ministerie van verkeer en waterstaat

koninklijk nederlands meteorologisch instituut

A primary experiment of statistical interpolation scheme used in sea waves data assimilation

Gao Quanduo

technical reports; TR-128

technische rapporten; TR-128

de bilt 1990

KNMI

A Primary Experiment of Statistical Interpolation Scheme Used In Sea Waves Data Assimilation

by

Gao Quanduo*

September 4, 1990

Abstract

This paper describes a statistical interpolation method which may incorporate the spatial information of forecast error of sea waves by analytical correlation functions. The basic assumptions are that the correlation of observed data is random and the cross-correlation of observation between wind and waves is independent. Based on observed data, the analysis shows that there exists the same statistical property for both hindcast error and forecast error. The least-square technique is used to fit isotropical and anisotropical functions, which may be used for sea wave data assimilation. The Fourier transform of correlation function,or wave number spectrum, is used to indicate the contribution of correlation to the interpolation and to determine the truncation length of the correlation.

*On leave of the National Research Center for Marine Environment Forecasting, Beijing, China

1 Introduction

For most operational real time forecast wave models, the initial wave fields is calculated from the wind fields. Except for the physical and numerical properties of wave model itself, the initial wave fields and its forecast results are mainly dependent on the accuracy of meteorological model. Data assimilation is one of the most important ways to improve the accuracy of the wave model. The optimum interpolation scheme developed by Gandin (1963) was one of the immediate insertion of combining observation and forecasts data, and the main emphasis of this scheme is on the determination of covariance or correlation by which the forecast error in the analysis point is best evaluated by the linear combination of observation minus forecast error in surrounding points.

One of the most serious problems in the wave data assimilation is to obtain covariance or correlation function which needs spacial structure of statistics of observation data of sea waves. The visual data from commercial ships, most of which concentrate on ship routes, has been verified unreliable and unfortunately, there is a few reliable observed data in offshore sea area and nearly no reliable data on the vast ocean area. This is the main difficulty of the sea wave data assimilation scheme based on the present observation method.

However, the launch of satellites such as ERS-1 in the near future will provide us a wealth of data, e.g. the waves and winds, and it will allows us to do operational ocean wave data assimilation.

The purpose of this paper is to obtain analytical correlation functions and a reasonable truncation length of correlation. The plan of the paper is as follows: in section two, a multivariate structure, which is of vital importance to estimate the forecast error of wave model, is derived, in section three, the attention is on the autocorrelation based on hindcast minus forecast error rather than observation minus forecast. In section four, analytical correlation functions with isotropy and anisotropy are fitted to a set of sample data, and in section five,the analysis of wave number spectrum and its relation with interpolation are discussed. The conclusions are presented in section six.

2 Data assimilation equation with multivariables

The main purpose of data assimilation is to seek an optimal estimation of forecast error at an analysis point by a linear combination of the forecast error in surrounding points, so that the analysed fields may be used as an initial fields for next forecast run. For wave model data assimilation, the wave height error of the model first guess not only relates to the error of wave height in surrounding points, but also relates to an inaccuracy wind forecast. The analysis can be improved by use of a multivariate assimilation equation, suggested by Rutherford (1973) and Thiebaux (1974), in which distinct but correlated variables, such as wave heights and friction velocity, were used simultaneously. If we only concentrate on the spatial interpolation problem at the same time, the equation may be written as follows:

$$\left\{ \begin{array}{c} h \\ u_* \end{array} \right\}_k^a - \left\{ \begin{array}{c} h \\ u_* \end{array} \right\}_k^f = \sum_{l=1}^{m} [a]_{k,l} \left[\left\{ \begin{array}{c} h \\ u_* \end{array} \right\}_l^o - \left\{ \begin{array}{c} h \\ u_* \end{array} \right\}_l^f \right] \tag{1}$$

where h is wave height and u_* is friction velocity of wind, the superscripts a, f and o denote analysis, forecast and observed variables, the subscripts k, l denote the analysis and observation points ($l = 1.2, ...m$) respectively, $[a]_{k,l}$ is a 2 x 2 matrix with four weights which are determined from the requirement of minimum analysis error. If we introduce the following notation :

$$Z = \left\{ \begin{array}{c} h \\ u_* \end{array} \right\} \tag{2}$$

where Z is 1 x 2 vector, using superscript t to denote 'true' value, and if we introduce the notations:

$$\begin{array}{lll} E_k = Z_k^t - Z_k^a & analysis\ error & (a) \\ De^f = Z^t - Z^f & forecast\ error & (b) \\ De^o = Z^t - Z^o & observation\ error & (c) \\ De_k = Z^a - Z^f & analysis - forecast\ error & (d) \\ De_l = Z^o - Z^f & observation - forecast\ error & (e) \\ A_{k,l} = [a]_{k,l} & & (f) \end{array} \tag{3}$$

where $A_{k,l}$ are a matrices, each of them has 2 x 2 weights, then Eq. (1) may be written as:

$$De_k = (A_{k,l})^T De_l \ , \tag{4}$$

where T denotes transpose of matrix. Since we don't believe that the observation is perfect, we need to introduce the "true" value in Eq (4) with aid of Eq (3a). After taking the square of it and then taking ensemble average, using Eq. (3), we have

$$E_k^2 =< (De_k^f)^2 > -2(A_{k,l}) < De_k^f De_l^f > +(A_{k,l})^T < De_l^f De_j^f + De_i^o De_j^o > A_{j,k} \quad (5)$$

The optimum interpolation requires the error E_k^2 to be minimum,which results in the condition that the derivative of E_k^2 with respect to $A_{k,l}$ is zero. One arrives at

$$A_{k,l} = M^{-1}[D_{k,l}]^T \ , \qquad (6)$$

where M is a l x l matrix, $D_{k,l}$ is 1 x l vector, each of them contain 2 x 2 weights,

$$M = \vec{C}_{i,j} \ , \qquad (a) \qquad (7)$$

$$D = \vec{C}_{k,l} \ . \qquad (b)$$

Here $C_{i,j}$ is error residual matrix of covariance of 'true' value minus forecast in observation point pairs,

$$C_{i,j} = \begin{pmatrix} < (h^t - h^f)_i \, (h^t - h^f)_j > & < (h^t - h^f)_i \, (u_*^t - u_*^f)_j > \\ < (h^t - h^f)_j \, (u_*^t - u_*^f)_i > & < (u_*^t - u_*^f)_i \, (u_*^t - u_*^f)_j > \end{pmatrix} \qquad (8)$$

$$+ \begin{pmatrix} \delta_{i,j} < (h^t \ - \ h^o)_i^2 > & 0 \\ 0 & \delta_{i,j} < (u_*^t - u_*^o)^2 > \end{pmatrix},$$

where δ is a dirac delta function. Finally $C_{k,i}$ is covariance between observation and analysis point pairs,

$$C_{k,i} = \begin{pmatrix} < (h^t - h^f)_k \, (h^t - h^f)_i > & < (h^t - h^f)_k \, (u_*^t - u_*^f)_i > \\ < (h^t - h^f)_i \, (u_*^t - u_*^f)_k > & < (u_*^t - u_*^f) \ (u_*^t - u_*^f) > \end{pmatrix} \qquad (9)$$

We have assumed here that the observation error is random and independent for different observation points, and the observation error of winds and waves are uncorrelated. Eq. (4) with the weights coefficient $A_{k,l}$ given in Eq. (6) may be represented by

$$De_k = C^T D_{k,l} \ , \qquad (10)$$

where $C = M^{-1} De_l$ is independent of the analysis point k and may be calculated once for all points. The covariance in the Eq. (8) and Eq. (9) only consider the wave height and friction velocity. However, direction of wave propagation has important relation with the wave height error of the first guess. For example, in case of swell, if there is a swell source at point k, wich propagates to the point i and j at the seem time, the error at point i and j then is strongly correlated and its magnitude will depends on the propagation direction of the swell.

Although many variants will contribute to the wave height error in the first guess field, for simplicity, the covariance between wind and wave will be neglected in order to obtained an analytical correlation function for wave heihgt. We have to point out that this approximation has an obvious shortcoming because for a reliable wave model, such as th third generation wave model, where the source terms in the wave evolution equation have been parameterized adequately, the forecast error of wind sea will be mainly caused by the error of forecast wind. The other approximation is that the covariances in Eq. (8) and Eq. (9) is not strict, we have to take $< De >^f = < h^t - h^f >= 0$ when we evaluate an analytic correlation function. This means that a statistic method is used to evaluate the covariance.

3 Statistical structure of the auto-covariance of wave forecast error

We denote that S is the sum of prediction and observation standard deviations. $r_{i,j}(x)$ is autocorrelation between two points i and j,

$$C_{i,j} = S_i r_{i,j}(x) S_j \ . \tag{11}$$

If there is only one observation point and this point is an analysis point, then the matrix C in Eq.(10) may be written as:

$$C = M^{-1} De_i = M^{-1}(h^o - h^f)_i \ , \tag{12}$$

where

$$M = (< (h^t - h^f)_i^2 > + < (h^t - h^o)_i^2 >) \tag{13}$$

Then Eq. (10) may be written in the form

$$(h^a - h^f)_i = \frac{< (h^t - h^t)_i^2 > (h^o - h^f)_i}{< (h^t - h^f)_i^2 > + < (h^t - h^o)_i^2 >} \tag{14}$$

$$= r_i(x)(h^o - h^f)_i$$

Because of the observation error introduced in the interpolation equation, the auto-correlation $r_i(x)$ in the analysis point is not equal to 1, but smaller than 1. Hence, we can express the correlation as follows: if i is equal to j,

$$r_i(x) = \frac{(S_p)_i^2}{(S_i)^2} = \frac{(S_p)_i^2}{(S_o)_i^2 + (S_p)_i^2} \tag{15}$$

and if i is not equal to j then

$$r_{i,j}(x) = \frac{1}{S_i \cdot S_j} < (De^f - < De >^f)_i \, (De^f - < De >^f)_j > \ . \tag{16}$$

Here S_p and S_o are forecast and observation standard errors respectively, and we have implicitly assumed that

$$< De >^f = 0 \ , \tag{17}$$

Eq. (3b) in section two may be written as:

$$De_i = (h^t - h^o)_i + (h^o - h^d)_i + (h^d - h^f)_i = Dh_i^o + Dh_i^d + Dh_i^f \ . \tag{18}$$

The analysis of forecast error covariance $< (h^t - h_i^f)_i (h^t - h^f)_j >$ using statistical method needs a sufficiently long period of observations data and long period of forecast run. Since there is a only few wave observation data, the more desirable solution to this problem is therefore to express the covariance in term of hindcast minus forecast error. Table 1 shows some results of statistical properties of wave hindcast and forecast, from which we may infer that hindcast and forecast bias are more or less the same, but the rms error and scatter index of forecast increase compared to the hindcast by a factor of 0.8 for + 24 forecasts on average. The analysis above shows that as the error of hindcast minus forecast is positive, the error sign of observation minus hindcast is the same, and the ratio of the both error is more or less the same in a statistical sense. Hence one may conclude that the statistical properties are the same for both hindcast and forecast errors, different windfields, such as analysed winds and forecast winds, don't change the statistical structure of wave forecast error. Then, as the wave model has stationary statistics, we may assume that

$$(Dh^d - < Dh >^d)_i + (Dh^f - < Dh >^f)_i = b \cdot \ (Dh^f - < Dh >^f)_i \tag{19}$$

Eq. 16 then may be written as:

$$r_{i,j}(x) = \frac{1}{S_i \cdot S_j \cdot N} \sum_{m=1}^{N} [(Dh_m^f - < Dh >^f)_i (Dh_m^f - < Dh >^f)_j]. \quad (20)$$

here N is total accumulation of individual hindcast and forecast runs, and for simplicity the constant b is taken as 1. $< Dh >^f$ in Eq. (20) is ensemble average of error accumulation of forecast deviated from hindcast, Dh_m^f is forecast error for every individual forecast and hindcast run. The observation error and hindcast errors contained in the standard deviation of forecast error are determined as follows: since we never know the 'true ' wave height , as shown in Fig. 1, we assume that the observed wave height oscilates around the 'true' value,hence a moving average in time is used.

$$\bar{H}_i = \frac{1}{m} \sum_{i=-3}^{i=+3} (H_i), \quad m = 7 \quad (21)$$

and that the measured value deviates from this moving average by

$$DH_i = (H_i - \bar{H}_i)/(\bar{H}_i), \quad (22)$$

and

$$< DH >= 0.0 \,, \quad (23)$$

The normalized standard deviation is

$$S_o = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (DH_i - < DH >)^2} \,, \quad (24)$$

here N is total wave record. This normalized standard deviation S_o is similar in statistical sense to the scatter index used in Table 1. From 286 successive wave records,the normalized standard deviation is 0.04, which is nearly one nineth of scatter index for +24 forecast. Here the calibration error of wave instrument is neglected. The standard deviation of wave forecast error obtained from several observation points in the North sea listed in table 1 is calculated by

$$S_p = \sqrt{< (DH_i - < DH >)^2 >} \quad (25)$$

where $< DH >$ is the ensemble average of forecast error and DH_i is the individual forecast error, the average value of standard deviation for South part of the North sea during Nov. 89 is about 50 cm.

4 Determination of analytical autocorrelation function

Because of nonlinearity of correlation sample data as shown in Fig. 2, an analytical function to fit such data usually has several adjustable parameters. Here a straightforward least-square error variance is used to fit desirable parameters. To this end, we minimized difference E,

$$E = \sqrt{\left(\frac{1}{N} \sum_{i=1}^{N} (R(x_i) - r(x_i))^2\right)} , \tag{26}$$

where $r(x_i)$ is a set of autocorrelation data calculated from Eq. (20), $R(x_i)$ is an analytical function with several adjustable parameters, x_i is separation between two grid points of wave model, N is total number of sample data. After some trial and error, we used the following correlation function:

$$R(x) = (A + B \cdot sin(w_o x)) \cdot exp(-\xi \cdot x) , \tag{27}$$

with the coefficients

$$A = \frac{(S_p)_i \cdot (S_p)_j}{S_i \cdot S_j} , B = 0.38 \cdot A , w = 0.4 , \xi = .225 , \tag{28}$$

The sample data used for empirical estimation is taken from hindcast and forecast runs of two months (Nov. and Dec., 1989, total 136 individual runs). Fig. 3 shows the analytical function and the sample auto-correlation data averaged over a subseparation of 0.2 GONO grid point. As shown in Fig. 4, there is usually no sample data between the analysis point and the nearby grid points. Data in a subseparation increases with the distance from the analysis point. Two points for correlation analysis are GONO grid points (11,9) and (8,25), the standard deviation errors between analytical function and the sample data for those two points are 0.074 and 0.094 respectively. Since the sample data of autocorrelation, as shown in Fig. 1, is quite scattered, the correlation might be anisotropical. We tried an analytical correlation function with anisotropy having the form,

$$R(x, \theta) = G(\theta, x) (A + B.sin(w_o \cdot x)) \cdot exp(-\xi_1 \cdot x) , \tag{29}$$

where

$$G(\theta, x) = cos^{\xi_2 x}(w_1 \cdot (\theta - \theta_o)) , \tag{30}$$

where A, B and w have the same value as in Eq. (27), and θ is direction counted from the model north. In addition, we take

$$\xi_1 = 0.165 , \theta_o = 1.75 , w_1 = 0.32, \xi_2 = 0.16. \tag{31}$$

The anisotropical function for four different directions is shown in Fig 5, Table 2 indicates the improvement of accuracy obtained by using an anisotropical correlation function.

5 Spectral properties of analytical correlation function

Equation (27) may be written by combination of two terms

$$R(x) = g(x)f(x) , \tag{32}$$

where

$$g(x) = (A + Bsin(w_o \cdot x)) , f(x) = exp(-\xi \cdot x) . \tag{33}$$

Then, Fourier transform pairs may be denoted as:

$$g(x) \iff G(w) , f(x) \iff F(w) , \tag{34}$$

and

$$R(x) \iff S(w) , S(w) \iff G(w) * F(w) .$$

Here $*$ is convolution of two separated Fourier transforms. Fourier transform of $g(x)$ gives

$$G(w) = \int g(x) \cdot exp(-iw \cdot x) \, dx , \tag{35}$$

where boundaries of the integration extend to infinity, which leaves an open area to determine the truncation length arbitrarily, and

$$w = \pi \cdot k/L , \tag{36}$$

Here k is wave number and L is truncation length of correlation. Performing the integration in Eq. (5.4) gives

$$G(w) = \delta(w) + iB \cdot \delta(w + w_o)/2 - iB \cdot \delta(w - w_o)/2 . \tag{37}$$

the fourier transform of damped exponential function is

$$F(w) = \int f(x) \cdot exp(-iw \cdot x)\, dx = (v - iw)/(v + iw) , \tag{38}$$

and the fourier transform of correlation $R(X)$ is the convolution

$$S(w) = Real \int G(w')\, F\,(w - w')\, dw' \tag{39}$$

$$= \frac{A.\xi}{\xi^2 + w^2} + \frac{B.(w + w_o)}{2 \cdot (\xi^2 + (w + w_o)^2)} - \frac{B.(w - w_o)}{2 \cdot (\xi^2 + (w - w_o)^2)} .$$

The boundaries of the integration of the fourier transforms extend to infinity in Eq (35) and Eq (38); however, the correlation fit to a set of sample data is restricted to a truncation distance. It is obvious that no data is available outside this limitation. Here, the Hamming window is used to impose this limitation:

$$H(x) = 0.54 + 0.46cos(x \cdot w^*) , \ |x| < L_m , \tag{40}$$

where

$$w^* = \pi / L_m , \tag{41}$$

L_m is truncated length, its Fourier transform is

$$Q(w) = 0.54\, \delta(w) + 0.23\delta(w + w^*) + 0.23\delta(w - w^*) . \tag{42}$$

The truncation spectrum $S(w)$ may be obtained by fitting the actual spectrum through this window by the convolution,

$$S_T(w) = \{\int_{-L_m}^{L_m} Q(w')\, S\,(w - w')\, dw'\} \tag{43}$$

$$= \{0.54.S(w) + 0.23.S(w - w^*) + 0.23S(w + w^*)\} .$$

From Eq. (34), it is possible to express the autocorrelation as the fourier transform of the spectrum $S(w)$,

$$R(x) = \frac{1}{2\pi} \int S(w).exp(iw.x)\, dw . \tag{44}$$

Since $S(w)$ can be separated into odd and even terms, the above equation may be expressed as an one-side integration:

$$R(x) = \frac{2}{\pi} \int (S_{even}(w) \cdot \ cos(wx) \ + \ S_{odd}(w) \cdot \ sin(wx) \ )dw \ . \qquad (45)$$

For zero separation x we have,

$$lim_{L_m \to \infty} \frac{2}{\pi} \int S_T(w) \ dw = \ R(0) \ = \ A \ , \qquad (46)$$

or integration of $S(w)$ is exactly the auto-correlation at the analysis point. However, the integration in Eq (46) represents spectrum between w and $w+dw$ which contributes to interpolation, the integral bounds is finite because truncation length can not extend to infinity. This means that the integration of Eq (46) is always an approximation to interpolation. The calculation of Eq. (46) indicated that a truncation length of 6 GONO grid points has good approximation to the interpolation.

6 Summary and Conclusion

In the least-square sense, the statistical method through which the spatial statistical information of variable was incorporated provides an optimal estimation of correlation or covariance used in sea wave data assimilation, without the need for assumptions of spatial homogeneity, or isotropy. The only assumption is that the observation error is random. The same spatial statistical properties for both forecast error and hindcast error are essential for our estimation of correlation based on hindcast minus forecast fields. An intuitive method was employed to determine the analytical function with isotropy and anisotropy. The isotropical function may fit the sample data quite well in sense of standard error, and the anisotropical function may improve the accuracy of correlation at a distance to 6 GONO grid points. Beyond this, it has no much effect because of exponential damped function. The sample data of correlation at GONO point (11,9) (corresponding to middle area of The North sea) and point (8,25) (corresponding to Norwegian sea) were used to verify the correlation functions. The correlation of forecast error of sea waves appears positive and has large horizontal scale. Usually, the positive definiteness of correlation is restricted to a limited area, and it is reasonable to assume that there is no appreciable contribution outside the limited area. We used the Hamming window to restrict the spectrum to a finite range, and the Fourier transform of correlation, or wave number spectrum, from which the accuracy of interpolation is attributed to the different truncated lengths, was discussed. We have found that a truncation length of 6 GONO grid points is a good approximation. Because the forecast error of wind and wave are strongly correlated, the accuracy of present scheme of wave data assimilation may be improved by employing a multivariate interpolation as described in section two.

7 Acknowledgements

The work presented in this paper was supported by KNMI, P.O. Box 201, 3730 AE De Bilt, The Netherlands. I am indebted to Dr. Peter A.E.M. Janssen for his review of the paper and many helpful suggestions. Also gratefully acknowledged is useful discussions with Dr. G. Burgers and E. Bouws. Finally, I thank Mr. R. van Moerkerken for his support of sea wave information, and thank Greet de Graaf and Brigitta Kamphuis for their careful preparation of the text.

References

[1] Gandin, L. S., 1963. Objective Analysis of Meteorological Fields. Gidrometeor., Leningrad. *[Israel Program for Scientific Translations, Jerusalem*, 1965, 242 pp.]

[2] Rutherford, I.D., 1973. Experiments on the updating of P.E.forecasts with real wind and geopotential data. Preprints Third Conf. Probablity and statistics in Atmospheric Science, Boulder, Colo., *Amer. Meteor.Soc.*, 198-201

[3] Thiebaux, H. J., 1974. Estimation of covariances of meteorological parameters using local-time averages. *J. Appl. Meteor.*, 13, 596-600.

Table 1:

Some statistical parameters of waves for 4 selected points, where N, N^+, N^-, SD and SI in the first row represent total observation data, positive bias number, negative bias number, rms error and scatter index, H is mean wave height. Hind., +12F and +24F in the first column denote hindcast, +12 hours and +24 hours forecast respectively. The unit of H, Bias and SD are in cm. Two months (October and November, 1989) observation data and the corresponding forecast results are listed here.

	N	H	Bias	SD	SI	N^+	N^-
Euro							
Hind.	87	145	-33	29	20	6	82
+12F	87	143	-30	37	26	10	73
+24F	87	143	-31	47	33	14	64
IJmuiden							
Hind.	87	152	-23	34	22	15	71
+12F	82	150	-18	44	29	21	61
+24F	77	150	-21	54	36	24	53
K13							
Hind.	88	171	-26	26	15	8	79
+12F	83	170	-19	40	23	14	64
+24F	78	171	-20	56	33	19	58
Mike							
Hind.	67	101	-6	14	14	21	46
+12F	67	101	-13	28	28	20	47
+24F	67	101	-13	34	34	24	43
Average							
Hind.	82	142	-22	26	18	13	70
+12F	80	141	-20	31	27	16	61
+24F	77	141	-17	48	34	20	55

Table 2:

Comparisons of standard errors between the sample correlation data and analytical correlation functions with isotropy and anisotropy, where k, R1(x), R2(x) represent direction bins, the standard errors of isotropical function at GONO point (11,9) and point (8,25); $R1(x,\theta)$ and $R2(x,\theta)$ correspond to the stardard errors of anisotropical function in those two points. The last column denote average errors, all value are multiplied by 100.

k=	1	2	3	4	5	6	7	8	9	10	11	12	Ave
R1(x)	5	4	2	4	7	5	6	5	3	13	23	12	7.4
R2(x)	10	16	2	5	3	10	13	5	14	18	7	5	9.4
$R1(x,\theta)$	5	5	2	4	7	6	7	8	7	10	11	6	6.0
$R2(x,\theta)$	5	12	3	5	3	11	14	5	9	10	14	8	8.2

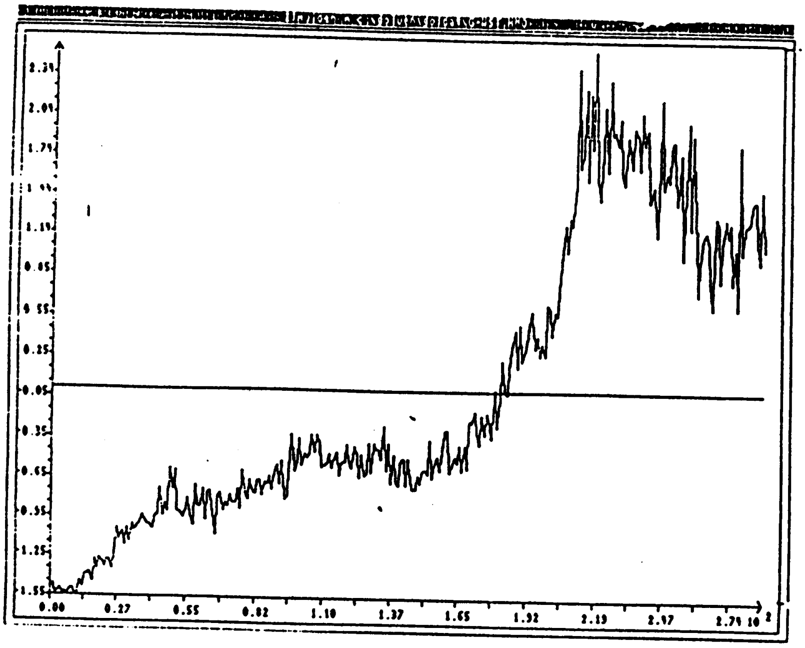

Fig.1. Contiunal data of significant wave height from Ijmuiden over two hrs period(07:00-09:00 Nov 09,1989), where abscissa is time and ordinate is wave height, each of which is calculated by spectral analysis over 10 mins wave record.

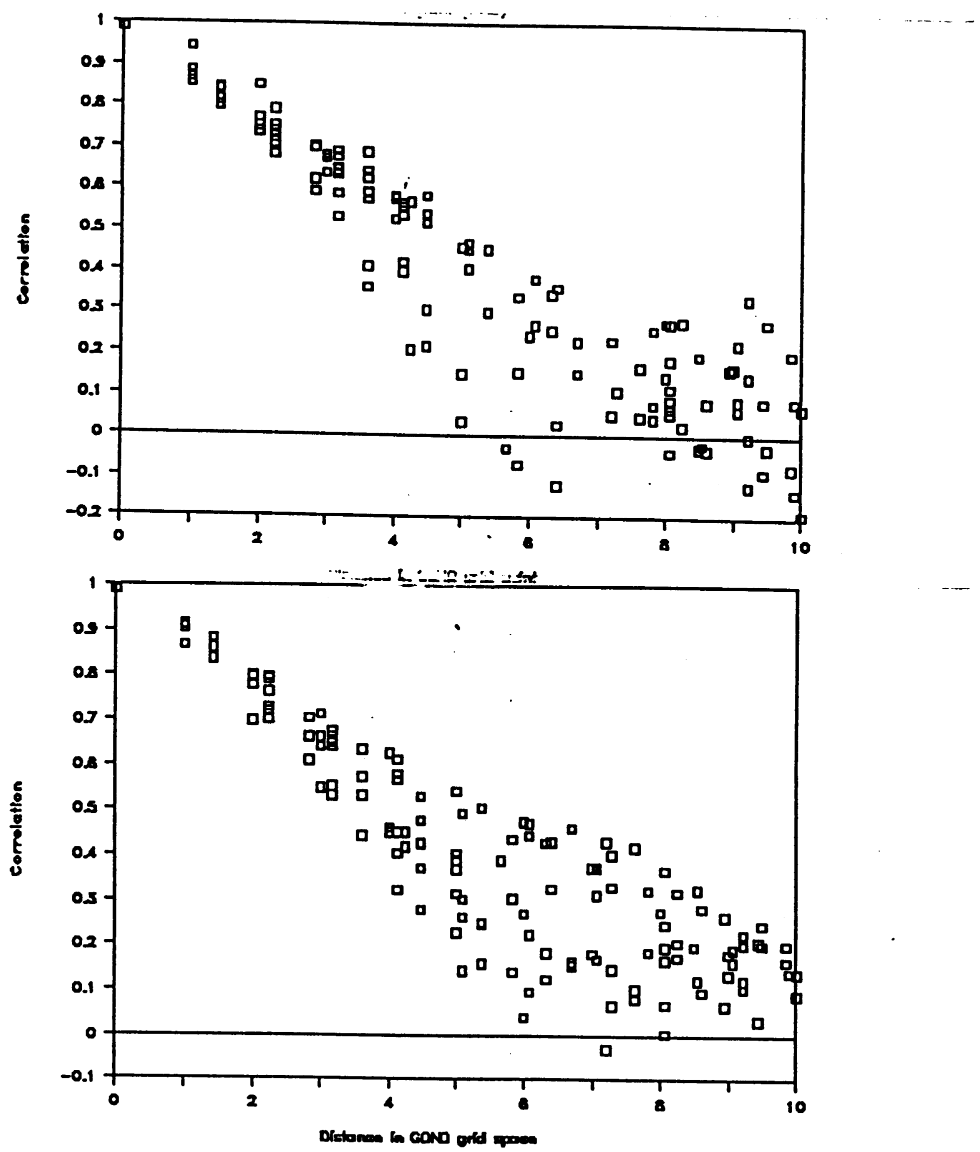

Fig.2 Data sample of auto-correlation for GONO point (8,25; Fig.2a) and (11,9; Fig.2b),where abscissa is distance from analysis point counted in GONO grid point and ordinate is correlation value.

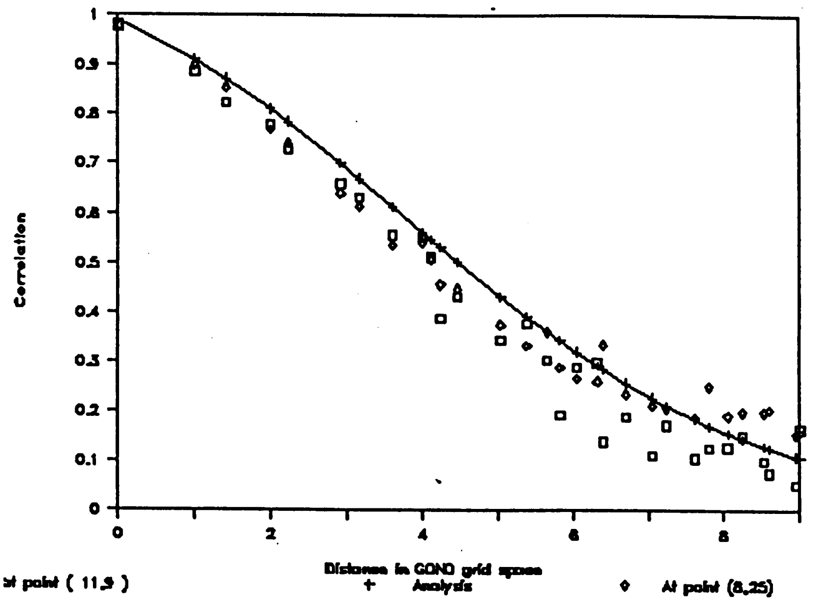

Fig.3.Analytical correlation function (line and +) and data sample of correlation at point (8,25) and (11,9),the data sample has been averaged over a 0.2 subseparation,number of data in the separation increase with the distance from the analysis point.

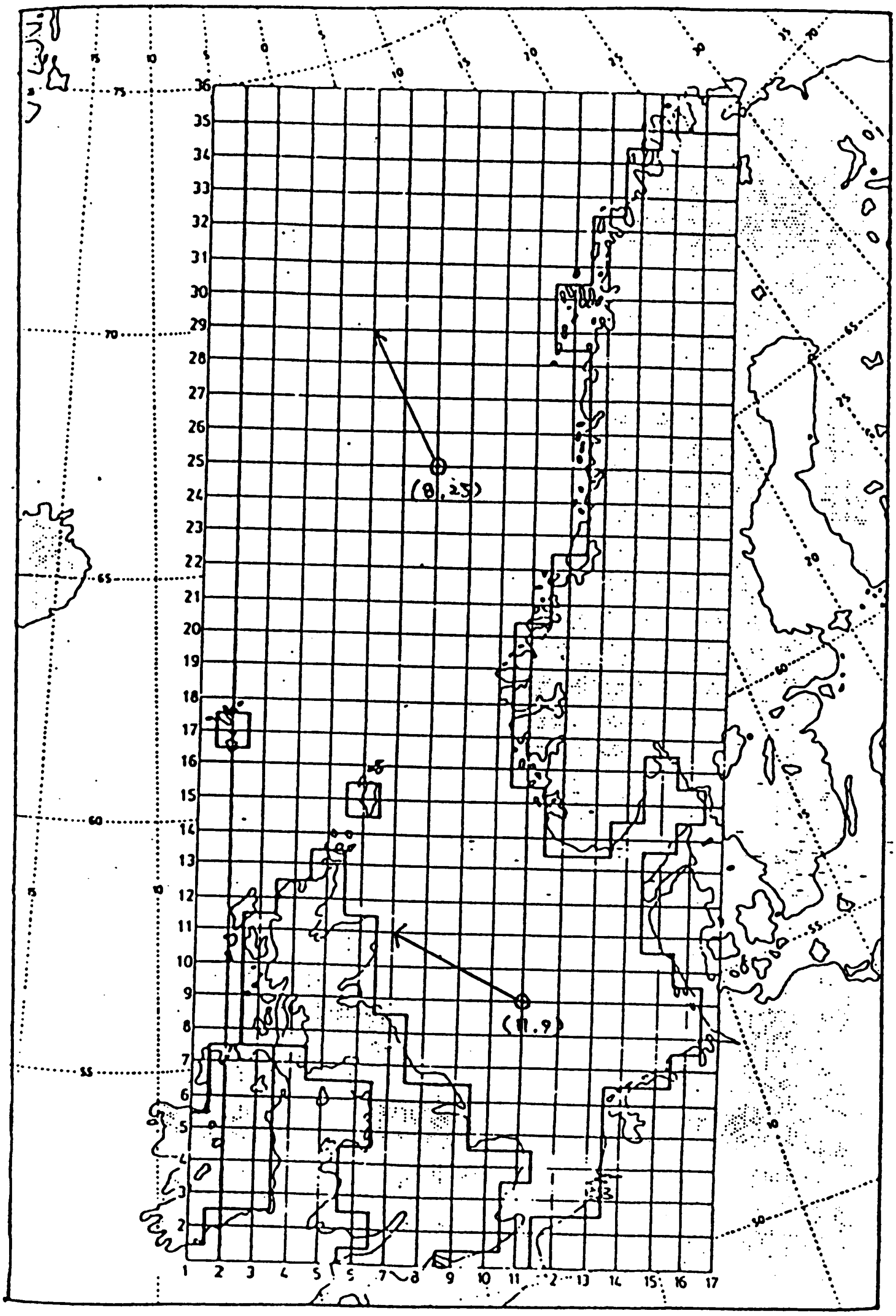

Fig.4.GONO grid points and the positions two points (circled) used for analysis of correlation function,the line with arrow marks the distance from the analysis point.

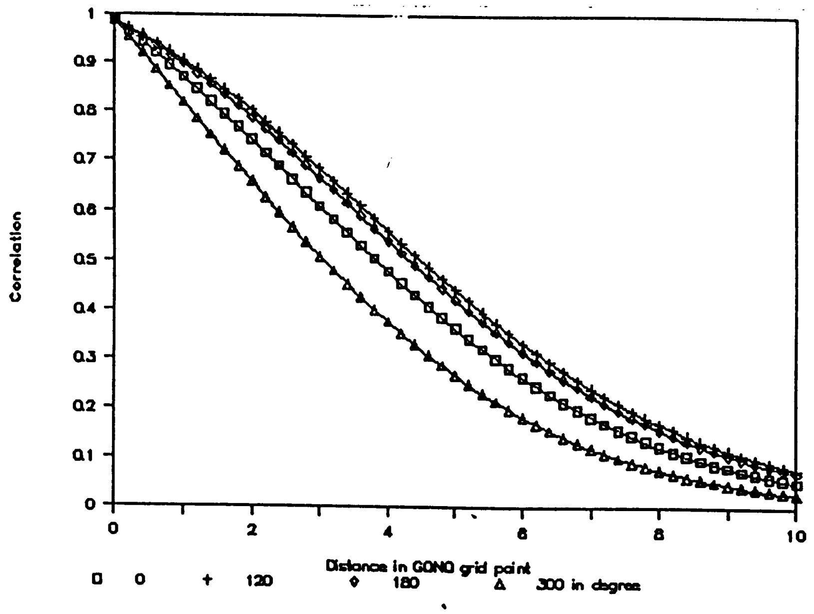

Fig.5. Analytical function of anisotropical correlation,the function is plotted for four different directions. The representation of coordinates is the same as indicated in Fig.2.