period, etc. A long term steady state for these properties is designed to be simulated in the techniques except for the simulation of definite wave profiles. However, those properties measured in a short duration are not always guaranteed to be the same as the designed ones but rather distribute around them statistically. Therefore, hydraulic tests which are carried out in durations in which no high wave run arrives may bring about an under-estimation of wave effects. Insufficient estimation from the test, especially under-estimation, may induce deficiencies in the design for prototype structures. Model tests always have a possibility of missing the temporal "severe sea state" if the test is not sufficiently long. However, there have, in fact, been no established criterion used in

Associate Prof. Tottori Univ., Koyama Minami 4-101, Tottori, Japan

the verification of test results regarding whether the severe sea state is realized during the test. In addition to ordinary criteria, then, a new criterion which can be a measure for the temporal state of irregular sea waves becomes necessary and the property of the high wave group, especially its maximum value, may be suitable for this. This study aims at clarifying the statistical properties of the maximum run of high waves which may be a good measure for the temporal "severe sea state". The statistical properties of the 2nd maximum, 3rd maximum, and further maximum runs of high waves are also clarified and confidence regions for their means are given and explained in the examination of the effectiveness of the model test in this study.

2. PROBABILITY DISTRIBUTION OF THE MAXIMUM RUN OF HIGH WAVES

The author has introduced the probability distribution of the maximum run of high waves to apply for "short term overtopping" (Kimura,1984). It can be summarized as follows:

Assuming the time series of wave height to be the Markov chain, the probability distribution of a run of waves which exceed the threshold wave height (h_t) consecutively is given as (Kimura,1980)

$$P_1(\ell)=p^{\ell-1}(1-p) \tag{1}$$

in which p is given by

$$p=\frac{\int_{h_t}^{\infty}\int_{h_t}^{\infty} P_0(h_1,h_2)dh_1dh_2}{\int_{h_t}^{\infty} Q_0(h_1)dh_1} \tag{2}$$

P_0 is the two dimensional Rayleigh distribution while Q_0 is the Rayleigh distribution. These are given as follows (wave height is normalized by the mean wave height throughout this study):

$$P_0(h_1,h_2)=h_1h_2/A\cdot I_0(h_1h_2\rho/A)\exp\{-(h_1^2+h_2^2)/\pi A\} \tag{3}$$

$$Q_0(h_1)=\pi h_1/2\ \exp(-\pi h_1^2/4) \tag{4}$$

$$A=4/\pi^2-\rho^2 \tag{5}$$

in which I_0 is the modified Bessel function of the 1st kind, and ρ is a correlation parameter between consecutive wave heights. This has the following relation with a correlation coefficient of consecutive wave heights.

$$\gamma_h=\{E(\pi\rho/2)-(1/2)(1-\pi^2\rho^2/4)K(\pi\rho/2)-\pi/4\}/(1-\pi/4) \tag{6}$$

K and E are the complete elliptic integrals of the 1st and 2nd kinds, respectively.

The probability distribution of the maximum run in a population of high wave runs of size N is introduced as follows:

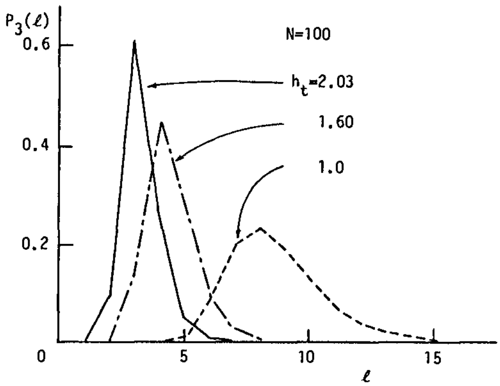
The probability that the run is equal to or smaller than $\ell-1$ is given as

$$P_2(\ell)=\sum_{\ell'=1}^{\ell-1} p^{\ell'-1}(1-p)=(1-p^{\ell-1}) \tag{7}$$

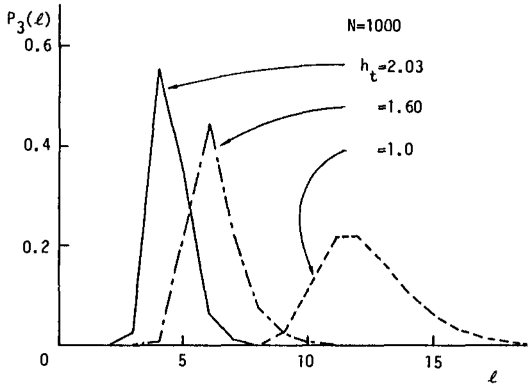
Therefore, all runs in the population are equal to or smaller than ℓ - 1 giving $(1-p^{\ell-1})^N$, assuming all runs to be independent. On the other hand, the probability that at least one or more runs in the population are equal to or greater than ℓ is given as $1-(1-p^{\ell-1})^N$. In the same way, the probability that at least one or more runs are equal to or greater than $\ell+1$ is given as $1-(1-p^{\ell})^N$. The probability that the maximum run becomes ℓ is, therefore, given as

$$P_3(\ell)=(1-p^{\ell})^N-(1-p^{\ell-1})^N=\exp\{N\cdot\ln(1-p^{\ell})\}-\exp\{N\cdot\ln(1-p^{\ell-1})\} \tag{8}$$

in which ln is the natural logarithm.
Equation (8) is shown in Fig.1; (a) N=100 and (b) N=1,000. The 1/10 maximum wave (solid line), the significant wave (chain line) and the mean wave (broken line) heights are used for h_t respectively.



(a) N=100



(b) N=1000

Fig.1 Probability distribution of the maximum run of high waves

Example 1 (duration and number of runs)

Giving a duration and wave properties, the probability distribution of the maximum run of high wave is given as follows:
An expected number of high wave runs N within a duration D is given by

$$N = D / \ell_t T_m \tag{9}$$

in which ℓ_t is a mean of the total run (recurrence interval of the runs) of high waves and T_m is a mean wave period. We assume here a steady sea state, its spectrum to be the Pierson-Moskowitz type (for simplicity this spectrum is used throughout this study ; the spectrum gives affect on the correlation parameter ρ in eq.3, Battjes and van Vledder,1984), h_t=1.6 (significant wave height) and T_m=10s. The averaged value of the total run is given by (Kimura,1980),

$$\ell_t = 1/(1-p') + 1/(1-p) \tag{10}$$

in which p is given by eq.(2) and p' is given by

$$p' = \frac{\int_0^{h_t} \int_0^{h_t} P_0(h_1,h_2)dh_1dh_2}{\int_0^{h_t} Q_0(h_1)dh_1} \tag{11}$$

Substituting ρ =0.247 (Pierson-Moskowitz spectrum; Kimura,1980) and h_t=1.6, we obtain ℓ_t=10.2. The expected number of high wave runs is calculated as

$$N = D / 10.2 \times 10 \tag{12}$$

If we use 1 week for the duration (D=604800s) we obtain N=5929. Substituting values of p and N into eq.(8), we obtain the probability distribution of the maximum run of wave height (h > significant wave height) which appears once a week. The relations between h_t and ℓ_t for the representative wave heights are listed in Table-1.

Table-1 h_t and ℓ_t

h_t	ℓ_t
1.0 (H_{mean})	4.7
1.60 ($H_{1/3}$)	10.2
2.03 ($H_{1/10}$)	30.1

3. MEAN OF THE MAXIMUM RUN OF HIGH WAVES

From eq.(8), the mean of the maximum run is introduced as

$$\sum_{i=1}^{N} {}_NC_i \frac{(-1)^{i+1}}{1-p^i} \tag{13}$$

Since this equation is difficult to carry out in practice when N and i are large, the following approximation is applied (Longuet-Higgins,

1952).
Applying ℓ_0, which is given by the equation $N \cdot p^{\ell_0}=1$, $(1-p^{\ell})^N$ in eq.(8) is written as

$$(1-p^{\ell})^N = (1-p^{\ell-\ell_0}/N)^N$$

Since p<1, for large values of N the above equation reduces to

$$(1-p^{\ell-\ell_0}/N)^N \fallingdotseq \exp(-p^{\ell'}) \tag{14}$$

in which $\ell' = \ell - \ell_0$

and $$\ell_0 = -\ln(N)/\ln(p) \tag{15}$$

The general form of eq.(14) is shown in Fig.2 (solid line). If N is sufficiently large, eq.(14) asymptotically approaches to

$$\exp(-p^{\ell'}) \fallingdotseq \begin{cases} 0 & \ell' < \ell_*' \\ 1 & \ell' > \ell_*' \end{cases} \tag{16}$$

in which ℓ_*' is a point which makes equal the upper and lower shadowed portions between eq.(14) and eq.(16) (Fig.2),

$$\begin{aligned} \ell_*' &= \int_0^{\infty} 1-\exp(-p^{\ell'})d\ell' - \int_{-\infty}^{0} \exp(-p^{\ell'})d\ell' \\ &= -\gamma/\ln(p) \end{aligned} \tag{17}$$

in which γ is Euler's constant(=0.5772...). In the same way, $(1-p^{\ell-1})^N$ in eq.(8) is given approximately as

$$(1-p^{\ell'-1}/N)^N \fallingdotseq \exp(-p^{\ell'-1}) \fallingdotseq \begin{cases} 0 & \ell' < \ell_*'+1 \\ 1 & \ell' > \ell_*'+1 \end{cases} \tag{18}$$

From the definition, the mean of the maximum run is given as

$$\ell_{1*} = \sum_{\ell=1}^{\infty} \ell P_3(\ell) \tag{19}$$

From eqs.(16) and (18),

$$\ell_{1*} = \sum_{\ell=\ell_0+\ell_*'}^{\ell_0+\ell_*'+1} \ell \tag{20}$$

Since $\ell_0+\ell_*'$ is not an integer in general, ℓ_{1*} is the integer between $\ell_0+\ell_*'$ and $\ell_0+\ell_*'+1$. Extending the definition for ℓ into real numbers and approximating the probability of ℓ between $\ell_0+\ell_*'$ and $\ell_0+\ell_*'+1$ with eqs.(16) and (18), we obtain an expected mean of the maximum run as

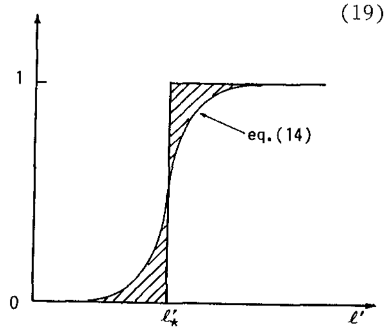


Fig.2 Approximation of eq.(14)

$$\ell_{1*} = \ell_0 + \ell_*' + 1/2 = -(\ln(N) + \gamma)/\ln(p) + 1/2 \tag{21}$$

If we apply eq.(1) for the probability of ℓ in the same interval instead, we obtain,

$$\ell_{1*} = \ell_0 + \ell_*' + \int_{\ell_0+\ell_*'}^{\ell_0+\ell_*'+1} \ell\, p^{\ell-1}(1-p)d\ell \Big/ \int_{\ell_0+\ell_*'}^{\ell_0+\ell_*'+1} p^{\ell-1}(1-p)d\ell$$

$$= -(\ln(N) + \gamma)/\ln(p) - 1/\ln(p) + p/(p-1) \qquad (22)$$

Figure 3 shows the relation between ℓ_{1*} and N. The approximation of eq.(22) is used in the calculation.

Example 2 (mean of the maximum run)

If we use the same conditions applied in Example 1 (h_t=1.6 ; significant wave height, T_m=10s), the expected mean of the maximum runs during 1 day, 1 week and 1 month are calculated as follows.

Table-2 D,N and ℓ_{1*}

duration	number of waves	number of runs	maximum run (eq.22)
1 day	8640	847	6.0
1 week	60480	5929	7.5
1 month	259200	25412	8.6

The number of runs is given by dividing the number of waves by the mean of the total run (10.2; h_t=1.6).

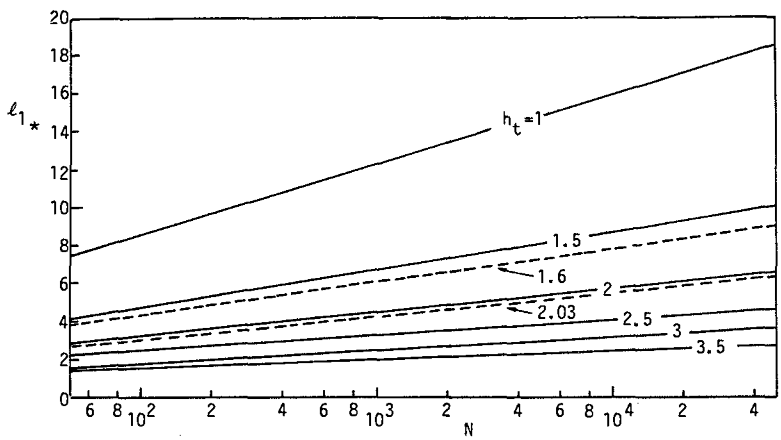


Fig.3 Expected mean of the maximum run

4. VERIFICATION OF THE EXPERIMENTAL RESULT

As explained in the introduction, high waves tend to arrive in groups. Therefore, if a short duration for the model test is applied, the possibility of missing high wave runs during the test may be large. Since a high wave or its group may actually have the greatest effect on coastal structures, the reliability of the model test is highly affected by the frequency of high waves and high wave runs which appear during the test. If a long duration is applied for the test, the statistical properties of the high wave run can be investigated by the ordinary theory (Kimura,1980). However, if a short duration is applied, the verification of the test is difficult due to an insufficient number of runs. When tests with 100 irregular waves are carried out, for example, we can investigate the ordinary properties such as the power spectrum, wave height distribution, etc. However, it may be difficult to investigate the results regarding the severity of the sea state for the model, since only 10 (roughly) high wave runs are expected if we apply the significant wave height for h_t. In this section, for verification of the efficiency of the model test which is short in duration, the ordered statistics of the run of high waves are investigated.

In the population of N independent high wave runs of $h > h_t$, the probability that i runs are equal to or greater than ℓ and that the other N-i runs are equal to or smaller than $\ell-1$ is given as

$$P_G(\ell)^i \cdot P_L(\ell)^{N-i} \tag{23}$$

in which P_G and P_L are the probabilities that a run is greater than or equal to ℓ and equal to or smaller than $\ell-1$ respectively,

$$\begin{aligned} P_G(\ell) &= \sum_{\ell'=\ell}^{\infty} p^{\ell'-1}(1-p) = p^{\ell-1} \\ P_L(\ell) &= \sum_{\ell'=1}^{\ell-1} p^{\ell'-1}(1-p) = (1-p^{\ell-1}) \end{aligned} \tag{24}$$

There are ${}_NC_i$ combinations to select i runs among N runs. Therefore, the probability that all i runs selected are equal to or greater than ℓ and that the remaining N-i runs are equal to or smaller than $\ell-1$ is given as

$${}_NC_i \cdot P_G(\ell)^i \cdot P_L(\ell)^{N-i} \tag{25}$$

The probability that at least n+1 (n=1,2,) runs are equal to or greater than ℓ is given as

$$1 - \sum_{i=0}^{n} {}_NC_i \cdot P_G(\ell)^i \cdot P_L(\ell)^{N-i} \tag{26}$$

In the same way, the probability that at least n+1 runs are equal to or greater than $\ell+1$ is given as

$$1 - \sum_{i=0}^{n} {}_NC_i \cdot P_G(\ell+1)^i \cdot P_L(\ell+1)^{N-i} \tag{27}$$

Consequently, the probability that the (n+1)-th maximum run (n=1,2, .) equals to ℓ is given as

$$P_{n+1}(\ell)= \sum_{i=0}^{n} {}_N C_i \{ p^{i\ell}(1-p^{\ell})^{N-i}-p^{i(\ell-1)}(1-p^{\ell-1})^{N-i}\} \tag{28}$$

The distribution function of P_{n+1} is given as

$$F_{n+1}(\ell) = \sum_{i=0}^{n} {}_N C_i \cdot p^{i\ell}(1-p^{\ell})^{N-i} \tag{29}$$

(The above equations also hold for n=0 ; maximum run.)
Applying the same approximation as in eqs.(16) and (18), the mean of the (n+1)-th maximum run (n=1,2,,,) is given as

$$\ell_{(n+1)*}= \ell_{1*} - \sum_{i=1}^{n} {}_N C_i \{ \sum_{\ell = \ell_{1*}}^{\infty} \left(\frac{p^{\ell}}{1-p^{\ell}}\right)^{i} \} \tag{30}$$

$$(n \geqq 1)$$

in which ℓ_{1*} is given by eq.(21) or (22). The definition for ℓ is extended into real numbers again in this equation. The summation with respect to ℓ is taken for the values at $\ell = \ell_{1*}$, $\ell_{1*}+1$, $\ell_{1*}+2$, and so on.
If measured values of $\ell_{(n+1)*}$ (n=1,2,3,...) fall within a significant part of the distribution (eq.28), the experimental data may be judged to be statistically significant. When we apply the 95%*) or the 99%*) confidence intervals, we may consider the data to be statistically significant if $\ell_{(n+1)*}$ falls within the intervals between $\ell_{(n+1)*}-\ell_c$ and $\ell_{(n+1)*}+\ell_c$. ℓ_c is a half of the confidence interval for the given confidence coefficient and is given by solving eq.(29) implicitly .
* (traditional standards are adopted, i.e. 95% and 99%.)

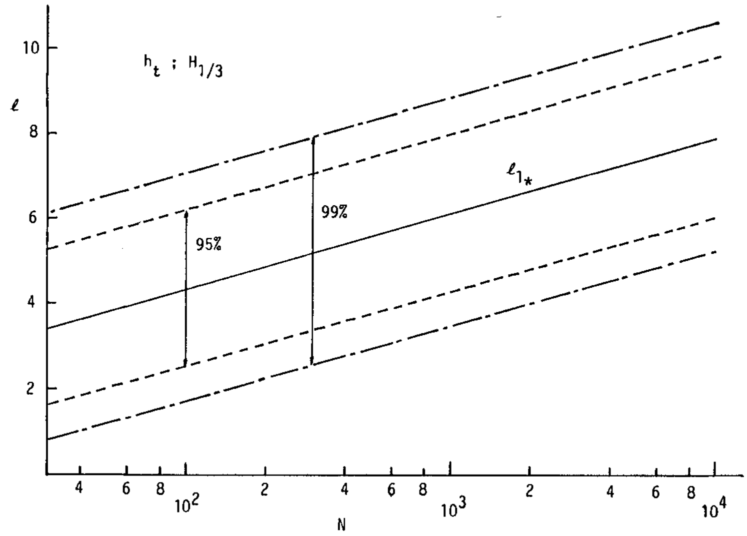
$$\left| \sum_{i=0}^{n} {}_N C_i \; p^{i\ell}(1-p^{\ell})^{N-i} \right|_{\ell_{(n+1)*}-\ell_c}^{\ell_{(n+1)*}+\ell_c} = 0.95 \text{ or } 0.99 \tag{31}$$

$$(n=0,1,2,..)$$

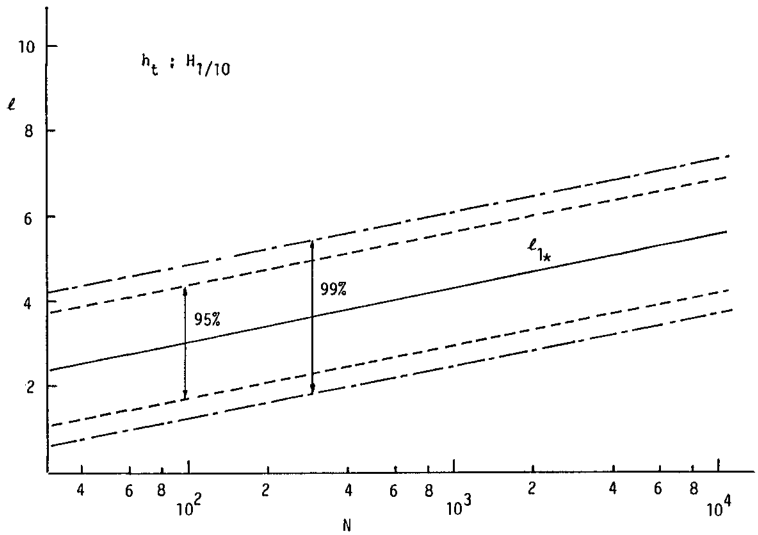
To have an exact value for ℓ_c in the above equation, the definition for ℓ has to be extended into the real numbers.
Figures 4 (a) and (b) show the changes in the confidence regions of the maximum run with respect to N when (a) the significant wave height and (b) the 1/10 maximum wave height are used for h_t respectively. The dotted line is for the 95% and the chain line is for the 99% confidence regions respectively. Figures 5 (a) and (b) show the changes in the confidence regions of the maximum run with respect to h_t when (a) N=50 and (b) N=500 are used respectively.

Example 3 (statistically significant run)

When N=100 and h_t=1.6, we can see from Fig.4(a) that the number of the maximum run in the data must fall between 3 and 6 if we apply the 95% confidence interval or 2 and 7 for the 99% confidence interval.

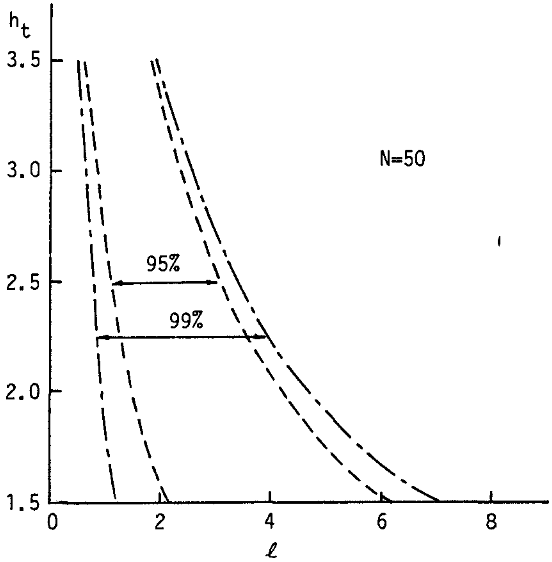


(a) h=1.60

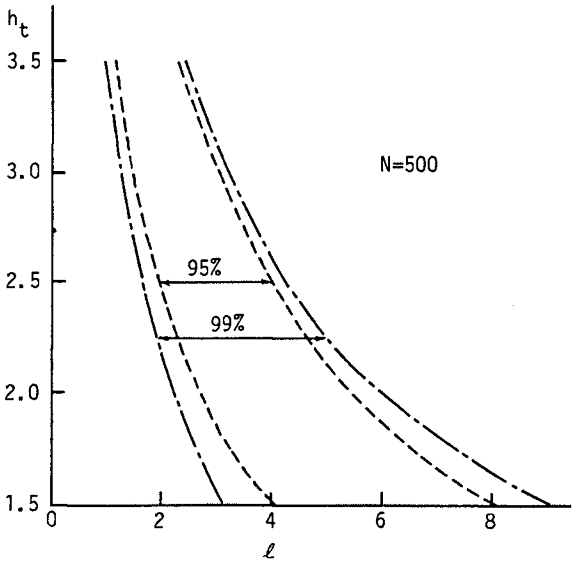


(b) h=2.03

Fig.4 Change in the confidence regions (maximum run)



(a) N=50



(b) N=500

Fig.5 Change in the confidence regions (maximum run)

Table-3 Measured runs of wave height in a numerical simulation ($h \geq 1$)

length	total	frequency	eq.(1)
1	22	0.44	0.467
2	10	0.20	0.249
3	6	0.12	0.133
4	3	0.06	0.071
5	4	0.08	0.038
6	1	0.02	0.020
7	0	0	0.011
8	3	0.06	0.006
9	1	0.02	0.003

Example 4 (verification of the experiment)

Table-3 shows the measured runs of wave height which exceed the mean wave height ($h_t=1$) obtained in the numerical simulation of the Pierson-Moskowitz spectrum. Figure 6 shows the comparison of data and the theoretical probability distribution of the run (eq.1). The agreement between the data and eq.(1) in the region $\ell \leq 4$ is fairly good. However, the verification of eq.(1) in the region $\ell \geq 5$ is difficult due to the fluctuation of data. This is mainly due to the insufficient amount of data. Instead of making a long measurement to obtain a sufficient amount of data, the following verification for the data is possible applying the ordered statistics explained so far. White circles in Fig.7 show the run of wave height in descending order from its 1st through 6th maximum values which are listed in Table-3. Two solid curves show the limit for the 95% confidence region given by eq.(31). Judging from the figure, the distribution has a statistical bias to the larger side.

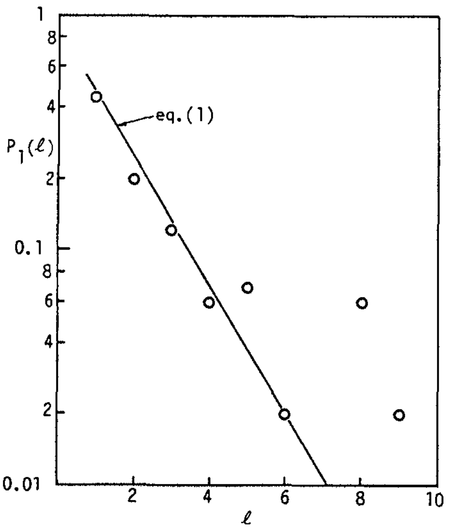


Fig.6 Probability distribution of the run of high waves ($h \geq 1$)

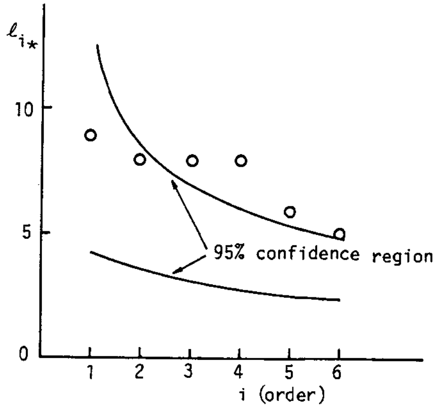


Fig.7 Ordered maximum runs and the limit for the 95% confidence regions

5. SUMMARY AND REMARKS

The probability distribution of the maximum run of wave height for the irregular sea waves is introduced theoretically. The mean of the run and its approximations are also introduced and the changes with respect to the number of runs for different h_t are shown in the figure. The probability distributions for the 2nd maximum, 3rd maximum and further maximum runs and their mean values are also introduced.
The confidence regions for the above means are introduced, although the definition for ℓ is extended into real numbers and the traditional confidence coefficients of 95% and 99% are applied. (We may have to introduce confidence coefficients which are effective in coastal engineering problems.)

REFERENCES

Battjes, J.A. and G. Ph. van Vledder (1984): Verification of Kimura's theory for wave group statistics, Proc. 19th ICCE, pp.642-648.

Kimura, A. (1980): Statistical properties of random wave groups, Proc. 17th ICCE, pp.2955-2973.

Kimura, A. and A. Seyama (1984): Statistical properties of short-term overtopping, Proc. 19th ICCE, pp.532-546.

Longuet-Higgins, M.S. (1952): On the statistical distributions of the heights of sea waves, Jour. Marine Res., Vol. IX, No.3, pp.245-266.