NOTES ON THE EFFECTS OF DATA LENGTH AND SAMPLING INTERVAL UPON THE SPECTRAL ESTIMATES FOR WIND WAVES

Rikiishi, Kunio
Research Institute for Applied Mechanics, Kyushu University : Lecturer

Mitsuyasu, Hisashi
Research Institute for Applied Mechanics, Kyushu University : Professor

Honda, Tadao
Research Institute for Applied Mechanics, Kyushu University : Research Associate

https://doi.org/10.5109/6610197

出版情報：Reports of Research Institute for Applied Mechanics. 23 (75), pp.131-147, 1976-02.
九州大学応用力学研究所
バージョン：

Reports of Research Institute for Applied Mechanics
Vol. XXIII, No. 75, 1976

NOTES ON THE EFFECTS OF DATA LENGTH AND SAMPLING INTERVAL UPON THE SPECTRAL ESTIMATES FOR WIND WAVES

By Kunio RIKIISHI*, Hisashi MITSUYASU** and Tadao HONDA***

The effects of data length and sampling interval on the spectral estimates have been investigated by analyzing a long equally-spaced wind-wave record of a stationary stage. It is shown that, although the autocovariance function obtained from a short wave data is considerably different from that obtained from sufficiently long wave data, qualitative features of the corresponding power spectra are not so different with each other owing to the effect of lag window. It is also shown that the folding effect for the spectral estimates of wind wave is practically negligible for the frequency range $0<f<f_N/2$ (f_N: Nyquist frequency). Finally it is concluded that a relatively short wave record (for about 100 times as long as the dominant wave period) can be reasonably substituted for a longer one and the sampling rates, approximately 8 data points per one dominant wave period, are practically sufficient for the ordinary spectrum estimation of wind wave.

1. Introduction

The spectral analysis of random time series has been an effective tool for studying the statistical feature of wind waves. In practical analysis of wind wave data, there are two important factors that may influence on the reliability of spectral estimates. These are data length of wave record, T_N, and sampling interval of successive data points, Δt. They are related to each other by an equation $T_N = N\Delta t$, where N is number of data points.

Regarding the effect of data length, it is obvious that longer data provide higher frequency resolution and make it possible to produce reliable spectral estimates with a high degree of freedom. However, it is not certain how qualitative features of autocovariance function and power spectrum are influenced by data length.

According to the sampling theorem, on the other hand, sampling interval should be determined so that Δt may be less than $1/(2f_M)$, where f_M is the

* Lecturer, Research Institute for Applied Mechanics, Kyushu University.
** Professor, Research Institute for Applied Mechanics, Kyushu University.
*** Research Associate, Research Institute for Applied Mechanics, Kyushu University.

frequency corresponding to the highest frequency component that the signal of wave record may contain. However, f_M is uncertain or may be infinitely large in the practical situation. Thus spectral estimates can not be free from the systematic error due to linear and non-linear folding effect (Rikiishi and Mitsuyasu, 1973). In studying the spectral structure for high frequency range, the order of magnitude of the folding effect should be evaluated carefully.

In this note, the effects of data length and sampling interval on the estimates of wave spectra are investigated by analyzing actual wind wave records measured in a laboratory tank. Besides, will be shown some statistical features of laboratory wind waves in the course of numerical examinations.

2. Wind wave data

The time series of wind waves used in this analysis is a part of wave records measured for studying the high frequency spectrum of wind waves (Mitsuyasu and Honda, 1974). The waves were generated by the wind with speed 12.5 m/s in a wind-wave channel, whose inside dimensions were 60 cm wide, 80 cm high and 851 cm long. The waves were measured by using a resistance-type wave gauge at a fetch of 8.25 m, and recorded on a magnetic tape for about 10 minutes in a statistically stationary state. To portray the actual wind wave data, a part of the wave records, digitized at equal sampling intervals of 1/25 sec, is reproduced in Fig. 1. It can be seen in the figure that wind waves measured at the station are roughly stationary in a statistical sense. In Fig.

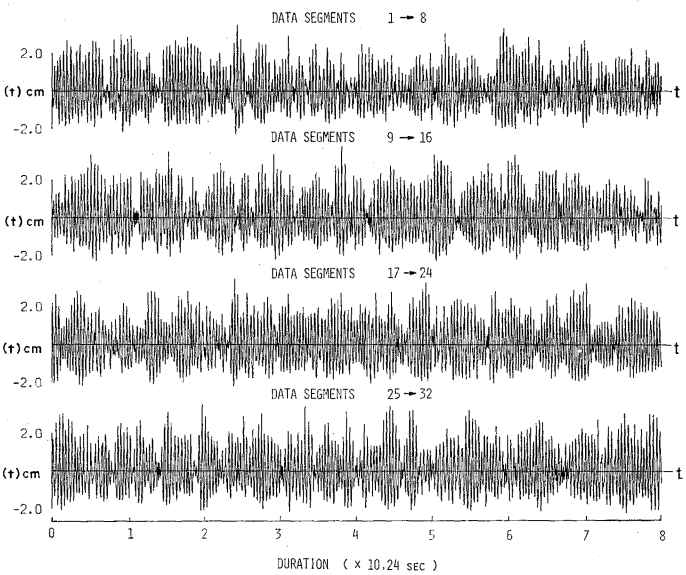

Fig. 1 Time history of wind waves.

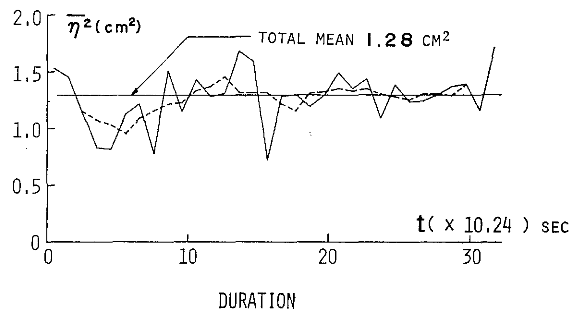

Fig. 2 The energy of wind waves η^2 versus duration t. A broken line indicates moving average of successive five points.

2, the energy (variance) of wind waves, defined by the mean square of surface displacement over 256 successive data points, are plotted versus duration time for every 10.24 sec (solid line). The figure shows that the energy varies irregularly, but there are no definite trend in the data. The broken line in Fig. 2 shows the moving average of successive 5 points for the solid line.

3. Effect of data length

In examining the effect of data length, the original data with length 327.68 sec (Δt: 1/25 sec, data points: 81926) are devided into 32 segments each of which consists of 256 data points. That is, one segment of wave record is 10.24 sec long or roughly 25 times as long as the dominant wave period ($\tilde{T} \approx$ 0.4 sec) in the present experiment.

First, these segments have been arranged in series of different length, and six sets of equally-spaced time series have been prepared by using the first 1, 2, 4, 8, 16, 32 segments. The numbers of data points of these time series are 256, 512, 1024, 2048, 4096, 8192 respectively. For these time series, autocovariance functions defined by

$$R(\tau) = \frac{1}{N}\sum_{t=0}^{N-1} \eta(t)\eta(t+\tau), \tag{1}$$

where $\eta(N+t) = \eta(t)$, have been calculated at $\tau = 0, 1, 2, \ldots$. And then the autocovariance functions have been multiplied by Hamming's lag window,

$$d_3(\tau) = 0.54+0.46 \cos\left(\frac{\pi}{M}\tau\right) \tag{2}$$

where M is the maximum lag number. The results have been Fourier cosine transformed under the condition that $M = 64$, and devided by the elementary frequency bandwidth Δf ($= 1/2\,M\Delta t$) to obtain the corresponding spectral densities.

In the upper portion of Figs. 3(a)~3(f), the power spectra obtained by the

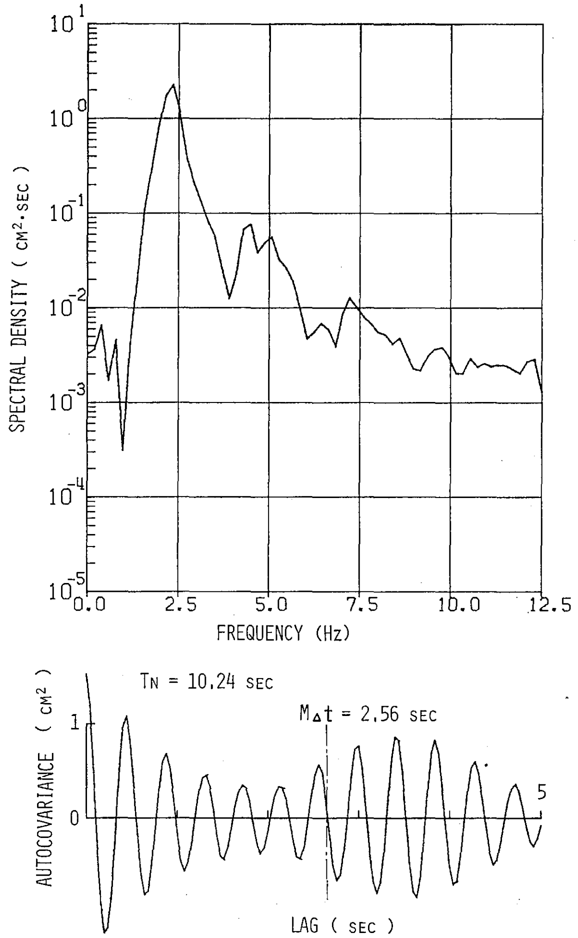

Fig. 3(a) Power spectra and autocovariance functions (data length 10.24 sec).

above procedures are shown with solid lines. The autocovariance functions are shown with solid lines in the lower portion of the same figures. Note that in these figures the autocovariance functions are shown beyond maximum lag ($\tau = 64$) up to $\tau = 125$ (= 5 sec in real time) in order to see their behaviours for large lag numbers.

It should be noted that, since the maximum lag numbers are assigned to $M = 64$ (frequency resolution: $1/M\Delta t$) equally for those six time series, the equivalent degree of freedom for those spectral estimates are given by 8, 16, 32, 64, and 128 respectively. However the frequency range over which the raw periodgrams are averaged is the same for all of them.

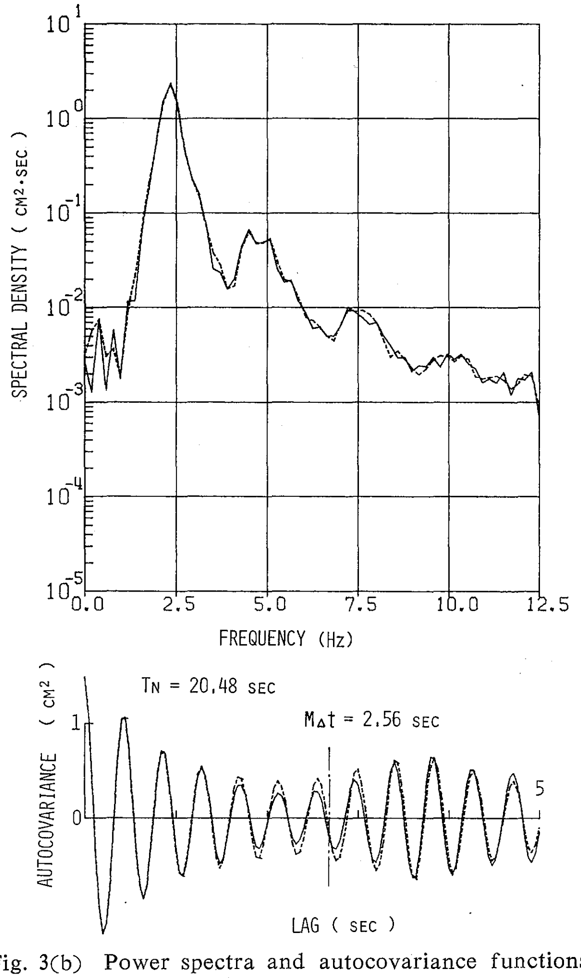

Fig. 3(b) Power spectra and autocovariance functions (data length 20.48 sec).

——: from one record with length 20.48 sec.
----: from two segments with length 10.24 sec.

Another way of analysis has been applied to the same data. In this case, the 32 segments of wave records, each of which contains 256 data points, have been arranged in parallel, and autocovariance functions $R_i(\tau)$, $i=1, 2, \ldots 32$, $\tau=0, 1, 2, \ldots, 125$ have been computed for each segment, and then the covariances have been averaged over the first n segments:

$$R(\tau)=\frac{1}{n}\sum_{i=1}^{n}R_i(\tau) \tag{3}$$

where n's are 1, 2, 4, 8, 16 and 32. After being weighted by Hamming's lag

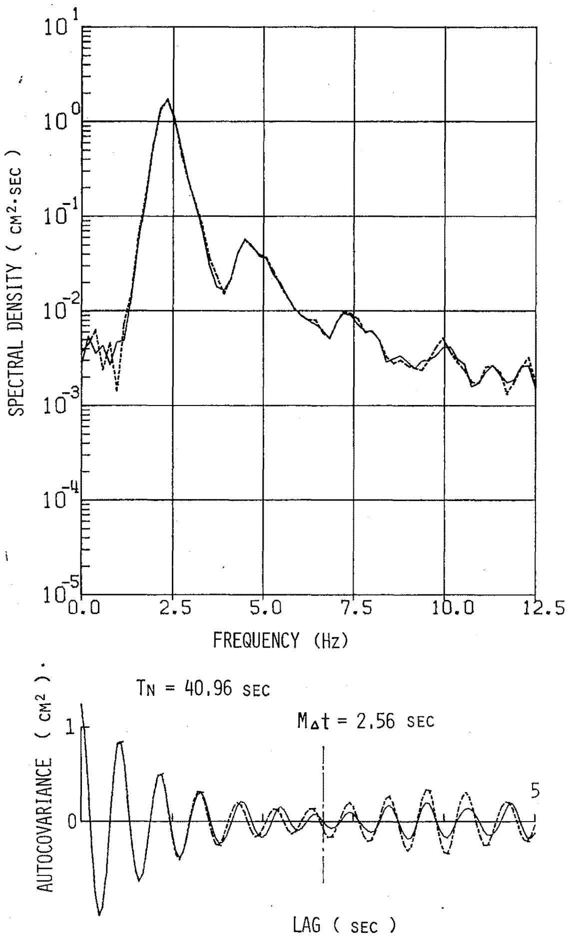

Fig. 3(c) Power spectra and autocovariance functions (data length 40.96 sec).
——: from one record with length 40.96 sec.
----: from four segments with length 10.24 sec.

window, these six autocovariance functions have been Fourier cosine transformed, and devided by the elementary frequency bandwidth to give spectral densities. The autocovariance functions and the spectral densities have been shown in Figs. 3(a)~3(f) with broken lines. Here again the autocovariance functions are shown up to $\tau=125$ ($=5$ sec).

It should be remembered that the both result shown with solid lines and broken lines in Figs. 3(a)~3(f) are obtained from exactly the same data. One may readily see that autocovariances and spectral estimates obtained through the two different procedures coincide with each other. The inspective com-

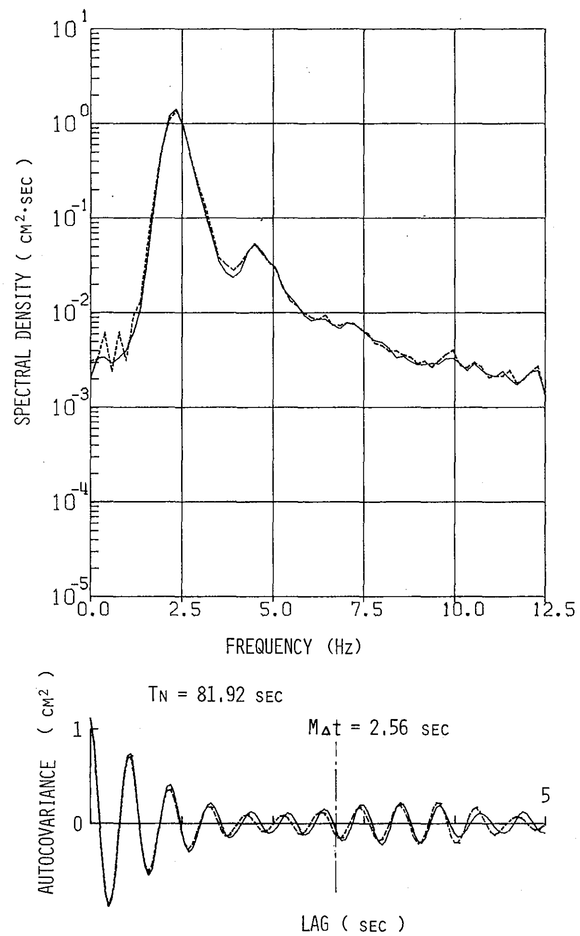

Fig. 3(d) Power spectra and autocovariance functions (data length 81.92 sec).

——: from one record with length 81.92 sec.
----: from eight segments with length 10.24 sec.

parison has shown that the differences between the two spectral densities are less than 10 % except for the lowest frequency range. Thus it can be concluded that the analysis of the ensemble of shorter segments, prepared by deviding a long time series, gives almost the same results with those obtained from the long time series, provided that both approaches have the same frequency resolution. The former method may be suitable for a small computer. Note that the data length of each shorter segments should be moderately long so as to retain an assigned frequency resolution.

From Figs. 3(a)~3(f), one can see the statistical feature of laboratory wind

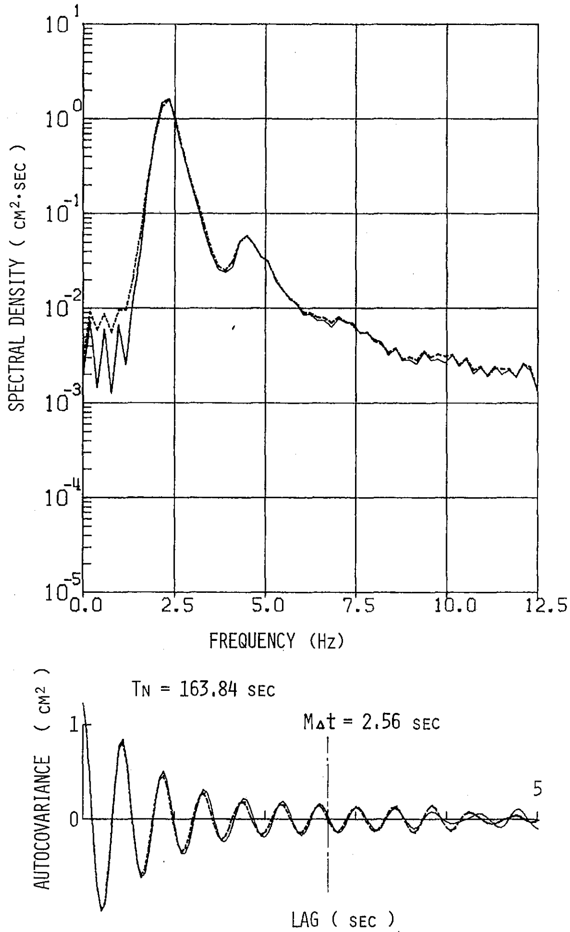

Fig. 3(e) Power spectra and autocovariance functions (data length 163.84 sec).
——: from one record with length 163.84 sec.
----: from 16 segments with length 10.24 sec.

waves that, although the autocovariance function estimated from small number of data points does not seem to tend to zero, it decreases rapidly to zero with increasing lag time when the data length is long enough. It is quite likely that autocovariance function, not only for laboratory wind waves but also for any kinds of random time series, is highly influenced by data length. As for spectral density, however, little differences are found among Figs. 3(c)~3(f). This is largely due to the effect of lag window or corresponding spectral window. Figure 3(g) show the comparison of the results obtained from relatively short data (broken line), and those from the longest data (solid line). The upper

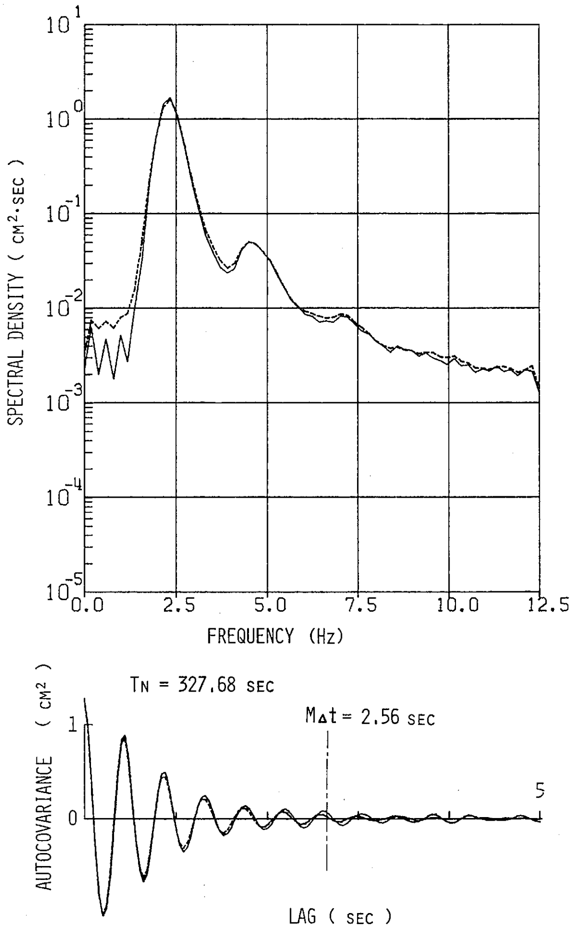

Fig. 3(f) Power spectra and autocovariance functions (data length 327.68 sec).

——: from one record with length 327.68 sec.
----: from 32 segments with length 10.24 sec.

portion of the figure shows the same spectra already shown in Figs. 3(c) and 3(f) with broken lines, and the lower portion of Fig. 3(g) shows the corresponding autocovariance functions multiplied by lag window (2). It can be seen in Fig. 3(g) that the modified autocovariance function of the relatively short wave data becomes very close to that of the longest wave data owing to the effect of the lag window (2). (Even in the case of the shortest wave data, the results are fairly consistent, though the higher order peaks in the spectrum are exaggerated too much.) These are the reason why the qualitative features of the power spectra have been not so sensitive to the data length as

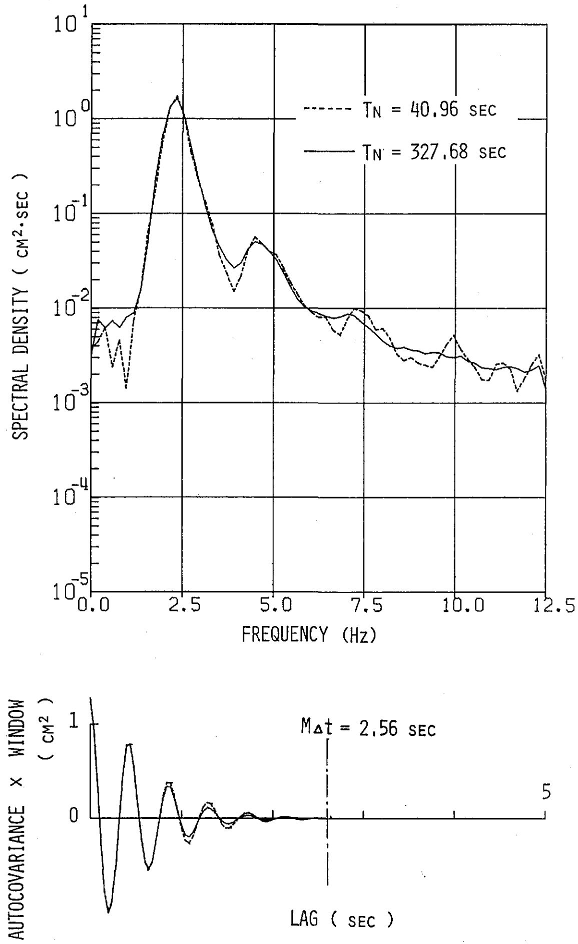

Fig. 3(g) Comparison of the results obtained from a relatively short record (broken lines which correspond to the broken lines in Fig. 3(c)) to those obtained from the longest record (solid lines which correspond to the broken lines in Fig. 3(f)). Autocovariance functions are multiplied by lag window (2).

compared with the autocovariance functions.

It may be reasonable to conclude that in describing the spectral structure of wind waves, a relatively short time series (for about 100 times as long as the dominant wave period) can be well substituted for a longer one. In Fig. 4 spectral densities obtained from eight relatively short time series ($N=1024$, $T_N=40.96$ sec$\approx 100\times 0.4$ sec$\approx 100\tilde{T}$) have been shown. Each of these spectra

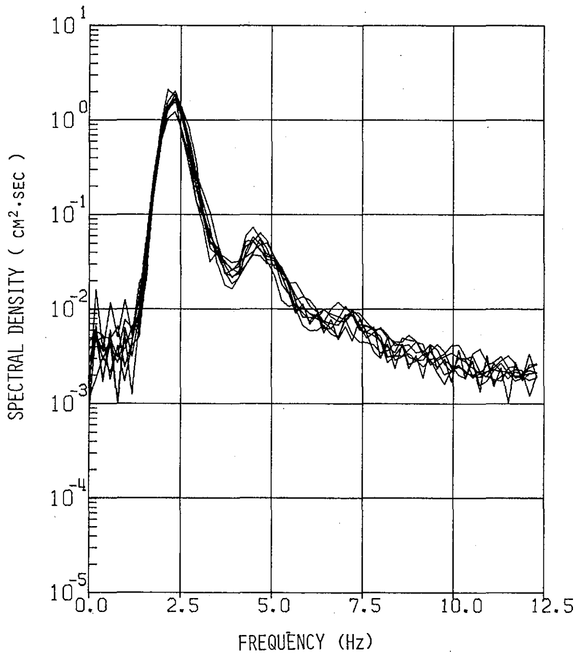

Fig. 4 Comparison of the power spectra obtained from eight samples of wave data with length 40.96 sec (= 10.24 sec×4).

seems to represent qualitative features that the spectral density from the longest time series has contained (see Fig. 3(f)).

Closer examination of Figs. 3(f) and 3(a) shows that the wave spectrum obtained from the sufficiently long wave record is smoother than that obtained from the very short wave data (note that both have the same frequency resolution). Since the former (Fig. 3(f)) is more reliable than the latter (Fig. 3(a)) in a statistical sense, it can be said that the power spectra for the wind waves have fairly smooth forms.

One more important feature of laboratory wind wave spectrum can be seen in Figs. 3(a)~3(f) that the spectral density is clearly marked by the primary peak, the secondary peak, the tertiary peak and so on. If we denote the frequency of the primary peak by f_m, the corresponding frequency to the secondary or tertiary peak is given by $2f_m$ or $3f_m$ approximately. This feature was also reported previously by one of the present authors (Mitsuyasu 1969). These secondary and tertiary peaks are still distinct in the case where the data length is the largest (Fig. 3(f)). It is probable that the higher order peaks do exist whether the data length is short or not.

4. Effect of sampling interval

In estimating wave spectra, sampling interval Δt is usually determined so that $f_m = (1/3 \sim 1/4) f_N$, where f_m is the frequency corresponding to the primary spectral peak and $f_N (= 1/2\, \Delta t)$ the Nyquist frequency. This means the continuous records are digitized at 6 or 8 points per one dominant wave. In order to study the spectral structure for higher frequency range, however, closer sampling intervals are required (see, for example, Mitsuyasu and Honda 1974). Besides, attention should be given to the influence of folding effect of frequency components higher than Nyquist frequency. The investigation of the effect of sampling interval on spectral estimates is mainly concerned with the examination of the folding effect.

According to the previous work of the present authors (Rikiishi and Mitsuyasu 1973) raw power spectrum obtained from a finite and equally-spaced time series is given by

$$P_k = \frac{1}{2}\left[a_k + \sum_{j=1}^{\infty}(a_{jN-k} + a_{jN+k})\right]^2 + \frac{1}{2}\left[b_k + \sum_{j=1}^{\infty}(-b_{jN-k} + b_{jN+k})\right]^2 \tag{4}$$

$$\left(k = 0,\ 1, \ldots\ldots, \frac{N}{2}\right)$$

and not by

$$P_k = \frac{1}{2}\left[a_k^2 + \sum_{j=1}^{\infty}(a_{jN-k}^2 + a_{jN+k}^2)\right] + \frac{1}{2}\left[b_k^2 + \sum_{j=1}^{\infty}(b_{jN-k}^2 + b_{jN+k}^2)\right] \tag{5}$$

$$\left(k = 0,\ 1, \ldots\ldots, \frac{N}{2}\right)$$

where a_k and b_k are Fourier coefficients obtained from the time series, and $jN \pm k (j = 1,\ 2, \ldots\ldots,\ \infty)$ correspond to higher frequencies outside the Nyquist frequency. Generally, the non-linear coupling between two components such as $a_k \cdot a_{jN-k}$, $b_{jN-k} \cdot b_{jN+k}$ can not be eliminated, though may be effectively reduced by the averaging process of these raw spectrum over the neighbouring frequencies. In order to evaluate the relative magnitude of this non-linear folding effect to the linear one represented by eq. (5), wind wave records have been analyzed in the following way.

First, an equally-spaced time series $\eta(t)$ ($N = 2048$, $\Delta t = 1/100$ sec) has been Fourier transformed by the use of the fast Fourier transform (FFT). Then Fourier coefficients for higher frequencies such that $f > 0.75\, f_N$ ($= 37.5$ Hz) has been reduced to zero, and the inverse Fourier transform has been applied to this modified Fourier coefficients producing a modified time series $\eta_1(t)$ ($N = 2048$, $\Delta t = 1/100$ sec) which may have no higher frequency components such that $f > 37.5$ Hz.

The modified time series $\eta_1(t)$ has been thinned out leaving one for two to

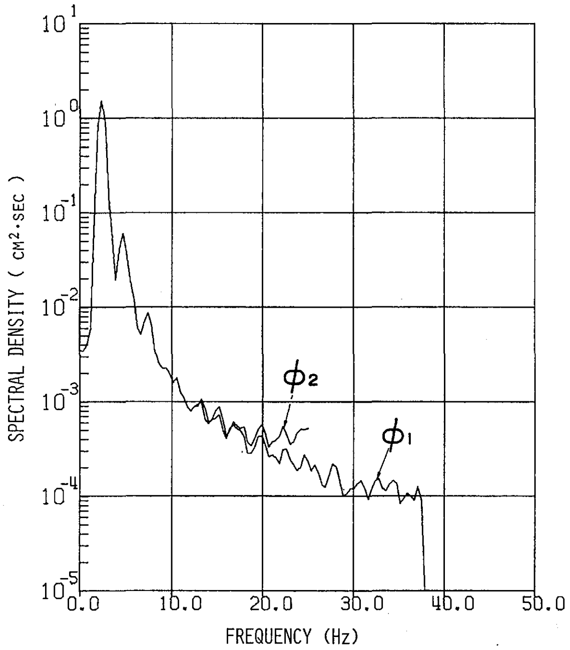

Fig.g. 5(a) Power spectra of different Nyquist frequencies. Data length; 20.48 sec for the both data.

ϕ_1; $f_N=50$ Hz ($\Delta t=1/100$ sec).
ϕ_2; $f_N=25$ Hz ($\Delta t=2/100$ sec).

produce an another time series $\eta_2(t)$ ($N=1024$, $\Delta t=2/100$ sec). Both $\eta_1(t)$ and $\eta_2(t)$ include frequency components such that $0<f\leqq 37.5$ Hz, but for $\eta_2(t)$ frequency components in the range 25 Hz $<f<37.5$ Hz are folded over the frequency range $0<f<25$ Hz and inseparable from intrinsic frequency components. Spectral densities $\phi_1(f)$ and $\phi_2(f)$ obtained from $\eta_1(t)$ and $\eta_2(t)$ respectively by using the FFT and a rectangular spectral filter (halfwidth: $4\Delta f$, frequency bandwidth: $4\Delta f$, $\Delta f=1/N\Delta t$) are shown in Fig. 5(a). Because the sampling time Δt is 2/100 sec for $\eta_2(t)$, spectral analysis of η_2 can not distinguish frequency components such that 25 Hz $<f<37.5$ Hz. Therefore spectral density $\phi_2(f)$ is subjected to the folding effect represented by eq. (4). On the other hand spectral density subjected to the linear folding effect, $\phi_3(f)$, can be calculated by using eq.(5), because Fourier coefficients for $\eta_1(t)$ are known. For comparison, both $\phi_2(f)$ and $\phi_3(f)$ are shown in Fig. 5(b). One may readily see in the figure that $\phi_2(f)$ and $\phi_3(f)$ coincide with each other approximately. Detailed comparison of spectal estimates between the two has shown that the ratio of $(\phi_2-\phi_3)$ to ϕ_3 is roughly 10 % and 25 % at most. The same procedure has been applied to some other time series, and the results

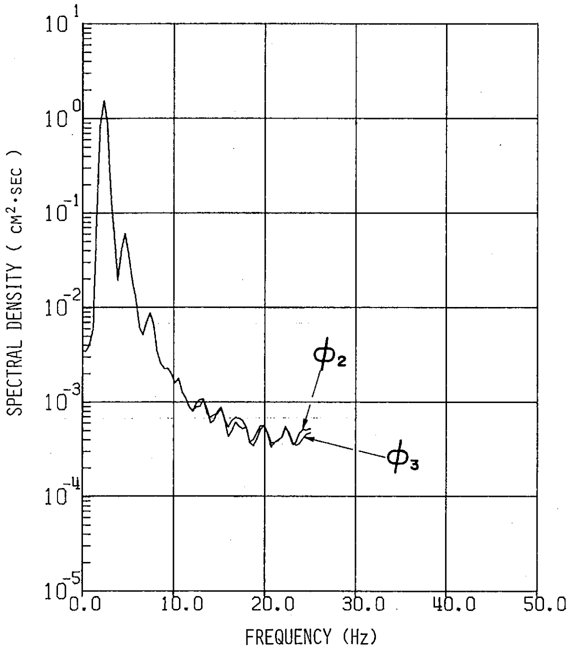

Fig. 5(b) Comparison of the power spectra ϕ_2 and ϕ_3. ϕ_2 is a power spectrum affected by linear and nonlinear folding effects. ϕ_3 is a power spectrum affected only by linear folding effect.

have been consistent with the above result. Thus, it may be concluded that the non-linear folding effect is negligible for wind wave records.

The above concluison makes it possible to evaluate roughly the folding effect by assuming the form of spectral density for higher frequency. As an example, we assume that true spectral density is represented by $\phi(f) = \alpha f^{-4}$ for $f > f_m$ (Toba, 1974). Owing to the folding effect, spectral density $\phi_0(f) = \alpha f^{-4}$ for $f > f_N$ is folded over the lower frequency domain. The folded spectral density is given by $\phi_0'(f) = \alpha(2f_N - f)^{-4}$. In the practical situation, of course, $\phi(f)$ and $\phi_0'(f)$ are unknown but $\phi'(f)$ $(=\phi(f)+\phi_0'(f))$ is obtained by the spectral analysis. The ratio of $\phi_0'(f)$ to $\phi(f)$, in the present case, is given by

$$\frac{\phi_0'(f)}{\phi(f)} = \left(2\frac{f_N}{f} - 1\right)^{-4}. \tag{6}$$

The spectral density to be modified, $\phi_0'(f)$, is 1.23 % of $\phi(f)$ at $f = 0.5\ f_N$, 12.96 % at $f = 0.75\ f_N$, 44.8 % at $f = 0.9\ f_N$, and 100.0 % at $f = f_N$. Mitsuyasu et al. (1973) previously proposed to improve the estimated spectrum by subtracting $\phi_0'(f)$ from $\phi'(f)$ on the assumption that $\phi_0(f) = \frac{1}{2}\phi'(f_N)\ f_N^5\ f^{-5}$ for $f \geqq f_N$.

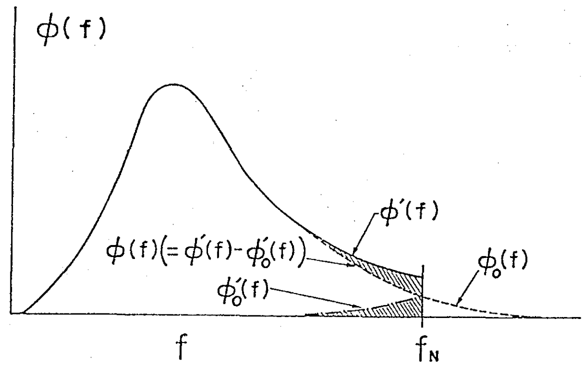

Fig. 6 Schematic representation of the linear folding effects.

Although the above discussion is based on the assumption of the linear folding effect and the "−4 power law" or the "−5 power law", it may represent the effects of sampling interval, particularly the order of the magnitude of folding effect on the spectral estimates in general. Qualitative support

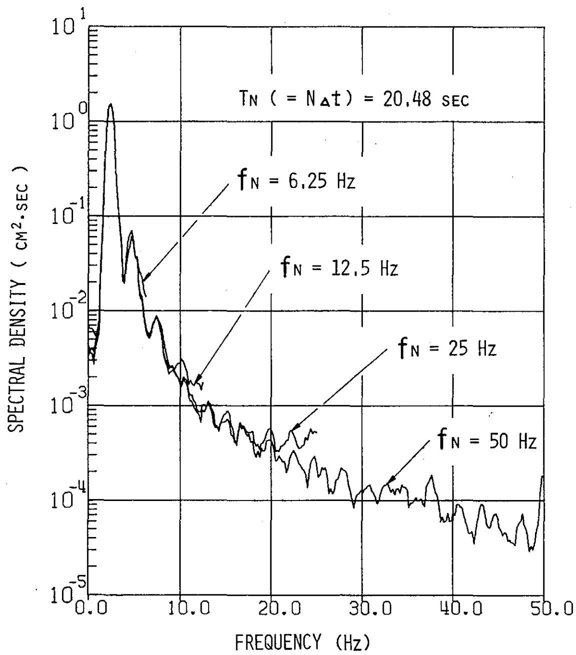

Fig. 7 Power spectra of different Nyquist frequencies. Data length; 20.48 sec for all data.

f_N; 50 Hz, 25 Hz, 12.5 Hz, 6.25 Hz.

is given by the following analysis.

Four sets of equally-spaced time series have been prepared from a closely sampled wave record by thining out the original data points. Their numbers of data points (N) and the sampling interval (Δt) are (1) $N=2048$, $\Delta t=1/100$ sec (2) $N=1024$, $\Delta t=2/100$ sec (3) $N=512$, $\Delta t=4/100$ sec (4) $N=256$, $\Delta t=8/100$ sec. Spectral densities have been obtained from these time series using the FFT and a rectangular filter (halfwidth: $4\Delta f$, frequency bandwidth $4\Delta f$, $\Delta f=1/N\Delta t$). The results shown in Fig. 7 indicate that spectral density for the lower frequency $0<f\leqq 0.5\ f_N$ are not influenced significantly by sampling interval, and that the sampling interval $\Delta t=8/100$ sec is practically sufficient for estimating the energy containing part of the present spectrum. Since $\tilde{T}\approx 1/f_m \approx 0.4$ sec for the present wave data, the above sampling condition corresponds roughly to 8 data points per one dominant wave period $\tilde{T}$.

5. Conclusion

The effects of data length and sampling interval on the spectral estimates have been investigated by using laboratory wind-wave data.

The results are summarized as follows:

1) The autocovariance function and the power spectrum obtained from an ensemble of shorter data segments differ little from those obtained from a long time series which consist of the same data segments in series.

2) The autocovariance function estimated from short data does not seem to tend to zero with increasing lag time. But, when the length of data is long enough, it decreases rapidly to zero.

3) Qualitative features of the power spectra are not influenced so much by data length, and this is largely due to the effect of lag window or corresponding spectral window. The results may lead to the conclusion that a relatively short wave data (for about 100 times as long as the dominant wave period) can be effectively used for estimating the qualitative feature of wind wave spectrum.

4) One-dimensional spectra of wind waves are fairly smooth and the irregular zigzag form overlapped some times on the dominant spectral form is largely attributed to the statistical inaccuracy due to the short length of the wave data.

5) The secondary and tertiary peaks of wind wave spectrum seem to exist whether the number of data points is small or not, though their order of magnitude decrease with increasing data points (length).

6) The non-linear folding effect is practically negligible for wind wave spectrum.

7) Spectral estimates for lower frequencies such that $0<f<0.5\ f_N$ are little influenced by sampling interval Δt if Δt is taken roughly as $1/8f_m$. Therefore, 8 data points per one dominant wave period are sufficient for the ordinary

spectrum estimation. That provides the reliable estimation of the wave spectra in a frequency range $0 < f < 2f_m$.

Acknowledgements

The authors are much indebted to Mr. K. Eto for the measurement of laboratory wind waves and Miss S. Koide for typing the manuscript. Numerical computations were done on the FACOM 270-20 system of the Research Institute for Applied Mechanics, Kyushu University.

References

1) Mitsuyasu, H.: *On the growth of the spectrum of wind generated waves* (II), Rep. Res. Inst. Appl. Mech. **17** No. 59 (1969) 235.
2) Mitsuyasu, H., et al: *Laboratory simulation of ocean waves*, Bull. Res. Inst. Appl. Mech. No. 39 (1973) 183 (in Japanese).
3) Mitsuyasu, H. and Honda, T.: *The high frequency spectrum of wind-generated waves*, J. Oceanogr. Soc. Japan. **30** (1974) 185.
4) Rikiishi, K. and Mitsuyasu, H.: *On the methods of computing power spectrum*, Bull. Res. Inst. Appl. Mech. No. 39 (1973) 77 (in Japanese).
5) Toba, Y.: *Local balance in the air-sea boundary process III*, J. Oceanogr. Soc. Japan. **29** (1974) 209.

(Received December 8, 1975)