Proceedings of The Thirteenth (2003) International Offshore and Polar Engineering Conference
Honolulu, Hawaii, USA, May 25−30, 2003

ISBN 1−880653-60−5 (Set); ISSN 1098−6189 (Set)

Automatic Quality Check Method on In-Situ Data from the Coastal Ocean Monitoring Net around Taiwan

Chia Chuen KAO *Dong Jiing DOONG* *Laurence Zsu Hsin CHUANG* *Beng Chun LEE*

Dept. of Hydraulic & Ocean Engineering, National Cheng Kung University, Tainan, Taiwan, ROC

Coastal Ocean Monitoring Center, National Cheng Kung University, Tainan, Taiwan, ROC

Dept. of Environment Design, Huafan University, Taipei, Taiwan, ROC

ABSTRACT

The Coastal Ocean Monitoring Center (COMC) was established to assist the government to develop and operate the hydrological monitoring network around Taiwan coast. Currently, the network consists of eight data buoys, one pile station, twelve coastal meteorological stations and seven tidal stations. Real time observation data are provided for coastal hazard warning system and coastal zone management applications. To assure the data correctness, a data quality-check (QC) program is developed to provide the systematical and timely examination on the measurements. This paper presents the automatic data quality check method for checking the wave statistical parameters.

KEY WORDS: data quality check, in-situ data, Markov process, wave measurement

INTRODUCTION

Taiwan Island locates in the subtropical region, where severe seas triggered by typhoons in summer seasons often result into terrible losses of the human life and property in the coastal areas. The Coastal Ocean Monitoring Center (COMC) was established within the National Cheng Kung University in 1997 to assist the government to develop and operate a hydrological monitoring network around Taiwan coast. Currently, the network consists of eight buoys stations, one pile station, twelve coastal meteorological stations and seven tidal stations. The location map of the stations is shown in Figure 1. The buoy station, a 2.5-meter wave-following discus buoy is deployed as shown in Figure 2. The buoy is equipped with a tri-axial accelerometer to measure surface wave particle movements for the estimation of directional wave spectrum. At shallow-water pile stations located in areas of mild slope and sandy seabed, an ultrasonic wave gauge array is installed to provide measurements of sea surface displacements. The in-situ meteorological and oceanographic observations from the network provide the government with critical information to prepare severe weather warnings. Long-term data from the network are used to calibrate and validate marine weather forecasting models and to develop design criteria for coastal structures.

During the operations of the network, human errors and malfunctions of equipments inevitably occur. They often cause incorrect or missing measurements, which, if not properly corrected, could significantly mislead the weather forecasting and the design conditions of constructions. The consequences of inaccurate observations may be more devastating than the lacking of observations. To ensure the data quality of the network, the COMC has developed and implemented a quality control program to provide a systematical and timely examination on the measurements from the network. The quality control program is composed of daily data quality check (QC), long-term data quality assurance (QA) and the research and development (R & D) of monitoring technologies, as shown in Figure 3. QC is the regulation of quality performance against set standards and acting on those whose performance is below default criteria. QA is the activity and proof showing that the quality operation is being carried out adequately and assuring user's confidence and satisfaction in using the data. QC and QA monitor the performances of measurement systems, which is useful for scheduling maintenances and calibrations. R & D is then the improvement of new monitoring technologies. QC, QA and R & D are highly correlated and complementary R & D, daily QC and long-term QA to improve the data quality.

In general, based on means of execution, the QC program is divided into automatic (referred to as AutoQC) and manual (referred to a ManuQC) procedures. The AutoQC use computer algorithms to test large amount measurements at real-time. The ManuQC is then applied to the suspicious data identified by the AutoQC to further check the validity of observation. The algorithms in the AutoQC are based on both objective criteria and subjective experiences. The use of algorithms by computers can significantly reduce the manpower that can be dedicated to the ManuQC. Based on the sequence of execution, the AutoQC consists of two stages. The first stage is to test raw time series data. The second stage is to test the statistical parameters derived from raw data. This paper presents the development of the AutoQC algorithms.

COASTAL OCEAN MONITORING NETWORK

Instrument Development

In order to have independent ability on ocean monitoring technologies,

COMC focuses on the development of structure and mooring design, measurement sensors, data acquisition and control, data processing and analysis, communication, power, and auxiliary sub-systems to be assembled to different measuring instruments (Kao et al., 1999). The data buoy, pile station, tidal station and meteorological station have been developed. For example, the disc type data buoy, which has a diameter of 2.5 m is designed to measure oceanographic and meteorological data in arbitrary water depth. The buoy hull consists of twelve foam flotation compartments surrounding a center payload compartment shell. A three-legged stainless steel mooring bridle beneath the buoy provides additional stability. Solar panels, a marker light, a radar reflector, antenna and sensors mount to the mast. There are two anemometers mounted on the mast at approximately 3 meters above the sea surface. Water and air temperature sensors are installed at 0.4 meters below and 2 meters above the surface water, respectively. Barometric pressure measurement is taken 2 meters above the sea level. The position of the buoy can be monitored by the GPS. Internally mounted electronics and batteries are installed on a removable aluminum rack in the central compartment. The electronics payload system is an automated, self-timed system that processes the data into required forms and transmits the formatted codes through the radio telemetry. The buoy's payloads and light are typically powered from secondary batteries with solar charging and primary-battery backup. The data buoy is used to measure wind speed, direction, gust wind, air and water temperature, barometric pressure, wave spectrum, significant wave height, wave period and direction and have high capability by its modular design.

Real-time Data Transmitting

The COMC monitoring systems are equipped with real-time data transmission system. For nearshore stations, field data is automatically transmitted to the ground station by radio telemetry or GSM modem and immediately relayed to the COMC via telephone modem after each measurement. However, as the radio telemetry reaches only a limited distance, the telecommunication link with satellite is chosen for far offshore stations. COMC systems have proven its reliability and survival ability in the severe environment especially during typhoons through the operation in the past years. The data buoys record large amount typhoon data from 1997. For example, data buoy measured the significant wave height up to 12 meter during typhoon BILIS in 2000, showing the performance of data transmitting system.

QUALITY CHECK ON STATISTICAL PARAMETERS

The quality check of statistical parameters use algorithms based on limitations imposed by measurement ranges, temporal variability, and correlations among parameters.

Range-Rationality Check (RRC)

Any statistical parameters cannot have its value exceed the range of sensors or the physical restrains of the marine environments. The range-rationality check is to assure that the magnitudes of statistical parameters are first within the limits. For example, due to the wave breaking induced in shallow waters, the measured significant wave height cannot exceed the breaking wave height imposed by the local water depth. In addition, the significant wave height should not larger than the range of wave measurement systems.

Variation-Continuity Check (VCC)

The variation-continuity check consists of time-continuity check (TCC) and space-continuity check (SCC), which are based on the concept that evolution of natural phenomenon in time or space should be gradual and smooth. The National Data Buoy Center (NDBC) of the United States has developed three time-continuity check algorithms for pressure, temperature, and other parameters (NDBC, 1996). In this study, we focus on the TCC algorithm on significant wave height. Field data show that the upcoming sea state has a leaving effect from sea states of preceding hours. This implies that the temporal change of significant wave height depend on the significant wave heights of previous hours. A sea-state independent fixed threshold for the TCC of significant wave height could underestimate the temporal variability in high seas and overestimate the variability in low seas. In this study, the Markov process theory is applied to develop the sea-state dependent thresholds for significant wave height quality check.

Markov process of significant wave height

Stochastic time processes can be ranked in increasing order of complexity, depending upon the degree of causality they embody as having a sort of "memory"' of its own past. That is, the random event which occurs at time n may be dependent upon that which occurred at time (n-1) or earlier stage. Markov process is a process with a short term memory, that means each random event is only influenced to some degree by its previous predecessors. Markov process has no direct memory of earlier events. We examine processes with simpler first order model for the operational QC program. That is, it may be possible to predict the probability of states of significant wave height at time n refer to its formal sea state at time (n-1). The acquiring additional information on time (n-2), time (n-3), etc. maybe not provide further useful information for making predictions at time n.

If a hydrologic data x at time n is affected by its previous state x_{n-1}, the stochastic process is modeled by the conditional probability $P[x_n | x_{n-1}]$. Furthermore, the transition probability of the data change from state i in stage X to state j in stage Y is expressed in the following equation:

$$p(j,i) = P[Y = y_j \mid X = x_i] \tag{1}$$

That is, the probability that the information in stage Y (i.e. time $n+1$), given the knowledge that it is in stage X (i.e. time n). We assume that the transition mechanism of the system, although random, remains constant over time, i.e. called the homogeneous Markov process. This collection of probabilities forms a transition probability matrix. Divide the historical data into i and j non-overlapping states respectively in the sequent stages, the transition probability matrix can then be expressed in the following equation.

$$P_{Y,X} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} \\ p_{21} & p_{22} & \cdots & p_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{i,j} \end{bmatrix} \tag{2}$$

$$p_{i,j} = P\left[Y_{n+1} = y_j \middle| X_n = x_i\right] = m_{i,j} \Big/ \sum_{k=1}^{j} m_{i,k} \tag{3}$$

In the above equation, $m_{i,j}$ is the sample number happened from state i of stage X to state j of stage Y.

Built and rebuilt of transition probability matrix

As an example for transition probability matrix analysis, significant wave height measurements at 2-hour interval from Longdong buoy in the year 2000 are selected. There are 4392 significant wave height data,

which are then divided into 10 states (0~30, 30~50, 50~80, 80~100, 100~150, 150~200, 200~300, 300~400, 400~500, >500cm). The resulted transition probability matrix is shown in Table 1. From the table, it is found that most of the high probability events occurred at the same states between two sequent time steps indicating the fact of large waves always come after large waves and vice versa. In order to fit this method to data quality check process, the second stage of Markov process is modified. As the observation interval of significant wave height is 2 hours, the significant wave height variation in 2 hours is defined as the variable for second stage. This newly calculated sea state variations are divided into 10 intervals, which are 0~10, 10~20, 20~30, 30~40, 40~50, 50~60, 60~70, 70~80, 80~90 and >90cm. The rebuilt transition probability matrix is calculated and shown in Figure 4. The critical limitations of the following stage are interpolated under a conference level of 95% (shadow area in Figure 4). These values will be used as the allowable range of sea state change in 2 hours of time-continuity check. The values are listed in Table 2 according to the previous sea state. For example, if the current significant wave height is 100~150 cm, there are 95% of probability that the change of significant wave height in the following hour is within 53.0 cm.

A comparison of the method on TCC presented with NDBC (1996) is made. The threshold of standard time-continuity check by NDBC is calculated by using the formula $\sigma_T = 0.58\delta\sqrt{T}$, where σ_T is the allowed difference after T hour. The parameter, δ is a non-dimensional parameter, its value varies with the change of observation objects. This value varies upon the local wave climate and 6.0 is used for significant wave height by NDBC. By applying above equation, the TCC threshold of wave height for 2 hours is 0.492 m. The given adjustable criteria of presented Markov method is listed in Table 2. Comparative result of time-continuity checks with TCC criteria and NDBC formula for significant wave height in the year 2000 is shown in Figure 5. Presented method filters out 156 records of wave height that does not comply with the continuity principles while NDBC filter out 344 suspicious data. The amount of suspicious data is overestimated by NDBC, which is twice of the proposed TCC threshold using Markov process. This difference could be even larger in severe sea states during winter monsoon seasons around Taiwan. Because the suspicious data by the AutoQC should be re-checked by ManuQC, the fewer suspicious data reduce the job of following manual quality checks.

Physical-Correlation Check (PCC)

Oceanographic and meteorological parameters provide measures of various physical elements of marine environment, which are often closely related. The correlation among various parameters can be used to develop algorithms for quality check (i.e. physical correlation check (PCC)). For example, sea surface winds are the major source of waves generated, a close relationship between wave energy and wind speed can be expected. Steele & Marks (1979) show that local wind is strongly correlated with wave energy in the frequency of 0.2~0.27Hz. Lang (1987) showed a better wind-wave correlation using the square of the mean wind speed four hours prior to the observation. In this study, 3725 simultaneous wind and wave data sets from Longdong buoy in the year of 2000 are analyzed. To assure the waves are actively generated by steady winds, the wind and wave data are selected based on three criteria (1) mean wind speed less than 25% of variance within a continuous eight hours period, (2) the difference between wind and wave directions are within 90 degrees, and (3)the wave spectral peak frequency is higher than the peak frequency based on the PM spectral model for the given wind speed.

There exists a good relation between wind and wave energy with the frequency bands of 0.257 Hz to 0.355 Hz as shown in Figure 6. Regression analysis indicates that there exists a linear relation between wind and wave energy within the frequency range of 0.257 to 0.355 Hz as shown in Fig. 7. The PCC procedure is therefore to check the correctness of wave energy by using confidence interval of the regression equation. The significant wave statistical parameters are calculated from wave spectra, such as significant wave height $H_s = 4.004\sqrt{m_0}$, which m_0 is total wave energy. Therefore, when the significant wave height is between 3.6 to 4.4 times of root total energy under a 90% confidence level, the data is viewed as a valid data. Otherwise, the wave height data will be marked by the PCC system.

In addition to the correlation between the wind and wave, we often can compare measurements from two collocated sensors measuring the same parameters. For example, two anemometers are often installed on the data buoy or pile station to assure the acquisition of wind data and reduce the probability of equipment malfunction. Quality of wind measurements of the two anemometers can be checked by the comparison between them, which can also be used to show the deviation caused by aging or damaged anemometers (Doong et al., 1997). The averaged wind speeds from the two anemometers appeared to be relatively synchronized, showing the stability of the observation system and increasing the reliability of the wind speed data. The upper and lower limits of the 95% confidence interval of the linear regression equation are the check thresholds of PCC on the wind speed. Wind speeds lies outside the band region will be treated as data failing quality check.

SUMMARY

Meteorological and oceanographic observations from a network of moored buoys and fixed platforms in the coastal waters of Taiwan are used to the validations and improvements of marine weather forecasting models and design criteria of engineering constructions. Erroneous measurements caused by severe seas, human errors and aging instruments could significantly decrease the values of measurements. To assure the quality of measurements, a quality check program is installed. In this paper, the development of automatic QC procedures on wave data is presented. The wave statistical parameters are examined by algorithms developed based on measurement ranges, continuity of temporal variations and correlations among wind and wave measurements. A sea-state dependent variation threshold is developed from Markov process for the time-continuity check of significant wave height. It is validated with effective results. The close correlation between the mean wind speed and the wave energy between 0.257 Hz ~ 0.355 Hz is shown in the study and used for the quality check of both wind and wave data.

The quality control of a data network requires the use of the power from both computer algorithms and human experiences. The AutoQC is not to simply reject measurements; instead it is to identify the suspicious measurements from large amount observations by the network for further manual check. The successful automatic quality control program is to balance the need of preserving both quality and quantity of observations.

ACKNOWLEDGEMENTS

This paper is supported by National Science Council (NSC) and Water Resources Agency (WRA) in Taiwan, R.O.C.. The in-situ data used in this study are provided by Central Weather Bureau. The authors would like to display their great thanks.

REFERENCES

Doong, D. J., Laurence Z. H. Chuang and C. C. Kao, "Development of Quality-Checking System on Oceanographic Observation Data", *Proceedings of 19th Conference on Ocean Engineering of Taiwan*, pp.477-484, 1997. (in Chinese)

Kao C. C., Laurence Z.H., Chuang, Y. P. Lin, and B. C. Lee, 1999 "An Introduction to the Operational Data Buoy System in Taiwan", *Proceedings of Int. MEDCOAST Conference*, Antalya, Turkey, pp. 33-39.

Lang, N. C., 1987 "An Algorithm for the Quality Checking of Wind Speeds Measured at Sea Against Measured Wave Spectral Energy", *IEEE Journal of Oceanic Engineering*, Vol. OE-12, No.4, pp. 560-567.

National Data Buoy Center (NDBC), 1996 *Handbook of Automated Data Quality Control - Checks and Procedures of the National Data Buoy Center*, NOAA, USA.

Steel, K. E. and G. E. Marks, 1979 "Detection of NDBO Wave Measurement Systems Malfunctions", *Proceeding Oceans 79*. New York: Marine Tech. Soc and IEEE, pp. 226-236.

Table 1 Transition probability of significant wave height between current and next sea states

Next states / Current states	State I	State II	State III	State IV	State V	State VI	State VII	State VIII	State IX	State X
	0~30	30~50	50~80	80~100	100~150	150~200	200~300	300~400	400~500	>500
State I 0~30	46.9	51.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
State II 30~50	7.6	76.5	15.6	0.3	0.0	0.0	0.0	0.0	0.0	0.0
State III 50~80	0.1	12.3	67.4	15.8	4.3	0.0	0.0	0.0	0.0	0.0
State IV 80~100	0.0	0.2	30.3	43.9	24.2	1.1	0.2	0.0	0.0	0.0
State V 100~150	0.0	0.0	2.3	16.2	63.7	16.1	1.7	0.0	0.0	0.0
State VI 150~200	0.0	0.0	0.0	0.4	26.2	52.1	21.0	0.4	0.0	0.0
State VII 200~300	0.0	0.0	0.0	0.0	1.5	23.5	68.0	6.3	0.7	0.0
State VIII 300~400	0.0	0.0	0.0	0.0	0.0	1.2	42.7	51.2	4.9	0.0
State IX 400~500	0.0	0.0	0.0	0.0	0.0	0.0	9.5	28.6	52.4	9.5
State X >500	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	66.7	33.3

Unit of wave height: cm

Table 2 Allowable variation of significant wave height under different current sea state

Current sea state (significant wave height, cm)	Allowable variation of significant wave height after 2 hours (cm)
0~30	16.1
30~50	18.2
50~80	31.3
80~100	35.2
100~150	53.0
150~200	68.5
200~300	85.2
300~400	169.5
400~500	188.9
> 500	157.8

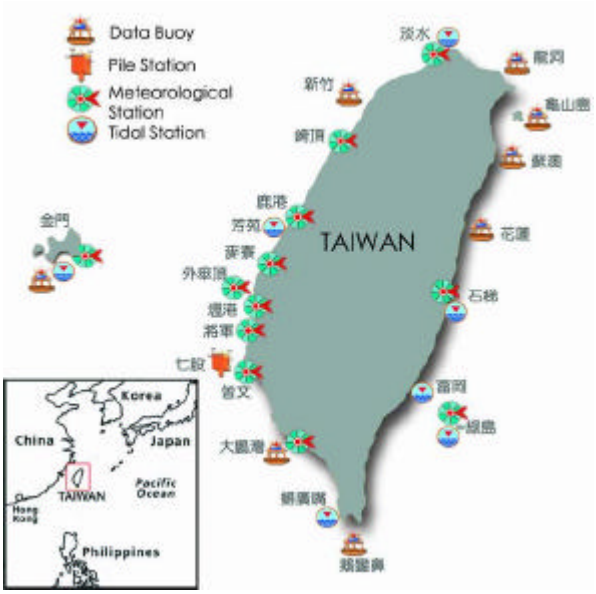

Fig. 1 Locations of COMC operational stations (before 2002)

Fig. 2 A photo of marine data buoy (supported by Central Weather Bureau of Taiwan; developed by COMC)

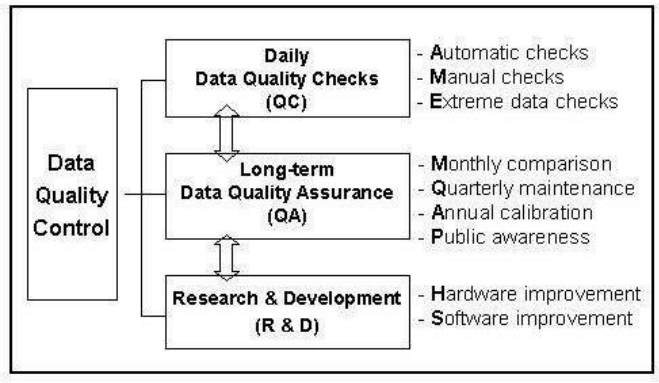

Fig. 3 Contents of data quality control

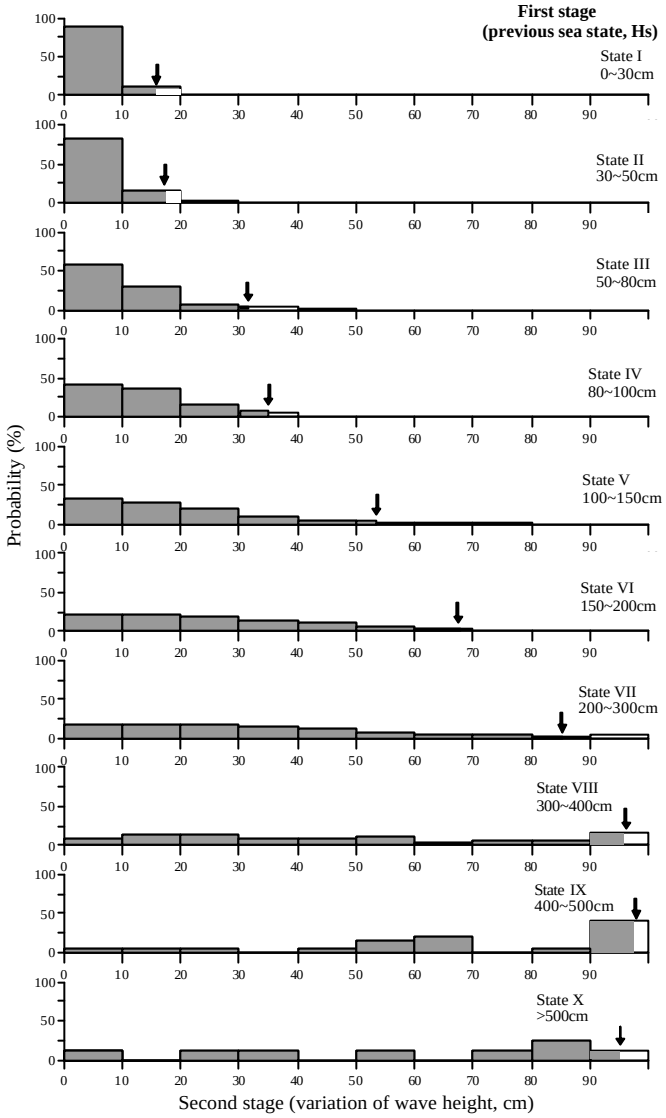

Fig. 4 Modified transition probability matrix of significant wave height

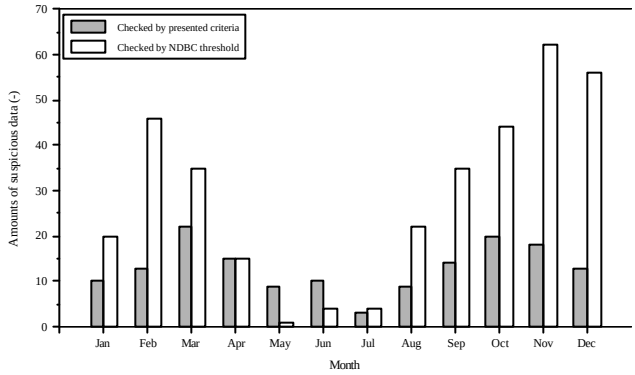

Fig. 5 Comparative result of time-continuity check between presented and NDBC thresholds

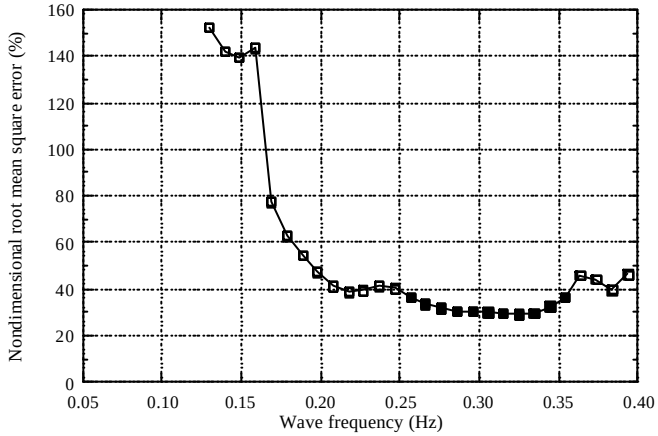

Fig. 6 Root mean square of wind-wave correlation

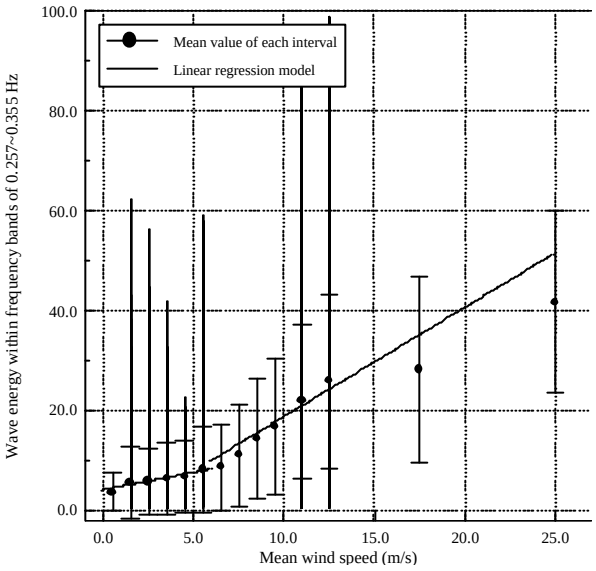

Fig. 7 Correlation between mean wind speed and wave energy within the frequency bands of 0.257 Hz to 0.355 Hz