FILE.

INTERNAL DOCUMENT No. 199

COMMERCIAL IN CONFIDENCE

I.O.S.

STATISTICAL ASPECTS OF THE DESIGN OF FIXED STRUCTURES

P.G. CHALLENOR

REPORT TO SIPM

NATURAL ENVIRONMENT RESEARCH COUNCIL
INSTITUTE OF OCEANOGRAPHIC SCIENCES

INSTITUTE OF OCEANOGRAPHIC SCIENCES

Wormley, Godalming,
Surrey, GU8 5UB.
(0428 - 79 - 4141)

(Director: Dr. A.S. Laughton)

Bidston Observatory,
Birkenhead,
Merseyside, L43 7RA.
(051 - 653 - 8633)

(Assistant Director: Dr. D.E. Cartwright)

Crossway,
Taunton,
Somerset, TA1 2DW.
(0823 - 86211)

(Assistant Director: M.J. Tucker

COMMERCIAL IN CONFIDENCE

STATISTICAL ASPECTS OF THE DESIGN OF FIXED STRUCTURES

Report to SIPM

P.G. Challenor

December 1983

CONTENTS

0. Summary

This report brings together several inter-related aspects of the use of meteorological and oceanographic data in the design of fixed structures. Although the majority of the work has been previously described in earlier publications, special attention has been given to their application to the determination of design wave criteria for engineering applications.

Consideration has been paid to the possible use of Bayesian statistics in re-evaluating design criteria. It is concluded that, except in the simplest cases, this approach is beyond our present level of capabilities, although in the future it may become feasible.

A rational procedure for the adoption of return periods for design criteria is proposed based on the concept of a design risk. Several examples are given to clarify the use of design risk in different situations applicable to the offshore engineering industry.

KEYWORDS:

DESIGN CRITERIA, DESIGN RISK, STATISTICS, STRUCTURAL DESIGN, WAVE HEIGHT.

1. Introduction

This report deals with some of the statistical problems encountered when designing structures subject to environmental forcing. Some of these problems are outlined in a paper by Vugts (1982).

The methods detailed in this report are intended to be general and applicable to any scalar environmental parameter, e.g. wave height, wind speed; therefore as far as is possible, no distributional assumptions will be made concerning these variables apart from assuming that they are positive. Section 4 however is concerned with the problems of obtaining return values of individual waves from significant wave height data. Otherwise the methods described are completely general and refer to the probabilities of exceeding return values of any positive variable be it mean wind speed, significant wave height, individual wave height or any response parameter.

Section 2 deals with the probability of an event with a particular return period occurring within given periods of time. Section 3 is concerned with the possibility of using additional information to update estimates of return values, and section 4 describes the problems specific to wave height. Section 5 introduces the concept of design risk and resolves the apparent paradoxes given in Vugts' paper.

2. The Probabilities of Rare Events

2.1. In this chapter the probability of a rare event being exceeded is discussed. This theory is not limited to a single environmental parameter, such as the individual wave height, but is general and may be applied to any long period return values, such as mean wind speed, gust speed, current speed or significant wave height.

To begin, define the n-year return value as that value that will be exceeded on average once in n years.

This definition tells us about the long term behaviour of exceedences of the return value, if we record data for m ($\gg$ n) years we would expect to get m/n exceedences of the n-year return value. However this does not tell us about the short term properties of these exceedences, for example the probability of the 50-year return value being exceeded in 5 or 10 years.

Let X be the environmental variable under consideration. As stated above this could be any such parameter, (e.g. mean wind speed, significant wave height or individual wave height). Assume that we can measure X directly and also assume that the probability of two exceedences of the return value (x_r say) within one year is negligible. This second assumption will be considered below.

Consider the event that x_r is exceeded once during a year and let the probability of this be p ($={}^1/r$ where r is the return period of x_r).

Then

P (x_r is exceeded in the first year) = p

P (x_r is exceeded in the second year but not in the first)

= P (x_r is not exceeded in the first year) x P (x_r is exceeded in the second)

= (1-p)p

P (x_r is exceeded in the kth year but not in the preceding (k-1) years)

$= \prod_{1=1}^{k-1}$ P (x_r is not exceeded in the i^{th} year) x P (x_r is exceeded in the k^{th})

$= (1-p)^{k-1}p$ (2.1)

Therefore

P (x_r is exceeded at least once in the first k years)

$= \sum_{i=1}^{k}$ P(x_r is exceeded in i^{th} year but not before)

$$= \sum_{i=1}^{k} (1-p)^{i-1} p$$

$$= p \sum_{i=1}^{k} (1-p)^{i-1}$$

This is a geometric progression which when summed gives

$$= 1-(1-p)^{k} \qquad (2.2)$$

where p is the probability that x_r is exceeded in any one year.

Equation 2.2 gives the probability that the return value will be exceeded in a given number of years but does not give the number of exceedences. The probabilities for the number of exceedences in a fixed period will be given below.

An alternative derivation, which may be easier to follow, is thus.

P (x_r is exceeded in k years) = 1-P (x_r is not exceeded in k years)

$$= 1-\prod_{i=1}^{k} P\,(x_r \text{ is not exceeded in year } i)$$

$$= 1-(1-p)^{k}$$

This distribution is the geometric distribution well known in statistics, Johnson and Kotz (1969). This problem is identical statistically to the number of throws of a die until a six is thrown.

Probabilities calculated from this formula (2.2) are given in Table 1 for p=0.02 (the fifty year return value) and p=0.01 (the hundred year return value) for differing values of k.

2.2. The two main assumptions involved in the above analysis are that each year's exceedence or non-exceedence is independent of every other year and that the probability that two or more exceedences occur in the same year is negligible. These two assumptions are linked. If we do not allow more than one event per year then these events are very likely to be independent, however if we allow several exceedences to happen in the same year this independence becomes less likely. As long as we are concerned with large return periods (of the order of tens of years) then the probability of two extreme events in one year will be very low indeed. For shorter return periods this is not so and the above analysis is not correct. However there is no necessity for us to use years; we could use shorter periods of time and thus effectively reduce the probability of two or more events within this period to zero again. This must be done with caution however. Within a year there is in general a seasonal variation which implies that p does not remain constant from trial to trial; this has to be ignored. In addition to this, as our periods get shorter the assumption of independence becomes increasingly less valid. To demonstrate that this approach is consistent, consider 6 month sections as opposed to years; the '100 year' return value is now the 50 year. From table 1 it can be seen that the probability of exceedence of the 50 year return value in k years is approximately the same as the 100 year return value being exceeded in 2k years.

If we are considering 6 monthly sections as opposed to years it is possible for the return value to be exceeded twice in one year. This explains the very small difference seen in the table.

The only other major assumption is that p does not alter with time. This has already been referred to above in connection with seasonal variation. There is a distinct possibility of climatic change that would make p a function of time. Such a variation has been suggested for the North Sea by Rye (1976) and for the eastern seaboard of the U.S.A. by Resio (1978). If such a change is in progress the probabilities given here are in error. However, since climatic changes such as these are slow in their effect, this error should be small for periods less than, say, ten years.

Returning to equation 2.2

$$P\ (x_r \text{ exceeded at least once in k years}) = 1-(1-p)^k$$

In general for small y,

$$(1-y)^{1/y} \simeq 1/e$$

Hence $(1-y)^k = [(1-y)^{1/y}]^{ky} \simeq e^{-ky}$

So for small p

P (x_r exceeded at least once in k years) $\simeq 1-\exp(-pk)$

But $p={}^1/_r$, so

P(x_r exceeded at least once in k years) $\simeq 1-\exp(-k/r)$ (2.3)

i.e. the probability that a return value is exceeded in a given number of years depends only on the ratio of this number to the return period. This therefore gives a theoretical justification for what was illustrated above using table 1. Table 2 gives a selection of ratios and their associated probabilities and Figure 1 gives a graph of this ratio vs. probability.

2.3. What is the probability of a given number of occurrences in a set number of years?

Let the period of interest be k years and consider P(x_r is exceeded exactly n times in k years). There must be n exceedences (i.e. a p^n term) and k-n non-exceedences (i.e. a $(1-p)^{k-n}$ term), also there are ${}^kC_n$ $\left(=\frac{k!}{n!(k-n)!}\right)$ ways of choosing the years with exceedences so

$$P(x_r \text{ is exceeded exactly n times in k years}) = {}^kC_n\, p^n(1-p)^{k-n} \qquad (2.4)$$

(This is the binomial distribution, for details see Johnson & Kotz (1969).

An approximation to this distribution, for $p \to 0$ and $k \to \infty$ as pk remains constant, is given by the Poisson distribution.

i.e. $P(x_r \text{ is exceeded exactly k times in n years}) \simeq \frac{e^{-kp}(kp)^n}{n!}$ (2.5)

Replacing p by 1/r as before gives

$$P(x_r \text{ is exceeded exactly n times in k years}) \simeq \frac{e^{-k/r}}{n!} \left(\frac{k}{r}\right)^n \qquad (2.6)$$

Again this probability depends only on the ratio of the period of interest to the return period. These probabilities are tabulated in table 3. For small values of n we would not expect this approximation to be good. Note that there is a finite probability of n being larger than k for all values of k but this is only applicable for small k. Table 4 shows the exact probabilities for p=0.02 and p=0.01, this clearly shows that the Poisson approximation is good down to very small k given that p is small.

3. Modifying design criteria with later data

3.1. The normal procedure used to estimate return values, and hence set design criteria, is to collect data over, usually, a year or two, but ideally, over a longer period. A distribution is then fitted to these data either empirically or by the analysis of extremes. Return values can then be estimated from this distribution. Once structures have been built, further data are normally collected and it might be useful if it were possible to modify the design criteria using these data. Data gathered in this way will vary in quality. Some of it will have been obtained from instruments and will be of high quality. On the other hand some will merely consist of the fact that certain return values have or have not been exceeded. Obviously these types of data must be dealt with in different ways.

One general method of dealing with all types of additional data would be to use the techniques of Bayesian inference. The statistical methods normally used in engineering and the physical sciences belong to another school of statistics concerned with so called classical inference. The differences between the classical, or frequentist, and the Bayesian schools are many and it is proposed to give only the brief details that are important in the present context. A fuller description of the arguments can be found in Barnett (1973).

It will be concluded however that the requirements of the Bayesian methods make them inapplicable to this problem at present.

3.2. In the classical framework parameters, such as the fifty year return value, are regarded as fixed, but unknown, constants, and the aim of statistical inference is to produce the best estimates of them from the data. Bayesian inference on the other hand regards these parameters as random variables and hence is concerned with the distribution of things such as the fifty year return value rather than estimators of them. (This point of view is regarded as unacceptable by many scientists). The basic tool of this Bayesian approach is our knowledge or degree of belief about the parameter in question. We have some initial beliefs about, for example, the 50 year return value. When we are presented with some data we change those beliefs (this is achieved, mathematically, via Bayes theorem). The initial distribution we hold to be true is the prior distribution, $f(\theta)$ say, and the final beliefs the posterior, $f(\theta|\underline{x})$ say (the vertical bar is used here in its statistical sense showing a distribution that is conditional - i.e. $f(\theta|\underline{x})$ means the distribution of the parameter θ given the data $\underline{x}$). Bayes theorem gives us

$$f(\theta|\underline{x}) \propto L(\underline{x}|\theta)f(\theta) \qquad (3.1)$$

where $L(\underline{x}|\theta)$ is the likelihood of the data, i.e. the probability of getting that data as a function of the parameter θ. The constant of proportionality is easily found since the left hand side must integrate to 1 over the range of definition of θ.

3.3. An example should make this clearer. Assume the random variable, X, has a negative exponential distribution with parameter θ, i.e.

$$f(x)=\theta e^{-\theta x}$$

We are interested in estimating θ; assume that before making any measurement of x we believe that θ also has an exponential distribution but with a parameter of 2

$$\text{i.e. } f(\theta)=2e^{-2\theta}$$

This is the prior distribution.

Assuming we take a sample size n, $x_1 \ldots x_n$, Bayes theorem gives

$$f(\theta|\underline{x}) \propto \prod_{i=1}^{n} \theta e^{-\theta x_i} \cdot 2e^{-2\theta}$$

$$\propto 2\theta^n e^{-\theta \Sigma x_i} e^{-2\theta}$$

$$\propto \theta^n e^{-\theta(2+\Sigma x_i)}$$

since $f(\theta|x)$ must integrate to 1 we can determine the constant of proportionality, and so find $f(\theta|x)$ exactly. In this example

$$f(\theta|\underline{x}) = \frac{\theta^n e^{-\theta(2+\Sigma x_i)}}{(2+\Sigma x_i)^{n+1} \; n!} \qquad (3.2)$$

which is a gamma distribution.

If we need a point estimate of the parameter, for instance for producing design criteria, then probably the best one would be to take the mean of the posterior distribution, however the mode and median are also obvious candidates.

3.4. What problems are involved in applying these techniques in practice? Apart from the philosophical objections that most scientists and engineers like to think in terms of return values being fixed natural constants, there are several practical problems. Firstly the mathematics can get very complicated. In the example of 3.3 we used probably the simplest distribution, the negative exponential, for both the prior and the likelihood but the integration necessary to get the posterior was not simple. Environmental variables are normally assumed to have distributions, such as the Weibull, which have two (or even three) parameters. This means that, in contrast to the example, the likelihood has two parameters, θ_1 and θ_2, say, and hence our prior and posterior distributions must also be bivariate. It is possible to parameterise the likelihood so that the return value of interest is one of these parameters. The second parameter which we are not really concerned with is called a "nuisance parameter", and has to be integrated out of the posterior distribution. These integrations are usually formidable and often have to be performed numerically. The hardest problem is specifying the prior distribution - and the selection of this distribution is as important as the data. Specifying the prior for only one parameter, as in the example, is not trivial. Where other, nuisance, parameters are involved these must be included in the prior. This implies that the prior must be a multivariate distribution.

For these reasons the Bayesian approach is at present not practicable for this problem. However there are significant advances being made in Bayesian statistics, in particular in the use of computers to aid in the production of priors, and within the next few years these methods may well become viable.

3.5. Having considered the possibilities of Bayesian inference we can now return to a more conventional approach. Consider first the situation where the new data are of a similar, or even better, quality than the original data. There are two possible approaches:

- combining both sets of data and analysing as one
- or analysing the new data separately.

Analysing both sets of data separately enables us to make a check on the analysis, both should give similar answers, and the results may throw some light on climatic variation. However the results obtained will be less accurate than using the whole data set, so for improved estimates of design criteria the new data should be added to the old and the whole set analysed. Since there will now be more data it is possible that a different analysis technique, such as the analysis of extremes, will be possible.

3.6. The treatment of good quality new data is reasonably straight forward; the problems arise when poorer quality data are produced. It is assumed here that data of this form are not measured directly but inferred from damage to the structure and therefore consist of the exceedence or non-exceedence of certain values. Normally these will be return values used in the original design, and hence the return periods will usually be quite large.

What then can we say about the fact that some return value has or has not occurred? Consider first the situation where a long period return value has happened, e.g. an oil rig has been hit by the '50-year wave'. This is an unlikely event, the probability of which was derived in chapter 2. Table 3 gives the probabilities of multiple occurrences. From this table it can be seen that for periods of time that are short compared to the return period it is unlikely that the return value will be exceeded at all, let alone more than once. Therefore if the estimated return value is exceeded more than once in a short period of time serious doubt must be thrown upon the estimate.

The above discussion assumes that we start to record whether the return value has occurred or not at a random time. Recording may start, however, at an extreme event. Because we assume that each recording interval is statistically independent of all others the point at which we start recording will make no difference to the results. However, if we are using the time between exceedences to check our estimate of the return value it is important to remember that in this case we have one more exceedence than we have intervals between exceedences. If we start at a random time there are the same number. Independence means that statistically the time from a random point to the next exceedence is the same as the time between events.

3.7. The opposite problem is now considered; the non-occurrence of a return value. Vugts (1982) suggests that structures with a service life of 20-30 years are designed for an event with a return period of 100 years. In table 3 this means that for such a structure n/r 0.3 and therefore the probability of the return value not being exceeded during its service life is 0.74. That the return value is not exceeded is only to be expected, if it were the design procedure would not be very satisfactory.

However over durations greater than the return period, it is possible to produce statistics of non-occurrence that do tell us something. For short return periods, 5 years say, this period is not long. Table 2 gives the relevant probabilities. After 75% of the return period there is a 50% chance of the return value being exceeded and after three return periods only a 5% chance of it not being exceeded. Taking 90% as a form of significance level if the estimated return value had not occurred in 2.3 (return periods) then there would be good evidence that the estimate was too high. Therefore when estimates of return values are obtained for design purposes it seems sensible to get some for shorter periods as well. If it were possible to detect the exceedence of these low return period return values these data could be used as a check on the methods employed to obtain the design criteria. It should be noted however that nothing can be found out about the design values directly in this way only about the methods used to obtain them.

3.8. So far we have stated only that 'serious doubt' or 'good evidence' exists indicating that estimates are wrong. Can anything more than this be said? Without actual data the answer is, unfortunately, no; knowing that an estimate is in error does not necessarily mean that a correction can be made. The information we have concerning the error is of a very basic kind. We know whether the estimate is too large or too small and, possibly, a crude idea of the magnitude of the error. For instance if the estimated 50 year return value had been exceeded twice in 5 years or a 5 year return value had not occurred in 25 years the estimates would obviously be in error. This sort of information, while useful, is not quantifiable and hence cannot be used to correct existing estimates or re-estimate the return values. For the production of new, better estimates, other data of a higher quality are required.

4. Special problems of wave height

4.1. For most environmental variables the return value required can be considered simply as a quantile of the variable measured. Wave height however is different. The parameter measured is normally significant wave height, H_s. To describe H_s as a wave height can be somewhat misleading. H_s is usually defined as four times the standard deviation of the sea surface (although it is still sometimes used to mean the average height of the third highest waves) and is thus a measure of the roughness or energy of the sea surface rather than an actual wave height. For some applications it is not the roughness of the sea surface that is required but the heights of the individual waves. Normally this is done by computing $H_{max,3\ hrs}$ (defined as the most likely highest wave in 3 hours assuming a constant H_s) from H_s and T_z. Return values of $H_{max,3hrs}$ are normally obtained by one of two methods:-

(a) Converting a return value of H_s to $H_{max,3hr}$

(b) Deriving a set of $H_{max,3hr}$ data from H_s data and obtaining a return value from this set.

Neither of these methods give a return value of individual wave height as they ignore any contribution from less extreme sea states to the distribution of individual waves. The estimate they do produce is the most likely highest wave associated with an extreme sea state. For some applications, e.g. the forces on a structure during a severe sea state, this is what is required. On the other hand, for some applications, e.g. the clearance height of a member, the return value of the individual wave height is required. A method for deriving this, based upon constructing the distribution of individual waves from H_s-T_z data will be described in this chapter.

4.2. Let the probability density of individual zero-crossing wave height, H, given H_s be

$$f(h|h_s)$$

and if the density of H_s is

$$f(h_s)$$

the unconditioned, or marginal, pdf of H is given by

$$f(h) = f(h|h_s)\ f(h_s) \qquad (4.1)$$

From assumptions of narrow-bandedness and linearity, it is normally assumed that $f(h|h_s)$ is approximately a Rayleigh distribution i.e.

$$f(h|h_s) = \frac{4h}{h_s^2} \exp\left\{\frac{-2h^2}{h_s^2}\right\} \qquad (4.2)$$

The problem then is to find the probability density function of H_s. So far no-one has suggested a theoretical distribution for H_s and therefore an empirical method must be used. The following method is due to Battjes (1970, 1972) and is further described in Carter and Challenor (1981).

The marginal distribution of H can be written

$$P(H<h) = \int P(H<h \mid h_s)\ f(h_s)dh_s$$

The pdf of h_s, $f(h_s)$, used here is not the usual distribution of significant wave height. The usual distribution, used, for example, for extrapolation to get return values, is the probability distribution associated with a randomly chosen time. The distribution given here on the other hand is associated with a randomly chosen wave. Since large waves, in general, have longer periods than small ones these two distributions are not the same.

$P(H<h \mid h_s)$ can be derived from, for example, the Rayleigh distribution given above and, if we change the integration into a summation, we can derive an approximation to $f(h_s)$ from the usual H_s-T_z joint distribution or "scatter diagram".

If n_{ij} is the number of occurrences of h_{si} and t_{zj} divided by the total number of 3hr periods in the diagram, the number of waves in one 3 hour period is estimated by $3hr/t_{zj}$ and the number of waves in this box is $(3hr/t_{zj}) \times n_{ij}$.

Therefore

$$f(h_s) = (\Sigma_j n_{ij} / t_{zj}) / (\Sigma_i \Sigma_j n_{ij}/t_{zj})$$

and

$$P(H\ h) = \frac{\Sigma_i \Sigma_j (1-\exp[-2(h/h_{si})^2])\, n_{ij}/t_{zj}}{\Sigma_i \Sigma_j n_{ij}/t_{zj}} \qquad (4.3)$$

4.3. This expression gives the distribution of zero-crossing wave heights within the scatter plot. Unfortunately it cannot be used directly to obtain return values as the scatter diagram does not usually include any of the higher h_s's.

In these circumstances, there are two possible approaches that could be followed:-

(a) Produce a scatter plot that would correspond to a very long data series (of the order of the return period required if not longer) and use this in conjunction with equation 4.3 to calculate the long term distribution of individual waves. Obviously the generation of the long term scatter plot is the crux of this problem, but it may be possible to use hindcasting to extend the measured scatter plot.

(b) Use the measured scatter diagram to calculate the distribution of individual waves, which is then extrapolated to the appropriate return value.

Both procedures will now be outlined

4.4. Procedure (a)

1. This method may be summarised as follows: Generate a long-term scatter diagram for the area (e.g. by hindcasting or by extrapolating the H_s distribution and the use of an empirical steepness relationship to produce the relevant T_z's).

2. Use equation 4.3 to obtain the required return value.

This method suffers from some disadvantages. One is that the H_s series has to be extrapolated much further than in any other technique for obtaining return values. To obtain good estimates of, say, the 100 year return value of individual waves at least ten occurrences would be needed so the extrapolation would have to be of the order of 1000 years. In most areas there can be little justification for extrapolating one or two years' data to these extremes without further information on rare sea states, for example from hindcasts.

Producing the periods for the whole scatter plot is also not straightforward. Simply using steepness criteria obtained from severe sea states will not give the values in the body of the table. However the periods only enter equation 4.3 as terms of the form $\sum_j n_{ij} t_{zj}^{-1}$ which is the mean value of T_z^{-1} for the i^{th} value of H_s. This statistic

resulting distribution is always negative exponential (a special case of the Weibull, with shape parameter of one). Although there is no theoretical justification for this (or for that matter for the Weibull), no station has yet been found where this distribution does not fit.

4.6. The use of the Rayleigh distribution for upcrossing waves entails certain assumptions. One of these is that the heights of the crests and following troughs are perfectly correlated. Except in the limiting case of narrow banded waves it is not known what the true distribution of zero up-crossing waves is. This lack of correlation would tend to reduce the probability of the occurrence of the highest waves. Even if we assume that the waves are perfectly correlated in this way there are still problems. Forristal (1978) and Longuet-Higgins (1980) have both suggested that non-linearities can cause high waves to have distributions other than the Rayleigh usually assumed. The effect of this on the analysis is not known but it is likely to be of secondary importance in most engineering applications.

4.7. Method (b) described in this section has worked well for UK waters where the extreme waves occur as part of the general population, but has not, as far as is known, been used for other areas of the world. Other parts of the world present difficulties, especially where there are separate populations of wave height corresponding to different causes, e.g. where hurricanes are important. Wave climates can in general be divided into 3 distinct groups.

I situations where the extreme waves come from the same population as the average conditions (e.g. the North Sea)

II situations where the extremes come from a different clearly defined population, i.e. hurricanes or typhoons (e.g. the Gulf of Mexico)

III situations where the extreme events can either belong to the same population as the ordinary waves or to some other distinct population (e.g. parts of the South China Sea).

The methods described above have been developed to deal with situation I. However it should be possible to adapt them to deal with the other two situations.

In situation II it should be possible, given sufficient data, to produce a scatter diagram for the extreme population, i.e. the typhoons or hurricanes, and use the methods above, together with the expected number of storms per year, to obtain return values. Situation III is rather more difficult. If it is possible to separate the data from the two populations then the methods for situations I and II could be applied to the relevant data sets and the worst result taken. If it is not possible to resolve the two data sets it has to be hoped that either during the analyses the populations would separate or that there was no significant difference between them.

4.8. Comparing procedures (a) and (b) it can be seen that <u>if</u> a reliable estimate of the long term scatter plot could be produced for a given area (by a combination of measurements, hindcasts and other techniques) then procedure (a) would be preferable. The advantage of procedure (a) is that it gives a scatter plot which can be used for other purposes and for this reason most engineers would prefer it. However in many situations it will not be possible to produce such a reliable estimate and in these cases procedure (b) which only involves the extrapolation of a single parameter (wave height) will still have to be used.

5. The Concept of Design Risk and some Applications

5.1. This section introduces the concept of the 'design risk' and shows how this can be used to solve some practical problems and apparent paradoxes which may be raised in setting return period criteria for structures with different service lives. The concepts derived in this chapter are, of course, equally applicable to all environmental parameters, e.g. significant wave height, mean wind speed, gust speed etc.

5.2. The design risk is defined to be the probability that a design criterion, expressed here as a return value, will be exceeded during the design life of a structure.

$$\text{i.e. } d = 1-(1-p)^s \qquad (5.1)$$

where d is the design risk

s is the service life

and $p={}^1/r$ is the probability that the r-year return value will be exceeded in one year.

There are obviously an infinite number of p and s combinations that will give the same design risk. These are given by

$$p = 1-(1-d)^{1/s}$$

$$\text{i.e. } \ln(1-p) = \frac{1}{s}\ln(1-d) \qquad (5.2)$$

This function of p, s is plotted for a variety of values of d in figure 2.

For small p, $\ln(1-p) \approx -p$

since $p = {}^1/r$

$$r \approx -s/\ln(1-d) \qquad (5.3)$$

Hence for a fixed design risk the return period is approximately proportional to the service life.

This approximation is only valid for large r but in practice, as can be seen from the figure, is excellent for $r > 10$ and reasonable over the whole range.

For this relationship it is easily seen that for combinations r_1 and s_1 and r_2 and s_2 to have the same design risk we must have

$$\frac{r_1}{s_1} = \frac{r_2}{s_2} \qquad (5.4)$$

and for r_1, $r_2 > 10$ this should be a reasonable approximation.

Some applications and examples of this theory will now be given.

5.3. Example 1. Suppose that the return period used for a service life of 25 years is 100 years. Then the design risk, d, is given by equation (5.1) as

$$d = 1-(1-{}^{1}/100)^{25}$$
$$= 0.22$$

NB this figure can also be obtained from table 2.

i.e. there is a 22% chance of the design criterion being exceeded in the service life of the structure.

Now suppose that another structure is being designed with a proposed service life of 5 years. The return period, r say, that should be used in its design for the same design risk is from equation 5.4 given by

$$\frac{r}{5} = \frac{100}{25}$$

i.e. r = 20 years.

(Alternatively from 5.3 we get $r = -s/\ln(1-d) = 20$).

5.4. Example 2. Consider the situation where, after a number of years, the structure is inspected and found to be in an "as new" condition. What return value should be used for the remainder of its life?

Suppose that the original service life of the structure was s years and that after s_1 years it has been inspected and found to be in its original condition. Let $s_2=s-s_1$ and let r be the return period for the original design. The original design risk is (eqn. 5.1):

$$d = (1-(1-p)^s)$$
$$(p= {}^{1}/r)$$

Assuming we keep the same design risk after the inspection,

then $\frac{r_2}{s_2} = \frac{r}{s}$

i.e. $r_2 = \frac{rs_2}{s}$

For instance a structure designed with a service life of 20 years using a return period of 100 years, would, assuming it passed a suitable inspection after 10 years, need to be checked for a 50 year return period for the rest of its design life (10 years).

5.5. Example 3. It may be contended, incorrectly, that using the arguments of example 2 one could design a structure for short return periods and inspect them regularly until the total service life had been reached. For instance, consider a structure that has been designed for a service life of 5 years with a return period of the design conditions of 30 years. After 5 years the structure is inspected and found to be "as new". Regarding the structure as equivalent to a newly installed one a service life of another 5 years is proposed. After this, another inspection is carried out until the structure has been in place for 20 years, which for the same design risk would demand a return period of 120 years!

This paradox is not difficult to resolve if one concentrates on design risk and the probability of exceedence, and takes into account that the inspections may be failed.

Consider the problem in general. Let the initial service life be s years and let the design risk and inverse of the return period be d and p respectively. Further let the relevant return value be x_r

$$P(x_r \text{ is exceeded in first s years}) = d = 1-(1-p)^s$$

Now assume that if x_r is exceeded the structure will not pass its inspection.

$$P(\text{inspection after s years is passed}) = 1-d$$

$$P(x_r \text{ is exceeded in next s years}) = d$$

Therefore $P(x_r$ is exceeded after passing first inspection)

$$= d(1-d)$$

Similarly

$$P(x_r \text{ is exceeded after passing m inspections})$$

$$= d(1-d)^m$$

i.e. the geometric distribution first seen in section 2.

The probability that the structure passes its m^{th} inspection is $(1-d)^m$, which means that the probability that it fails an inspection up to, and including, the m^{th} is $1-(1-d)^m$. If we take this as the design risk, d', then from eqn. 5.2 it corresponds to a return period for design criteria, for a service life of ms years, of

$$r = (1 - (1-d')^{1/ms})^{-1}$$

but $d' = 1-(1-d)^m$

i.e. $r = (1-(1-1+(1-d)^m)^{1/ms})^{-1}$

$$= (1-(1-d)^{1/s})^{-1}$$

but from eqn. 5.2 this is the original return period. Hence the total design risk is identical regardless of whether the structure was designed for ms years or for m periods of s years.

5.6. Example 4. A second similar paradox will now be considered. A structure is designed which has a drilling phase, s_1, of 2 years and is designed for extreme environmental conditions with a return period of 10 years. During drilling it becomes apparent that the drilling phase will have to be extended to 4 years and so should have originally been designed for conditions with a return period of 20 years. It is proposed to consider these 4 years as two periods of length 2 years with an inspection in between and thus argue that the return period of 10 years will suffice.

This situation is, of course, almost identical to that given above. The design risk for the original two year drilling phase, d, is given by eqn. 5.1.

$$d = 1-(1-p)^s$$

$$= 1-0.9^2$$

$$= 0.19$$

The probability that the inspection between drilling phases will be passed is

$$1-d = 0.81$$

Therefore the probability that the design conditions will be exceeded during both phases is

$$1- (1-d)^2$$
$$= 1- 0.81^2$$
$$= 0.34$$

The design risk for a service life of four years with a design criterion of the ten year return value is from equation 5.1 also 0.34. Hence these two procedures are equivalent.

The problem inherent in both these examples is that it is assumed that the design conditions do not occur and therefore all the inspections are passed. Although the probability of occurrence is small over the service life of a structure it cannot, by any stretch of the imagination, be regarded as negligible.

It is nevertheless valid to reason that <u>if</u> the inspection after the first two years shows the structure to be "as new", then the probability of this inspection being passed becomes 1 (having already happened and thus certain) rather than 0.81. In this case, the structure need only be designed for its remaining service life and thus we are in the situation discussed earlier in example 2 of section 5.4.

5.7. It may be argued that for manned structures a different procedure may be required. The method proposed (Vugts 1982) is to limit the probability of exceedence for any one year to a certain value, p_m, say. For a given service life an upper bound can then be obtained for the design risk. This can be translated into a lower bound for the return period used for design. Given p_m and a service life of s years the bound on the design risk is given by

$$d < 1-(1-p_m)^s$$

but from equation 5.1

$$d = 1-(1-p)^s$$

i.e. $1-(1-p_m)^s < 1-(1-p)^s$

$$p < p_m$$

i.e. $r > r_m$

where r_m is p_m^{-1}

This means that the return value used for designing manned structures should never be less than the return value giving the maximum allowable risk.

For instance, if the maximum allowable risk, p_m, is put at 1%, i.e. r_m=100 years, then all manned structures, regardless of their service life would have to be designed with a return value greater than 100 years.

6. Conclusions

By the use of well-established statistical methods, it has been possible to associate probabilities of exceedence to the occurrence of rare events often used in the design of structures.

Although Bayesian methods could permit the modification of estimates of design criteria based on original hypotheses and subsequent measurements, the theoretical and numerical procedures available at present do not make this a practical proposition. It is concluded that the best use of subsequent good quality measured data is in its application, together, with any earlier collected data, in the re-derivation of the required criteria. In the future, a Bayesian methodology may become a practical approach.

It is concluded that, depending on the engineering requirements, the most satisfactory method for determining extreme individual wave heights, is by use of the wave scatter diagram for the appropriate area together with a suitable extrapolation procedure.

It has also been possible to quantify the design risks associated with the use of varying return periods and associated service lives of structures. A straightforward procedure presenting a rational and consistent approach is presented, based on the adoption of an acceptable design risk.

7. REFERENCES

Barnett, V. 1973 Comparative statistical inference. J. Wiley & Sons.

Battjes, J.A. 1970 Long-term wave height distribution at seven stations around the British Isles. NIO Internal Report No. A44 (unpublished).

Battjes, J.A. 1972 Long-term wave height distribution at seven stations around the British Isles. DHZ, 25 (4), 179-189.

Carter, D.J.T. and Challenor, P.G. 1981. Estimating return values of wave height. IOS report 116 (unpublished).

Cartwright, D.E. & Longuet-Higgins, M.S. 1956. The statistical distribution of the maxima of a random function. Proc. Roy. Soc., A237, 212-232.

Forristall, G.Z. 1978 On the statistical distribution of wave heights in a storm. J. Geophysical Res., 83(C5) 2353-2358.

Johnson, N.L. and Kotz, S. 1969 Discrete distributions. Houghton Mifflin.

Longuet-Higgins, M.S. 1980 On the distribution of the heights of sea waves: some effects of non-linearity and finite band width. J. Geophysical Res., 85(C3), 1519-1523.

Resio, D.T. 1978 Some aspects of extreme wave prediction related to climatic variations. Offshore Technology Conf. OTC 3278, Houston, 1978.

Rye, H. 1976 Long-term changes in the North Sea wave climate and their importance for extreme wave predictions. Marine Science Communications, 3(6), 419-448.

Vugts, J.H. 1982 Design environmental conditions for fixed structure design. SIPM Internal Document.

8. List of Symbols

d	Design risk of a structure
$F(x)$	Probability distribution function, $=P(X\ x)=\int_{-\infty}^{x} f(y)dy$
$f(x)$	Probability density function
H	Random variable individual zero upcrossing wave height
h	A numerical value of zero upcrossing wave height
h_s	Significant wave height
i	Dummy integer used in summations etc.
k	Number of years until a certain value of X is exceeded a given number of times
m	A large number of years over which data is collected; number of inspections of a structure
n	Number of exceedences of a certain value of X in a given number of years; in Bayesian analysis the number of new data points
$P(\text{event})$	The probability that the specified event will occur
p	The numerical value of a probability
p_m	The maximum probability of exceedence of the design level allowed in any one year for manned structures
r	A return period (normally $= {}^{1}/p$)

r_m	The return period associated with $p_m (= 1/p_m)$
s	Service life of a structure
X	A random variable described any environmental variable e.g. significant wave height, gust speed etc.
x	A numerical value of X
$\underline{x}$	A vector of sample values; the new data in a Bayesian analysis
x_r	The r-year return value of X
θ	A vector of unknown parameters.

9. Tables and Figures

Table 1 Probabilities that the 50 and 100 year return design waves will be exceeded within a given number of years.

Table 2 Probabilities of a design wave being exceeded during given proportions of its return period.

Table 3 Probabilities of 0, 1, 2 or 3 exceedences of the r year return value in a given period k (using the Poisson approximation). (For exact probabilities with low k, refer to Table 4).

Table 4 Exact probabilities for the number of exceedences.

Figure 1 Relationship between the probability of exceedence of rare events as a function of the exposure in years (k) and the return period (r) of the event.

Figure 2 Graph showing the relationship between the return period (r) and service life (s) for values of the design risk (d) in the range 0.1 to 0.395 (ref equation 5.2).

No. of years	P(x_{50} being exceeded)	P(x_{100} being exceeded)
1	0.020	0.010
2	0.040	0.020
3	0.059	0.030
4	0.078	0.039
5	0.096	0.049
6	0.114	0.059
7	0.132	0.068
8	0.149	0.077
9	0.166	0.086
10	0.183	0.096
12	0.215	0.114
15	0.261	0.140
18	0.305	0.165
20	0.332	0.182
25	0.397	0.222
30	0.455	0.260
40	0.554	0.331
50	0.636	0.395
100	0.867	0.634

Table 1. The probability that the 50 and 100 year return design waves will be exceeded within a given number of years.

Ratio of number of years to return period ($^k/r$)	Probability of exceedence
0.01	0.010
0.02	0.020
0.03	0.030
0.04	0.039
0.05	0.049
0.10	0.095
0.15	0.140
0.20	0.181
0.25	0.221
0.30	0.260
0.50	0.394
0.75	0.528
1.00	0.632
1.25	0.714
1.50	0.777
1.75	0.826
2.00	0.865
2.50	0.918
3.00	0.950

Table 2. Probability of a design wave being exceeded during a given proportion of its return period.

Ratio of exposure in years (k) to return period (r)	No of exceedences n: 0	1	2	3
0.01	0.99	0.01	0.00	0.00
0.02	0.98	0.02	0.00	0.00
0.03	0.97	0.03	0.00	0.00
0.04	0.96	0.04	0.00	0.00
0.05	0.95	0.05	0.00	0.00
0.06	0.94	0.06	0.00	0.00
0.07	0.93	0.07	0.00	0.00
0.08	0.92	0.08	0.00	0.00
0.09	0.91	0.08	0.00	0.00
0.10	0.90	0.09	0.00	0.00
0.12	0.89	0.11	0.01	0.00
0.15	0.86	0.13	0.01	0.00
0.18	0.84	0.15	0.01	0.00
0.20	0.82	0.16	0.02	0.00
0.25	0.78	0.19	0.02	0.00
0.30	0.74	0.22	0.03	0.00
0.40	0.67	0.27	0.05	0.01
0.50	0.61	0.30	0.08	0.01
1.00	0.37	0.37	0.18	0.06

Table 3. The probability of 0, 1, 2 or 3 occurrences of the r year return value in a given period k, using the Poisson approximation. (For exact values with low k see Table 4).

a) p=0.02

number of years	Number of exceedences 0	1	2
1	0.98	0.02	-
2	0.96	0.04	0.00
3	0.94	0.06	0.00
4	0.92	0.08	0.00
5	0.90	0.09	0.00
6	0.89	0.11	0.01
7	0.87	0.12	0.01
8	0.85	0.14	0.01
9	0.84	0.15	0.01
10	0.82	0.17	0.02

b) p=0.01

number of years	Number of exceedences 0	1	2
1	0.99	0.01	-
2	0.98	0.02	0.00
3	0.97	0.03	0.00
4	0.96	0.04	0.00
5	0.95	0.05	0.00
6	0.94	0.06	0.00
7	0.93	0.07	0.00
8	0.92	0.07	0.00
9	0.91	0.07	0.00
10	0.90	0.09	0.00

Table 4. Exact probabilities for the number of exceedences of the 50 (p=0.02) and 100 (p=0.01) year return values.

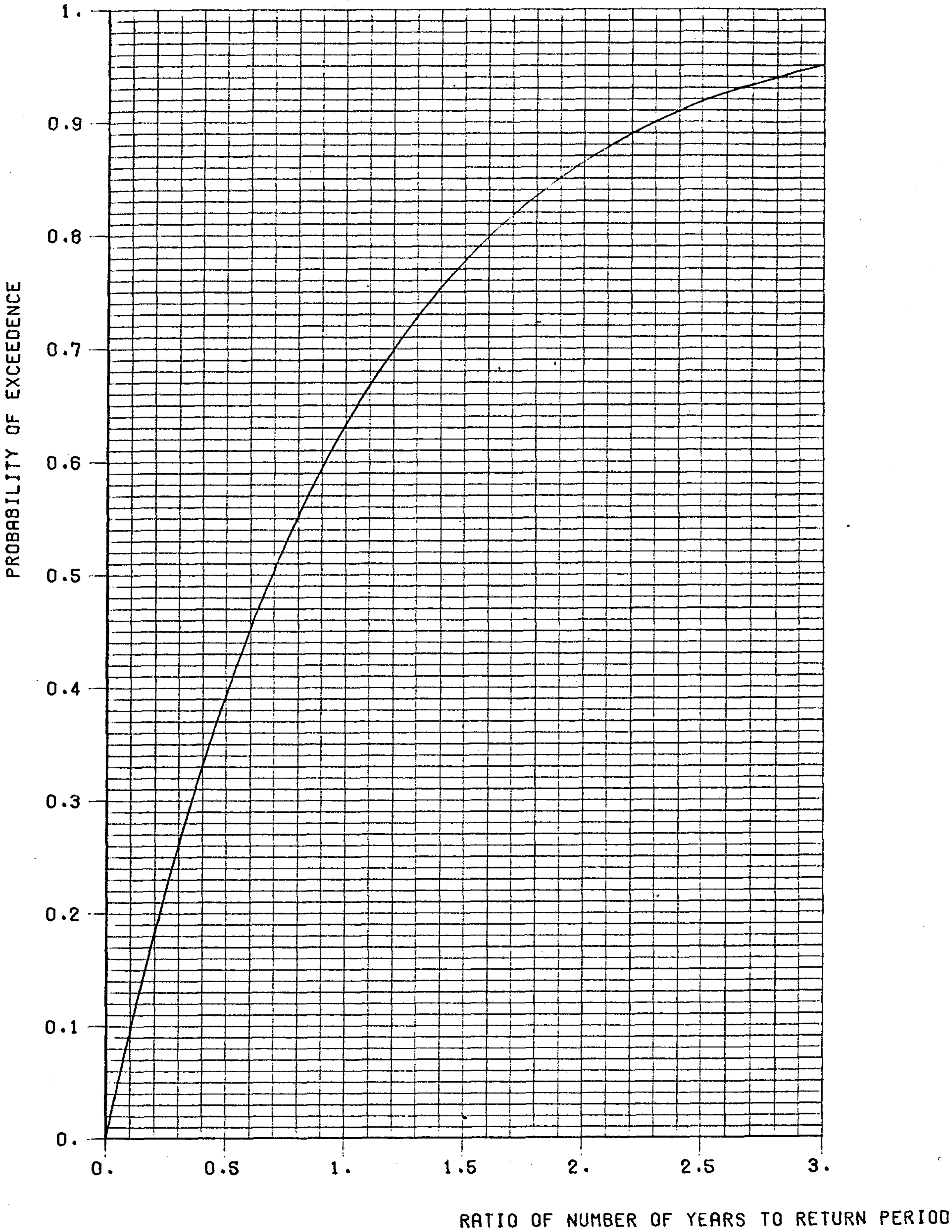

Figure 1 Relationship between the probability of exceedence of rare events as a function of the exposure in years (k) and the return period (r) of the event.

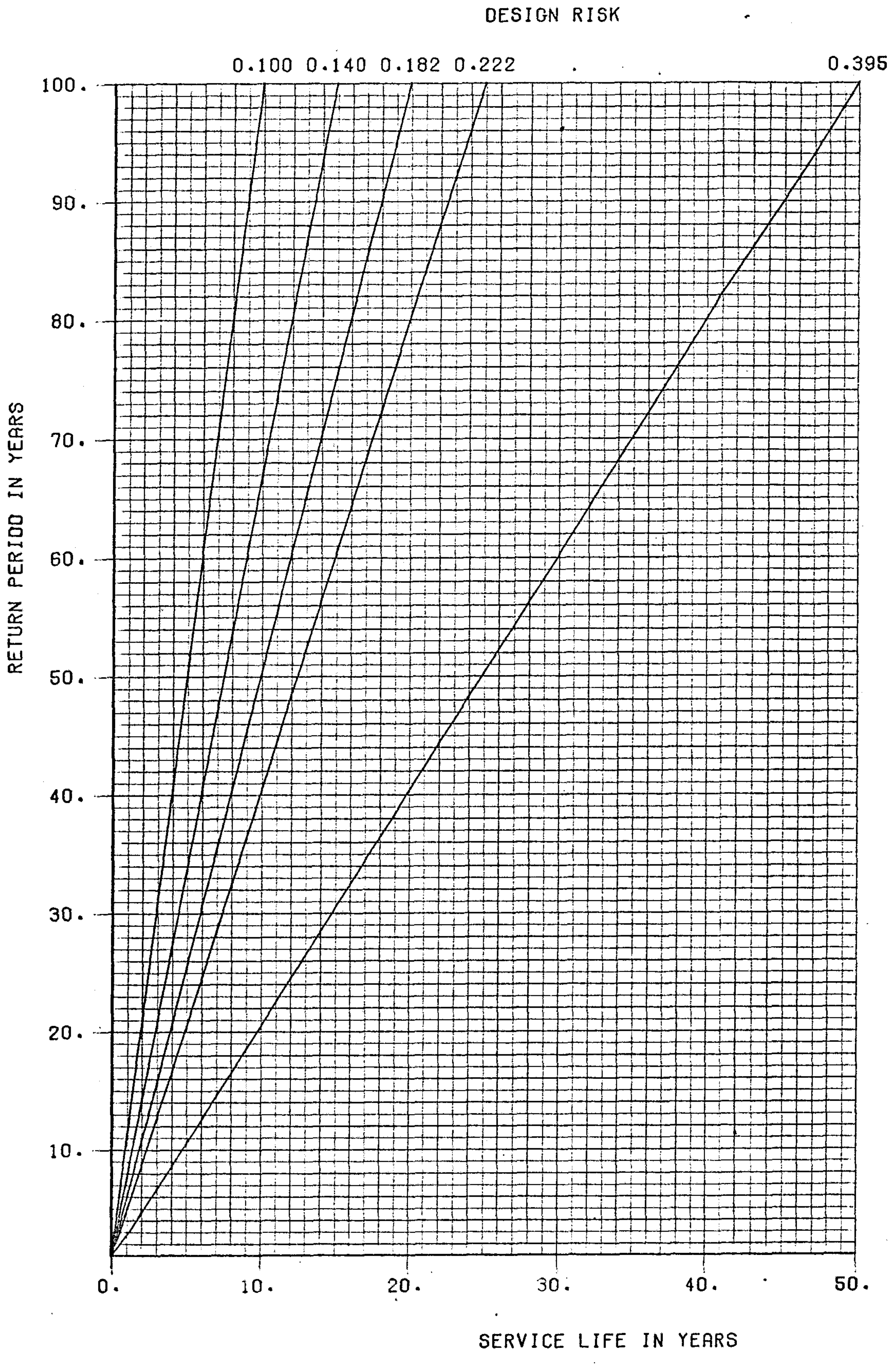

Figure 2 Graph showing the relationship between the return period (r) and service life (s) for values of the design risk (d) in the range 0.1 to 0.395 (ref equation 5.2).