CHAPTER 42

AN EFFICIENT STATISTICAL METHOD OF ESTIMATION OF EXTREME MARITIME EVENTS USING TWO SETS OF RELATED INFORMATION

by

B. MANOHA, J. BERNIER and M. GRAFF *

ABSTRACT

The construction of marine works necessitates a good knowledge of the extreme events which can affect it, such as the storm waves or the water levels. A statistical method of estimation of these extreme events has been developped at L.N.H. The principle of this method, based on the theory of renewal processes, is presented hereafter, as well as the results that it can provide. Some applications of the method are described in the paper ; in particular the application to a site where the available data are insufficient, and where the method enables to precise the results by using complementary numerous data on a neighbouring site, is presented.

1. INTRODUCTION

The important nuclear program developped by Electricité de France since many years, has necessitated, for reasons of safety and reliability, to develop efficient statistical methods enabling to estimate, on the basis of necessarily limited information, extreme natural phenomena such as :

- extreme storm surges (coastal plants) or extreme river flows (river plants), as to fix the highest possible water levels and consequently the corresponding platform levels, thus preventing from flood problems

- extreme low water levels in order to secure a continuous feeding of the intake pumping stations

- extreme waves as to define correct protections of the breakwaters.

The extreme events need to satisfy safety criteria related to return periods associated to these events : for coastal nuclear plants, these return periods are usually 100 years for waves and water level setdowns, 1000 years for water level setups and for river floods (river plants).

* Laboratoire National d'Hydraulique, Electricité de France, Chatou, FRANCE

The method presented hereafter is based on statistical methods previously developped at L.N.H. for problems of floods in rivers ; these methods, derived from the theory of renewal processes, have been adapted to the study of extreme waves, setups and setdowns. It enables a large variety of developments by comparison with the classical method of annual maxima ; in particular this method enables to precise the results by taking into account the complementary information available on a neighbouring site.

We shall only present here examples concerning extreme waves. After a short recall of the notion of return period and of the main other method used, we shall describe the principle of the "partial duration series" method, its practical use, its particular application to the use of two sets of related information, and we shall end with some examples of other applications which can prove useful.

2. THE NOTION OF RETURN PERIOD

It is certainly not useless to recall the important notion of return period and its practical meaning :

The return period T associated with the value H_T of a random variable H (wave height) is defined as the average number of years between two successive occurrencies of the event : $H > H_T$. This definition implicitly assumes a given time unit, the year, at which scale the variable H is defined, for example :

. annual maximum wave height

. significant wave height, for 20 minutes recordings, over-topped every year during 1 day, 2 days, ...

This definition does not imply any regularity of the occurrencies $H > H_T$. The basic assumptions concerning the annual successive occurrencies are :

a/ their independance

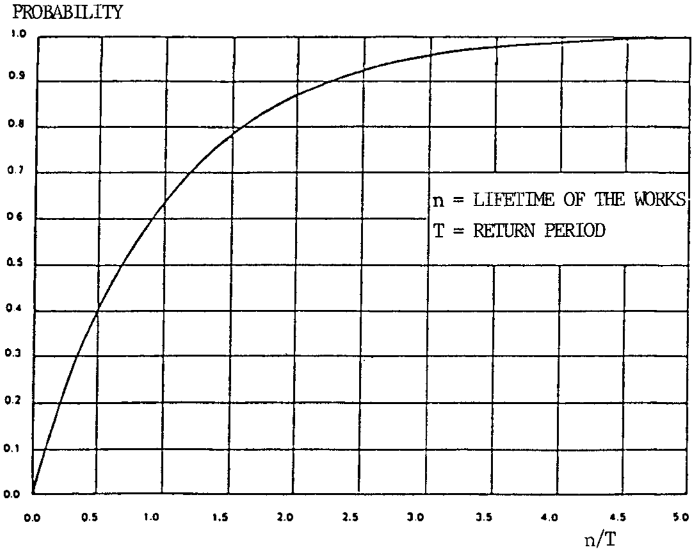
b/ the stationarity : invariance of their probabilities of occurrency during the time.

The number of years N separating two successive occurrencies of the studied event is a random variable such as :

$$\text{Prob}\left[N \leqslant n\right] = 1 - (1 - \frac{1}{T})^n \simeq 1 - e^{-\frac{n}{T}}$$

(valid for T big enough)

In practice, this formula gives, for a project whose lifetime is n, the probability of occurrence of an event whose return period is T (figure 1) :



- Figure 1 -

For example, a plant whose lifetime is 50 years (n = 50) has approximately 10% chances to see, at least once in its life, an event associated with a return period of 500 years (T = 500) ; this occurrency reaches 40% for T = 100 years, and 63% for T = 50 years, which is far from negligible.

This emphasizes the point that the return period of a given project must be chosen equal to several times the lifetime of the works in order to limit the risk of damages to an acceptable level.

3. THE STATISTICAL METHODS

The classical method of estimation of the risks associated to extreme geophysical events uses the notion of annual random variable : maximum annual wave height, maximum annual rainfalls or floods, ... : it is known as the method of annual maximas.

Considering that the risk is based on an annual time scale, it seems logical to base the estimation on series of annual maximum values derived from complete sets of data.

This method, widely used in many geophysical problems, can be applied only if the number of years of available data is sufficiant to make a reasonnable extrapolation to extreme events.

In fact, very often, the available data for marine phenomena are related to short periods of measurements, and this method of annual maxima gives information which can lead to a very important uncertainty.

Another approach enables to use a more complete information : not necessarily the complete set of basic data, which would lead to a too complex probabilistic structure, but the most "significant" information constituted of the maximum values of each storm (for sea states), these values being chosen higher than a given threshold Ho as to eliminate small and non significant values.

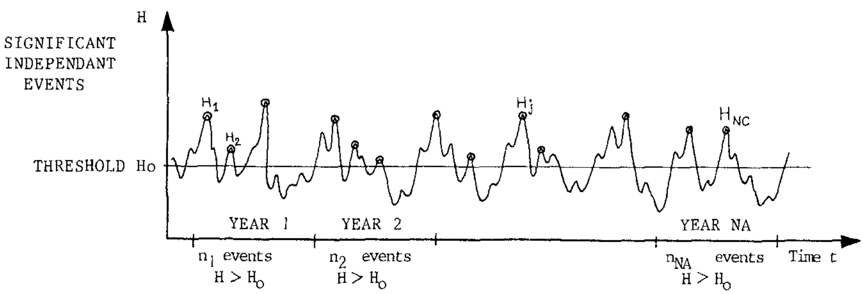
This approach, based on the theory of renewal processes (by analogy with probabilistic models used in problems of reliability of components in plants), is called the "partial duration series" method.

As we shall see hereafter, this method, compared to other classical methods, enables to use a more complete set of information (bigger than a given threshold), and is then much more efficient in the very frequent case of short periods of available data. Moreover, the method enables to provide a "confidence interval" giving the range of reliability of the obtained results.

4. THE PARTIAL DURATION SERIES METHOD

4.1. Principle of the method

Given a chronological series of wave height data (figure 2), one can define significant independant events Hj higher than a given threshold Ho as follows :



- Figure 2 - Chronology of waves

The notion of significant events, related to the threshold Ho, means that the small values corresponding to weak storms do not need to be considered in this kind of problems where we are in fact only interested in strong storms and their associated probabilistic laws enabling to extrapolate to extreme events.

The other notion of independant storms is very important and corresponds to an assumption of the method. Various definitions can be used : the simplest one consists in considering that two successive storms, characterized by their maximum wave heights Hj, are independant if the wave height between those two storms comes lower than the threshold Ho (figure 2) ; other more sophisticated, and more adapted, definitions use different thresholds (one H_1 for defining significant storms, another lower one Ho, related to H_1 as a percentage of it, defining the end of the storm), or integrate the notion of duration of calm (H < Ho) between two successive storms. In fact this notion of independance is rather difficult to appreciate automatically and must be considered very carefully and cautiously.

From this chronology of significant independant events measured during NA years, one can then define two series :

- Series $\{n_i\}$ i = 1, NA

  where n_i is the number of events H > Ho during year i

  and P (k) is the associated probability of trespassing k times Ho during year i

- Series $\{H_j\}$ j = 1, NC

  series of independant wave heights H_j > Ho (NC values)

  with F^k (H) associated probability that H > height > Ho during the year.

The selected storms are then interpreted as the combination of two stochastic processes : one of occurrencies (n_i , P (k)), the other of distribution of selected wave heights (H_j , F^k (H)).

For annual risks, the two processes must be combined to obtain the distribution of the annual maximum wave (same kind of procedure for decennial, centennial or other return periods) :

$$\text{Prob } (H_{max} > H) = 1 - \sum_{k=o}^{\infty} P(k) . F^k (H)$$

4.2. Statistical laws

. Series of occurrencies $\{n_i\}$:

In practice, when the selection of independant storms has been correctly made and the threshold Ho chosen high enough, the series of occurrencies most of the time follows a Poisson process of intensity

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Generally, one assumes that the storm phenomenon is stationary, i.e. its statistical characteristics are independant of the season, which means λ = constant.

However it happens sometimes that this Poisson law does not appear realistic ; the computer program developped at LNH enables then to use also the binomial negative law for the series of occurrencies :

$$P(k) = \frac{\Gamma(\gamma + k)}{k!\,\Gamma(k)} \; p^{\gamma} (1 + p)^{k}$$

$$\text{with } \Gamma(\alpha) = \int_0^{\infty} e^{-x} \, x^{\alpha - 1} \, dx$$

. Series of wave heights $\{H_j\}$:

The series of wave heights can follow various statistical laws which have also to be adjusted in the method. The computer program tests automatically 4 different laws which, from experience, proved generally to be the best fitted with the wave data :

- Exponential : Prob $(h < H) = 1 - e^{-\rho (H - Ho)}$
- Weibull : Prob $(h < H) = 1 - e^{-\rho (H - Ho)^{p}}$
- Log-exponential : Prob $(h < H) = 1 - e^{-\rho (\mathrm{Log}\, H - \mathrm{Log}\, Ho)}$
- Squares : Prob $(h < H) = 1 - e^{-\rho (H^2 - Ho^2)}$

In practice, for estimate of extreme wave heights, the Weibull law proved, on the french coasts, to be generally the best fitting law. It has to be noted that the exponential law is a particular application of Weibull law with p = 1.

4.3. Adjustment of the statistical laws :

For all the statistical laws tested for the series of occurrencies and heights, the various parameters characterizing the theoretical expressions (λ, ρ, p in particular) are estimated through the classical statistical method of Maximum Likelihood, except for the binomial negative law (series $\{n_i\}$) where the Momentum method is used.

4.4. Validation tests

The validation of the assumptions of the model and of the adjusted distribution laws (occurrencies and wave heights series) is performed through statistical tests which are systematically applied for various thresholds Ho. These tests are :

- Serial independance test of the successive selected maximum wave heights higher than a given threshold Ho. This test uses the classical transformation $T = \frac{\sqrt{V}\,R}{1-R^2}$, approximatively distributed according to a Student law with V degrees of freddom, where R is the autocorrelation coefficient and V the number minus 2 of couples of storms used for the calculation of R. The calculated value T is then compared with the threshold of probabilistic significance equal to 0.95.

- Classical X^2 test. The determination of the K classes necessary to the calculation of :

$$X^2 = \sum_{i=1}^{K} \frac{(n_i - v_i)^2}{v_i}$$

is made so that the theoritical absolute frequencies v_i, to be compared with the observed frequencies n_i, are equal to the constant value n/K (n is the size of the studied sample).

4.5. Confidence intervals

On the contrary of semi-empirical methods, the partial duration series method, which uses a rigorous modelisation of the probabilistic properties of the wave heights process, enables to calculate confidence intervals associated to wave heights H_T corresponding to a given return period T.

The confidence interval, associated to a confidence threshold $1-\alpha$ (70%, 90%, ...), covers the real, but unknown, value of the extreme wave height H_T with a probability equal to $1-\alpha$.

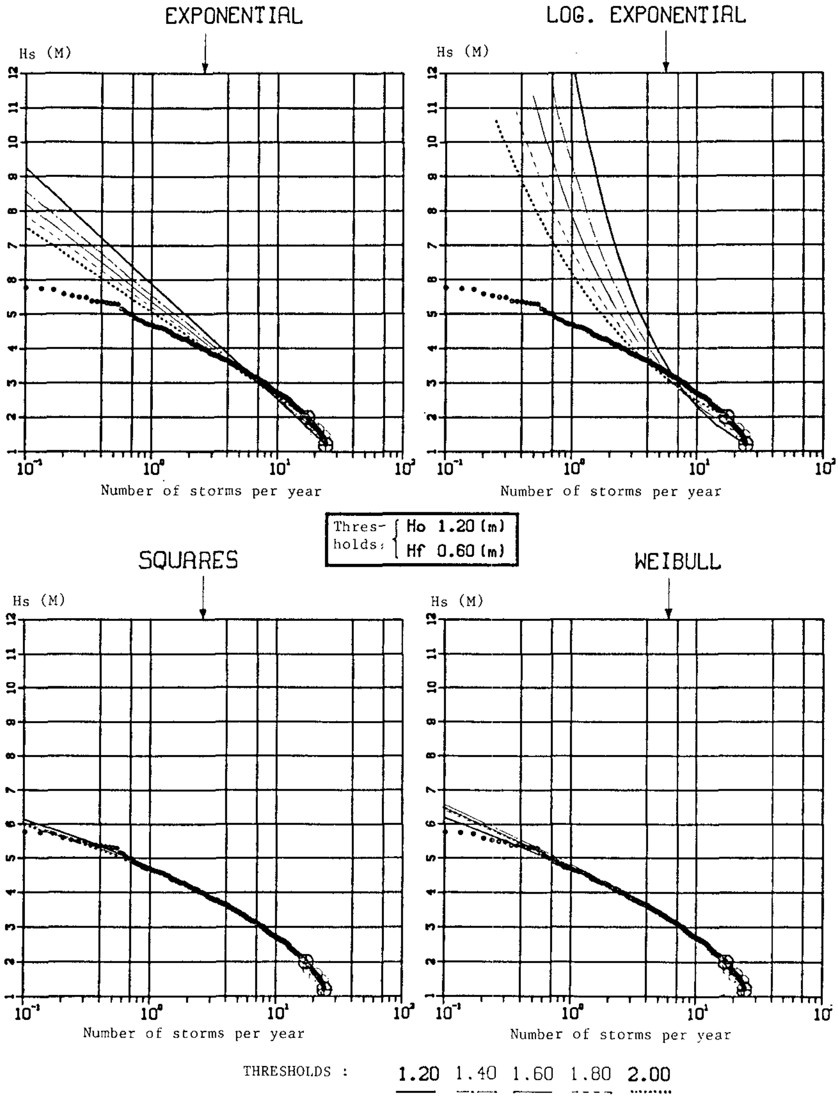
The amplitude of this confidence interval enables to measure the uncertainties on H_T related to the sampling. It is of course all the bigger as the size of the sample is smaller. The interval is calculated for each fixed statistical law, and then it does not take into account the uncertainty related to the choice of the law, this choice being made through the validation tests previously described (cf. 4.4), and through graphical adjustments (see hereafter).

The method of calculation of these confidence intervals, automatically integrated to the computer program which calculates the 70% confidence interval, uses the approximate properties of the estimations of the maximum likelikoods.

5. PRACTICAL USE OF THE METHOD

Given a sample of wave heights during a certain period of recordings, the use of the method implies different steps before leading to the final result, i.e. decennial, centennial, ... wave height with its corresponding confidence interval :

- Choice of the series of independant significant storms whose maximum wave heights H are bigger than a given threshold Ho.



- Figure 3 - Adjustment to the different laws

These wave heights can correspond to mean, or significant H 1/3, or maximum Hmax wave heights derived from the recordings, depending on the problem which is considered. As mentionned previously, different definitions of "independant" events can be chosen.

- From this sample H>Ho, the adjustments to the 4 statistical laws are tested automatically : exponential, Weibull, Log-exponential and squares law (see example of results on figure 3).

- For each statistical law, different thresholds H'o over a basic threshold Ho are tested. As a matter of fact, the threshold which will be finally chosen will have to be :

  . not too low as to represent sufficiently well the impact of the extreme observed events (we are interested here in extreme events corresponding to big return periods)

  . not too high in order to keep a sufficiently large number of events as to be able to make correct statistics on a big enough sample.

- The final choice (statistical law and threshold to be finally considered) will then be made on the basis of the three main following criteria (example of final result on figure 4) :

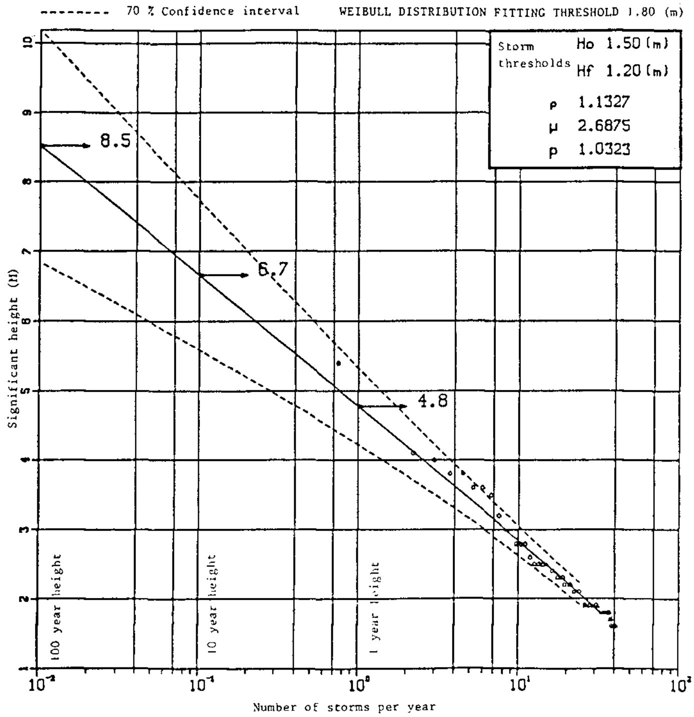


Fig.4 - Annual, decennial and centennial significant waves at BIARRITZ

- graphic adjustment : for each law and threshold, one can see visually if the theoretical adjustment fits or not with the observed values : for the example on figure 3 one can see that the squares law gives the best adjustment with the observations

- X^2 statistical tests : the method provides automatically the X^2 value for each law and threshold and gives an advice, based on statistical criteria, on wheter the adjustment should be accepted or not.

- stability with the threshold : usually, if the adjustment is good, the results on the extreme events do not depend very much on the choice of the threshold H'o.

On the basis of these criteria, one can then make his choice, which is not always very easy, especially for short periods of measurements. An example of final choice is given on figure 4 : on the site of Biarritz, south of France, where some measurements where available from may 1972 to february 1977, the Weibull law was finally chosen with a threshold H'o = 1.80 m (basic threshold Ho = 1.50 m for the definition of independant storms) ; the centennial significant H1/3 wave height (return period of 100 years) was found to be 8.5 m with a 70% confidence interval between 6.8 m and 10.2 m, showing that in this case the uncertainty on the results remains rather large (many holes in the period of measurements).

The method has been used systematically up to now, for 10 sites on the French coast (about 20 more applications are currently under progress), and shows in particular that we need, if we want to get sufficiently reliable results on a site for return periods of at least 10 years, an average period of wave measurements of about 4 years minimum, taking into account the fact that, for many reasons, on average only half of this period (about 2 complete cumulated years) will really provide reliable and useful data.

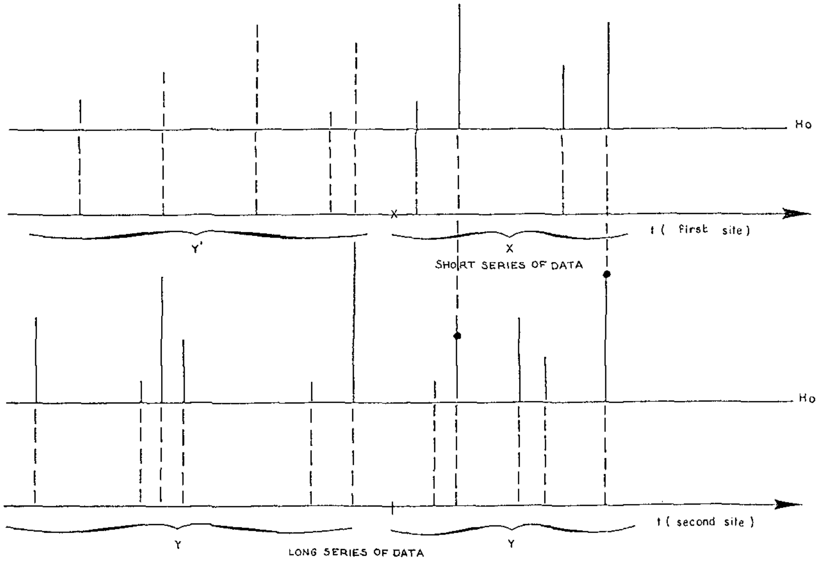
6. USE OF TWO SETS OF RELATED DATA

6.1. Principle

One of the interests of the method is that it is easy to take into account complementary information (punctual historical data, or information on a neighbouring site) in order to precise the results.

A fresent case concerning the wave recordings is that of a site where only a short series of data X is available, which can lead to a big uncertainty on the estimation of extreme events.

If there exists, on a neighbouring site, a longer series of available data Y, this complementary information can be used to precise the results on site X if there exists a good correlation between the data recorded during the concomitant period (figure 5) :



- Figure 5 - Example of concomitant and non concomitant data on the two sites

The idea is then, in this case, to complete the existing information X, Y on the two sites by a fictive information Y' corresponding to missing data on the first site.

The problem is now to estimate this missing information Y' (Y,λ,ρ, p, r) ; as a matter of fact this estimation depends, not only on the complementary existing information Y, but also on the parameters of the partial duration series (λ,ρ, p), and on the correlation parameter r between the two sites.

The calibration of the model will then consist in estimating the parameters λ ,ρ, p and r on the basis of the real observations X completed by the rebuilt complementary fictive data Y' (Y,λ,ρ, p, r), which necessitates a certain preliminary knowledge of these parameters : this approach is then necessarily based on an iterative process. The E.M. algorithm (E : expectation, M : maximization) developped by Dempster and al-1977 was adopted to solve the problem.

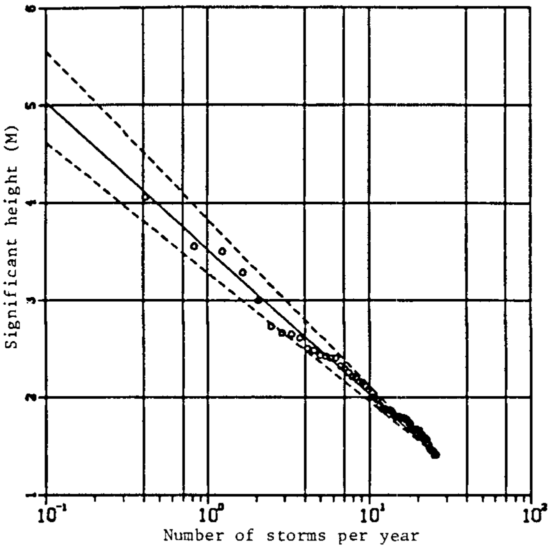
6.2 Example

An example of application of this method is presented on figure 6 on the case of Paluel, where 5 year's are available, completed by the 10 years of available data on the neighbouring site of Antifer. It shows the estimation, at Paluel, of annual, decennial and centennial significant wave heights obtained without and with the complementary Antifer information. One can see that with the exponential law chosen as the best adjustment, the decennial wave height changes from 5 m to

EXPONENTIAL DISTRIBUTION

Fitting threshold 1.40 m

2σ - 0.94 m



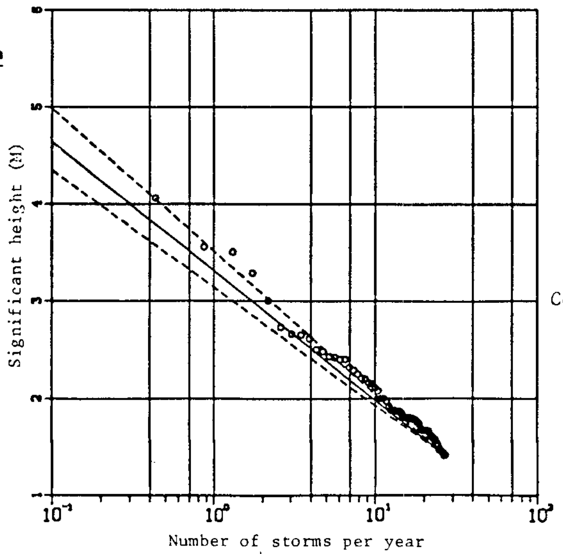
Storm thresholds:

Ho 1.20 m
Hf 0.60 m

with the short set of data on PALUEL (datawell)

13/01/77 - 11/11/81
(PALUEL)

2σ - 0.64 m



CONCOMITANCE

Threshold: 1.20 m

Correlation: 0.8553

with the fuller set of data on ANTIFER

3/ 7/72 - 11/ 5/82
(ANTIFER)

- Figure 6 - Estimation of the design wave at PALUEL

4.6 m with this complementary information, the confidence interval being moreover reduced from 0.94 m to 0.64 m. In this case the correlation coefficient corresponding to the concomitant data was 0.86.

6.3. Further developments

The general problems to be thought about now are to try to determine which is the minimum duration of the short series, and wich is the minimum value of the correlation in order to be sure to obtain good results.

These problems are linked and depend very much on the quality and homogeneity of the data, and on the proximity of the site where the complementary data are available, considering that we are only interested here in extreme events corresponding to strong storms which, if the neighbouring site is not too far, should reasonnably occur rather similarily on both sites.

Up to now, only the application to Paluel - Antifer has been made and no general rule (if any can be found...) can be deduced from the results. Moreoever the method was applied with various lengths of the measuring period at Paluel (6, 9, 12, 21 and 24 months), showing that, in this case, at least 21 months of cumulated available data were necessary, the correlation coefficient varying from 0.61 (with 6 months) to 0.84 (21 months).

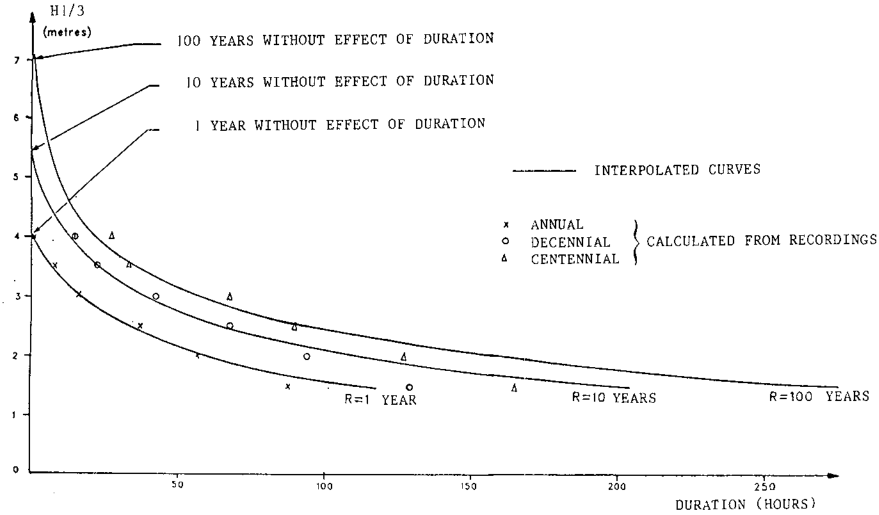
Some applications to other sites are planned and should provide interesting information on these particular points. It seems nevertheless that the degree of correlation between the two sites should be at least 0.8 as to be able to get reliable results on the first site.

7. OTHER POSSIBLE APPLICATIONS

Without going into detail in the various applications which have been made with the partial duration series method, we can mention some of them and some further developments associated with the method :

- Integration of historical data : the same kind of methodology as described for two sets of related data can be used in the case where punctual historical data are available. This is often the case for setups or river floods, and can also, but less often, happen for waves ; this complementary information can be introduced in the data in order to precise the results concerning extreme events.

- Duration of the storms : for a given sample of wave recordings $H>Ho$, one can associate a complementary third series of data $\{Dj\}$ corresponding to the durations of the storms. Then a statistical analysis of these durations, being made for various increasing thresholds $H'o$ through the partial duration series method, can be performed in order to associate to a given return period, a series of wave heights with their corresponding durations ; this gives then a much more complete information than only the classical extreme wave height associated to a "zero" duration. This has been made for different sites on the french coast and gives very interesting results which can be used for many purposes (example of height - duration diagram at Antifer on figure 7).



- Figure 7 - Height - duration diagram at ANTIFER

- Effect of the sampling interval : very often wave measurements are made every 2 or 3 hours, sometimes rather continuously (every 20 or 40 minutes for example). Then a study has been made to see the effect of the chosen sampling interval on the results ; this study is in fact very connected with the previous one on the durations of storms.

- Effect of the definition of independant storms : different definitions, more or less sophisticated, have also been tested and gave some interested information which, as for the other applications hereabove mentionned, cannot unfortunately be described here in detail.

8. CONCLUSIONS

The partial duration series method has been systematically applied to extreme wave heights on a certain number of sites on the french coast, through an automatic procedure developped in a computer program. Similar versions applied to setups, setdowns and river floods have been used for the setting of nuclear platforms along the coast or rivers. The basic philosophy of elaboration of the method was to develop practical tools enabling a complete calculation of extreme waves, from the preparation of data, the determination of analysed events, the statistical validation of the assumptions, the adjustments of the laws, to the final presentation (graphic in particular) of the results.

The need of promoting methods adapted to the most common applications encountered in practice led to develop also procedures and methods enabling in particular the treatment of various types of incomplete information.

Further useful developments can be envisaged concerning in particular the adjustment of a non-stationnary model (effect of the seasons in particular) and of a model of treatment of the missing data.

REFERENCES

J.BERNIER, M.GRAFF - Méthodes d'estimations de la houle de projet avec données incomplètes - Rapport EDF-DER HE/42/83.34.

A.R.DEMPSTER, N.M.LAIRD, D.B.RUBIN - Maximum likelihood from incomplete data via the E.M. algorithm. J.R Stat. Soc - B - n° 39-1-1977.

J.FEUILLET, J.BERNIER, Y.COEFFE, B.CHALOIN - Le dimensionnement des digues à talus - Eyrolles - 1985.

M.GRAFF, Y.COEFFE, J.BERNIER, B.DELANNOY - Notice d'utilisation du programme EVENAL - Estimations d'évènements extrêmes (houle) par la méthode du renouvellement - Rapport EDF-DER HE/42/84.35.

M.G.KENDALL, A.STUART - Advanced theory of statistics - Vol 2 - Griffin - 1961.

J.MIQUEL, J.BERNIER - Sûreté des centrales et états de la mer - Bulletin EDF-DER série A n° 2 - 1981 - pp 73-78.

J.MIQUEL - Guide pratique d'estimation des probabilités de crue - Eyrolles - Collection EDF-DER n° 53 - 1984.