UDC 539.3

CALCULATION OF RANDOM WAVE PARAMETERS USING REGRESSION METHOD

Dang Xuan Truong[1]

[1]*PhD. Candidate - MSc. Civil engineering Faculty – University of transport Ho Chi Minh city, PhD. Candidate – HCMC University of Technology – VNU Ho Chi Minh city , Viet Nam*

Abstract: This research introduces a new method to calculate the parameters of random waves. Here we calculate the wave height and period. The process observation data from multiple waves of different conditions is a complex problem for the designers of offshore structures. The regression method is proposed as a new method with high reliability to solve this problem.

Key words: random waves, regression method.

1. Definition

In fact, the wave is usually non-uniform waves and random transformation (random). It is a combination of waves component from different areas on the sea and spread to observe location. Sothat height and wave period is fixed time to be continuous alternation. Currently, we are using two analytical random wave method which are statistical and spectral analysis method.

In this research, we used statistical methods to determine the wave parameters for design and calculations process. Regression function is used to analyze and process of observed figures from the wave parameters such as period, season, wind direction, frequency etc. Regression analysis was used to field in success and we want to propose new method in order to solving on random wave of problem.

2. Regression function

The analysis is the basis for the theory the analysis factor and prepare the ground for problem of experimental plan.

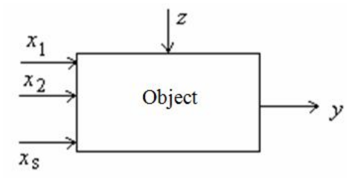

Fig. 2.1. Multidimensional non-linear object

Regression analysis method create the condition to determine the mathematical description of the objects that have characterized the unknown on basis of monitoring the values input and output (figure 2.1).

We consider an object input $x_1, x_2, ..., x_S$ and output *y*. Object is disturbance *z* (2.1). Subject specific nonlinear to be unknow:

$$y = f(x_1, x_2, ..., x_S, z). \quad (2.1)$$

Factor caused constantly disturbance *z* change. As value specified by the input $x_1, x_2, ..., x_S$, to be able to value of output *y*.

We are interest in dependence (2.1) of output y and in put $x_1, x_2, ..., x_S$, it did not depend on function that dependent caused disturbance.

The concept of the relationship functional dependence is used in field of science and other technology, in which the mathematical analysis. In fact this functional dependence is change y and change x to be one-valued (figure 2.2).

Each value of x may be define a value y as following:

$$y=f(x). \quad (2.2)$$

In the statistical analysis, the concept is very narrow $y=f(x)$ this, because the people provided general conception which is dependent relation to cause disturbance. In fact when the observational function (2.2) is rarely because we did not exclude the effect factor to be random. In fact, the observational dependence of relation are dependent relation to disturbance, therefore dependent relation is multi-value (figure 2.3).

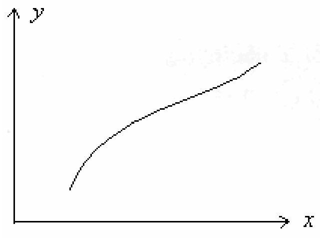

Fig. 2.2. Functional Dependence

In fact, dependent relation of disturbance is dependent relation of random variable y and random x, which is single-valued dependence relation of probability distribution of random variable y depends on value x to be selected through random variable x.

It said that random variable Y is relation of disturbance on random variable X if distribution F of random variable Y with condition $X=x$ are a value function of the random variable Y that of value x

$$F(y/x)=P(Y<x/X=x)=f(y/x), \quad (2.3)$$

$$E(Y/X=x) = f(x). \quad (2.4)$$

Now, we said that correlation are random variables X and Y

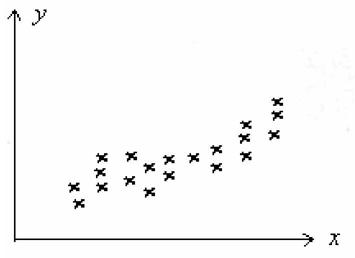

Fig. 2.3. Dependent relation of disturbance

When you have point n, know the value of input $x_1, x_2, ..., x_S$ and corresponding with value of output y as aboved (figure 2.1), we group data into observation as follows:

$$\left.\begin{array}{l} x_{11}, x_{12}, \cdots, x_{1S}, \quad y_1 \\ x_{21}, x_{22}, \cdots, x_{2S}, \quad y_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_{N1}, x_{N2}, \cdots, x_{NS}, y_N \end{array}\right\} \quad (2.5)$$

x_{NS} is value of input x_S at point n, y_N is the value of *output* at point n.

The purpose is identified specific alternative as selected the arbitrary:

$$\widehat{y} = f(x_1, x_2, ..., x_S; b_0, b_1, ..., b_k). \quad (2.6)$$

In (2.6) there is $k+1$ factor of unknown b_k (k = 0,1,2,... $K < N$-1C).

For some given point n, the function has a value (2.6)

$$\widehat{y}_n = f(x_{n1}, x_{n2}, ..., x_{nS}; b_0, b_1, ..., b_K). \quad (2.7)$$

Specific (2.7) approximately characterized (2.1) that characterized (2.1) unknow because of disturbulance z. It is continuous time n (n=1,2,…., N) it did not know usual natural type of non-linear object (2.1).

Specific (2.6) approximately characterized (2.1) often called the mathematical model of the subject drawing on figure (2.1).

To determine the optimum factor of b_0, b_1, ..., b_k of mathematical model (2.6) we chose this model below the function:

$$S_R = S_R\left[\{y_1, y_2, ..., y_N\}; \{\widehat{y}_1, \widehat{y}_2, ..., \widehat{y}_N\}\right]. \quad (2.8)$$

It determine distance between group of output value $\{y_1, y_2, ...y_N\}$ and object (2.1) and the output $\{\widehat{y}_1, \widehat{y}_2, ..., \widehat{y}_N\}$ of the model (2.7).

In fact it is the simplest to Ơclit distance in space direction n, it took mathematics to determine optimum factor b_0, b_1, ..., b_k of the model (2.7), bring to math minimum expression

$$S_R = \sum_{n=1}^{N}(y_n - \widehat{y}_n)^2 = \sum_{n=1}^{N}\left[y_n - f(x_{n1}, x_{n2}, ..., x_{nS}; b_0, b_1, ..., b_K)\right]^2. \quad (2.9)$$

For a long time, the math called square minimum math to entered by Lagrang 1806, the math is the foundation of regression analysis.

After the minimize formula (2.9) for the value of parameter b_0, b_1, ..., $b_{k.}$ we determine the optimal parameters through b_0, b_1, ..., b_k, that way we get the function:

$$\widehat{y} = f(x_1, x_2, ..., x_S; b_0, b_1, ..., b_K). \quad (2.10)$$

Often called the regression function.

We called regression function which is linear for unknown parameter:

$$\widehat{y} = \widehat{y}(x; b_0, b_1, ..., b_K) = b_0 + b_1 f_1(x) + f_2(x) + ... + b_K f_K(x) \quad (2.11)$$

Sign:

$$x = [x_1, x_2, ..., x_S]. \quad (2.12)$$

With the function $f_k(x)$ and k=1,2,..., K knows and it is independent linear. Generally, it maybe non-linear [1], [2].

Calculation of factor of regression function

Firstly, we study as a separate case determine factor of linear regression function in directions.

We find object to research on figure 2.1 at the time N to determine the input of value S: $x_1, x_2, ..., x_S$ and correspond with output y, we group results into a table of output monitoring (2.5). We find regression function with directions:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + ... + b_S x_S. \qquad (2.13)$$

Mean:

$$\hat{y}_n = b_0 + b_1 x_{n1} + b_2 x_{n2} + ... + b_S x_{nS}, \quad n = 1, 2, ..., N. \qquad (2.14)$$

The parameters are unknown the regression function (2.13) how to make the best approximated y with meaning is the sum of squared deviation of the observation data (2.5) have the smallest value.

$$S_R = \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 = \sum_{n=1}^{N} \left[y_n - f(x_{n1}, x_{n2}, ..., x_{nS}; b_0, b_1, ..., b_K) \right]^2 = \min. \qquad (2.15)$$

The math selected optimum factor of regression function in way of matrix as followed. So, we present result of results into a table of output monitoring (2.5) as form of matrix

Input X:

$$X = \begin{bmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1N} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N0} & x_{N1} & x_{N2} & \cdots & x_{NN} \end{bmatrix}. \qquad (2.16)$$

We insert to be fixed column elements No1 among them:

$$x_{10} = x_{20} = \cdots = x_{N0} = 1. \qquad (2.17)$$

For the purpose described in the form of matrix which included the factor b_0. We write the observation results of input object and the output model to present of model vectors, which means value of regression function in the following:

$$y = [y_1, y_2, ..., y_N], \qquad (2.18)$$

$$\hat{y} = [\hat{y}_1, \hat{y}_2, ..., \hat{y}_N]. \qquad (2.19)$$

We write unkown parameter to form of thier vector as following

$$b = [b_0, b_1, ..., b_S]. \qquad (2.20)$$

Then, value of regression function (2.14) shall present form:

$$\hat{y} = Xb. \qquad (2.21)$$

Pay attention to $(A+B)^T = A^T + B^T$ and $(AB)^T = B^T A^T$ math for minimum the total squared deviation of (2.15) form shall be followed:

$$S_R = \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 = (y - \hat{y})^T (y - \hat{y}) = (y - Xb)^T (y - Xb) =$$

$$= y^T y - y^T Xb - b^T X^T y + b^T X^T Xb = y^T y - 2b^T X^T y + b^T X^T Xb = \min. \qquad (2.22)$$

Optimum conditions of partial derivative S_R (2.22) with b

$$\frac{\partial S_R}{\partial b} = -2X^T y + 2X^T Xb. \tag{2.23}$$

And it is zero as we have set of equation as following:

$$X^T Xb = X^T y. \tag{2.24}$$

It is multiplied by two member of (2.24) matrix $(X^T X)^{-1}$ we have vectors b:

$$b = (X^T X)^{-1} X^T y. \tag{2.25}$$

The experimental plan

We studied the object to figure 2.1 as following:

$$y = f(x_1, x_2, ..., x_S, z). \tag{2.26}$$

Including $x_1, x_2, ..., x_S$ effect on the subject, z is a factor disturbance. Form of the function (2.26) unknow, assuming the function y is continuous and a point called only extremity. Functional assumption (2.26) is non-linear and we can get it around the given point $x_1^0, x_2^0, ..., x_S^0$ by regression function in the linear:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_S x_S. \tag{2.27}$$

Firstly, we prepared experimental to determine the value of input $x_{n1}, x_{n2}, ..., x_{nS}$ with n=1,2,…,N of a series of experiments. We shall make next experiment to receive the value of input the corresponding $y_1, y_2, ..., y_n$ of the subject. Basing on this experiment we had data as following:

$$\left.\begin{array}{l} x_{11}, x_{12}, \ldots, x_{1S}, y_1 \\ x_{21}, x_{22}, \ldots, x_{2S}, y_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_{n1}, x_{n2}, \ldots, x_{nS}, y_n \end{array}\right\}. \tag{2.28}$$

We calculate coefficent of regression function: $b_0, b_1, ..., b_S$, coefficent determined the direction of Gradient and find extreme in this direction [3], [4].

The two level of experimental plan based on the receiving of value input x_s, in which s=1,2,..., S in two levels:

$$x_S^0 - \Delta x_S;\ \ x_S^0 + \Delta x_S; \tag{2.29}$$

s=1,2,..., S may be two cases:

- full extreme 2^S testing;
- particular extreme $2^{2\text{-}S}$ testing.

To improve reliability of the study, we consider a simple example to prepare the planning with two levels to illustrate the method optimization. We find out of the subject S = 3 which, it was written by non-linear function:

$$y = f(x_1, x_2, x_3). \tag{2.30}$$

It is limited in the area

$$\left.\begin{aligned} 0 \le x_1 \le 100, \\ 0 \le x_2 \le 500, \\ 0 \le x_3 \le 100. \end{aligned}\right\} \tag{2.31}$$

Now we study found initial locally max with first point begin

$$\left.\begin{aligned} X_1^0 = 30, \\ X_2^0 = 250, \\ X_3^0 = 50. \end{aligned}\right\} \tag{2.32}$$

At nearby points (2.32) find nearby function in the following:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3. \tag{2.33}$$

We have increments

$$\left.\begin{aligned} \Delta x_2 = 3, \\ \Delta x_2 = 20, \\ \Delta x_2 = 2. \end{aligned}\right\}$$

Carried out the experiement 2^3= 8 time of testing as following:

The results of experiments Boxa-Wilsona

Factor x_S	X_1	X_2	X_3	y
Facility x_S^0	30	250	50	160
Increment x_S	3	20	2	
High level: $x_S^0 + \Delta x_S$	33	270	52	
Low level: $x_S^0 - \Delta x_S$	27	230	48	
Data testing	27	230	48	169,4
Data testing 2	33	230	48	182,6
Data testing 3	27	270	48	129,4
Data testing 4	33	270	48	142,6
Data testing 5	27	230	48	178,6
Data testing 6	33	230	52	189,4
Data testing 7	27	270	52	138,6
Data testing 8	33	270	52	149,4

As result of matrix of input and output of object y as following:

$$X = \begin{bmatrix} 1\ 27\ 230\ 48 \\ 1\ 33\ 230\ 48 \\ 1\ 27\ 270\ 48 \\ 1\ 33\ 270\ 48 \\ 1\ 27\ 230\ 52 \\ 1\ 33\ 230\ 52 \\ 1\ 27\ 270\ 52 \\ 1\ 33\ 270\ 52 \end{bmatrix}; \qquad y = \begin{bmatrix} 169,4 \\ 182,6 \\ 129,4 \\ 142,6 \\ 178,6 \\ 189,4 \\ 138,6 \\ 149,4 \end{bmatrix}. \tag{2.34}$$

Coefficients b_0, b_1, b_2, b_3 of model (2.33) defined on general basic of regression analysis (2.25). When multiplication of matrix $(X^TX)^{-1}X^{-1}$ with vector y we will receive the coefficient vector of regression function is:

$$b = \left(X^T X\right)^{-1} X^T y = \begin{Bmatrix} 250,0 \\ 2,0 \\ -1,0 \\ 2,0 \end{Bmatrix}. \tag{2.35}$$

Mathematical model (2.33) explicit form is:

$$\hat{y} = 250,0 + 2,0x_1 + 1,0x_2 + 2,0x_3. \tag{2.36}$$

Meaning of regression function

Regression function $\hat{y}$ considered full range of many factors x_i impact to y – object is considering. Example in economical $\hat{y}$ is price of product, in biological $\hat{y}$ is growth height of plant. In our case, H is the height of wave and depend on many factors as cycle, season, wind direction, frequency, etc.

Upon having regression function, we can easily take input x_i depend on the choice, at the same time it demonstrates the dominant of any factor at x_i to $\hat{y}$. Specifically in this case, seasonal factors are most obvious effectively to the height of wave. So here we have chosen in June to calculate for H.

Apply regression function method to calculate the height of wave

View the height of wave is a fuction, with 3 variables as frequency, direction and month have been observational [5]

Period T	Direction (frequency =50)	Month	H (T)	H (direction)	H (month)
10,7	8,9	1	4,5	9,2	7,5
11,3	6,5	2	5,3	6,9	7,9
12,0	3,6	3	5,9	4,3	4,4
12,6	5,0	4	6,4	5,6	5,1

13,4	6,1	5	7,2	6,3	5,3
4,1	5,9	6	7,9	6,1	5,0
14,5	6,7	7	8,3	7,3	6,0

H is the height of wave was defined by average of 3 final columns:

$$H= (H(T) + H(\text{direction}) + H(\text{month}))/3.$$

Period (*T*)	Direction (50)	Month	*H*
10,7	8,9	1	7,07
11,3	6,5	2	6,7
12,0	3,6	3	4,87
12,6	5,0	4	5,7
13,4	6,1	5	6,27
4,1	5,9	6	6,33
14,5	6,7	7	7,2

In case data is being in original form of observing data [5] with 50 years direction (column 2^{nd}). Since there are 3 unknowns, so we should add a photogrammetry of the 8^{th} (that is need to conduct $2^S = 8$ observations or experiment). There are 7 experiments in the table above, so we should add the 8^{th} experiment from the 2^{nd} experiment in the table above. Since we have two full matrixes with input (*X*) and output {*Y*}:

$$(X)=\begin{pmatrix} 1 & 10,7 & 8,9 & 1 \\ 1 & 11,3 & 6,5 & 2 \\ 1 & 12,0 & 3,6 & 3 \\ 1 & 12,6 & 5,0 & 4 \\ 1 & 13,4 & 6,1 & 5 \\ 1 & 4,1 & 5,9 & 6 \\ 1 & 14,5 & 6,7 & 7 \\ 1 & 11,3 & 6,5 & 8 \end{pmatrix}; \qquad \{Y\}=\begin{Bmatrix} 7,07 \\ 6,7 \\ 4,87 \\ 5,7 \\ 6,27 \\ 6,33 \\ 7,2 \\ 6,7 \end{Bmatrix}.$$

Putting 2 matrixes into the program to find a linear function $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3$ (inside $\hat{y} \equiv H$ is the height of wave) according to 3 variables:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \ (\hat{y} \equiv H).$$

Period $T(x_1)$, direction (x_2), month (x_3).

Search from the computer programs [6], [7]:

$$b_0=2{,}75,\ b_1=0{,}018,\ b_2=0{,}478,\ b_3=0{,}103.$$

So regression function explicit form is:

$$\hat{y} = 2{,}75 + 0{,}018x_1 + 0{,}478x_2 + 0{,}103x_3.$$

Selected point T=6 ($4,1 \le T \le 14,5$); direction =3 ($3,6 \le direction \le 8,9$); month = 6 ($1 \le month \le 8$), from there:

$H_{\max}$=4,9 *m* with T=6 *sec.*

The 8th Experiment can be defined by average of any 2 month (input is (X) and output is {Y}:

$$(X)=\begin{pmatrix} 1 & 10,7 & 8,9 & 1 \\ 1 & 11,3 & 6,5 & 2 \\ 1 & 12,0 & 3,6 & 3 \\ 1 & 12,6 & 5,0 & 4 \\ 1 & 13,4 & 6,1 & 5 \\ 1 & 4,1 & 5,9 & 6 \\ 1 & 14,5 & 6,7 & 7 \\ 1 & 17,6 & 7,8 & 8 \end{pmatrix}; \qquad \{Y\}=\begin{Bmatrix} 7,07 \\ 6,7 \\ 4,87 \\ 5,7 \\ 6,27 \\ 6,33 \\ 7,2 \\ 7,13 \end{Bmatrix}.$$

Conclusions and recommendations

From the observed data as the height of wave H and wave period T depend on factors of month, season, wind direction, frequency, etc, regression analysis can define the height of wave H and the wave period T with the high reliability to make input basis for the calculation process and design.

Random wave theory becomes more actual when using regression function to handle the observation data about the real waves on sea. It can be applied to analyse other parameters of wave when having results of observation from many different conditions.

REFERENCES

[1] Nguyen Huu Bang - Tran Van Ban, *Basis designed of works ocean to serve the oil and gas industry*. Publishing house of Science and Technology, Hanoi, Vietnam 2009

[2] Nguyen Huu Bang, *Theory of structures*, Publishing house of Science and Technology, Hanoi, Vietnam 2005

[3] G.S. Pisarenko and Coworker, *Strength of materials*, Publishing house of Science and Technology, Kiev 1967

[4] B.P. Đêmiđôvitrơ - I.A Maroon, *Basis Mathematical calculations,* Publishing house of Science, Maskva 1970

[5] Dai Hung complex metocean & environmental design criterria. Petrovietnam Exploration and production company (PVEP). 2006.

[6] Gunther F. Clauss, *Offshor structures volum 1*, Edition London, Verlag. 1992.

[7] Design guides for offshore structures. Clarom Edition Technip 1993.

[8] James F. Winson, *Dynamics of Offshore structures*, John Wiley & Sons Inc. Canada 2003.