STATISTICAL MODELS OF ENVIRONMENTAL DATA FOR MARINE STRUCTURE DESIGN

R.Prince-Wright, B Sc.

Thesis submitted for the Degree of Doctor of Philosophy

Department of Naval Architecture and Ocean Engineering
University of Glasgow

November 1992

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Statistical models of environmental data
for
marine structure design

Table of contents

Chapter 3: modelling univariate environmental data

Chapter 4: modelling multivariate environmental data

Chapter 5: modelling directional seas

Chapter 6: loading and response model of a tension leg platform

Chapter 7: reliability analysis of a TLP

Chapter 8: conclusions

TABLES AND FIGURES

CHAPTER 2

Acknowledgements

I would like to thank the Staff of the University of Glasgow's Department of Naval Architecture for their help and advice throughout my research contracts. In particular, I am most grateful for the support of:

my supervisors, Professor D. Faulkner and Dr. A Incecik, who encouraged me and secured the necessary funding;

Dr. B. Torsney of the Statistics department at the University of Glasgow who made useful suggestions on some of the work in Ch's 4 & 5;

Mr. D Percival for his help in writing some of the computer code during the early stages of the work;

and Dr. V Zanic who made useful comments on Ch's 3 & 4 .

This work was funded by the Science and Engineering Research Council through the Marine Technology Directorate. I thank them for having provided the funds to complete this thesis.

Finally, the support of my parents and Ann Robertson throughout the past four years has to be rewarded by my dedicating this work to them.

Declaration

Except where reference is made to the work of others, this thesis is believed to be original.

SUMMARY

In this thesis univariate and multivariate statistical inference is examined and then used to model the joint distributions of the environmental winds, waves and currents recorded by the DB1 data buoy. This model is then used to examine the return period responses of a tension leg platform using a linearised frequency domain solution.

The thesis is arranged into eight chapters each of which has its own nomenclature, conclusions, tables, and figures. The references use a name and year system and are given at the end of the thesis.

Chapter 1 reviews the contents of the thesis and outlines the analysis methodologies used to synthesise a joint probabilities model for wind, wave, and current magnitudes and directions. The use of this model in a level III, time-variant reliability analysis is then discussed to illustrate the two different design philosophies used by the American Petroleum Institute and the United Kingdom Department of Energy.

Chapter 2 summarises the wind, wave and current data recorded by the United Kingdom Offshore Operators (UKOOA) DB1 data buoy. This data has been assembled into a multivariate dataset and screened to assess if there is any underlying structure in the data. The marginal distributions of the population and monthly componentwise maxima are then examined to assess if the data result from the mixture of more than one population.

Chapter 3 reviews both parametric and intrinsic estimators for univariate samples of data. The desirable characteristics of an estimator are examined and then used to select maximum likelihood (ML) as the best estimator for this project. One major advantage of this method is that the sampling covariance matrix for the model parameters can easily be estimated from the sample information matrix. The ML estimators and sample information matrices for the Weibull and Generalised extreme value distributions are then developed and applied to both the DB1 data and a sample of structural response time series. A comparison of population and extreme asymptotic methods is then made to determine which approach is most suited to environmental datasets. The results indicate population modelling is reasonable when the

correct model is used and that the asymptotic approach can lead to poor estimators in the small sample case.

Chapter 4 examines both intrinsic and parametric estimators for multivariate samples of data. The multivariate kernel density estimator is discussed and then used with the bivariate pairs of DB1 data to confirm the multivariate sample is unstructured in the statistical sense. The transformation of the marginal DB1 data to near-Normal distributed variates is then examined and extended to the multivariate case using the method of maximum likelihood. This method is applied to bivariate and multivariate sets of the DB1 data and the results are then used to select the best set of transformation parameters. The selection criterion for the best parameters is the accuracy of the extreme value predictions from the population model. The results demonstrate the transformed Normal estimator has margins that give accurate predictions for the 50 year return period values. In addition, when the modal value of say zero-up-crossing period conditioned on significant height is checked against the scatter plots it is found the results are in close agreement. The chapter then concludes with a review of the currently available multivariate extreme value models.

Chapter 5 deals with the modelling of directional probabilities and in particular uses circular statistical theory with standard directional wave analysis theory to infer the parameters of cosine and von Mises models of directional spreading. The robustness of simply equating the angular moments of the data to the angular moments of a model is examined using simulation. The results indicate the second angular moments are more robust to noise in the buoy response. Consequently they are used with the directional spectra recorded in seastates with significant wave heights greater than 6.0 metres to determine if the spreading is more narrow in extreme seas than predicted by the Hasselmann and Mitsuyasu models. This comparison indicates that the Hasselmann study is applicable to extreme seas.

Chapter 6 describes the frequency domain model of a tension leg platform that is used in a subsequent reliability study. The stochastic wind, stochastic first and second order wave, and steady current loading calculations are explained and then a series of parametric sensitivity studies are discussed. This identifies the winds, and waves as the primary causes of the response of the platform. The response calculation considers all six degrees of freedom and allows for the coupling of some modes of motion.

Chapter 7 brings together all of the previous chapters into a time-variant reliability analysis of the tension leg platform developed in chapter 6. The effects of spectral shape, wind speed, and directional spreading on the within seastate exceedance probabilities for a variety of thresholds are examined to assess which parameters have a significant influence on the levels of structural reliability. The multivariate transformed normal model for the DB1 data

is then used in the reliability calculation to determine the motion and tendon stress response levels with return periods of 50 years. These values are then compared with the responses that result from a design event approach in which the wind speed, wave height, and current speed are set at their 50 year return period values.

Chapter 8 contains the final discussion and conclusions and ends with some recommendations for future research work.

Chapter 1

INTRODUCTION

1. INTRODUCTION

The philosophy used to design an offshore structure can influence the lifetime levels of operational risk and structural reliability (see Madsen et al (1986)) owing to the large uncertainties in the loading, strength, and modelling. These uncertainties have been examined by Miller (1987) who concludes the only rational way of setting acceptance levels for safety is to use reliability based methods. An early attempt at using probabilistic methods to set environmental criterion for the design of a tension leg platform is described by Leverette et al (1982). They examined the joint behaviour of wind speed and significant wave height using simple methods of analysis which unfortunately do not generalise to the complete multivariate case. Other models for joint probabilities have been proposed: for example, Pugh & Vassie (1980) examined the joint distribution of tidal and storm induced surge elevations, and Mathisen & Bitner-Gregersen (1990a) examined the joint distribution of significant wave height and wave zero-up-crossing period (which as we shall see is one of the most difficult pair of variates to model). Finally, it is worth noting the work reported in the Exploration & Production Forum (1985).

The rules and recommendations of several classification societies and regulatory authorities are currently used to design mobile or compliant offshore structures: most of them have been critically reviewed in Miller (1987) from the point-of-view of loading uncertainties. The United Kingdom, Department of Energy (1990) guidance for designing offshore installations implicitly uses a design event approach in which the environmental wind, wave, and current loads are determined using concurrent 50 year return period extreme values for the wind speed, wave height, and current speed. The wave zero-up-crossing period used in the analysis can be determined from wave steepness limits. Whilst it is simple, this approach will lead to populations of structures with widely varying levels of structural reliability since it is the combined wind, wave and current *load effect* which is of interest.

If the use of design Guidelines or Recommendations are to result in populations of structures with near uniform levels of reliability then the structure must be designed so that the component and system reliabilities for each design check (displacements, yield strength, interaction equations, collapse mechanism etc.) are the same. The American Petroleum Institute: Recommended Practice for tension leg platforms - RP2T (1987) - recognises this fact and states:

> *Environmental criteria should be associated with a recurrence interval of the response of the structure.*

This criterion is stipulated because RP2T recognises there may be different design events which give rise to the worst responses in different parts of the structure. For example, the return period tendon stresses may be induced by one of several combinations of wind speeds, wave heights, zero-up-crossing periods, tide levels and so on. The disadvantages of the criterion however are that it requires a detailed joint probabilities model for the environmental variables; and the analysis required for the conceptual and detailed design is more complex. An example of how environmental criterion can be set for floating structures is given by Leverette et al (1982) for the Hutton tension leg platform.

For Tension leg platform design several cases must be checked to ensure fitness-for-purpose during the operational life. For example, RP2T recommends: project, system condition, environment, and safety criteria all be examined. This thesis concentrates on modelling methodologies for the environmental criteria which are classified as: extreme, reduced extreme, normal, and calm. A joint probabilistic model for these cases must therefore apply both to the population as well as the extremes whilst being sufficiently general to deal with large numbers of random variables. The aim of this thesis is to define a methodology for the statistical inference of a probabilistic model of the environmental parameters used in a reliability study.

There are several approaches to reliability calculations which are generally classed as levels I, II or III, see Thoft-Christensen & Baker (1982). This work examines how a level III joint probabilities model can be synthesized; using data measured by the United Kingdom Offshore Operators data buoy DB1 as an example. This dataset was chosen because it contains hourly ten minute mean wind, hourly five minute mean current, and three hourly heave, pitch and roll measurements, see Freathy et al (1982). These values were recorded continuously for a period of four years, thus they provide sufficient data for a detailed statistical model.

In conventional deterministic design 'fitness-for-purpose' is checked using design codes or recommendations in which thresholds are set for the responses. For example, members sustaining tension forces are designed so that the peak expected stress during the design life does not exceed the yield stress divided by some factor of safety. In fact most structural reliability problems can be formulated in terms of loading and response variables.

The loading on an offshore structure arises from the combination of several correlated, time varying, environmental processes. The historical development of data measurement programmes, see Tucker (1991), has resulted in the development of a seastate based analysis methodology for marine and offshore structures in which the wind, wave and current stochastic processes are assumed to be stationary for some finite duration. Clearly none of the processes are stationary in the strict sense, see Prince-Wright (1991a), however most seastates can be assumed to be weakly stationary (in their mean, and variance) for durations of say three hours. This point was confirmed recently by Labeyrie (1990).

The adoption of a seastate based model enables steady-state solutions for the responses of compliant systems to be calculated using frequency, or time, domain analysis procedures. The problem of calculating exceedance probabilities, or its complement survivor probability, can then be formulated as a short-term/long-term problem which is expressed in integral form as

$$P_F = \int \ldots \int_{\forall x} P_f(x) f(x) dx \qquad [1.1]$$

In this *time-variant* integral the within seastate (short-term) probability of the response exceeding some level ξ is given by $P_f(\xi)$ where x is the vector of environmental loading variables. The second term in the integral, which is also used by Bjerager et al (1988), is effectively a weighting function for the probability of occurrence of the set of variables x . The integral is similar to the Battjes (1970) summation used for the determination of extreme wave height and enables the long-term survivor or exceedance probabilities to be calculated for the threshold level. In Ch.7 it is shown how this method can be generalised using the time-invariant reliability method, see Madsen et al (1986), to obtain failure probabilities for systems with uncertain strength.

One disadvantage of reliability methods is the frequent lack of suitable parametric models for a given set of loading and strength variables. The statistical modelling problem is in fact more difficult for the environmental variables since it is extremes of the combined loading

process which result in crossings of the design thresholds. On the other side of the equation the strength models generally have some well defined lower bounds imposed by the use of quality control in the fabrication yard. Models for the strength uncertainty do not therefore need to be accurate in the upper tails of the distribution since at the time of a threshold exceedance, or failure, the strength will be at or near its mean value. Another reason why strength uncertainty is more simple to model is the variables are generally independent. One complication which does arise for some structures is time degrading strength. This is particularly important for cargo ships which are often poorly maintained and susceptible to damage during operation. In such cases a time-variant strength model is preferable, see Shi (1991).

This thesis does not examine the modelling of strength and fatigue uncertainties in any detail. However, it should be borne in mind that the statistical methods used for the environmental winds, waves, and currents, are also applicable to any set of variables, and in fact the transformation modelling method discussed below is in many ways better suited to strength uncertainty.

Equation [1.1] can be reduced to three separate problems. The first is estimation of the short-term exceedance probability for the threshold level, conditioned on the occurrence of the set of variables $\boldsymbol{x}$. The second is modelling the joint density for the environmental parameters $f(\boldsymbol{x})$. And third, the solution of the integral, which for most practical purposes will have several dimensions. Although the intention was to concentrate on modelling the joint density $f(\boldsymbol{x})$ for use in the reliability integral, a new approach to solving the integral developed out of the use of data transformation methods. The importance of this new approach is most easily understood by comparing it with the contemporary methods used to model $f(\boldsymbol{x})$ and the solution methods usually used to solve the integral.

Several joint probability models for winds, wave and currents have been proposed. For example, Bitner-Gregersen & Haver (1991), Haver & Winterstein (1990), and Turner & Baker (1988) all give models for the North Sea. The formulations for these models were to some extent determined by the use of the Rosenblatt transformation (see Ch.4, Appendix C) in the level III reliability method outlined in Ch.7. In this method the joint density function is defined by a marginal distribution and then a series of conditional density functions, each of which are modelled using parametric functions that have been fitted to the conditioned data using some form of estimation process. Whilst this approach is simple it requires the data be conditioned and *then* used to fit empirical functions to relate, for example, the parameters for the wind speed distribution to the level of the seastate.

The primary disadvantage of the conditional modelling approach is that there is no theoretical method for selecting the best way of defining the joint density. For example, do we use wind speed or significant wave height as the marginal distribution off which all the other variables are conditioned? A second disadvantage of the method is that there is no way of assessing the shape, scale and location parameter uncertainties for the fitted models; this makes it difficult to test whether one model is more appropriate than another. Finally, the last disadvantage is the Rosenblatt method itself which is computationally expensive and usually requires a series of numerical integrations.

Before attempting to model the multivariate sample, univariate estimation methods were reviewed. Univariate statistical estimators using intrinsic kernel density methods, and parametric maximum likelihood methods were developed with a view to their extension to the multivariate case. The normally subjective nature of comparing models was reduced by first defining those characteristics which exemplify a good estimator and then by selecting the 'best' estimation process for the problem. The method of maximum likelihood is selected for both its overall performance and the ease with which it can be generalised to the multivariate case. In addition, it is shown how the information matrix can be used to quantify the sampling uncertainty of the estimated model parameters. This theory is of importance to designers wishing to specify confidence limits or partial safety factors for design event environmental parameters.

Population and asymptotic methods were then compared as estimators of design return period values for winds, waves, and currents. This comparison shows that the population modelling approach can be used reliably given the correct population model: this is significant because the statistical uncertainty estimated using the information matrix is much smaller in the large sample case. Furthermore, the assumption of asymptotic Normality for the parameter sampling distribution is only accurate for large samples.

In this work two parametric approaches to modelling the multivariate density were examined. The first used the asymptotic multivariate models of Tiago de Oliveira (1980), and Tawn (1990). In their work they use the concept of marginal M-ordering in which componentwise maxima are extracted from the vector observations. This approach is suitable for some applications, like for example the hydrology problem of predicting the probability that no floods occur at any of several sites during a year. However, because the events are virtual events, some of which may be physically inadmissible, it is questionable whether

these models are suited for use with compliant structures. This is currently the subject of research by J.A Tawn at Sheffield University.

The second approach is to use population models for the joint data: however, unlike the conditioned models outlined above, a transformation approach was used in which the parameter estimation is performed using the full sample. There are several good estimation procedures for a model's parameters: moments, least squares, maximum likelihood are examples. In this work, the method of maximum likelihood was selected as the best overall estimator because in the large sample case: its estimates have the optimal sampling variance obtainable from any estimator; and the estimates are generally unbiased. Two further advantages are that it can be generalised easily to the multivariate case and furthermore estimates of the parameter covariance matrix can be obtained simply from the Hessian matrix of the log-likelihood at the maximum likelihood point.

The heave pitch, and roll time series recorded by the DB1 data buoy were converted to co- and quadrature spectral densities to enable the directional characteristics of the wave process to be identified. During the initial stages of this study in 1988 it was unclear whether or not the spreading models developed by Hasselmann et al (1980), and Mitsuyasu et al (1975) were satisfactory for use with extreme waves. One oil company stated their doubts that any significant amount of spreading was present and consequently it was decided to examine the directional spectra recorded by the DB1 in those seastates with significant wave heights greater than 6.0 m. The directional wave analysis mostly follows the work of Cartwright (1963), Longuet-Higgins et al (1963), and kuik et al (1988). However, the use of Mardia's (1972) detailed account of circular statistics resulted in a solution for the parameters of a von Mises distribution which has been presented as an alternative to the conventional cosine half angle model.

The repeated use of the time-variant reliability method of probabilistic analysis enables us to calculate the response levels with a given return period, as required by the API Recommendations. This method has been used with: a transformed Normal model of the joint density for the winds, waves, and currents recorded by the DB1 data buoy; and a tension leg platform defined by Tan & de Boom (1981).

The first order wave loading on the TLP was modelled using a combination of a Morison loading model for the pontoons and Chakrabarti (1987) closed form solutions for the columns. Second order slowly varying wave drift forces were calculated only for the columns using the simple MacCammy and Fuchs method as modified by Chakrabarti (1984). The responses of the

TLP were calculated for all six degrees of freedom taking into account all coupling between modes of motion. The first order wave force and motion transfer functions were then compared with the results from a comparative study in which several organisations had analysed the same structure, see Eatok-Taylor & Jeffereys (1986). This comparison demonstrated the loading and response model was accurate for all but the high frequency waves which are diffracted by the structure.

The wind and current forces on the TLP were also modelled. The Ochi (1988) wind gust spectrum was used to model the stochastic wind forces and a simple single degree of freedom oscillator was used to determine the in-line responses of the platform. The current forces on the platform were taken as constant during the three hour duration seastates. Sensitivity studies were performed to assess the effects of varying the significant wave height, zero-up-crossing period, wind speed, and current speed. This demonstrated the responses were small for currents, and significant for the gusting wind and wave drift. The influence of currents in the reliability analysis was found to be small, as expected for this type of structure, however it should be borne in mind that the current forces on shallow water fixed structures are significant, see Department of Energy (1988).

The motion and tendon stress responses of the TLP corresponding to a 50 year return period were examined in both unidirectional and directional seas. These response levels were calculated using the time-variant reliability method and the DB1 data joint probabilities model. The results were then compared with a notional design event approach in which the responses were calculated for concurrent 50 year wind speed, significant wave height, and current speed. Two zero-up-crossing periods were used based on the upper and lower wave steepness limits recommended by the Department of Energy Guidance notes. The results of the comparison show the design event approach results in response estimates which are some 15~25% higher than the responses predicted by the reliability approach. This comparison is however rather artificial since like is not being compared with like, the real importance of the reliability method is as a calibration tool which can be used to quantitatively assess the importance of variations in one or more random variables.

1.1. MATHEMATICAL NOTATION

The mathematical notation used in this thesis is defined at the beginning of each chapter, however, some general rules apply. To unambiguously identify a quantity in an equation we must differentiate between:

- deterministic values
- randomly selected samples from a population
- specified variables for a population
- vectors of deterministic values
- vectors of samples
- vectors of random variables
- expected values
- sample means
- maximum likelihood estimates for model parameters
- kernel estimates
- sample estimates of statistics and model parameters

The conventions adopted in the subsequent chapters are as follows:

- a deterministic value or sample is written in plain, lower case
- a random variable is written in plain,uppercase
- a vector of samples or deterministic values is written in bold,lowercase
- a vector of random variables is written in bold,uppercase
- a maximum likelihood or kernel estimate has the caret ($\hat{x}$) symbol on top
- a sample estimate has the tilde ($\tilde{x}$) symbol on top

Random variables which are independent, identically distributed are denoted *iid* and the condition if and only if is denoted as *iff* .

We use the terms *survivor* probability for $Pr(X \leq x) = F_X(x)$ and *exceedance* probability for $Pr(X > x) = 1 - F_X(x)$.

Chapter 2

ANALYSIS OF THE DB1 DATASET

NOMENCLATURE

D_w	*wind direction*
D_c	*current direction*
$H(f)$	*transfer function*
H_s	*significant wave height*
m	*component maxima event*
n	*number of samples*
p	*number of variables*
$S(f)$	*spectral density*
T_z	*zero – up – crossing period*
U_w	*mean wind speed*
U_c	*mean current speed*
X	*random vector*
x	*observed sample on vector X*
x_{ij}	*ith sample of the jth variable*

2. INTRODUCTORY REMARKS: ANALYSIS OF THE DB1 DATASET

This chapter summarises an analysis of the joint wind, wave and current data recorded by the United Kingdom Offshore Operators Association (UKOOA) data buoy DB1 during its operation in the South Western approaches to the United Kingdom. The primary objective here is to describe the multivariate data set used in the statistical modelling chapters of this thesis. The ultimate aim is to produce a multivariate parametric model suitable for use in a probabilistic analysis of a tension leg platform. This results in two requirements: first, a *descriptive* joint statistical model of the observed wind, wave and current populations; and second, a joint model for estimating the wind, wave and current extremes, or return period values.

The first stage of any multivariate analysis is to assess if the data are structured in the statistical sense. By structured we mean there is some functional relationship between one or more of the random variables. The simplest way of identifying structure in a multivariate sample is to examine the bivariate scatter plots for each pair of variables. The scatter plots also give an indication of the dependence between the pairs which provides a qualitative check on the results from the multivariate analysis in Ch. 4.

The DB1 data buoy was deployed at location 48° 42' 55" N, 8° 58' 15" W during the period from June 1978 until March 1982. In total the buoy was operative for a period of 45 months during which time the data return was 'fair' for the heave displacement and wind speed, and poor for the pitch angle, roll angle, and current speed. The dataset was selected after consultation with the Institute of Oceanographic Sciences (IOS); essentially, it is one of the few datasets which is not confidential and contains wind, wave and current measurements taken in the same location and at approximately the same time. The raw (unprocessed) data

was analysed using a FORTRAN time and frequency domain program which uses some of the standard time series analysis theory discussed in Tucker (1991). A more detailed account of the methods used for this work is given in Prince-Wright (1991a).

2.1 VARIABLES MEASURED BY THE DB1 DATA BUOY.

In total, the buoy recorded both meteorological and oceanographic environmental data for a period of four years, during which time it recorded 32981 hourly summaries of:

Meteorological -

10 minute mean wind speed (8.7m elev.)	(2)
3 second gust wind speed	(2)
Wind direction	(2)
Barometric pressure	(2)
Air temperature	(2)

Oceanographic -

Sea surface temperature	(2)
Surface current (3m below surface)	(1)

And 9034 three hourly 20 minute duration time series of:

Oceanographic -

Heave	(1)
Pitch	(1)
Roll	(1)
Compass heading	(1)

The numbers in brackets indicate the numbers of sensors used to record the data, and hence the level of redundancy in the system. A more detailed description of the specification of the recording instruments is given in Table 2.1, and Freathy et al (1982).

2. 1.1 SPECIFICATION OF THE DB1 RECORDING INSTRUMENTS

The general arrangement of the DB1 data buoy is shown in Fig. 2.1. The hull and mooring arrangements were designed to provide wave following characteristics and the ability to survive a 50 year return period wave. The specification of the meteorological and oceanographical sensors is reproduced from an IOS report in Table 2.1 . Of particular note is the

field accuracy of ± 0.2m for the heave elevation, and $\pm 1^{\circ}$ for the pitch and roll. Note also that the currents are recorded as a five minute mean taken about the half hour with a field accuracy of 0.02 m/s. This means the current data is recorded approximately ten minutes after the wave record has been recorded. However, visual examination of the time series for heave, and checks on the stationarity using autocorrelation show this is sufficiently close for them to be taken as concurrent. Likewise the wind speeds, which are recorded with a field accuracy of 1m/s, are practically concurrent being recorded as an average during the ten minutes preceding the 20 minute wave recordings.

2.2 BUOY RESPONSE TRANSFER FUNCTION: SURFACE FOLLOWING QUALITIES

The integrated heave, pitch and roll accelerations give the response time series for the buoy displacements. Model testing was performed by the Institute of Oceanographic Sciences to determine the surface following characteristics of the buoy; they found that for frequencies less than 0.2 Hz the transfer function was approximately 1.0. At higher frequencies the transfer function was less ideal with a value of 0.75 at a frequency of 0.33 Hz, however, the scatter plots indicate the minimum observed wave period was approximately 4.0 seconds.

For this study we are primarily concerned with the higher seastates in which the zero-crossing period would be greater than 5 seconds, consequently, any bias in the buoy response at the higher frequencies is unlikely to significantly affect the estimated significant height and zero-up-crossing period. On the other hand it would be worth checking the bias if the buoy response transfer function were available since the true surface elevation spectrum $S_X(f)$ can be recovered from the measured response spectrum $S_R(f)$ using

$$S_X(f) = S_R(f) / \left(H_R(f)\right)^2$$

For simplicity, in this study the seastate characteristics have been calculated using the moments of $S_R(f)$ since the precise form of the buoy transfer function was not known.

2.3. CHECKS ON THE DATA ANALYSIS

The first objective of the data analysis is to provide a reliable database of jointly occurring winds, waves and currents. To ensure the results from this analysis are correct a sample of this study's (denoted GU) frequency domain results is compared with the results from the original frequency domain analysis (denoted UKOOA). The results are shown in Fig. 2.2 and 2.3 which compare estimates of significant wave height and zero-crossing period. In general,

the results for the significant wave height are in close agreement with a few exceptions, however, the results for the zero-up-crossing period have larger scatter and a number of outliers.

The second objective of the data analysis is to provide a sample of multivariate random variables corresponding to the extreme winds, waves and currents. To ensure the monthly extremes are error free the results from this study with seastates greater than or equal to 5.0 metres significant height were cross-checked visually against the original frequency domain results supplied by the Marine Information Advisory Service (MIAS), Bidston. This check shows that in general the estimates agree but there are again some differences, in many cases the cause was corrupted observations.

2.4 OBSERVED VECTOR EVENTS FOR THE POPULATION MODEL

The data recorded by the DB1 can be treated as independent, identically distributed vector observations drawn from a multivariate process X where the set of observed environmental variables is

$$x_i^T = \{H_s, T_z, U_w, U_c, D_w, D_c\} \qquad i = 1,\ 9034$$

In this work, H_s is the significant wave height, T_z is the zero-up-crossing period, U_w is the 10 minute mean wind speed (8.7m elev.), U_c is the surface current 3m below surface, and D_w and D_c are the associated directions for the wind and the currents, respectively. Note the winds and the waves are assumed to be colinear in this work. In matrix notation the full sample of vector observations is written as $x_{ji}; i = 1,..,n, j = 1,..,p$. Here n is the number of complete vector observations - that is observations in which there are no missing values - and p is the number of variables. Writing the matrix long-hand then we have

$$x^T = \begin{bmatrix} x_{11} & x_{12} & .... & x_{1p} \\ x_{21} & . & . & x_{2p} \\ . & . & . & . \\ x_{n1} & . & . & x_{np} \end{bmatrix}$$

where the columns of x^T represent the marginal sample of data, and the rows represent each jointly occurring vector observation. In fact, after eliminating the vectors with missing values the number of samples left in x was 5673. This complete sample has been used in all of the statistical inference in Ch. 3 and Ch. 4.

2.5 COMPONENTWISE MAXIMA FOR THE MULTIVARIATE EXTREMES MODEL

Multivariate extreme value theory provides a framework for modelling the joint distribution between *component maxima*. Denoting all of the environmental data recorded by the DB1 in some fixed period of time, or number of samples t, as

$$\boldsymbol{x}_i = \{H_s, T_z, U_w, U_c\}_i \qquad i = 1, \ldots, t$$

the component maxima are then the random vectors

$$\boldsymbol{m} = \max \boldsymbol{x}_i = \left(\max x_{1:t}, \ldots, \max x_{p:t} \right)$$

In terms of the DB1 data the vectors of componentwise maxima $\boldsymbol{m}_i$ might correspond to the monthly largest values of the components of $\boldsymbol{x}$, that is

$$\begin{aligned} \boldsymbol{m}_k = \{ & \textit{largest } U_w \textit{ in month } k, \\ & \textit{largest } H_s \textit{ in month } k, \\ & \textit{largest } T_z \textit{ in month } k, \\ & \textit{largest } U_c \textit{ in month } k \} \end{aligned}$$

Here we instantly see the limitation of current asymptotic multivariate theory which deals with the distribution of the random vector $\boldsymbol{m}_k$. The problem is that $\boldsymbol{m}_k$ primarily consists of virtual events, for example, the largest wind speed U_w does not necessarily occur at the same time as the largest current speed U_c observed in the same month! The consequences of this limitation are generally ignored in the work of Tawn, Pickands and others, and further research is needed to identify the effects of this limitation of the theory.

2.6 RELIABILITY OF THE TIME SERIES ANALYSIS

Problems with the time series were frequently encountered when analysing the heights and periods using the time domain method outlined in Section [2.6.1] below. Overall some 20% of the data is corrupted by spikes and missing values which cause significant bias in the estimated mean level and resulted in the time domain analysis giving inaccurate results for the characteristic heights and periods. The sensitivity of time domain analysis to errors in the time series was identified by the original data collectors who eventually came to regard the time domain height and period results as so unreliable, compared to the spectral estimates, that they were never quality controlled. Their findings are therefore largely validated by our

experience, however, it is essential to realise that the wave records used for this project's time domain processing are the same as those used by their spectral analysis software. It is therefore important to ensure the frequency domain results are not biased by corrupted data and consequently it was processed both in the time and frequency domain to compare one with the other.

2.6.1 TIME DOMAIN ANALYSIS

The time domain analysis of the wave 'burst' recordings follows the recommendations of Draper (1963), Tucker (1963), and Goda (1985).

The datum level for each wave record was determined by fitting a parabolic mean level to each sample using a simple least squares method, Goda (1985). Once the datum level for the heave displacements was found the record of discrete points is examined in order to estimate the location of: zero up-crossings; local maxima at the crests; and local minima at the troughs. The results were then used to calculate the zero-crossing and crest periods. The procedure adopted was an approximate one and is only satisfactory when the sample rate is sufficiently high to define the occurrence and location of the peaks and crossings. A sample period of not greater than one tenth the significant wave period is recommended by Goda (1985). However, the DB1 sample period for the heave, pitch and roll signals was 1.2 seconds suggesting a minimum significant wave period of 12 seconds. Clearly, this is too high for the bulk of the recorded data and we could only ignore this advice.

A summary of the time domain estimated wave statistics calculated by the program is illustrated in Fig. 2.4, which also shows how the heights and periods are defined. Using the wave-by-wave statistics the significant height is defined as the average of the highest one third zero-up-crossing waves, and the zero-up-crossing period is defined as the average period of the wave zero level up-crossings.

2.6.2 SPECTRAL ANALYSIS

Harmonic or spectral analysis of data provides a powerful and robust means of analysing the composition and statistics of a random process given finite duration samples. There are several classical, and contemporary texts on Fourier analysis, perhaps the best known early works are those of Blackman & Tukey (1958) and Cooley & Tukey (1965). More recently Chatfield (1991) and Newton (1988) presented useful and theoretical texts and Tucker (1991) wrote a detailed practical and theoretical book on wave measurement and analysis.

Meteorological and oceanographical time series are generally difficult to record reliably and inevitably some data will be corrupted. Spectral methods, however, rarely fail to

produce an estimate for the characteristics and are therefore more robust. However, they can often be too robust since they will transform just about any signal, no matter how corrupt it is. This often results in spectral estimates of statistics which are at best biased, and sometimes completely wrong . On this project, the policy adopted when analysing the heave time series was to process the signal in both the time and frequency domain and then compare the results to determine if it was likely the data were corrupted. In most cases the results were reasonably close and the data was included in the database. However, when estimates of variance and mean period were not in agreement the data was rejected.

2.6.3 COMPARISON OF TIME AND FREQUENCY DOMAIN ESTIMATORS

As a further check on the quality, and accuracy of the results from 'DB1' 50 pairs of significant wave height, and zero-crossing period were compared with results supplied by MIAS in listings from the original spectral analysis. Fig. 2.5 shows the GU time and GU frequency domain results for the significant wave height. The GU time domain values of significant height are calculated as the average height of the highest one third crest to subsequent trough waves, ie $H_{1/3}$. The spectrally calculated significant height H_{m0}, based on $4\sqrt{M_0}$, is consistently higher than the time domain estimate, with the difference appearing as a gradient of 1.12 in the linear regression equation shown on the graph. The scatter of the estimated Hs is relatively low giving a correlation coefficient 'R ' of 0.987.

A comparison of the GU time domain and UKOOA frequency domain significant wave heights is shown in Fig. 2.6. The original analysis produces estimates with slightly lower bias but larger scatter. Overall the trend is the same as for the GU time and frequency domain results with the original frequency domain estimator overestimating the significant heights by some ten per cent. The GU time and GU frequency domain estimates for the average zero-up-crossing period are compared in Fig. 2.7. Clearly, there is a linear trend with larger scatter than is the case for the significant height. The frequency domain estimates of period are generally shorter than the time domain estimates, however, this is a well known characteristic of spectral estimates which has been examined by Goda (1974), he found the ratio $T_z(\text{freq}) / T_z(\text{time})$ varies with $\Delta t / T_p$, where Δt is the sample rate in seconds and T_p is the spectral peak period.

Goda *loc cit* suggests a correction factor $T_z(\text{freq}) = 0.83\ T_z(\text{time})$ and attributes the differences to both non-linear harmonic components in the high frequency range, and aliasing. This value is close to the 0.86 correction suggested by Fig. 2.8, however, if we examine the example time series and its heave spectra, shown in Fig. 2.9, it is clear the effects of aliasing

are small for the DB1 data. Tucker (private communication) disagrees with Goda however and believes the differences are due to the thresholds of the recording instruments.

The effect of non-linearity has been examined by Longuet-Higgins (1963), and Bitner-Gregersen (1980) who adopt the Gram-Charlier perturbation of the Gaussian distribution to model the effects of mildly non-Gaussian surface elevation. These methods were not used since the wave buoy tends to act as a linear filter on the surface process time series.

2.7 SCATTER DIAGRAMS FOR PAIRS OF VARIABLES

The scatter diagram for significant wave height and zero-crossing period is the basic environmental input for a long-term, compliant systems response analysis. The information in a scatter diagram, which is effectively a bivariate histogram, suffers from the same drawback as the univariate histogram in that some of the information is lost by binning the data. Nonetheless they do provide a convenient way of summarising pairs of random variables and are simple to interpret. The data recorded by the DB1 during its four year operation has been sorted and grouped to form the scatter diagrams shown in Tables 2.2 to 2.10. Note the rows and columns have been summed in each table to give the marginal histograms and that the cells contain the actual number of observations in the range, the upper limit of which is given for each cell.

In Ch. 3 and Ch. 4 both parametric maximum likelihood methods and non-parametric kernel methods are outlined for the univariate and multivariate case and then applied to the data.

2.7.1 SIGNIFICANT HEIGHT AND PERIOD

Scatter diagrams for the time and frequency domain estimates of significant wave height and zero-crossing period are shown in Tables 2.2 and 2.3. The correlation between the frequency domain and time domain estimators is illustrated in Table 2.4 which shows the results agree very well for the lower wave heights, albeit with some scatter about the mean, but that the bias appears to increase with increasing wave height. At the extreme sea states the frequency domain estimator overestimates the significant height giving a bias of approximately 10 per cent.

2.7.2 SIGNIFICANT WAVE HEIGHT AND WIND SPEED

The ten minute mean wind speed and significant wave height scatter diagram is given in Table 2.5. The degree of correlation is difficult to assess since there is considerable scatter about the mean. This is perhaps surprising since we might expect the wind and the waves to be

strongly correlated, one reason for the low correlation is the short duration of the 10 minute wind time average which is reasonable for the design of fixed and compliant systems. It has been suggested (Dr. C. Graham, personal communication) that the correlation is stronger if a longer time average is chosen, however, it is the short duration average (3 seconds up to 10 minutes) which is of interest for design since this is consistent with the size, and motion periods of typical offshore structures.

In fully developed seas the Pierson-Moskowitz surface elevation spectrum predicts the mean significant height to be $Hs = 0.0227 * U^2$, where the wind speed U is the 10 metre reference height, hourly averaged, wind speed which is closely related to the ten minute average wind speed recorded by the DB1 buoy. If we superpose this on the scatter plot the quadratic mean model follows the mode of the bivariate data at the lower wind speeds; at higher wind speeds it is difficult to assess where the mode lies because of the small number of samples recorded in extreme conditions, however, it is clear the quadratic model overestimates the significant height by a large amount. This is most likely because the fetch length is usually too short for the seas to develop fully and suggests that even the $\{H_s, U_w\}$ pair are unstructured - or at least can be treated as such.

2.7.3 TEN MINUTE MEAN AND THREE SECOND GUST WIND SPEEDS

The scatter diagram for the ten minute mean wind speed and the three second gust speed is shown in Table 2.6. The two wind speeds are closely correlated over the whole range of speeds with a linear mean of the form $U_{10}(10m) = 1.33U_{10}(3s)$ which corresponds closely to the recommendations of the Department of Energy (1990a) ,Table 11.5.

2.7.4 SIGNIFICANT WAVE HEIGHT AND CURRENT SPEED

Currents are generated by a number of mechanisms for example, tidal and storm induced surge, density differences, large scale eddies, and ocean circulation. The tidal component of a current results from the daily, seasonal, and 'nodal' changes in the earth's gravitational field caused by the motion of the moon and the sun, see Pugh (1987). This imposes a periodicity with two dominant components and a number of sub-harmonics, each with random phase and amplitude. Storm induced currents are caused by the shear forces at the air-water interface and pressure gradients over the storm system. Unlike tidal currents storm induced currents are not periodic but random in occurrence (though of course they are seasonal). The result of summing each component is a current with both periodic, and random components, as seen in Fig. 2.10.

In the analysis of the DB1 data no distinction is made between the separate components and the total current will be assumed to be the variable of interest for the design of offshore structures. The scatter diagram for significant wave height and current speed is shown in Table

2.7. The quadrant of the scatter plot defined by the marginal extremes tends to suggest that the extreme values of significant wave height and current speed are asymptotically independent, with low probability of occurrence for simultaneous high waves and current speed. This indicates that the conventional design approach of combining extreme seastates with extreme currents will result in an over-conservative model for the South Western Approaches site; though of course the currents at other sites may be more highly correlated with the waves.

2.7.5 WIND SPEED AND CURRENT SPEED

The 10 minute mean wind speed and the 5 minute mean current speed show little or no correlation in the scatter diagram shown in Table 2.8. This may be due to the currents being dominated by the tidal component and suggests a refined analysis - in which the tidal and storm currents are de-convolved - would give an improved model. This could be done in a future investigation using the methods discussed in Pugh (1987).

2.7.6 DIRECTIONAL DISTRIBUTION OF SIGNIFICANT WAVE HEIGHT AND ZERO-UP-CROSSING PERIOD

The directional distributions of the time domain significant height and zero-up-crossing period with wind direction are shown in Tables 2.9 and 2.10. Note that the conditional, directional distribution $P(\theta \mid Hs)$ can be inferred by summing the number of occurrences in the required range of heights and that a more rigourous analysis of the directionality is given in Ch. 5., where the within seastate, and long-term directional distributions are examined using Fourier methods.

2.8 POPULATION MODELS FOR THE MARGIN

The marginal histograms for the sample x can be inferred from the scatter plots: however, a better non-parametric estimator is the kernel density estimator. The theory for this method is discussed in Ch. 3 but it is worth looking at the kernel estimates for the marginal densities of the significant wave height, zero-up-crossing period, wind speed and current speed. The results are shown in Fig. 2.11 in which a quadratic (Epanechnikov) kernel with optimal window width has been used for each density. Note that each distribution is unimodal and that none of them seem to be the result of a mixture of random variables. This suggests that, providing the correct model can be found, a parametric modelling approach is reasonable and may be acceptable for the extremes.

2.9. MONTHLY MAXIMA

The need for componentwise maxima is outlined in Section [2.5] with a view to the requirements of the multivariate extreme value model discussed in Ch. 4. The vector of all 11000 or so environmental random variables z was used to determine the vector of monthly virtual events (or componentwise maxima) m. The result is given in Table 2.11 which shows the 42 monthly maxima recorded by the DB1 buoy. The table includes both the time and frequency domain estimators for the seastate heights and periods, and missing wind and current values appear as zeros. Ideally, annual maxima should be used in the extreme value models since the component maxima are seasonal and therefore periodic, however, the DB1 was only operational for four years and no other multivariate dataset was available. One possibility would be to use the model for dependent extremes given by Tawn (1988a), this method would be preferable if the results were for use in a real design but in this work the dependence is ignored since ultimately a population modelling approach is used in the reliability analysis.

2.9.1 UNIVARIATE MONTHLY EXTREMES

The marginal monthly extreme value data given in Table 2.11 has been analysed using the theory and software described in Ch. 3. The results are summarised in Table 2.12 which gives:

(i) the sample mean, variance, skewness, and kurtosis

(ii) the shape, scale, and location parameters for the Weibull and GEV estimators

(iii) the covariance matrices for the parametric models

Both the 3-parameter Weibull and the GEV models were fitted to the extremes to enable a comparison of the two estimators.

The Weibull and GEV models (see Ch. 3.) fit the time domain and frequency domain estimates of significant height very well, with the negative GEV shape parameter indicating the population of extremes is Frechét distributed with a lower bound. It is interesting to note that the Weibull parameter covariance matrix contains negative terms for both the variance of the location parameter, and the covariance of the scale and location. This suggests the GEV model is a better conditioned solution. The Weibull and GEV models both fit the time and frequency domain estimates of zero-crossing period, however, the error residual plots show the GEV model has slightly smaller residuals, furthermore the covariance matrix for the GEV indicates the model parameters have lower uncertainty associated with them. However, we shall see in Ch. 3 that in some cases the GEV model gives poor estimates of the extremes corresponding to design return periods.

2.10 CONCLUSIONS

Two vector samples of wind, wave, and current data have been created from the DB1 data; one for the population of three hourly observations and the other for the monthly, componentwise, maxima.

Inspection of the scatter diagrams for the data indicate that it can be treated as unstructured in the statistical sense for the multivariate modelling. The only evidence of structure appears in the significant wave height and zero-up-crossing period density. This is the result of the breaking wave criterion which results in a steep rise of the density on the forward face of the distribution.

The marginal kernel densities show no evidence of bimodality or mixing of different statistical populations in the modal region or indeed for moderately large values. However, the irregularity of the tails needs to be examined using an adaptive kernel (see Ch. 3) in which the degree of smoothing is matched by the local density.

The data return for complete vector observations with no missing values was only 50%, however, there was no systematic reason for the missing values which occurred in both low and high seastates. We can conclude therefore the missing observations occurred randomly for the univariate and multivariate statistical inference.

The lack of strong structure in the data suggests a population modelling approach is reasonable, providing the correct tail behaviour can be guarantied by the use of a suitable population model and estimation process.

Measured variable	Sensor type & manufacture	Range	Location on buoy	Time, averaging period or sampling rate	Calibration accuracy	Printed resolution	Field accuracy
Wind speed	1 Cup counter Vector instruments A100R 2 Cup counter Vector instruments A100R	0 to 77 m/s (150 kn)	6.0m above sea level 8.7m above sea level	mean value during last 10 minutes of hour	±0.26 m/s (0.5 kn)	0.1 kn	±1 m/s (±2kn) or ±5%
Wind direction	1 Wind vane -self referencing Vector instruments SRW1G 2 Weather measure W102 Referencing from digital compass	0 to 360°	6.0m above sea level	mean value during last 10 minutes of hou	±2°	1°	±10°
Air pressure (one static pressure head)	(2) Aneroid capsule KDG8190	925 to 1050 mb	Head at 6.0m but measure-ment made at sea level	mean value during last 45 seconds of hour	±0.2 mb	0.1 mb	±1mb
Air temperature	(2) Platinum resistance Rosemount Ltd E13418	-10 to 40°C	Stevenson's screen 5.8m above sea level	mean value during last 45 seconds of hour	±0.1°C	0.1°C	±0.2°C
Relative humidity	(2) Chemical Hygrometer Phys-Chemical Res. Corp. PCRCII	0 to 100%RH	Stevenson's screen 5.8m above sea level	mean value during last 45 seconds of hour	±3%	0.01%	±5% 0 to 85% ±3% over 85%
Sea surface temp. (1) Sea surface temp. (2)	Platinum resistance Platinum resistance	-10 to +40°C -5 to +20°C	Base of hull Base of hull	45 secs. mean every hour	±0.1°C ±0.001°C	0.1°C 0.06°C	±0.2°C ±0.06°C
Heave amplitude Pitch and roll	Accelero-meter Datawell HIPPY Gravity stabilised platform Datawell PIRO M402	-20m to +20m 0 to 60°C	Centre well Main battery compartment	continuous record for 20 minutes at the start of each hour	3% up to 15s period 1% up to 30°C 2.5% up to 60°	0.1m 0.1°	±0.2m ±1°
Surface current magnitude EW&NS Direction	Acoustic pulse velocity AERE Magnetic compass Colnbrook Instruments Ltd	0 to 2.55 m/s (no lower threshold) 0 to 360°	3 metres below sea level 5.2 metres above sea level	5 minutes mean about the half hour	±0.01 m/s ±2°	0.01 m/s 0.4°	±0.02 m/s ±2°

Table 2.1 Specifications of the recording instruments on-board the DB1 data buoy

Hs																		
11.5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0.000
11	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	3	0.000
10.5	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0.000
10	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	3	0.000
9.5	0	0	0	0	0	0	0	0	0	0	1	2	1	1	0	0	0	5	0.001
9	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	4	0.001
8.5	0	0	0	0	0	0	0	0	0	2	5	2	2	0	0	0	0	11	0.001
8	0	0	0	0	0	0	0	0	0	2	3	4	0	0	0	0	1	10	0.001
7.5	0	0	0	0	0	0	0	0	0	4	8	3	1	0	0	0	0	16	0.002
7	0	0	0	0	0	0	0	0	0	7	9	8	0	1	1	0	0	26	0.003
6.5	0	0	0	0	0	0	0	0	4	16	12	9	2	0	1	0	0	44	0.006
6	0	0	0	0	0	0	0	0	9	30	20	11	3	1	1	0	0	75	0.010
5.5	0	0	0	0	0	0	0	1	39	50	19	8	10	1	0	0	0	128	0.016
5	0	0	0	0	0	0	0	16	73	63	26	6	1	0	0	0	0	185	0.024
4.5	0	0	0	0	0	0	0	40	119	64	43	10	4	0	0	0	0	280	0.036
4	0	0	0	0	0	0	5	94	167	94	43	14	4	0	0	0	0	421	0.054
3.5	0	0	0	0	0	0	28	212	237	94	33	3	2	1	0	0	0	610	0.078
3	0	0	0	0	0	0	121	324	212	127	38	11	1	1	1	0	0	836	0.107
2.5	0	0	0	0	0	31	354	411	188	92	33	6	1	0	0	0	0	1116	0.143
2	0	0	0	0	0	154	559	394	209	83	9	0	0	0	0	0	0	1408	0.180
1.5	0	0	0	0	7	379	650	358	129	34	6	1	0	0	0	0	0	1564	0.200
1	0	0	0	0	51	458	350	95	19	6	0	0	0	0	0	0	0	979	0.125
0.5	0	0	0	0	2	45	20	7	1	1	0	0	0	0	0	0	0	76	0.010
	0	0	0	0	60	1067	2087	1952	1406	769	308	107	35	7	4	0	1	7803	<-TOTAL
	0.00	0.00	0.00	0.00	0.01	0.14	0.27	0.25	0.18	0.10	0.04	0.01	0.00	0.00	0.00	0.00	0.00		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Tz	

TABLE 2.2 SCATTER DIAGRAM FOR THE TIME DOMAIN ESTIMATES OF SIGNIFICANT WAVE HEIGHT (M) AND ZERO CROSSING PERIOD (S)

Hs																		
12.5	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	2	0.000	
12	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2	0.000	
11.5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0.000	
11	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2	0.000	
10.5	0	0	0	0	0	0	0	0	0	0	2	3	0	0	0	5	0.001	
10	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	3	0.000	
9.5	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	5	0.001	
9	0	0	0	0	0	0	0	0	0	1	3	1	0	0	0	5	0.001	
8.5	0	0	0	0	0	0	0	1	0	4	9	2	0	0	0	16	0.002	
8	0	0	0	0	0	0	1	0	1	3	9	1	0	1	0	16	0.002	
7.5	0	0	0	0	0	0	0	0	0	11	7	3	2	1	0	24	0.003	
7	0	0	0	0	0	0	0	1	7	21	10	7	2	1	0	49	0.006	
6.5	0	0	0	0	0	0	1	1	13	37	15	3	2	0	0	72	0.009	
6	0	0	0	0	0	0	2	5	38	39	16	8	0	0	0	108	0.014	
5.5	0	0	0	0	0	1	2	12	74	41	11	6	1	0	0	148	0.019	
5	0	0	0	0	0	2	11	38	96	47	22	2	0	0	0	218	0.028	
4.5	0	0	0	0	0	7	7	93	139	64	19	4	1	0	0	334	0.043	
4	0	0	0	0	1	10	33	170	157	67	22	2	1	0	0	463	0.059	
3.5	0	0	0	0	5	11	91	254	168	64	12	5	0	0	0	610	0.078	
3	0	0	0	0	6	13	270	338	120	59	15	0	3	0	0	824	0.106	
2.5	0	0	0	0	4	100	468	321	142	56	9	1	0	0	0	1101	0.141	
2	0	0	0	0	2	362	549	286	112	26	3	0	0	0	0	1340	0.172	
1.5	0	0	0	0	42	585	504	199	71	8	0	0	0	0	0	1409	0.181	
1	0	0	0	0	206	472	256	47	6	0	0	0	0	0	0	987	0.126	
0.5	0	0	0	0	24	30	5	0	0	0	0	0	0	0	0	59	0.008	
	0	0	0	0	290	1593	2200	1766	1144	548	192	54	13	3	0	7803	<-TOTAL	
	0	0	0	0	0.04	0.2	0.28	0.23	0.15	0.07	0.02	0.01	0	0	0			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Tz		

TABLE 2.3 SCATTER DIAGRAM FOR THE FREQUENCY DOMAIN ESTIMATES OF SIGNIFICANT WAVE HEIGHT (M) AND ZERO CROSSING PERIOD (S)

Hsfreq														
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
13	0	0	0	0	0	0	0	0	0	0	1	1	2	0.000
12	0	0	0	0	0	0	0	0	0	1	4	0	5	0.001
11	0	0	0	0	0	0	0	1	1	4	0	0	6	0.001
10	0	0	0	0	0	0	0	1	12	1	0	0	14	0.002
9	0	0	0	0	0	0	2	19	2	0	0	0	23	0.003
8	0	0	1	0	0	0	44	5	0	1	0	0	51	0.007
7	0	0	0	1	6	124	24	0	0	0	0	0	155	0.022
6	0	1	1	2	227	79	0	0	0	1	0	0	311	0.043
5	0	0	4	419	232	0	0	0	0	0	0	0	655	0.091
4	1	3	624	606	0	0	0	0	0	0	0	0	1234	0.172
3	2	560	1291	3	0	0	0	0	0	0	0	0	1856	0.258
2	457	2386	31	0	0	0	0	0	0	0	0	0	2874	0.400
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
	460	2950	1952	1031	465	203	70	26	15	8	5	1	7186	<-TOTAL
	0.06	0.41	0.27	0.14	0.06	0.03	0.01	0	0	0	0	0		
	1	2	3	4	5	6	7	8	9	10	11	12	Hs time	

TABLE 2.4 SCATTER DIAGRAM FOR TIME AND FREQUENCY DOMAIN ESTIMATES OF THE SIGNIFICANT WAVE HEIGHT

Hs																	
11.5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0.000	
11	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	3	0.000
10.5	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	2	0.000
10	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	3	0.000
9.5	0	0	0	1	0	1	0	0	0	0	1	1	0	0	0	4	0.001
9	0	0	0	0	0	1	0	1	1	0	0	1	0	0	0	4	0.001
8.5	0	0	0	2	3	0	0	2	2	2	0	0	0	0	0	11	0.002
8	0	0	1	1	0	4	0	3	1	0	0	0	0	0	0	10	0.001
7.5	0	0	0	3	4	1	2	1	2	0	0	0	0	0	0	13	0.002
7	0	0	2	3	2	2	3	6	2	4	1	0	0	0	0	25	0.004
6.5	0	2	3	7	4	3	6	4	5	1	2	0	0	0	0	37	0.005
6	0	3	4	8	8	8	6	12	5	1	1	0	0	0	0	56	0.008
5.5	1	3	5	14	14	14	19	14	6	2	1	0	0	0	0	93	0.014
5	0	10	11	11	23	25	25	23	14	2	2	0	0	0	0	146	0.021
4.5	1	11	21	32	30	43	35	20	11	1	1	1	0	0	0	207	0.030
4	2	17	30	60	70	67	40	28	12	4	0	0	0	0	0	330	0.048
3.5	5	16	49	87	117	96	90	27	14	1	0	0	0	0	0	502	0.073
3	11	33	93	118	173	172	71	24	10	2	1	2	0	0	0	710	0.104
2.5	15	65	118	189	238	219	81	33	12	2	1	0	0	0	0	973	0.142
2	24	110	216	311	313	196	53	20	6	2	1	0	0	0	0	1252	0.183
1.5	55	158	349	420	299	103	38	14	9	4	2	2	0	0	1	1454	0.213
1	48	176	291	278	93	19	10	5	7	3	0	0	0	0	0	930	0.136
0.5	5	24	24	14	2	0	0	0	1	0	0	0	0	0	0	70	0.010
	167	628	1217	1559	1395	975	480	238	121	31	16	7	0	1	1	6836	<-TOTAL
	0.02	0.09	0.18	0.23	0.2	0.14	0.07	0.03	0.02	0	0	0	0	0	0		
	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	Uw	

TABLE 2.5 SCATTER DIAGRAM OF THE TIME DOMAIN ESTIMATES OF SIGNIFICANT HEIGHT (M) AND THE TEN MINUTE MEAN WIND SPEED (M/S)

U3

40	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0.000
38	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0.000
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
34	0	0	0	0	0	0	0	0	0	1	0	1	0	0	2	0.000
32	0	0	0	0	0	0	0	0	0	0	3	1	0	0	4	0.001
30	0	0	0	0	0	0	0	0	1	2	2	3	0	0	8	0.002
28	0	0	0	0	0	0	1	0	3	5	2	2	0	0	13	0.002
26	0	0	0	0	0	1	0	3	8	9	5	0	0	0	26	0.005
24	0	0	0	0	0	0	0	9	35	8	1	0	0	0	53	0.010
22	0	0	0	1	0	0	8	40	39	4	0	0	0	0	92	0.018
20	0	0	0	0	0	10	41	73	23	0	0	0	0	0	147	0.028
18	0	0	0	0	5	37	147	74	0	0	0	0	0	0	263	0.050
16	0	0	0	2	37	201	178	6	0	0	0	0	0	0	424	0.081
14	0	0	0	16	194	412	34	0	0	0	0	0	0	0	656	0.126
12	0	1	12	147	644	127	0	0	0	0	0	0	0	0	931	0.179
10	0	5	71	626	245	0	0	0	0	0	0	0	0	0	947	0.182
8	1	29	466	400	0	0	0	0	0	0	0	0	0	0	896	0.172
6	8	208	266	0	0	0	0	0	0	0	0	0	0	0	482	0.093
4	74	164	0	0	0	0	0	0	0	0	0	0	0	0	238	0.046
2	26	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0.005
	109	407	815	1192	1125	788	409	205	109	29	13	7	0	1	5210	
	0.02	0.08	0.16	0.23	0.22	0.15	0.08	0.04	0.02	0.01	0	0	0	0		
	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	Uw

TABLE 2.6 SCATTER DIAGRAM OF THREE SECOND GUST AND TEN MIUTE MEAN WIND SPEEDS

Hs

11.5	0	0	0	0	0	0	0	0	0	0	0	0.000
11	0	1	0	1	0	1	0	0	0	0	3	0.000
10.5	1	0	0	1	0	0	0	0	0	0	2	0.000
10	0	0	0	0	1	1	0	0	0	0	2	0.000
9.5	0	1	0	0	1	1	0	0	0	0	3	0.000
9	0	2	0	0	0	1	0	0	0	0	3	0.000
8.5	0	0	0	1	4	1	0	0	0	0	6	0.001
8	1	1	3	1	1	0	0	0	0	0	7	0.001
7.5	1	4	2	3	2	0	1	0	0	0	13	0.002
7	4	3	7	6	2	3	0	0	0	0	25	0.004
6.5	5	2	8	8	5	1	0	0	0	0	29	0.004
6	6	9	10	13	7	1	1	0	0	0	47	0.007
5.5	6	19	15	19	11	6	3	0	0	0	79	0.012
5	10	21	36	28	22	13	1	1	0	0	132	0.020
4.5	16	29	52	50	30	6	4	1	0	0	188	0.029
4	14	54	76	61	47	32	3	1	0	0	288	0.044
3.5	31	106	110	101	68	37	8	4	0	0	465	0.071
3	41	127	185	152	108	51	14	1	1	1	681	0.103
2.5	44	145	222	265	167	73	34	13	6	0	969	0.147
2	57	172	282	299	216	133	54	14	8	1	1236	0.188
1.5	46	189	322	355	266	158	54	21	12	0	1423	0.216
1	34	129	217	216	170	95	40	7	1	0	909	0.138
0.5	7	7	14	17	15	9	7	0	0	0	76	0.012
	324	1021	1561	1597	1143	623	224	63	28	2	6586	<-TOTAL
	0.05	0.16	0.24	0.24	0.17	0.09	0.03	0.01	0	0	Uc	
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1		

TABLE 2.7 SCATTER DIAGRAM FORTHE TIME DOMAIN ESTIMATE OF SIGNIFICANT WAVE HEIGHT (M) AND CURRENT SPEED (M/S)

Uc																	
1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	2	0.000
0.9	0	6	6	7	7	0	2	0	0	0	0	0	0	0	0	28	0.004
0.8	3	11	9	12	11	12	4	2	0	0	0	0	0	0	0	64	0.010
0.7	5	35	47	59	43	19	10	2	1	1	0	0	0	0	0	222	0.035
0.6	22	68	111	153	110	73	37	12	14	7	4	0	0	0	0	611	0.096
0.5	29	120	214	232	219	153	68	41	27	2	4	3	0	0	0	1112	0.174
0.4	41	144	274	356	323	220	110	41	17	8	2	0	0	0	1	1537	0.240
0.3	30	122	269	349	321	230	100	56	20	7	2	0	0	0	0	1506	0.235
0.2	17	47	167	235	212	160	79	47	22	1	3	2	0	1	0	993	0.155
0.1	6	26	53	63	57	49	33	19	9	3	1	1	0	0	0	320	0.050
	153	579	1150	1466	1304	917	443	220	110	29	16	6	0	1	1	6395	<-TOTAL
	0.02	0.09	0.18	0.23	0.2	0.14	0.07	0.03	0.02	0	0	0	0	0	0		
	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	Uw	

TABLE 2.8 SCATTER DIAGRAM OF TEN MINUTE MEAN WIND SPEED (M/S) AND FIVE MINUTE MEAN CURRENT SPEED (CM/S)

Hs														
11.5	0	0	0	0	0	0	0	0	0	1	0	0	1	0.000
11	0	0	0	0	0	0	0	0	1	2	0	0	3	0.000
10.5	0	0	0	0	0	0	0	0	2	0	0	0	2	0.000
10	0	0	0	0	0	0	0	0	1	2	0	0	3	0.000
9.5	0	0	0	0	0	0	0	0	2	2	0	0	4	0.001
9	0	0	0	0	0	0	0	1	1	2	0	0	4	0.001
8.5	0	1	0	0	0	0	1	1	3	4	1	0	11	0.002
8	0	0	0	0	0	0	0	2	4	3	0	0	9	0.001
7.5	0	0	0	0	1	0	2	3	3	6	0	1	16	0.002
7	0	2	0	0	0	1	4	3	7	7	0	2	26	0.004
6.5	1	1	1	0	0	2	3	4	9	17	4	1	43	0.006
6	2	1	2	0	0	5	2	15	11	16	9	8	71	0.010
5.5	4	1	1	0	4	8	10	14	19	30	18	7	116	0.016
5	8	5	5	2	6	8	18	29	24	29	21	12	167	0.023
4.5	13	4	11	3	7	13	39	37	44	38	21	12	242	0.034
4	17	23	11	7	12	28	47	61	58	50	39	16	369	0.051
3.5	29	16	7	19	19	50	73	109	60	66	71	40	559	0.078
3	50	43	23	9	25	73	102	123	70	77	103	69	767	0.107
2.5	58	44	48	22	43	75	107	149	114	130	124	108	1022	0.143
2	104	59	76	45	61	100	143	150	110	165	191	124	1328	0.185
1.5	163	97	89	66	66	95	114	106	113	189	210	139	1447	0.202
1	119	74	64	70	47	79	68	47	44	81	116	78	887	0.124
0.5	2	4	14	16	4	6	6	1	2	4	8	3	70	0.010
	570	375	352	259	295	543	739	855	702	921	936	620	7167	<-TOTAL
	0.08	0.05	0.05	0.04	0.04	0.08	0.1	0.12	0.1	0.13	0.13	0.09		
	30	60	90	120	150	180	210	240	270	300	330	360	Dw	

TABLE 2.9 SCATTER DIAGRAM OF SIGNIFICANT WAVE HEIGHT (M) AGAINST THE WIND DIRECTION (DEGREES)

Tz														
17	0	0	0	0	0	0	0	0	0	1	0	0	1	0.000
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
15	1	0	0	0	0	0	0	1	1	0	0	1	4	0.001
14	0	0	0	0	1	0	0	1	1	3	0	1	7	0.001
13	1	0	0	0	1	1	1	6	5	10	3	3	31	0.004
12	0	0	0	2	2	2	11	10	27	24	15	6	99	0.014
11	12	7	8	3	5	12	34	53	40	51	27	26	278	0.039
10	37	36	21	19	19	69	89	111	79	97	81	37	695	0.097
9	84	62	57	27	48	102	141	196	153	161	148	83	1262	0.176
8	134	77	93	58	76	143	207	230	164	211	222	164	1779	0.248
7	169	98	95	68	89	151	171	203	178	258	292	193	1965	0.274
6	124	85	72	77	51	60	81	44	54	99	144	99	990	0.138
5	8	10	6	5	3	3	4	0	0	6	4	7	56	0.008
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000
	570	375	352	259	295	543	739	855	702	921	936	620	7167	<-TOTAL
	0.08	0.05	0.05	0.04	0.04	0.08	0.1	0.12	0.1	0.13	0.13	0.09		
	30	60	90	120	150	180	210	240	270	300	330	360	Dw	

TABLE 2.10 SCATTER DIAGRAM OF ZERO CROSSING PEROD (S) AGAINST THE WIND DIRECTION (DEGREES)

index	year/month	Hs(TIME)	Hs(FREQ.)	Tz(TIME)	Tz(FREQ.)	Uw	Uc
1	7806	4.1	4.2	9.4	8.4	15.5	89.0
2	7807	3.2	3.3	10.9	9.7	13.15	76.0
3	7808	2.6	2.8	10.5	9.7	11.6	87.0
4	7809	5.5	5.9	14.5	12.9	17.75	82.0
5	7810	3.9	4	12.4	11.4	14.5	88.0
6	7811	5.3	5.5	11.8	14.3	19.5	77.0
7	7812	11.5	12.3	13.9	12.3	11.9	70.0
8	7902	6.4	7.5	14.5	14	17.9	0.0
9	7903	6.9	7.8	14.5	13.5	0	0.0
10	7904	4.7	5	12.2	11	0	55.0
11	7905	3.5	3.6	8	7.5	0	42.0
12	7906	3.3	3.4	9.9	9.6	12.7	71.0
13	7907	2.4	2.5	9.8	9.3	14.3	81.0
14	7908	5.7	5.9	9.4	9.2	18.4	88.0
15	7909	4.3	4.4	11	9.8	15.1	87.0
16	7810	5.2	5.6	12.6	12.1	19.25	59.0
17	7911	4.4	4.6	10.3	10.1	13.9	47.0
18	7912	10.9	11.7	12.8	11.9	26.4	67.0
19	8002	10.2	10.5	11.8	11.1	22.7	66.0
20	8003	11	12.4	12.9	12.3	22.7	71.0
21	8004	3	3.3	10	11.1	15.2	71.0
22	8005	3	3.1	10.4	9.5	14.05	68.0
23	8006	5.4	5.6	10	9.5	15.15	65.0
24	8007	3.5	3.7	9.2	8.6	15.6	68.0
25	8008	2.8	3	9.6	13.1	12.35	96.0
26	8009	4.2	4.5	9.9	9.2	14.25	61.0
27	8010	7.9	8.4	12.8	12.6	20.25	78.0
28	8011	5.6	6.5	11.3	11.5	18.1	58.0
29	8012	8.4	9.3	13	11.6	17.6	61.0
30	8102	7.3	8.2	11.3	10.6	17.5	65.0
31	8103	2.3	2.5	8.1	10	20.3	36.0
32	8104	3.9	4.4	11	10.2	12.45	65.0
33	8105	5	5.7	10.1	9.7	17.2	77.0
34	8106	4	4	9.7	9.2	14.2	68.0
35	8107	3.5	3.6	9.7	9	14.75	75.0
36	8108	3.1	2.9	10.7	10.1	12.45	81.0
37	8109	6.1	6.3	10.6	9.7	21.6	83.0
38	8110	6.6	7.5	13.2	12.5	21.65	76.0
39	8111	5.6	6.5	11.3	11.5	14.7	73.0
40	8112	3.1	2.9	10.7	10.1	28.55	65.0
41	8202	5.8	6.4	12.3	12	0	0.0
42	8203	9.5	8	16.9	11.5	16.2	51.0

TABLE 2.11 VECTORS OF MONTHLY COMPONENTWISE MAXIMA FOR THE UKOOA-DB1 DATA SET

TABLE 2.12 STATISTICAL MOMENTS AND MAXIMUM LIKELIHOOD PARAMETERS FOR THE WEIBULL AND GENERALISED EXTREME VALUE DISTRBUTIONS

VARIABLE	no. of	statistical moments				weibull model			GEV model		
	samples	mean	variance	skewness	kurtosis	shape	scale	location	shape	scale	location
Hs time	42	5.348	6.1118	1.005	0.1009	1.39	3.622	2.053	-0.2448	1.532	4.052
Hs freq.	42	5.695	7.201	0.9577	0.107	1.319	3.767	2.232	-0.267	1.665	4.255
Tz time	42	11.31	3.497	0.7051	0.3022	2.086	4.163	7.616	0.06684	1.554	10.5
Tz freq.	42	10.78	2.56	0.2933	-0.7414	2.611	4.315	6.952	0.1889	1.463	10.16
current velocity	39	70.36	176.4	-0.4565	-0.05562	6.983	82.83	-7.098	0.4231	13.96	66.63
wind speed	38	16.88	15.98	0.973	0.4973	1.473	6.284	11.2	-0.1324	2.747	14.91

WEIBULL ->

time domian significant height

scale	shape	location
0.030162	0.009145	0.002944
	0.010847	-0.00647
		-0.00554

frequency domain significant height

0.02832	0.00843	0.001737
	0.00863	-0.006
		-0.0042

time domain zero crossing period

0.021099	0.006918	0.0084899
	0.009097	-0.0055377
		-0.010937

WEIBULL ->

frequency domain zero crossing period

0.011732	0.006652	0.016534
	0.07569	-0.00602
		-0.0196

current speed

-0.0436	0.003266	0.067762
	0.00354	-0.00324
		-0.07167

wind speed

0.030162	0.009145	0.002944
	0.010847	-0.0064721
		-0.0055395

GEV ->

time domian significant height

scale	shape	location
0.008607	0.003753	0.006468
	0.029348	0.008003
		0.012552

frequency domain significant height

0.00971	0.006148	0.007704
	0.038834	0.010873
		0.013681

time domain zero crossing period

0.010687	0.004105	0.0050878
	0.011838	0.005884
		0.021142

GEV ->

frequency domain zero crossing period

0.015052	0.010023	0.007656
	0.018348	0.011689
		0.028676

current speed

0.023575	0.016778	0.017813
	0.016757	0.020904
		0.045311

wind speed

0.010214	0.005017	0.006856
	0.027224	0.0088122
		0.016568

COVARIANCE MATRICES FOR THE ENVIRONMENTAL DATA MARGINAL DISTRIBUTIONS

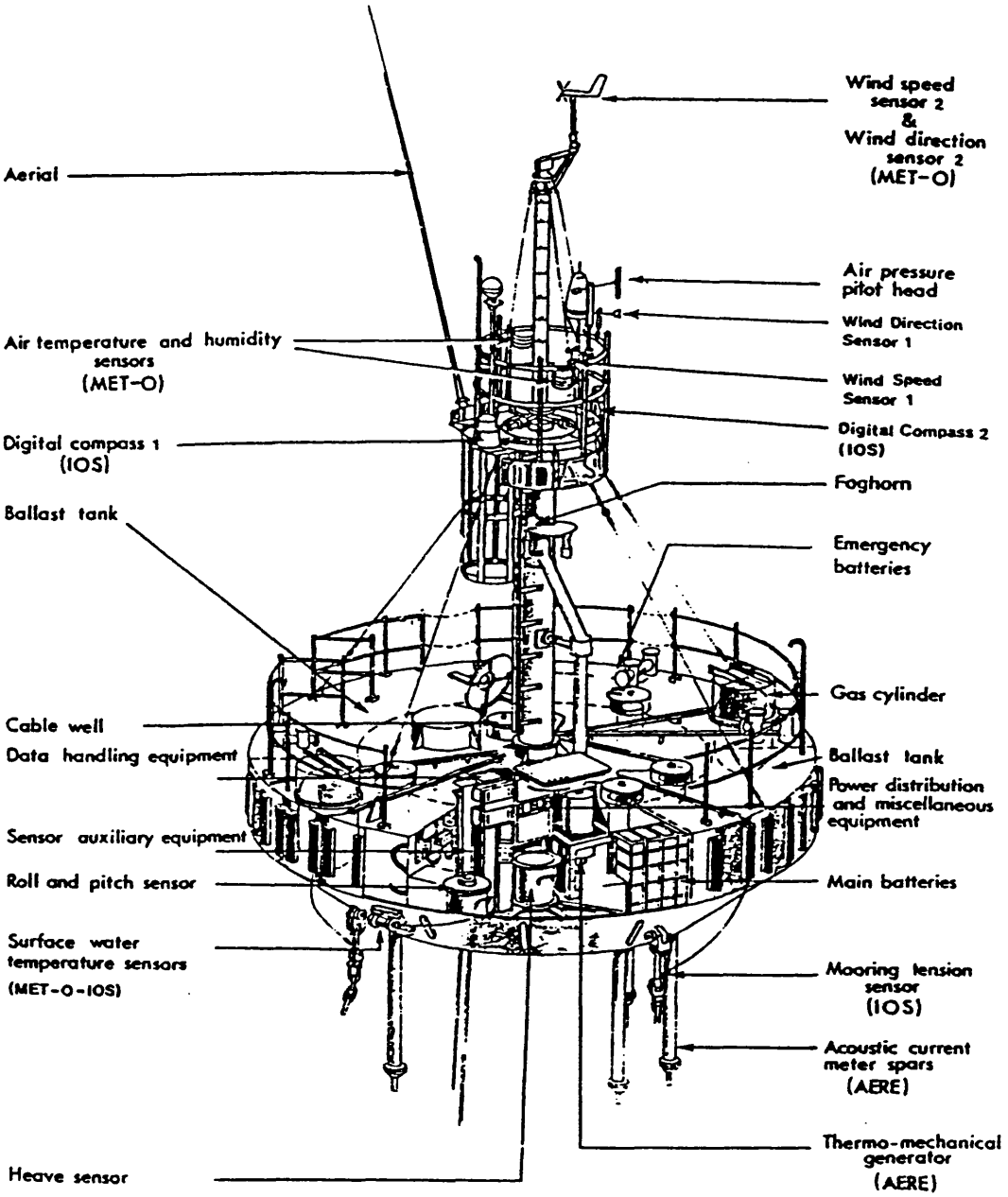


FIGURE 2.1 GENERAL ARRANGEMENT OF THE UNITED KINGDOM OFFSHORE OPERATORS BUOY DB1

Figure 2.2 Comparison of the UKOOA frequency domain and GU frequency domain significant wave height

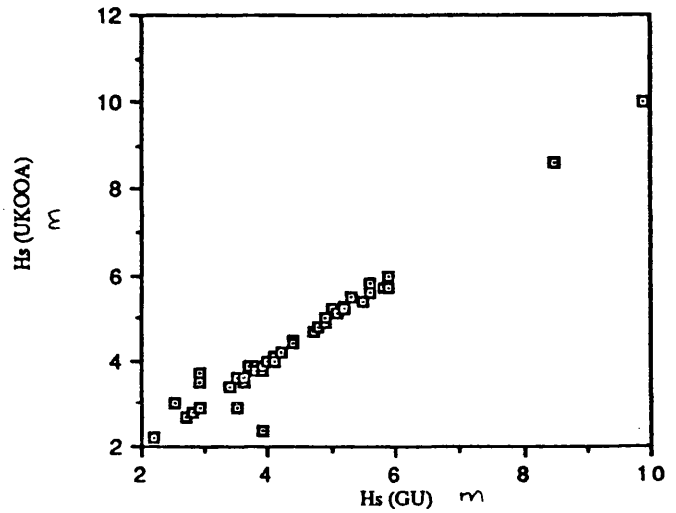
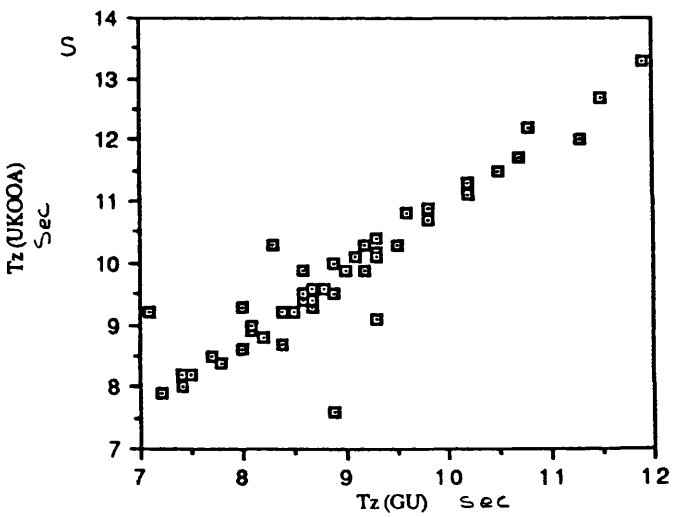


Figure 2.3 Comparison of UKOOA frequency domain and GU frequency domain zero crossing period



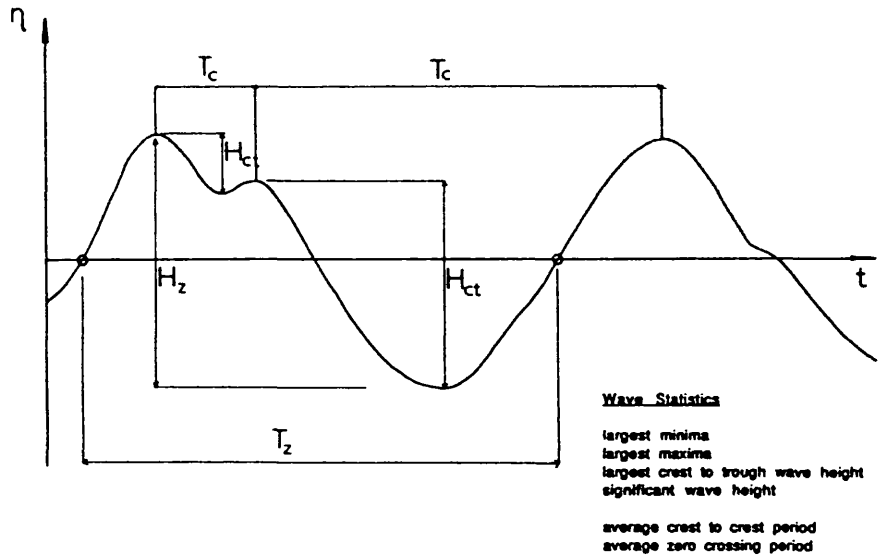


FIGURE 2.4 DEFINITION OF WAVE STATISTICS

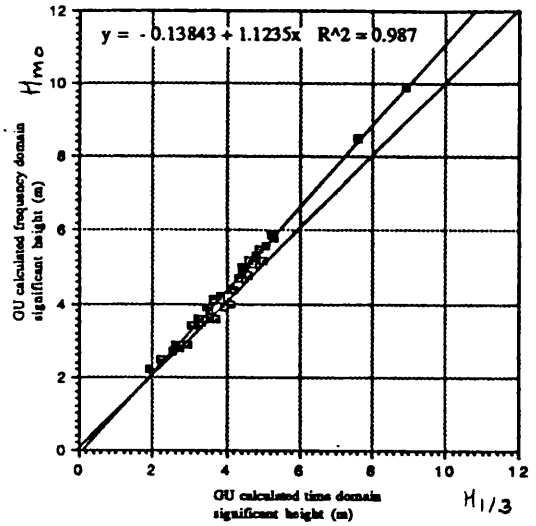


Figure 2.5 Comparison of GU calculated time and frequency domain estimates for significant wave height

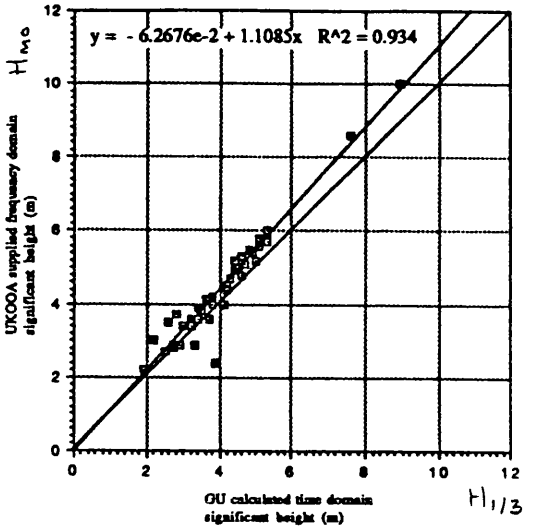


Figure 2.6 Comparison of GU calculated time domain and UKOOA supplied frequency domain estimates for significant wave height

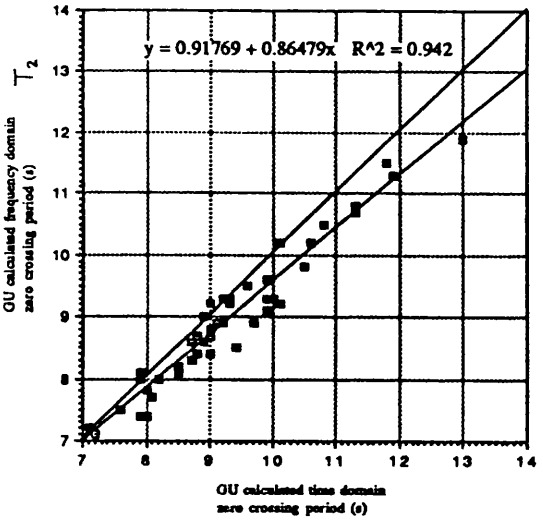


Figure 2.7 Comparison of GU calculated time and frequency domain estimates for zero crossing period

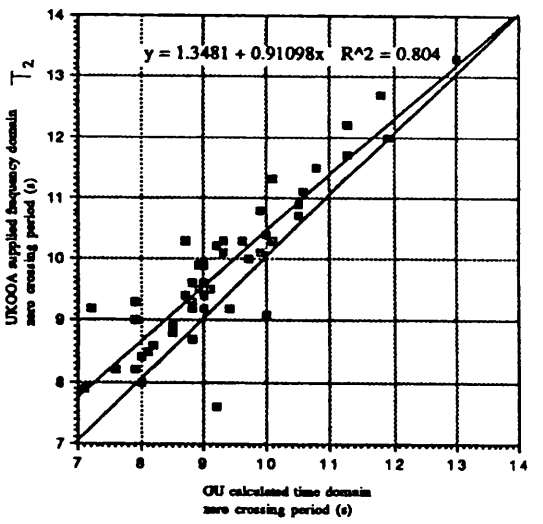


Figure 2.8 Comparison of GU calculated time domain and UKOOA supplied frequency domain estimates for zero crossing period

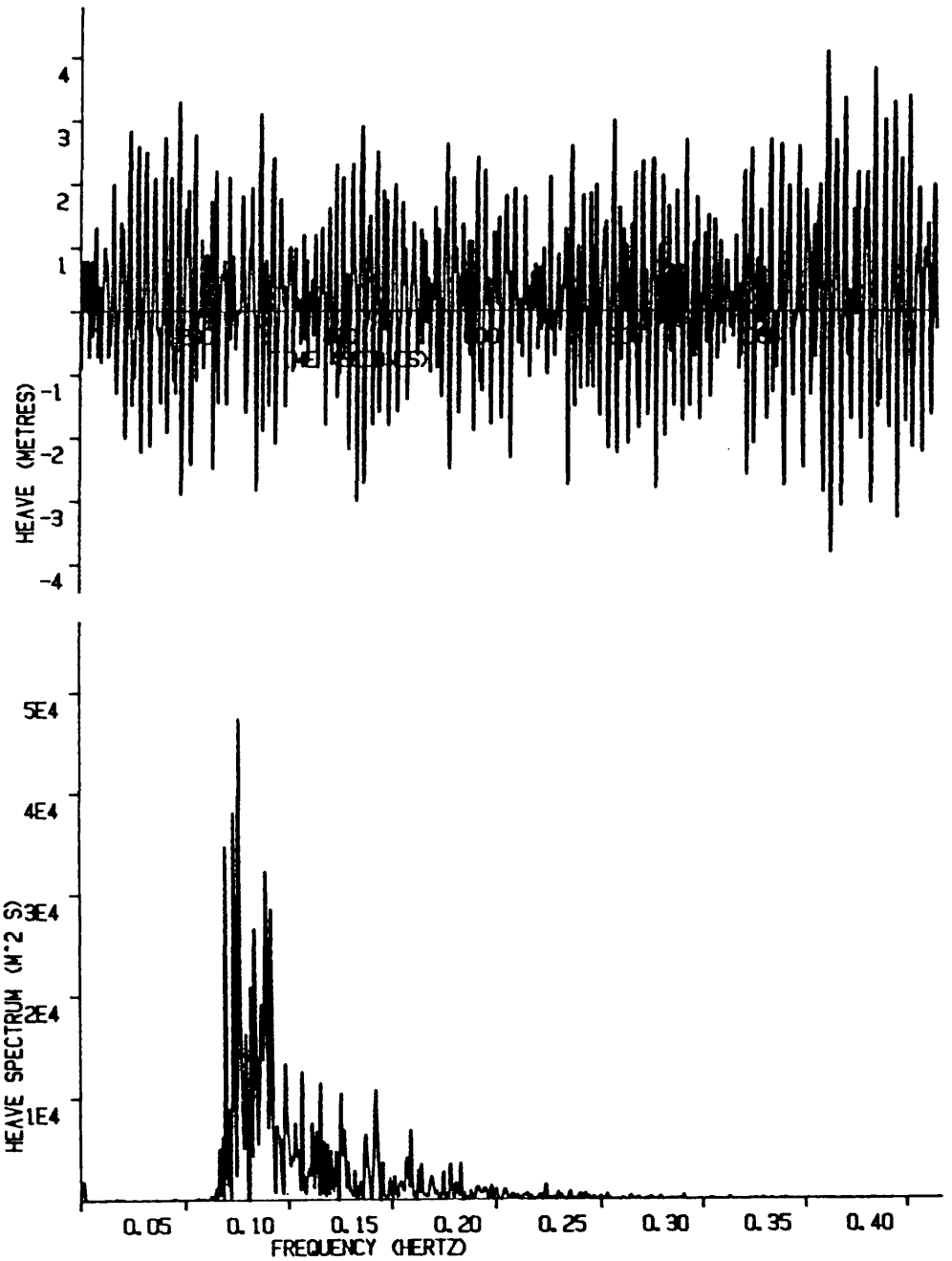
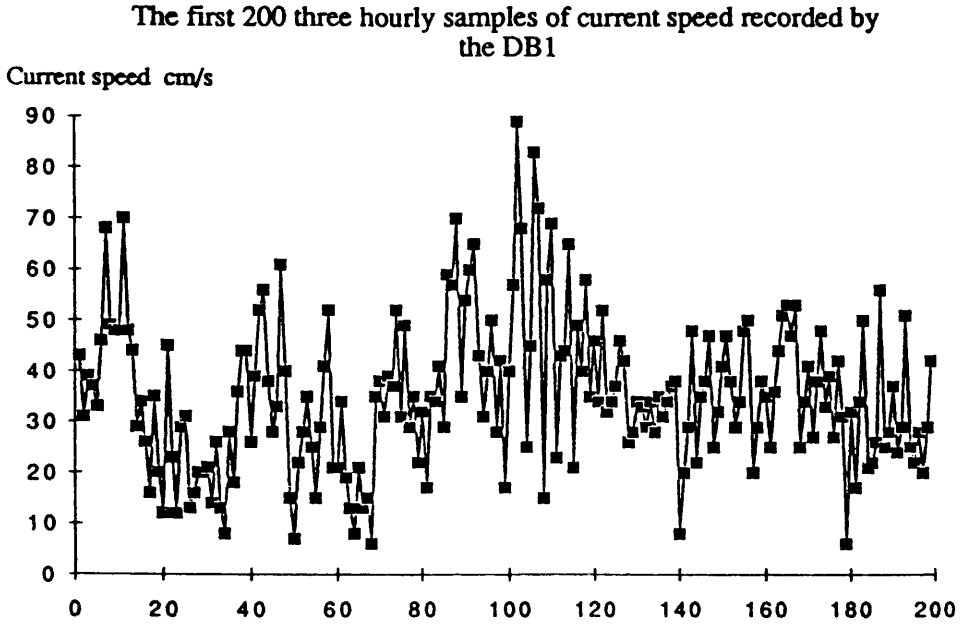


FIGURE 2.9 HEAVE TIME SERIES AND ITS PERIODOGRAM
(no smoothing)



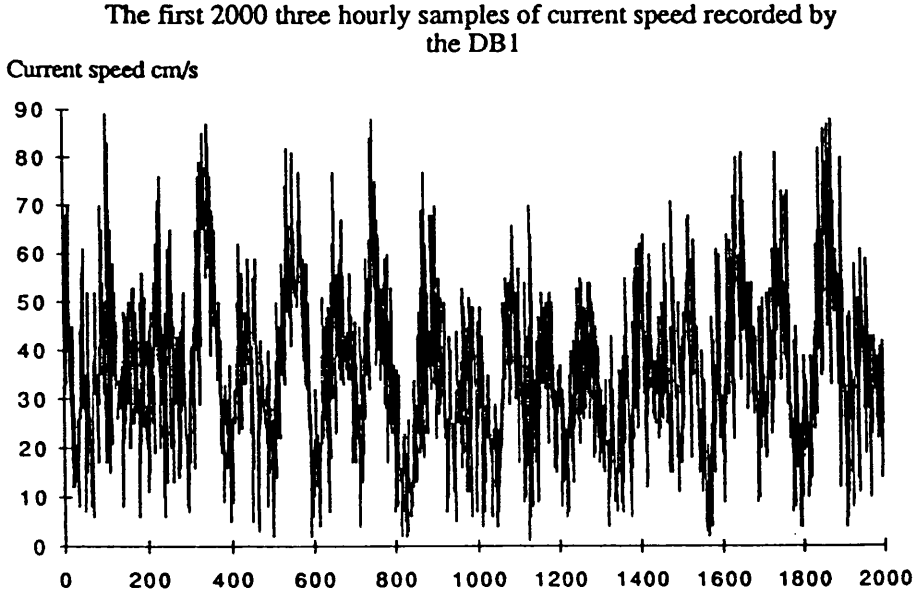


FIGURE 2.10 CURRENT TIME SERIES

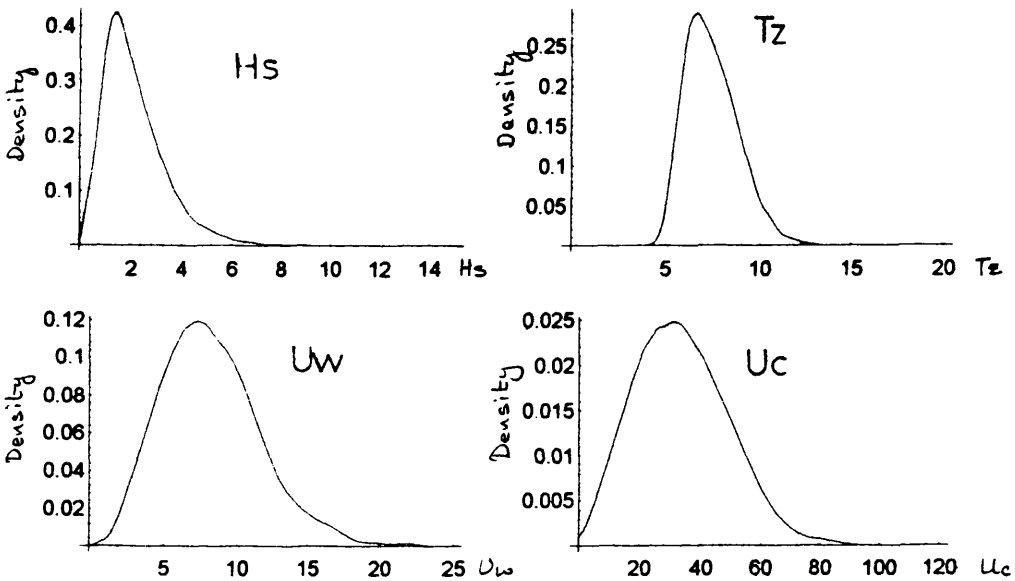


Figure 2.11 Kernel density plots using an optimal window for the populations of significant wave height (m), zero-up-crossing period (s), wind speed (m/s), and current speed (cm/s).

Chapter 3

MODELLING UNIVARIATE ENVIRONMENTAL DATA

NOMENCLATURE

a_i	*coefficients for a linear function*
a_k	*un-centred sample moments*
b	*bandwidth*
$b_n(\theta_i)$	*bias in model parameter* θ_i
c	*Weibull shape parameter*
$cov[.]$	*covariance*
$C(\boldsymbol{\theta})$	*expected covariance matrix for model parameters* $\boldsymbol{\theta}$
$c(\boldsymbol{\theta})$	*observed information matrix*
$E[.]$	*expected value operator* $E[y^k] = \int_{-\infty}^{\infty} y^k f(x)dx$
E_r	*the event r*
$f(x)$	*probability density function*
$F(x)$	*cumulative distribution function* $F(x) = Pr(X \le x)$
$F^{-1}(x)$	*inverse distribution function*
$g(\boldsymbol{x};\boldsymbol{\theta})$	*sample joint density with parameters* $\boldsymbol{\theta}$
$G(.)$	*extreme value distribution OR gradient vector*
$H(\boldsymbol{\theta};\boldsymbol{x})$	*entropy function on the sample* $\boldsymbol{x}$
k	*GEV shape parameter*
$k(x)$	*kernel density function*
$\boldsymbol{l}$	*vector of lower bounds*
$I(\boldsymbol{\theta})$	*Fisher's information matrix*
$L(\boldsymbol{\theta},\boldsymbol{X})$	*likelihood random variable*
$\ell(\boldsymbol{\theta},\boldsymbol{x})$	*likelihood function on the sample* $\boldsymbol{x}$
m	*maximum amplitude*
m_n	*largest in sample of size n*
$\hat{m}_r$	*rth centred sample moment, ie* $\hat{m}_1 = 0$
$M_{r,s,t}$	*probability weighted moment*
n	*number of samples*
P	*probability* $P = Pr\{X \le x\}$
p	*number of model parameters or counter for order stat's*
$Pr(.)$	*probability*
T_R	*return period*
$\boldsymbol{u}$	*vector of upper bounds*
$v[.]$	*variance*
w_j	j^{th} *weight*

$\boldsymbol{x}$	*sample vector*
x_i	*ith sample in the vector* $\boldsymbol{x}$ *of observations on the variable X*
X	*random variable*
$\boldsymbol{X}$	*random vector*
α	*Weibull Scale parameter*
$\hat{\delta}$	*threshold estimated probability of exceedance*
γ	*Euler's constant - 0.57721 5664....*
$\Gamma\{.\}$	*gamma function*
$\Delta\theta_i$	*finite difference interval*
$\theta_i, \hat{\theta}_i$	*ith model parameter and maximum likelihood estimate*
$\boldsymbol{\theta}$	*parameter vector*
μ	*GEV location parameter*
$\boldsymbol{\Omega}$	*set of possible parameters for* $\boldsymbol{\theta}$
ξ_0	*Weibull lower bound*
π	*Pi - 3.14159 2654....*
$\prod$	*multiplication operator*
ρ	*correlation coefficient*
$\rho(\boldsymbol{\theta})$	*error function*
σ	*GEV scale parameter or standard deviation*
$\Phi(.)$	*function of the ML parameters*

3. INTRODUCTORY REMARKS: MODELLING UNIVARIATE ENVIRONMENTAL DATA

Statistical estimation is dominated by large sample asymptotic methods which aim to describe data using optimal measures of, say, location and spread. The synthesis of a model can either be parametric, or intrinsic: however, as Hoaglin (1983) points out the distinction between the two classes is not clear. Parametric estimation involves fitting a functional form to the data under the hypotheses that the data are drawn from some known distribution and usually that the distribution of errors is Gaussian. Intrinsic (or nonparametric) estimation makes less rigid assumptions about the sample and seeks to estimate the statistics for the population using the data directly: examples are order statistics and kernel density estimators. In this chapter, both parametric and intrinsic estimation of univariate data is examined and applied to modelling environmental data. Although most of the theory in this chapter is intended for use in modelling the DB1 data discussed in Ch 2. it should be borne in mind the methods and conclusions are general and equally applicable when modelling strength data.

Predictive and descriptive statistical models of environmental data are required in: the calibration of design codes; the formulation of reliability methods; and the prediction of 'design events'. In most cases, estimation must be performed using small samples with missing values. There are therefore three main sources of uncertainty, which Thoft-Christensen & Baker (1982) define as physical, statistical, and modelling uncertainty. Physical uncertainty arises from the basic randomness of the variable and is quantified by the parametric or intrinsic model; statistical uncertainty arises from the use of finite sample sizes and is characterised by the bias and variance of the model parameters; and finally modelling uncertainty arises from the need to hypothesise a parametric form in the estimation (even kernel density estimators and order statistic estimators suffer from this uncertainty since they require a parametric smoothing function, and quantile estimation formula, respectively).

For design, a descriptive environmental model is required in fatigue damage calculations and operations analysis. On the other hand, predictive models are used to specify design events, which are usually taken as the combination of marginal extremes corresponding to some return period like 50 or 100 years. Two approaches are therefore used to model the data. The first is population modelling in which the full sample of data is used in the estimation; this approach benefits from a large sample which reduces the statistical uncertainty in the estimated model parameters and allows us to use the asymptotic properties of the parameter uncertainty, Silvey (1975). The population method is ideal for descriptive analysis but can result in biased estimates when used as a predictive model for the most probable largest extreme value in a large number of observations N. This is because N is usually large and therefore $F(x)^N$, Fig. 3.1, is only significant in the region of the largest few data, Galambos (1984). In consequence, if the wrong population model is used for $F(x)$ then it will give poor estimates for the 50 or 100 year return period values.

The second modelling method is extreme value estimation in which the largest values occurring in a fixed time period, or sample size, are modelled using one of the extreme value distributions, Gumbel (1958). The problem with modelling only the largest values (annual maxima ect) is the sample size is generally very small. This results in large statistical uncertainty in the estimated parameters which in turn results in wide confidence limits for the model. The motivation of extreme value theory is to reduce the modelling uncertainty in the model by identifying the set of distributions from which the extremes must have been drawn. Fisher & Tippet (1928) demonstrated that, under the assumption of independent and identically distributed events, the distribution of the largest drawn from any population $F(x)$ is attracted to one of only three distributions.

The assumption of independence is reasonable when using annual maxima but real datasets like the DB1 recordings are generally only recorded for a few years. With such small samples it is necessary to consider modelling say the monthly maxima, in which case the assumption of independence and identical distribution is questionable. Carter & Challenor (1981), for example, examine the error in estimates of return period values resulting from the *iid* assumption. They model the monthly maxima separately and then obtain the annual maxima from the individual distributions (assuming independence between the months). Challenor (1982) later improved the method by modelling the seasonal variation in the location parameter of the monthly maxima models. However, we cannot use the method with the multivariate asymptotic methods reviewed in Ch. 4 since they require that the marginal data first be transformed to unit exponential survivor functions using, for example, the generalised extreme value (GEV) distribution.

There are therefore three choices to be made when modelling the data:

whether to model the population or the extremes
whether to use intrinsic or parametric methods
the type of estimation process

In *Section 1* those properties which characterise a 'best' estimator are defined using classical statistics and practical limitations. *Section 2* examines how to choose amongst the various estimation procedures using the criterion for best. This involved a comparison of results from a number of Monte Carlo simulations reported in various statistical journals. The maximum likelihood method was selected as the best procedure because of its good overall performance despite some methods performing better under certain conditions. *Section 3* shows how nonparametric kernel density estimators can be used to aid data visualisation, which is used as the basis for parametric model selection. *Section 4* introduces the theoretical basis of maximum likelihood estimation and shows how statistical uncertainty can be estimated using the information matrix. It is then shown how this information can be used to calculate the uncertainty in functions of the model parameters using level II reliability methods. *Section 5* outlines the modelling procedure adopted in this thesis and *Section 6* then develops the likelihood function and information matrix for the Weibull model. *Section 7* introduces the generalised extreme value distribution, its likelihood and information matrix. *Section 8* discusses some of the numerical difficulties associated with the likelihood method and then *Section 9* applies the theory to simulated datasets in order to check the likelihoods and derivatives derived for this study. *Section 10* introduces a simple peak over threshold (POT) approach which is intended to overcome the problems associated with using the Weibull distribution to model extremes. The POT method is then applied in *Section 11* to a sample of structural response time series to illustrate the generality of the likelihood method. Finally, in *Section 12* both population and monthly maxima models are used with the DB1 wind, wave and current data.

3.1 CHARACTERISTICS OF A GOOD ESTIMATOR

Before looking at the principal methods used to estimate a model's parameters we must define those characteristics which exemplify a good estimator: unfortunately, it is not straightforward to define what is meant by good. In the statistical context Ledermann (1984) defines four desirable characteristics for an estimator which are summarised below:

Consistency - The estimation procedure should produce an estimate which is accurate. That is, if a sample replicates the population the estimated parameters $\hat{\theta}$ should be close to the population parameters and, moreover, the estimates should improve as the sample size increases

$$\left.\begin{aligned} E\left[\hat{\theta}\right] &\rightarrow \theta \\ v\left[\hat{\theta}\right] &\rightarrow 0 \end{aligned}\right\} \quad as \quad n \rightarrow \infty$$

To satisfy these conditions the estimated parameters must have a high probability of being close to the population parameters. A problem is encountered when modelling extremes because the assumption of multivariate normality for the distribution of the model's parameters is invalid when the sample size is too small; but, how small is too small? We shall see later that for practical estimation less than 20 is too small for the model parameters to be taken as multivariate Normal distributed.

Sufficiency - some procedures enable more information to be extracted from the sample than others do. A sufficient estimator is one which can extract all the information from the sample which is relevant to the parameter.

Bias - The estimate may differ from the population parameter θ_i, due to a bias $b_n(\theta_i) = \theta - E\left[\hat{\theta}_i\right]$. An unbiased estimate is not necessarily the most important property for an estimation procedure since an unbiased estimate $\hat{\theta}_i$ will not result in an unbiased estimate for some quantile $F\left(x_0; \hat{\theta}\right)$ if the model parameters have non-zero variance.

Low sampling variance - Each parameter derived using an estimator with a finite sample size will have some statistical uncertainty usually characterised by its bias and variance. Furthermore, models with more than one parameter will have multivariate distributed parameters. A good estimator will provide parameter estimates with the minimum sampling variance attainable for the sample size.

In addition to consistency, sufficiency, low bias and low sampling variance, two further characteristics should be added for the method to be of practical use by engineers and scientists:

Reliability - From the practical point of view the methods must be robust, that is, it should be possible to obtain solutions to the model parameters for nearly all samples of data without prior knowledge of the model parameters.

Simplicity - The process of estimating the model parameters, and their sampling covariance matrix, should be simple to apply and interpret.

Several estimators have one or more of these desirable characteristics: moments, least squares, and maximum likelihood are a few. The final choice of estimator must allow for the objectives of the modelling process, which in our case is the modelling of extreme quantiles with low bias and low variance.

3.2 ESTIMATING THE MODEL PARAMETERS

Five methods are commonly used to estimate the parameters of a model given a sample of data $x = \{x_1, \ldots x_n\}$. The simplest method is to calculate the integrated moments of the theoretical distribution, a_k where

$$a_k = \int_{-\infty}^{\infty} (x)^k f(x)dx \equiv E\left[(x)^k\right] \qquad [3.1]$$

and then equate them to the sample moments of the data $\tilde{a}_k$, where

$$\tilde{a}_k = \frac{1}{n}\sum_{j=1}^{n}\left(x_j\right)^k \qquad [3.2]$$

In the case of a Weibull distribution with three parameters, the first three moments are required, i.e. $k=1,2,3$, to solve for the shape, scaling, and location parameters. The parameters are then obtained by solving three nonlinear simultaneous equations $a_k = \tilde{a}_k$. This solution procedure is relatively simple since closed form gamma functions are available for the moments of a Weibull model.

The second method is a least squares fit based on percentile estimation and order statistics. This approach is intuitively appealing since the process can be performed graphically when the model has two parameters, in which case the minimisation of the sum of

squares error is done by visually fitting the distribution to the data on logarithmically scaled paper.

If the model has more than two parameters the process can be automated by using some form of weighted, orthogonal distance, regression routine to minimise an error function $\rho(\boldsymbol{\theta})$ which is a function of the model parameters $\boldsymbol{\theta}$

$$min\left\{\rho(\boldsymbol{\theta}) = \sum_{j=1}^{n} w_j\left(\tilde{F}(x_j) - F(x_j;\hat{\boldsymbol{\theta}})\right)^2\right\} \qquad [3.3]$$

Here, the expected probability $E\left[Pr(X < x_p)\right]$ for the observation x_p is given by ordering the data from the smallest x_1, to the p^{th} largest x_p, up to the largest x_n. The expected probability for the quantile x_p is then found using an order statistic plotting position like

$$\tilde{F}(x_p) = E\left[F(x_p)\right] = \frac{p}{n+1} \qquad [3.4]$$

Other plotting positions are given in Gumbel (1958), and Crowder et al (1991).

The third method is parametric maximum likelihood estimation in which the model parameters are estimated directly from the data. In common with the method of moments, and the method of least squares, the realisation $\boldsymbol{x}$ is assumed to be a sample of independent, identically distributed random variables drawn from the population random vector $\boldsymbol{X} = \{X_1,.....X_n\}$ with sample pdf denoted by

$$g(\boldsymbol{x};\boldsymbol{\theta}) = g(x_1,.....x_n;\theta_1,....\theta_p) \qquad \boldsymbol{\theta} \in \boldsymbol{\Omega}$$

The set $\boldsymbol{\Omega}$ is the set of possible parameters, and the form of $g(\boldsymbol{x};\boldsymbol{\theta})$ is guessed at after graphical and numerical examination of the data. The likelihood function of $\boldsymbol{\theta}$ on the sample $\boldsymbol{x}$ is then defined as

$$\ell(\boldsymbol{\theta};\boldsymbol{x}) = a(x)g(\boldsymbol{x};\boldsymbol{\theta})$$

where $a(\boldsymbol{x})$ is any multiplier solely a function of $\boldsymbol{x}$ and independent of $\boldsymbol{\theta}$. It may be chosen arbitrarily to aid estimation of the parameter vector. If the sample $\boldsymbol{x}$ is assumed to comprise *iid* Weibull random variables then the sample likelihood is given by

$$\ell(\theta;x)=\prod_{j=1}^{n} f(x_1,\ldots x_n) \quad : \theta=\{\alpha,c,\xi_0\}$$

where $f(x)$ is the Weibull probability density function, and θ is the vector of Weibull model parameters. An estimate for the parameters $\hat{\theta}$ is found by maximising $\ell(\theta;x)$ over the range of feasible parameters.

The fourth method of estimation is the probability weighted moments (PWM) estimator examined by Hosking et al (1985). The PWM are defined by

$$M_{r,s,t}=E\left[x^r\{F(x)\}^s\{1-F(x)\}^t\right] \quad ;r,s,t=1,2,3...$$

Where r,s,and t are chosen to give a sufficient number of equations to solve for the model parameters. The solution procedure is similar to the ordinary moments method in that the expected moments are compared with the sample moments. At first sight the additional complexity of this estimation process might seem unwarranted, however, the behaviour of PWM for the Generalised Extreme Value distribution is shown to be superior to the ordinary method of moments in Hosking et al (1985) for cases where the sample has a high COV.

The fifth method is entropy estimation, Jaynes (1982). This method uses entropy as a measure of the amount of information in a sample and has proven popular with physicists. Given a continuous distribution $f(x)$ the entropy is defined as

$$H(\theta;x)=\int_{-\infty}^{\infty} f(x;\theta)ln\{f(x;\theta)\}dx$$

The solution for the model's parameters θ is then found by maximising the entropy over the feasible region on the sample of data x.

3.2.1 CHOICE OF ESTIMATION PROCEDURE

Five methods for estimating the parameters of a model have been discussed briefly. Amongst engineers least squares and moment methods are favoured for their simplicity, robustness and ease of implementation, moreover, both methods generally provide good model fits to both large and small samples. Statisticians on the other hand favour the method of maximum likelihood and scientists often use entropy methods which were developed for statistical physics applications. In order to choose the best method for modelling the design events of ships and offshore structures it is necessary to define what is meant by best. In Section [3.1] six desirable characteristics are defined for an estimation procedure:

- consistency
- sufficiency
- low bias
- low sampling variance
- reliability
- simplicity

These six characteristics can be used to assess the relative performance of an estimator, but the process is not straightforward since low bias can be less desirable than low sampling variance. This is because we are generally interested in estimating the quantiles of a random variable (that is, values which correspond to a particular probability level or return period) and in general the quantiles will be biased, even when the parameters are not. This is dealt with in more detail in Section [3.4.2] which shows how confidence limits can be estimated for both a model's parameters and arbitrary, non-linear functions of the model's parameters.

Monte Carlo simulation can be used to assess the parameter bias and sampling error associated with an estimation procedure by generating a large number of samples from a predefined model with known parameters. Singh et al (1990) present the results of their extensive computer experiments using a two parameter Weibull model. They simulated seven Weibull populations having a range of COV from 0.3 to 3.0. For each case they used the simple inversion

$$x(F) = \alpha\left[-log_e(1 - P(X < x))\right]^{\frac{1}{c}}$$

to create a large number of random samples with sizes 10, 20, 30, 50, 75, 100, 500, 1000. Each sample was then used to determine the Weibull shape and scale parameters using five different estimation methods:

• moments	MOM
• maximum likelihood	ML
• entropy	ENT
• probability weighted moments	PWM
• least squares	LSQ

The results from each method were used to determine the bias in each parameter and their root mean square errors (RMSE). Furthermore, they calculated the bias and RMSE for a range of quantiles corresponding to low probability events like annual maxima for winds and currents. A summary of their major conclusions follows

Bias in shape, and scale parameters - ML, LSQ, and ENT performed consistently for each case and sample size when estimating the shape parameter - though it may be significant LSQ had a negative bias. MOM did not perform well and PWM performed very poorly for the small sample COV cases giving a large negative bias. On the other hand PWM performed very well

for data having high COV (this suggests the PWM method would be a poor estimator for extremes which generally have low COV)

Bias in quantile estimates - As expected, bias in the quantile estimates reduces as the size of sample increases, and for a given sample size the bias increases as the quantile (or equivalent return period) increases. ML and ENT both perform well though they have a negative bias for the small COV cases, this would result in an under-estimate for the value of a random variable corresponding to some return period. MOM performed the best for the sample with the lowest COV suggesting it is well suited to the estimation of extremes in samples with COV's of less than 0.3. In general, all methods performed better with the data having larger COV's.

RMSE of shape, and scale parameters - As before both ML and ENT perform well, however, the best result for the shape parameter of the smallest sample was obtained with LSQ. The best estimator for the scale parameter was always ML. MOM performed poorly for small samples over the whole range of COV's; furthermore, the error increased significantly as the size of sample increased. All five methods showed a deterioration in their estimates of the shape parameter as the sample COV increased.

RMSE of quantile estimates - MOM and ML performed best over the whole range of COV's especially for the small samples. PWM and LSQ performed poorly in all cases with their estimates particularly poor for the larger COV's.

Based on the above summary of Singh's results we can conclude both ML and ENT are the *best* estimators of the Weibull parameters if the statisticians criterion are taken. However, it is worth noting MOM gave good unbiased results for the extreme quantiles, albeit with the highest sampling error. In the remainder of this work the ML method is adopted as the primary estimation method. At first sight this might seem at odds with practicality since the maximum likelihood method requires the derivation of complex likelihoods which must then be optimised using multivariate function minimisation routines. A reasonable question is then why not use a simpler estimator for the problem at hand, for example, the method of moments for fitting a model to small samples with low COV. Previously this argument has won in engineering and, as Crowder et al (1991) point out, in consequence the statistical analysis performed by engineers has fallen behind the methods adopted by for example the medical and clinical sciences. The principal advantage of maximum likelihood over the simpler alternatives is the rigourous formalisation of likelihood estimation, and its wide degree of application. By adopting the method we avoid having to use various *ad hoc* estimators like weighted least squares to determine the 'optimal' parameters. In addition, the method provides a general procedure which allows the statistical uncertainty of functions of the

models parameters to be found using the information matrix. This avoids having to perform time consuming and expensive Monte Carlo simulations to determine the levels of uncertainty in a statistic derived from a fitted model.

Before discussing likelihood estimation in detail a brief introduction to kernel density estimation will be given to highlight some of the problems with parametric estimation and demonstrate the differences between parametric and non-parametric methods of density estimation. Non-parametric kernel methods have been popular in the medical and clinical sciences for many years since for large samples they can provide accurate, low bias estimates of the density without pre-supposing a parametric form. The advantage of this method is that it enables the presence of multi-modality to be identified, for example when analysing data which result from the mixing of more than one random process. For our purposes, this might be of use when analysing wave zero-up-crossing period populations which result from the mixing of swell and wind driven seas.

3.3 KERNEL DENSITY ESTIMATORS: NON PARAMETRIC ESTIMATION

In this section, three procedures which allow data to be explored graphically are briefly examined. These procedures can be used subjectively to identify the nature of the distribution function underlying a sample of data with distribution function $f(x)$. Engineers are familiar with histogram and quantile plots as methods for examining the distribution of a sample $x_i; i = 1,....,n$, but these methods can only give a subjective measure of the central tendency, skewness and kurtosis for the sample. Furthermore, neither histograms nor quantile plots can be used reliably to visually identify the family of underlying distribution functions when the sample is small (<50). Normally, a quantile plot is more informative than a histogram, however, because it is a plot of probability level and not density against variate it is difficult, without experience, to identify the underlying distribution function.

For this work it is useful to determine non-parametric estimates for the density function which can be used to select the correct model for use in a parametric study. Rosenblatt (1971) shows that for large samples the kernel density estimator gives low bias and low variance estimates of density. Kernel estimators are used primarily as graphical tools, however, they can also be used to determine robust estimates of density for use in, for example, hazard rate estimation.

Perhaps the simplest estimator is the Nearest Neighbour Estimate for which

$$\hat{f}(x)=\frac{\hat{F}(x_i)-\hat{F}(x_{i-1})}{\{x_i-x_{i-1}\}}=\frac{1/n}{\{x_i-x_{i-1}\}} \quad iff \quad x_{i-1}<x<x_i$$

This estimator is of little use for practical applications however it does illustrate how the density $\hat{f}(x)$ can be estimated from the samples. The kernel density estimator uses a similar principal but instead of using only two 'adjacent' samples it uses a weighted average over the samples within the bandwidth,or window, with width b.

The two principal advantages of kernel estimators are that they assume no particular functional form for the density and that their behaviour is well defined mathematically in the large sample case, Silverman (1986). A further advantage of the method is that it can be extended to the multivariate case (see Ch.4) to explore the underlying structure of joint probability distributions. The general form of the univariate kernel estimator is

$$\hat{f}(x) = \frac{1}{nb}\sum_{i=1}^{n} k\left\{\frac{x-X_i}{b}\right\} \qquad [3.5]$$

where n is the number of samples, b is the bandwidth for the smoothing, and $k(.)$ is a kernel function which must satisfy

$$\int_{-\infty}^{\infty} k(x)\,dx=1; \qquad k(x)\le 1; \qquad \int_{-\infty}^{\infty} x\ k(x)\ dx=0$$

The kernel is therefore a density function with zero mean. Several types of kernel have been proposed but most users agree that the choice of kernel is less important than the selection of an optimal bandwidth, consequently for this work we use the well known Epanechnikov (1969) kernel for which

$$k(y)=\tfrac{3}{4}(1-y^2) \quad ,|y|\le 1 \quad ; \quad k(y)=0, \ otherwise \qquad [3.6]$$

This kernel has been studied extensively and simulation shows that it provides asymptotically unbiased estimates with low variance given the correct choice of bandwidth. The optimal smoothing width is dependent on sample size and sample standard deviation, as a rough guide it should be taken as $b=2\sigma n^{-1/5}$, however, if a more accurate result is needed, for example when using the estimate in hazard rate calculations, the method of maximum likelihood should be used. It is worth noting however that the results may be biased and noisy when the distribution is long tailed, in such cases adaptive kernel estimation (in which the bandwidth varies) should be used.

Applications for these procedures using populations of wind wave and current data were given in Chapter 2 which examines the environmental time series recorded by the IOS DB1 data buoy. To demonstrate the method, a sample of 1000 Weibull distributed random numbers with shape, scale and location $\theta = \{2.0, 1.463, 10.16\}$ is generated and then examined using the fixed bandwidth, Epanechnikov, kernel estimator outlined above. The results are shown in Fig. 3.2 which shows the population model plotted against two kernel estimates each having different bandwidth. Overall, the optimal window width kernel estimates agree well with the population model with good fit in the body of the distribution.

3.4 MAXIMUM LIKELIHOOD ESTIMATION

The likelihood random variable for the vector of observed values $X = \{X_1, ..., X_n\}$ is defined as the joint density for the observation. Assuming the observations are *iid* then the joint density is

$$f(X) = f(X_1; \theta) \times f(X_2; \theta) \times ..... \times f(X_n; \theta)$$

the likelihood random variable is then defined as

$$L(\theta ; X) = \prod_{i=1}^{n} f(X_i; \theta)$$

For a sample of *iid* data the sample likelihood is defined by

$$\ell(\theta ; x) = \prod_{i=1}^{n} f(x_i; \theta)$$

where the lowercase ℓ denotes the quantity is the *sample* likelihood, that is an observation on the random variable L. The sample likelihood is simply the product of the density for each of the samples with the model parameters as the unknowns. The principal of the maximum likelihood method is that the 'best explanation' for the sample of data x is provided by the value $\hat{\theta} \in \Omega$ which maximises the the sample likelihood $\ell(\theta ; x)$, that is

$$\ell\left(\hat{\theta} ; x \right) \geq \ell(\Omega ; x)$$

In real applications there are theoretical and practical advantages gained from working with the natural logarithm of the likelihood, in fact it is the log-likelihood which is fundamental

to the derivation of the parameter covariance matrix. The solution for the most likely set of parameters $\hat{\theta}$, for the chosen distribution function $f(x;\theta)$, is obtained by maximising θ on the parameter space Ω. For the case when the parameters of the model are unbounded the solution $\hat{\theta}$, is then defined by the turning points given by

$$\frac{\partial \ell(\theta;x)}{\partial \theta_j} = 0 \qquad j = 1,...,p$$

where, p is the number of parameters in θ; for example, six in the case of a bivariate Normal and three in the case of the univariate Weibull. The solution for $\hat{\theta}$ will usually involve a non-linear minimization of the negative likelihood $-\ell(\theta;x)$, since

$$max[\ell(\theta\,;x)] \equiv min[-\ell(\theta\,;x)]$$

However, in practice, the solution for $\hat{\theta}$ is sought using the natural logarithm of the likelihood since the log-likelihood is used in the calculation of model parameter uncertainty. In this work the *Numerical Algorithms Group* (NAg) gradient based, minimisation routine E04VDF was chosen for the optimisation. This routine has proven ideal since it is capable of handling large numbers of model parameters, and can deal with non-linear, and linear inequality constraints. The initial position of the search for the maximum must be chosen with care to avoid the location of local and not global minima. However, the starting value of the parameters can generally be estimated by graphical inspection of the marginal distribution.

3.4.1 THE INFORMATION MATRIX

Each parameter vector θ estimated using a finite realization of a random vector $X = \{X_1,.....,X_n\}$, will itself be a random vector drawn from a sampling distribution. For large samples it is well known that maximum likelihood estimates are asymptotically multivariate normal distributed with minimum bias and accuracy not far from the theoretical optimum. Furthermore, under certain conditions on the joint density, and assuming the estimator is unbiased, the symmetric pxp Hessian matrix of the negative log-likelihood, called the information matrix, is related to the sampling distribution covariance matrix C where

$$C = (I(\theta))^{-1} \qquad [3.7]$$

that is

$$C = \begin{bmatrix} -E\left[\frac{\partial^2 \log_e L}{\partial\theta_1\partial\theta_1}\right], \ldots\ldots, -E\left[\frac{\partial^2 \log_e L}{\partial\theta_1\partial\theta_p}\right] \\ \cdot \qquad\qquad \cdot \\ -E\left[\frac{\partial^2 \log_e L}{\partial\theta_p\partial\theta_1}\right], \ldots\ldots, -E\left[\frac{\partial^2 \log_e L}{\partial\theta_p\partial\theta_p}\right] \end{bmatrix}^{-1} \qquad [3.8]$$

The proof of this important result, is given in Ledermann (1984). In practice, the expected values of each second derivative in Equation [3.8] can be difficult to determine and in this work the covariance matrix is approximated by the observed information matrix

$$c = \begin{bmatrix} -\frac{\partial^2 \log_e \ell}{\partial\theta_1\partial\theta_1} & \ldots\ldots & -\frac{\partial^2 \log_e \ell}{\partial\theta_1\partial\theta_p} \\ \cdot & \cdot & \cdot \\ -\frac{\partial^2 \log_e \ell}{\partial\theta_p\partial\theta_1} & \ldots\ldots & -\frac{\partial^2 \log_e \ell}{\partial\theta_p\partial\theta_p} \end{bmatrix}^{-1} \qquad [3.9]$$

The joint sampling distribution of the random vector θ, is then multivariate normal, with zero expectation vector, and covariance matrix given by the second derivatives of the log-likelihood function. The asymptotic form for the distribution of the maximum likelihood set of parameters is then approximated by

$$f(\theta) = \frac{1}{(2\pi)^{p/2}\sqrt{|c|}} exp\left\{-\tfrac{1}{2}\ \theta^T\ c^{-1}\ \theta\right\} \qquad [3.10]$$

A knowledge of $f(\theta)$ enables the confidence regions for the parameters to be constructed. This information is important since for design purposes it means we can identify characteristic upper and lower bounds on the form of the distribution function and hence allow for the uncertainty in the maximum likelihood model. Of course we often have small sample sizes (<50) in which case the assumption of asymptotic normality will be approximate, or may not hold.

3.4.2 CONFIDENCE INTERVALS AND REGIONS

There are two possible ways of using the model parameter covariance matrix. In the simplest case each parameter θ_i is considered independently of the other model parameters, which are assumed to be deterministic. We have seen that the model parameters are approximately multivariate normal, and consequently the random variable θ_i is approximately normal with zero expectation, and variance approximated by

$$v[\theta_i] \approx -\left[\frac{\partial^2 \log_e L}{\partial \theta_i^2}\right]^{-1} \qquad [3.11]$$

using the properties of the standard Normal distribution confidence intervals can be calculated for the individual model parameters, for example, the approximate 95% confidence limits for θ_i are

$$\theta_i - 1.96\sigma(\theta_i) \leftrightarrow \theta_i + 1.96\sigma(\theta_i) \qquad [3.12]$$

In practice, the single parameter case is of little use other than to indicate the degree of uncertainty in the parameter, thereby indicating the goodness of fit. For the two parameter case the confidence limits become an ellipse and therefore rather more difficult to compare between models and fits. Ultimately, we are interested in the uncertainty in some function of the parameters like the quantiles for a given design return period. In this case the contribution of each parameter to the uncertainty in the quantity of interest must be allowed for using a method like the Rao (1973) delta method.

3.4.3 FUNCTIONS OF MLE PARAMETERS

In most cases, we are interested in functions of the maximum likelihood estimates of a model's parameters. We have seen the model parameters can be assumed to be approximately multivariate Normal, with zero mean vector, and variance-covariance matrix determined by the second derivatives of the log-likelihood. Assuming asymptotic normality holds for small samples we can generally apply the delta method, Rao *op cit*, to determine the mean value and variance for functions of the model parameters. One obvious class of functions which are of interest are the quantiles x_P corresponding to a design life with survival probability P.

Assuming $\Phi(\boldsymbol{\theta}) = x_P(\boldsymbol{\theta})$, then the mean value of the quantile x_P is given by substituting the maximum likelihood set of parameters into the function $\Phi(\boldsymbol{\theta})$ i.e. $\Phi\left(\hat{\boldsymbol{\theta}}\right) = x_P\left(\hat{\boldsymbol{\theta}}\right) = \hat{x}_P$. The variance is given by

$$v[\hat{x}_P] = v\left[F^{-1}(P)\right]$$

Providing the inverse distribution function F^{-1} can be found the variance of a quantile can be approximated by assuming the parameters are multivariate normal. In practice, this enables the mean value of a design parameter to be determined for a given level of probability or return period, together with an estimate of its statistical uncertainty. This information is of considerable use since it could directly be used to assess the magnitude of a partial safety factor required to give a defined level of reliability.

In the simplest case the quantile is a simple linear function of the model parameters $\boldsymbol{\theta}$ with coefficients a_i

$$x_P = \sum_{i=1}^{p} a_i \theta_i$$

the mean is then given by

$$E[x_P] = \sum_{i=1}^{p} a_i E[\theta_i]$$

and the variance is

$$v[x_P] = \sum_{i=1}^{p} a_i v[\theta_i] + \sum_{i=1}^{p} \sum_{j=1}^{p} a_i \, cov[\theta_i, \theta_j] \qquad [3.13]$$

One example of the simple linear case which is of importance to the design of ships and offshore structures is the Gumbel model, $F(x) = exp\left(-exp-\left(\frac{x-\mu}{\sigma}\right)\right)$, which has quantiles

$$x_P = \mu - \sigma log(-log P) \qquad [3.14]$$

with mean $\hat{x}_P$ and variance

$$\left\{1 + 6/\pi^2 \left(1 - \gamma - log(-log P)\right)^2\right\} \frac{\hat{\sigma}^2}{n} \qquad [3.15]$$

In this case maximum likelihood estimates of the mean quantile, and its variance, can be calculated very simply using the estimates $\hat{\mu}$, $\hat{\sigma}$.

If $\Phi(\boldsymbol{\theta})$ is not a simple linear function of the model parameters but a non-linear, continuous, differentiable function of the probability model parameters $\boldsymbol{\theta}$ then the mean and variance of $\Phi(\boldsymbol{\theta})$ can be found to a first approximation using the Taylor expansion

$$\Phi(\boldsymbol{\theta}) = \Phi(\hat{\boldsymbol{\theta}}) + \sum_i \left(\theta_i - \hat{\theta}_i\right) \left[\frac{\partial \Phi}{\partial \theta_i}\right]$$

The mean value of the function Φ is then approximated by

$$E[\Phi(\boldsymbol{\theta})] = \Phi(\hat{\boldsymbol{\theta}}) + \sum_i \left(\theta_i - \hat{\theta}_i\right) \left[\frac{\partial \Phi}{\partial \theta_i}\right] = \Phi(\hat{\boldsymbol{\theta}})$$

that is the mean value is given by substituting the maximum likelihood parameters in to the function $\Phi\left(\hat{\theta}\right)$. The variance of the function of the model parameters is given by

$$v[\Phi(\theta)] = \sum_i v\left[\hat{\theta}_i\right]\left[\frac{\partial \Phi}{\partial \hat{\theta}_i}\right]^2 + \sum_i \sum_j cov\left[\hat{\theta}_i, \hat{\theta}_j\right]\left[\frac{\partial \Phi}{\partial \hat{\theta}_i}\right]\left[\frac{\partial \Phi}{\partial \hat{\theta}_j}\right] \qquad [3.16]$$

This equation can be written in matrix notation giving a more compact form which is easily programmed using standard library routines

$$v[\Phi(\theta)] = G^T(\Phi)\, C\, G(\Phi) \qquad [3.17]$$

Where C is the parameter covariance matrix, and $G(\Phi)$ is the function gradient vector. In the simplest case, the model has a single parameter θ and the variance of the function $\Phi(\theta)$ at the maximum likelihood point is

$$v[\Phi(\hat{\theta})] = v[\hat{\theta}]\left[\frac{\partial \Phi}{\partial \hat{\theta}}\right]^2 \qquad [3.18]$$

For the two parameter case we use the summation Equation [3.16] to obtain

$$v[\Phi(\theta)] = v[\hat{\theta}_1]\left[\frac{\partial \Phi}{\partial \hat{\theta}_1}\right]^2 + 2cov\left[\hat{\theta}_1, \hat{\theta}_2\right]\left[\frac{\partial \Phi}{\partial \hat{\theta}_1}\right]\left[\frac{\partial \Phi}{\partial \hat{\theta}_2}\right] + v[\hat{\theta}_2]\left[\frac{\partial \Phi}{\partial \hat{\theta}_2}\right]^2 \qquad [3.19]$$

Some applications of these techniques are given later using the Weibull and GEV models developed in Sections [3.6&3.7].

3.5 MODELLING PROCEDURE: APPLIED MAXIMUM LIKELIHOOD

A summary of the most common univariate distributions is given in Lewis & Orav (1989) and a more complete discussion is given in the Johnson & Kotz (1970). For multivariate distributions see Johnson & Kotz (1972). The procedure for modelling a distribution using maximum likelihood involves eight primary steps:

1 graphical examination of the marginal distributions and scatter plots
2 checks for the *iid* assumption using autocorrelation of the time series
3 hypothesis of the distribution function for the data

4 formulation of the sample log-likelihood for the chosen model
5 maximization of the sample log-likelihood
6 calculation of the information matrix if the problem is regular
7 estimation of the confidence regions if possible
8 assessment of the uncertainty in the quantity of interest to the designer

If the random variable can be transformed into a normal variate then the problem of estimation is considerably simplified. However, if this is not the case then prior to statistical inference it is often advantageous to use a simple transformation which produces a zero mean process, with normalised variance

$$y_j = \frac{(x_j - E[x])}{\sqrt{v[x]}} \qquad [3.20]$$

This transformation to random variable y_j scales the data and in some cases improves the numerical conditioning of the log-likelihood, thereby improving the solution obtained using the optimisation. The probability density for the random variable $X, f(x)$ is related to the probability density for $Y, f(y)$ through the simple relationship

$$f_Y = \left|\frac{dx}{dy}\right| f_X \quad = \quad \sigma_x \, f_X(x)$$

This is simply a probability preserving transformation such that $f_Y |dy| = f_X |dx|$.

3.6 THE WEIBULL DISTRIBUTION

A popular generalisation of the exponential distribution for extreme value and reliability calculations is the Weibull (1951) distribution. The popularity of the Weibull model stems from the flexibility of the scaling and shape transformation

$$x = \left\{\frac{z - \xi_0}{\alpha}\right\}^c$$

Before developing the Weibull sample log-likelihood functions it will be useful to carefully examine the properties of the Weibull distribution. The cumulative distribution function and density function are given by

$$F(x;\theta) = 1 - exp\left\{-\left[\frac{(x-\xi_0)}{\alpha}\right]^c\right\} \quad [3.21]$$

and

$$f(x;\theta) = \frac{c}{\alpha}\left\{\frac{x-\xi_0}{\alpha}\right\}^{c-1} exp\left\{-\left[\frac{(x-\xi_0)}{\alpha}\right]^c\right\} \quad [3.22]$$

where, the bounds on the model shape, scale and lower bound parameters are given by

$$x > \xi_0;\ 0 < \alpha < \infty;\ 0 < c < \infty$$

Inverting the distribution we get the level of variate corresponding to survival probability P_s

$$x_s = \alpha log\left\{\frac{1}{1-P_s}\right\}^{1/c} + \xi \quad [3.23]$$

This relationship enables us to calculate the return period values for the DB1 wind speed, significant wave height, and current speed. It also serves as a check on the fit of the marginal models in the extremes.

The first four moments of are given by the gamma functions

$$E[x] = \xi_0 + \alpha\Gamma\left\{1+\frac{1}{c}\right\} \quad [3.24]$$

$$var[x] = \alpha^2\left[\Gamma\left\{1+\frac{2}{c}\right\} - \Gamma^2\left\{1+\frac{1}{c}\right\}\right] \quad [3.25]$$

$$E\left[(x-\bar{x})^3\right] = \alpha^3\left[\Gamma\left\{1+\frac{3}{c}\right\} - 3\Gamma\left\{1+\frac{1}{c}\right\}\Gamma\left\{1+\frac{2}{c}\right\} + 2\Gamma^3\left\{1+\frac{1}{c}\right\}\right] \quad [3.26]$$

$$E\left[(x-\bar{x})^4\right] = \alpha^4\left[\Gamma\left\{1+\frac{4}{c}\right\} - 4\Gamma\left\{1+\frac{1}{c}\right\}\Gamma\left\{1+\frac{3}{c}\right\} + 6\Gamma^2\left\{1+\frac{1}{c}\right\}\Gamma\left\{1+\frac{2}{c}\right\} - 3\Gamma^4\left\{1+\frac{1}{c}\right\}\right] \quad [3.27]$$

The lower limit on the range of the variate results in the distribution commonly being referred to as the Type III lower bound distribution since its form corresponds to the third Fisher Tippet distribution when X is replaced by $-X$. The physical interpretation of the lower bound is that the variate cannot be less than some threshold value ξ_0 which very often can be taken as 0.0,

in which case estimation of the model parameters is relatively simple since only two unknowns exist.

The three parameters in the Weibull distribution each control a different aspect of its shape. The 'c' value is a shape parameter which influences the skewness and peakedness, and the alpha parameter is the scale factor which controls the spread along the variate axis. When c = 2 and $\xi_0 = 0$ the distribution becomes the Rayleigh distribution, and when c=3.25 the Weibull distribution is very nearly distributed as a Normal variate, Dubey (1967). Plots of the Weibull distribution and density for a range of integer shape parameters from 1 (exponential) to 4 are given in Fig. 3.3.

3.6.1 WEIBULL LIKELIHOOD FUNCTION

The sample log-likelihood function for the three parameter Weibull distribution is given by

$$log_e(\ell(\boldsymbol{\theta};\boldsymbol{x})) = \sum_{i=1}^{n} log_e\left(f_{X_1,\ldots X_n}(x_1,\ldots x_n)\right)$$

which on algebraic manipulation gives

$$log(\ell(\theta;x)) = nlog_e c - cnlog_e \alpha + (c-1)\sum_{i=1}^{n} log_e(x_i - \xi_0) - \frac{1}{\alpha^c}\sum_{i=1}^{n}(x_i - \xi_0)^c \qquad [3.28]$$

For the two parameter Weibull (i.e. $\xi_0 = 0$) the solution for the maximum likelihood set of parameters $\hat{\theta} = \{\hat{c}, \hat{\alpha}\}$ is then given by

$$\frac{\partial l(\theta;x)}{\partial \theta_r} = 0 \quad ; \quad r = 1,2$$

Solving for the first derivatives of the log-likelihood gives two non-linear simultaneous equations which are a function of the unknown parameters $\hat{\theta}$ and the sample $\boldsymbol{x}$:

$$\frac{\partial log_e \ell}{\partial \alpha} = \frac{c}{\alpha^{c+1}}\sum_{i=1}^{n}(x_i - \xi_0)^c - \frac{cn}{\alpha} = 0 \qquad [3.29]$$

$$\frac{\partial log_e \ell}{\partial c} = n\left\{\frac{1}{c} - log_e \alpha\right\} + \sum_{i=1}^{n} log_e(x_i - \xi_0) + \frac{log_e \alpha}{\alpha^c}\sum_{j=1}^{n}(x_j - \xi_0)^c - \frac{1}{\alpha^c}\sum_{i=1}^{n}(x_i - \xi_0)^c log_e(x_i - \xi_0) \qquad [3.30]$$

Solving this problem is relatively simple when compared to the three parameter case since an iterative scheme can easily be implemented to search for the solution of the two equations. Unfortunately, for the three parameter case the first derivative of the log-likelihood with respect to the lower bound is not zero

$$\frac{\partial log_e \ell}{\partial \xi_0} = -(c-1)\sum_{i=1}^{n}\frac{1}{(x_i - \xi_0)} + \frac{c}{\alpha^c}\sum_{i=1}^{n}(x_i - \xi_0)^{c-1} \neq 0 \qquad [3.31]$$

owing to the bound on the parameter ξ_0. Solving Equations [3.29 & 3.30 & 3.31] therefore requires classical, constrained optimisation procedures to minimise the negative log-likelihood, i.e.

$$max[log_e \ell(\theta; x)] \equiv min[-log_e \ell(\theta; x)]$$

In this work, the Numerical Algorithms Group (NAg) Fortran subroutine E04AJF was adopted. This routine uses a double precision, 'downhill', finite differencing algorithm to solve problems of the form

$$min\{F(x)\} \quad subject\ to \quad l < \{x\} < u$$

where, $F(x)$ is a smooth nonlinear function, x is the vector of function variables, and l, u are the vectors of upper and lower bounds on the model parameters.

3.7 THE GENERALISED EXTREME VALUE DISTRIBUTION

The three Fisher Tippet distributions can be generalised as a single equation usually referred to as the Generalised Extreme Value (GEV) distribution, Jenkinson (1955) . The cumulative distribution function for the GEV is

$$G(x; \mu, \sigma, k) = exp\left[-\left\{1 - k\left(\frac{x-\mu}{\sigma}\right)\right\}^{\frac{1}{k}}\right] \qquad [3.32]$$

where the scale coefficient $\sigma > 0$, and the range of x is determined by the inequality $1 - k(x-\mu)/\sigma > 0$. Inverting the distribution we get the level of variate corresponding to survival probability P_s

$$x_s = \mu + \frac{\sigma}{k}\left(1 - \left(-log\{P_s\}\right)^k\right) \qquad [3.33]$$

The three fundamental forms of the GEV are controlled by the shape parameter k which can either be zero, negative or positive:

$k \to 0$ **Fisher Tippett Type I** - in the limit as the shape parameter tends to zero the GEV becomes the Gumbel (or double exponential) distribution with unbounded variate, and distribution function given by

$$G_I(x;\mu,\sigma) = exp\left[-exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\} \right] \qquad [3.34]$$

$k < 0$: **Fisher Tippett Type II** - when the shape parameter is negative the GEV has a Frechet distribution with lower bound on the range of x defined by the equality $x_B = \sigma/k + \mu$

$k > 0$: **Fisher Tippett Type III** - when the shape parameter is positive the GEV has a upper bound on the range of x defined by $x_B = \sigma/k + \mu$

The form of the GEV distribution for $-x$ with positive shape coefficient is Weibull with the shape, scale and lower bound coefficients given by $c = 1/k$; $\alpha = \sigma/k$; $\xi_0 = \sigma/k + \mu$. For comparison, the three forms of the GEV are plotted in Fig. 3.4 using equal scale and location parameters.

3.7.1 THE GEV LIKELIHOOD FUNCTION

Estimation of the parameters of the GEV distribution is discussed in Prescott & Walden (1980) in relation to the estimation of extremes in hydrological data. They use an iterative scheme involving the Fisher information matrix to derive the maximum likelihood estimates for the shape, scale and location parameters. In this work, the problem is solved more directly by minimising the negative log-likelihood using a non-linear function minimisation routine. This approach was adopted successfully by Tawn (1988a) who examines in detail the fitting of a GEV model to data having non-stationary mean. Differentiation of Equation [3.32] gives the GEV probability density function

$$f(x)=\frac{1}{\sigma}\left[1-k(x-\mu)/\sigma\right]^{\frac{1}{k}-1}exp\left\{-\left[1-k(x-\mu)/\sigma\right]^{\frac{1}{k}}\right\} \qquad [3.35]$$

In the normal manner, the natural log of the sample likelihood is given by

$$log_e\left(\ell(\theta;x)\right)=\sum_{i=1}^{n}log_e\left(f_{x_1,\ldots x_n}(x_1,\ldots x_n)\right)$$

whereupon making the substitution for the density of each sample we get

$$log_e\left(\ell(\theta;x)\right)=-nlog_e\,\sigma+$$
$$\left(\frac{1}{k}-1\right)\sum_{i=1}^{n}log_e\left\{\left[1-k(x_i-\mu)/\sigma\right]\right\}-\sum_{i=1}^{n}\left[1-k(x_i-\mu)/\sigma\right]^{\frac{1}{k}} \qquad [3.36]$$

The maximum likelihood solution $\hat{\theta}=\left\{\hat{\sigma},\hat{k},\hat{\mu}\right\}$ is then found by minimising the negative sample log-likelihood, subject to the linear and non-linear constraints

$$0<\sigma<\infty$$
$$-\infty<k<1$$
$$1-k(x-\mu)/\sigma>0$$

The first two constraints are simple bounds on the range of the parameters. The upper limit on the shape parameter k corresponds to a limit on the so called regularity space for the log likelihood, that is no MLE exists for values of k greater than unity. During the initial stages of this work it was assumed that the FT III was totally equivalent to the upper bounded form of the Weibull distribution, however, the limit on k means the GEV cannot model Weibull populations for which the shape parameter c is less than 1.0. This suggests both the GEV and Weibull models should both be fitted to a sample to test which model fits the data best.

The third constraint makes the GEV more difficult to solve than the individual Fisher Tippet distributions since it is non-linear in the scale parameter. Finding the maximum likelihood parameters therefore requires minimisation of a non-linear function subject to both bounds on the parameters and non-linear constraints. When programming the solution of the GEV maximum likelihood parameters two non-linear constraints are required to allow for all feasible values of the variate

$$1 - k(x_{min} - \mu) / \sigma > 0$$
$$1 - k(x_{max} - \mu) / \sigma > 0$$

where, x_{min} and x_{max} are the smallest and largest values in the sample. The first constraint is necessary when the current iteration has a negative value for k corresponding to a lower bounded variate with Frechet distribution. The second constraint is necessary when the current step has a positive value for k in which case the variate has an upper bound with FT III distribution.

When solving the MLE's for the GEV it was not possible to use the same NAg function minimisation routine that was used with the Weibull distribution since it is not capable of dealing with the non-linear constraints. It was therefore necessary to use the more general minimisation routine, E04VDF, which solves problems having the form

$$min\{F(x)\} \qquad subject\ to \quad u < \begin{Bmatrix} x \\ Ax \\ c(x) \end{Bmatrix} < l$$

Where:

	$F(x)$	is a smooth nonlinear function;
	x	is the vector of function variables;
	(u,l)	are the vectors of upper and lower bounds on the range, linear and non-linear constraints;
	Ax	are the linear constraint equations;
	$c(x)$	are the non-linear constraint equations.

NAg routine E04VDF requires analytical expressions for the derivatives in the log likelihood gradient vector $G(\hat{\theta})$ in order to perform the minimisation

$$G(\hat{\theta}) = \left\{ \frac{\partial \log_e l}{\partial \sigma}, \frac{\partial \log_e l}{\partial k}, \frac{\partial \log_e l}{\partial \mu} \right\}^T$$

This gradient vector is used by the routine to determine the next step length and direction, each term (given by the first partial derivative of the likelihood w.r.t. the model parameters) is given in Appendix A.2.

Experience with the GEV distribution shows the correct choice of starting point is essential for the optimisation to converge to the maximum likelihood point. In this work the moments estimator for the Gumbel scale and location parameter are chosen as the starting point, with a shape factor set to 0.1. The Gumbel model has a mean and variance given by

$$mean = \mu + \gamma\sigma$$
$$variance = \frac{1}{6}\pi^2\sigma^2 \qquad [3.37]$$

where γ is Euler's constant. Equating these to the sample moments we get the solutions

$$\mu^* = \left\{\frac{1}{N}\sum_{i=1}^{N} x_i\right\} - 0.58\sqrt{\frac{6}{\pi^2 N}\sum_{i=1}^{N}(x_i - E[x])^2} \qquad [3.38]$$

$$\sigma^* = \sqrt{\frac{6}{\pi^2 N}\sum_{i=1}^{N}(x_i - E[x])^2} \qquad [3.39]$$

$$k^* = 0.1 \qquad [3.40]$$

Using these values usually ensures that the correct maximum is found: however, in some cases the solution for k is zero indicating the domain of attraction is the Gumbel model. In this case, the $1/k$ terms in the likelihood and its derivatives tend to infinity and a trap is required to prevent numerical ill-conditioning. In such cases the results in Tiago de Oliveira (1989a) are useful since he shows the location and scale parameters are bivariate normally distributed with:

$$E[\mu] = \hat{\mu}$$

$$E[\sigma] = \hat{\sigma}$$

$$E\left[(\mu - E[\mu])^2\right] = \left(1 + \frac{6}{\pi^2}(1-\gamma)^2\right)\frac{\sigma^2}{N}$$

$$E\left[(\sigma - E[\sigma])^2\right] = \frac{6}{\pi^2}\frac{\sigma^2}{N}$$

$$E[(\mu - E[\mu])(\sigma - E[\sigma])] = \frac{6}{\pi^2}\frac{\sigma^2}{N}(1-\gamma)$$

where $\gamma = 0.5772....$ is Euler's constant, and the asymptotic correlation coefficient is

$$\rho = \left(1 + \frac{\pi^2}{6(1-\gamma)^2}\right)^{-1/2} = 0.313$$

These results can be compared with the covariance matrix estimated either analytically or numerically from the information matrix (Appendix A.2).

3.8 ESTIMATION OF THE COVARIANCE MATRIX: NUMERICAL PROBLEMS

Confidence intervals, and standard errors can be estimated using the Fisher information matrix, which is constructed from the second derivatives of the log-likelihood with respect to the model parameters. In most cases, the equations for the second derivatives (which are given in, Appendix A) were cumbersome and prone to error in their derivation and implementation in FORTRAN, consequently a check was made on each result by calculating the derivatives using finite differences. Initially, it was hoped that finite difference solutions would be sufficiently accurate for the method to be used routinely . However, in practice, a problem was encountered in a few instances. The gradients at the maximum likelihood point are, rightly, near zero and this caused a number of numerical problems because the central and forward difference estimates calculated by the NAg routine were not sufficiently close to allow the routine to accept the results. To overcome this problem the gradients and Hessian were calculated at a point very close to, but not at, the maximum likelihood point

$$\theta' = F\hat{\theta} \quad ; \quad F \geq 1.0$$

The size of the perturbation factor F varied with model type and sample of data but generally a value of 1.0001 was found to be acceptable. The introduction of the factor F gave biased estimates for the gradient vector and the Hessian matrix which are of course already biased by the error residuals $O(.)^2$ of the Taylor approximation.

A further problem was encountered when calculating the parameter uncertainty. In some instances, negative estimates appeared on the leading diagonal of the covariance matrix, which is clearly incorrect since only the covariance terms can be negative. The cause of the negative terms was not clear in some cases, however, in the case of the Weibull model the

problem was related to the fact the lower bound parameter acts as a constraint in the optimisation. Very often the solution for the maximum likelihood point occurred at the boundary constraint. In such cases, the asymptotic properties of the likelihood (which are the basis of the derivation of the information matrix) do not hold.

One 'solution' to this problem was to re-run the analysis using a two parameter Weibull model in which the lower bound was set at $-\xi$. This resulted in acceptable covariance estimates but as pointed out by Dr. B. Torsney, University of Glasgow Statistics Department, this makes the results for the scale and shape parameters conditional on the choice of lower bound. One proper solution to this problem would be to include the order statistics estimator for the smallest in the sample in the likelihood function, see Galambos (1978); this could be included in any extension to this work or a related subject.

3.9 FORTRAN MAXIMUM LIKELIHOOD PROGRAM

The Weibull and GEV models discussed so far have been programmed into a FORTRAN 77 program called MAXLIK. This program uses *Numerical Algorithms Group* (NAg) subroutines for the optimisations and inversions required by the maximum likelihood method. The core of the program is NAg routine E04VDF which is a general minimisation routine which can handle functions with large numbers of variables and linear and non-linear boundary constraints. All computer code has been written in a modular format to enable new likelihood functions (ie new distributions) to be added without having to alter the structure of the program.

3.9.1 CHECKS WITH SIMULATED DATA

The Weibull code in the program has been checked by comparing its results with a dataset analysed by Menon (1963), and Cohen (1965). The original problem was defined by Menon loc cit and consisted of maximum likelihood estimation of the parameters for a Weibull population with zero lower bound. The population scale parameter was $\theta = 1.649$ (note Cohen writes $\theta = \alpha^c = 1.649$) and the shape parameter was $c = 0.5$. Two years later Cohen loc cit examined the data using maximum likelihood and showed how the information matrix could be estimated and then used to calculate the covariance matrix for the model parameters. The results of moment, Menon, Cohen, and MAXLIK estimators are summarised in Table 3.1 together with the 20 samples drawn from the population. This table shows that the method of moments gives the poorest estimate for the shape parameter and the best estimates are obtained with the Cohen, and MAXLIK likelihood estimators which are in error by only 1 per cent. The estimates for the scale parameter are more biased than the shape parameter with the worst solution given by the moment method.

In addition to the maximum likelihood estimates of the parameters Cohen loc cit estimates the information matrix from the sample log-likelihood and then derives the shape and scale parameter covariance matrix. Based on the results in Section [3.4] we can calculate the 95% confidence intervals for the shape and scale parameters using the information matrix. These values, reported by Cohen, are compared with those calculated by the MAXLIK program below

estimator	Cohen	MAXLIK
scale - θ	$0.68 < \theta < 2.05$	$0.62 < \theta < 1.81$
shape - c	$0.30 < c < 0.70$	$0.4 < c < 0.7$

The MAXLIK results are a slight improvement over those obtained by Cohen but this is probably due to the higher precision of the NAg routines used for the optimisation. Note the width of the 95% confidence limits is quite large for the scale parameter indicating the high level of uncertainty caused by the small sample size. The shape parameter has a narrower set of limits, which is encouraging since the quality of fit for extreme level quantiles is more sensitive to bias and uncertainty in this parameter than the scale parameter. If we examine the covariance matrix for the parameters it appears the scale and shape parameters are modeately correlated

$$c = \begin{bmatrix} var(\alpha) & cov(\alpha,c) \\ cov(\alpha,c) & var(c) \end{bmatrix} = \begin{bmatrix} 0.7275 & 0.022 \\ 0.022 & 0.0066 \end{bmatrix}$$

As a further check on the fit of the estimators, the parametric form of the exact, moment, Menon, Cohen, and MAXLIK estimates can be plotted against the quantile estimate

$$E\left[Pr(X \le x_k)\right] = \frac{k}{N+1}$$

This has been done in Fig. 3.5 and 3.6 which have been plotted on a linearising scale. Note the moment estimator is clearly shifted from the data and has the wrong slope whereas the MAXLIK estimator fits the data well and is very close to the exact model.

The GEV estimation subroutines have been checked by using two simulated data sets provided by S. Coles at Sheffield University. Each data set contains 30 samples drawn independently from a GEV population with location, scale and shape parameters:

sample	location	scale	shape
dataset 1	-0.20106	1.03457	0.14165
dataset 2	-0.04722	0.88554	-0.14655

Both datasets given in Table 3.2 have been analysed using the MAXLIK program and its' results compare favourably with those reported by Coles, suggesting the likelihood and optimisation routines are correct.

To check the information matrix terms two approaches have been programmed: the first uses analytical expressions for the second derivatives of the sample log-likelihood (Appendix A); and the second uses the finite difference method. The results from each method were found to be identical to the first four digits of accuracy and no perturbation was required for either data set. The results from the finite difference method are presented in Table 3.3. Based on the covariance matrices the 95% confidence limits are then:

GEV parameter	dataset 1	dataset 2
shape	0.7378 <-> 1.332	0.6112 <-> 1.160
scale	-0.1160 <-> 0.3991	-0.4140 <-> 0.1203
location	-0.625 <-> 0.2227	-0.4044 <-> 0.3101

Note in this case the bounds are more narrow than was the case for the Menon Weibull data due to the larger sample size. On the other hand in this case the 95% confidence limits change the sign of the scale and location parameters suggesting the sample may not be sufficiently large to establish which Fisher-Tippet model is the correct one for the data.

3.10 PEAK OVER THRESHOLD PROBABILITY

When fitting a population model to a sample of data it is often the case that either: the order statistic quantiles do not fit the estimated model in the tails of the data; or the model tail does not behave as expected when predicting return period levels. Since the $F(x)^N$ approach to calculating the distribution of the largest or smallest extremes is sensitive to the fit in the tails of the distribution it is important to ensure the fit is good for the largest or smallest values in the data. One method commonly used to improve the fit in the tails is to use a weighted least squares regression, Isaacson (1981). This method suffers from a lack of consistency in that the choice of weights is arbitrary, resulting in different estimates of extremes given different weights. A better procedure is to censor the data by selecting only the r largest in the data sample and then fit the model to these data. This approach is can be used successfully with the maximum likelihood method and is useful when data are only recorded

when the values exceed a threshold value. For example, a data buoy can be designed to record heights which are in excess of say 5 metres significant, thereby reducing the amount of information which must be stored and analysed.

Assuming independence, the distribution of the largest L in sample of N amplitudes A is given by

$$F(L \le l) = [F(A \le a)]^N \qquad [3.41]$$

However, a model $\hat{F}(m)$ fitted only to the r largest in a sample of size k cannot be used in the conventional way for estimating the distribution of the maxima since it is not an estimate for the population model $F(A \le a)$ used above. The model $\hat{F}(m)$ must therefore be used with the theorem of total probability to calculate the distribution of the largest in N independent, identically distributed amplitudes.

Let E_r be the event that there are r amplitudes greater then some threshold ξ_0 in k trials. The probability that there are r values greater than ξ_0 in a sample of size k is given by the binomial distribution, where

$$P(E_r) = \binom{k}{r} \delta^r (1-\delta)^{k-r} \qquad [3.42]$$

here

$$\binom{k}{r} = \frac{k!}{r!(k-r)!}$$

Now, given that r of the amplitudes are greater than the maximum likelihood threshold ξ_0 in k samples, the conditional probability of the event that the maximum amplitude M will be less than a value m $(M \le m / R = r)$ is

$$P(M \le m / R = r) = \left[\hat{F}(m)\right]^r$$

Here, $\hat{F}(m)$ is the model fitted to the r largest amplitudes using the method of maximum likelihood. For our problem, the probability sought is the unconditional probability $P(M \le m)$ which, by the theorem of total probability, is given by

$$P(M \le m) = \sum_{r=0}^{N} \left(\hat{F}(m)\right)^r \binom{N}{r} \hat{\delta}^r \left(1-\hat{\delta}\right)^{N-r}$$

where $\hat{\delta}$ is the probability that the observed amplitude will be greater than the threshold. This simplifies to

$$P(M \le m) = \left(\hat{\delta}\, \hat{F}(m) + 1 - \hat{\delta} \right)^N \quad m \ge \hat{\xi}_0 \qquad [3.43]$$

Estimates of the quantities $\hat{F}(m)$ and $\hat{\delta}$ can be obtained using maximum likelihood or, alternatively, $\hat{F}(m)$ may be found using maximum likelihood and $\hat{\delta}$ can be estimated using $\hat{\delta} = r/k$ For the case $r=k$, we have $\hat{\delta} = 1.0$ and Equation [3.43] reduces to the correct result $F_l(M \le m) = [F(A \le a)]^N$. The value of r used for censoring the data must be chosen carefully, since if the value is made too small - less than 10 say - the maximum likelihood solution will be biased to such a degree that the advantage of the optimal sampling variance associated with maximum likelihood methods will be lost. On the other hand if r is too large the fit $\hat{F}(m)$ may be poor for the largest values in the sample which are most important for the estimation of extremes.

3.11 EXTREME RESPONSES OF A FLARE TOWER

The maximum likelihood procedures discussed previously are so general they can be applied to a variety of statistical modelling problems. One useful application is estimation of the most probable maximum structural response, in say three hours, given non-linear time domain response simulations which last for say 2000 seconds. This problem was of interest to an Oil Major who commissioned a study into the use of maximum likelihood for predicting design levels of response. Full details are given in Prince-Wright (1990), and a summary of the study is included below.

Four time series of the forces in a flare structure brace element, Fig. 3.7, were examined to extracted the maxima from each time series. In the linear response case, we would expect the maxima to be Rayleigh distributed when the sea is narrow banded. Consequently, the maxima were assumed to be Weibull distributed since it includes the Rayleigh model as a special case when the shape parameter is 2.0. However, the non-linearity of the structural response means this is not so and therefore the shape parameter was included in the likelihood maximisation.

3.11.1 RESPONSE MAXIMA POPULATION

The sample of response maxima was extracted from the four time series. The maxima populations were then modelled using both two and three parameter Weibull models which are shown in Fig. 3.8 and 3.9. Overall, the three parameter model gave the best fit to the populations, which is to be expected, with little systematic curvature in any of the plots: this

suggests the Weibull model is reasonable as a descriptive model for the maxima. One problem with the Weibull model is that it decays rapidly in the upper tail which results in under-estimation in the case of wave heights. For this reason, a simple threshold approach was developed, Prince-Wright (1991b), so that a Weibull model could be fitted to the r largest response maxima.

The obvious question is what value should be used for r ? The previous literature review, Section 3, indicated that the likelihood approach could be used with samples with as few as 20 observations. Furthermore, the tests with the simulated Menon data also indicated the model parameters' bias and variance were still acceptable for such small samples. Consequently, r was taken as 20.

The results are shown for two of the time series in Fig. 3.10. Again the model fit is good, however, the small sample size results in a large increase in the statistical uncertainty of the Weibull model parameters, which are given in Table 3.4. A further disadvantage of this method was that in some cases the best results were obtained when a two parameter Weibull was used. This makes it difficult to program the estimator in a 'black box' routine since, ideally, both models must be fitted and then compared; when in fact goodness-of-fit tests are unreliable in the small sample case. This example therefore tends to suggest it is preferable to model the population using a distribution which has the correct tail behaviour.

3.12 MODELS OF THE DB1 WIND, WAVE AND CURRENT DATA

Two approaches were used to model the marginal distributions of the DB1 wind speed, significant wave height, and current speed. The first is summarised in ch. 2 where maximum likelihood was used to fit both Weibull and GEV models to the marginal monthly maxima given in Table 2.11: Ch. 2. The quality of the fits to the monthly maxima was generally good as shown in Fig. 3.11 and 3.12. Unfortunately it is difficult to quantitatively compare two models - which both seem to linearise the data - so their residuals were plotted. The results are shown in Fig. 3.13 and 3.14 but again both appear to be similar.

In the second approach both Weibull and GEV models were fitted to the full population of 3-hourly records for each variable (in fact the wind and current data was recorded every hour but we only use the values taken simultaneously with the seastate parameters). The results from modelling the population are given in Table 3.5 which shows in general the likelihood estimator's mean and variance agree well, the skewness is reasonable, and the kurtosis is poor.

The primary purpose of a marginal model for the wind, wave and current data is to estimate the level of a variable exceeded on average only once in a specific return period like 50 or 100

years. It is therefore essential that the marginal models be capable of modelling the tails of the data: this suggests the third and fourth moments of the estimator must agree with the sample moments. Using the definition of return period given in Department of Energy (1990b) then the survival probability corresponding to the 50 year return period is

$$1 / (50 * 365.25 * 8)$$

for the population models, and

$$1 / (50 * 12)$$

for the monthly maxima models. Note 8 is the number of 3-hourly observations in one day. The return period level for the Weibull estimator is then given by Equation [3.23] and for the GEV model the value is given by Equation [3.33]. The results for the 50 year return period wind speed, significant wave height, and current speed are given in Table 3.6 for both the population and the monthly maxima models obtained using maximum likelihood. In both cases the population and monthly maxima models for the significant wave height give obviously poor estimates. On the other hand the results for the wind and current speeds are reasonable.

Although these results alone do not indicate population models are definitely superior to the monthly maxima models in the small sample case it should be noted only two distributions are examined. In fact, Carter & Challenor (1983), and Carter (1987) have shown the FT-1 model gives good unbiased results when fitted to wave data using moments. Using their method the 50 year return period value for the DB1 data is given by

$$x_{50} = E[x] + 8.821\sigma[x]$$

This gives

$$H_{s50} = 13.3m$$
$$U_{w50} = 38.3m/s$$
$$U_{c50} = 1.7m/s$$

Comparing these values with the Department of Energy Guidance notes, which give $H_{s50} \approx 15.5m$ and $U_{w50} \approx 38.5m/s$, the significant height is seen to be too low, and the wind speed is in agreement - the current speed cannot be compared since we have not deconvolved the tidal and storm surge components.

In the next chapter an alternative method for modelling the populations is given in which the data are transformed to a normally distributed variate using the Box transformation. This approach also gives good estimates for the extremes and has the further advantage of generalising to the multivariate case.

3.13 CONCLUSIONS

The desirable characteristics of a good estimator were stated as: consistency, sufficiency, low bias, low sampling variance, reliability, and simplicity. Intrinsic and parametric analysis of marginal data has been examined and kernel density and maximum likelihood analysis were selected for special attention since they satisfy most of the characteristics.

Univariate kernel density methods are an ideal visualisation tool for large samples (>100) and can be generalised to the multivariate case.

Maximum likelihood is an optimal estimator in the large sample case (>100) and can be used with population data and extreme value data. ML estimates of a model's parameters are efficient with low bias and the parameter variance is close to the optimum attainable from any estimator.

The ML method can easily be generalised to the multivariate case providing the likelihood can be optimised over the parameter space. Furthermore, the parameter uncertainty can be deduced directly from the information matrix if the ML solutions can be found. However, in some cases the solution may lie on the bounds to the parameter space and in such cases the asymptotic properties of the information matrix will not hold. This problem was encountered for the lower bound when using the Weibull model. Two solutions to the problem are: to set the lower bound to zero, in which case we are fitting a two parameter model; or, to set the lower bound of the model to the smallest observed value, which makes the solution conditional on this value.

Likelihood software has been developed to fit Weibull and GEV distributions and estimate the parameter uncertainty using the information matrix. The code has been tested using simulated data and checked against results reported in the statistical literature.

Both Weibull and GEV population and monthly maxima models have been compared as estimators of return period values. Both the population and monthly maxima models gave poor results for the significant wave height, with the Weibull model under-estimating and the GEV model overestimating the value. By comparing only the return period estimates it is not clear if the population method or the extreme value method results in the lowest *overall* modelling uncertainty. However, if the model *parameter* uncertainty is included in the comparison it is clear the population approach results in an estimator with lower statistical uncertainty. This suggests that the population modelling method is best when only a few years of data are available - providing the correct model can be found.

The 50 year return period estimate of significant wave height obtained from fitting a GEV model to the monthly maxima is inaccurate and has a negative lower bound. The cause is the attraction to a lower bounded FT-II (Frechet) model which has a long upper tail. Two additional constraints in the likelihood optimisation may result in an improvement. The first is to restrict the shape parameter to be greater than or equal to zero which effectively constrains to the model to be either an FT-I (Gumbel) or FT-III (Weibull) model; the second is to constrain the lower bound to be zero or positive. This should be examined in a future study.

APPENDIX A

A.1 WEIBULL PARTIAL DERIVATIVES

The second partial derivatives of the Weibull sample log-likelihood are required for estimating the Fisher information matrix. For the simplest case the samples are *iid* in which case the second derivatives are given by

$$\frac{\partial^2 \log_e \ell}{\partial \alpha^2} = \frac{-c(c+1)}{\alpha^{c+2}} \sum_{i=1}^{n} (x_i - \xi_0)^c + \frac{cn}{\alpha^c}$$

$$\frac{\partial^2 \log_e \ell}{\partial c^2} = \frac{-n}{c^2} - \frac{1}{\alpha^c} \sum_{i=1}^{n} (x_i - \xi_0)^c \left\{ \log_e (x_i - \xi_0) \right\}^2 -$$

$$\frac{\left\{ \log_e \alpha \right\}^2}{\alpha^c} \sum_{i=1}^{n} (x_i - \xi_0)^c + \frac{2 \log \alpha}{\alpha^c} \sum_{i=1}^{n} (x_i - \xi_0)^c \log_e (x_i - \xi_0)$$

$$\frac{\partial^2 \log_e \ell}{\partial \xi_0^{\,2}} = \frac{-c(c-1)}{\alpha^c} \sum_{i=1}^{n} (x_i - \xi_0)^{c-2} - (c-1) \sum_{i=1}^{n} (x_i - \xi_0)^{-2}$$

$$\frac{\partial^2 \log_e \ell}{\partial \alpha \partial c} = \frac{-n}{\alpha} + \frac{c}{\alpha^{c+1}} \sum_{i=1}^{n} (x_i - \xi_0)^c \log_e (x_i - \xi_0) + \frac{1}{\alpha^{c+1}} (1 - c \log_e \alpha) \sum_{i=1}^{n} (x_i - \xi_0)^c$$

$$\frac{\partial^2 \log_e \ell}{\partial \alpha \partial \xi_0} = -\frac{c^2}{\alpha^{c+1}} \sum_{i=1}^{n} (x_i - \xi_0)^{c-1}$$

$$\frac{\partial^2 \log_e \ell}{\partial c \partial \xi_0} = -\sum_{i=1}^{n} (x_i - \xi_0)^{-1} + \frac{1}{\alpha^c} (1 - c \log_e \alpha) \sum_{i=1}^{n} (x_i - \xi_0)^{c-1} +$$

$$\frac{c}{\alpha^c} \sum_{i=1}^{n} (x_i - \xi_0)^{c-1} \log_e (x_i - \xi_0)$$

A.2 GEV PARTIAL DERIVATIVES

Writing $z=(x-\mu)/\sigma$ the first derivatives of the Generalised Extreme Value distribution sample log-likelihood are given by

$$\frac{\partial \log_e \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{k}{\sigma}\left(\frac{1}{k}-1\right)\sum_{i=1}^{n} z_i(1-kz_i)^{-1} - \frac{1}{\sigma}\sum_{i=1}^{n} z_i(1-kz_i)^{\frac{1}{k}-1}$$

$$\frac{\partial \log_e \ell}{\partial k} = -\left(\frac{1}{k}-1\right)\sum_{i=1}^{n} z_i(1-kz_i)^{-1} - \frac{1}{k^2}\sum_{i=1}^{n} \log_e(1-kz_i) +$$
$$\frac{1}{k}\sum_{i=1}^{n}(1-kz_i)^{\frac{1}{k}}\left\{z_i(1-kz_i)^{-1} + \frac{1}{k}\log_e(1-kz_i)\right\}$$

$$\frac{\partial \log_e \ell}{\partial \mu} = \frac{k}{\sigma}\left(\frac{1}{k}-1\right)\sum_{i=1}^{n}(1-kz_i)^{-1} - \frac{1}{\sigma}\sum_{i=1}^{n}(1-kz_i)^{\frac{1}{k}-1}$$

The second partial derivatives of the GEV are given by

$$\frac{\partial^2 \log_e \ell}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{2(1-k)}{\sigma^2}\sum_{i=1}^{n} z_i(1-kz_i)^{-1} + \frac{2}{\sigma^2}\sum_{i=1}^{n} z_i(1-kz_i)^{\frac{1}{k}-1} -$$
$$\frac{k(1-k)}{\sigma^2}\sum_{i=1}^{n} z_i^2(1-kz_i)^{-2} - \frac{k}{\sigma^2}\left(\frac{1}{k}-1\right)\sum_{i=1}^{n} z_i^2(1-kz_i)^{\frac{1}{k}-2}$$

$$\frac{\partial^2 \log_e \ell}{\partial k^2} = \frac{2}{k^3}\sum_{i=1}^{n}\log_e(1-kz_i) + \frac{2}{k^2}\sum_{j=1}^{n} z_i(1-kz_i)^{-1} - \left(\frac{1}{k}-1\right)\sum_{i=1}^{n} z_i^2(1-kz_i)^{-2}$$
$$-\frac{2}{k^3}\sum_{i=1}^{n}\log_e(1-kz_i)(1-kz_i)^{\frac{1}{k}} - \frac{2}{k^2}\sum_{i=1}^{n} z_i(1-kz_i)^{\frac{1}{k}-1} -$$
$$\frac{1}{k^4}\sum_{i=1}^{n}\log_e(2-2kz_i)(1-kz_i)^{\frac{1}{k}} - \frac{1}{k}\left(\frac{1}{k}-1\right)\sum_{i=1}^{n} z_i^2(1-kz_i)^{\frac{1}{k}-2} -$$
$$\frac{2}{k^3}\sum_{i=1}^{n} z_i\log_e(1-kz_i)(1-kz_i)^{\frac{1}{k}-1}$$

$$\frac{\partial^2 \log_e \ell}{\partial \mu^2} = -\frac{k(1-k)}{\sigma^2}\sum_{i=1}^{n}(1-kz_i)^{-2} - \frac{(1-k)}{\sigma^2}\sum_{i=1}^{n}(1-kz_i)^{\frac{1}{k}-2}$$

$$\frac{\partial^2 \log_e \ell}{\partial \mu \partial \sigma} = -\frac{(1-k)}{\sigma^2}\sum_{i=1}^{n}(1-kz_i)^{-1} + \frac{1}{\sigma^2}\sum_{i=1}^{n}(1-kz_i)^{\frac{1}{k}-1} - \frac{k^2(1-k)}{\sigma^2}\sum_{i=1}^{n}z_i(1-kz_i)^{-2} - \frac{k}{\sigma^2}\left(\frac{1}{k}-1\right)\sum_{i=1}^{n}z_i(1-kz_i)^{\frac{1}{k}-2}$$

$$\frac{\partial^2 \log_e \ell}{\partial \mu \partial k} = -\frac{1}{\sigma}\sum_{i=1}^{n}(1-kz_i)^{-1} + \frac{k}{\sigma}\left(\frac{1}{k}-1\right)\sum_{i=1}^{n}z_i(1-kz_i)^{-2} + \frac{1}{\sigma k^2}\sum_{i=1}^{n}\log_e(1-kz_i)(1-kz_i)^{\frac{1}{k}-1} + \frac{1}{\sigma}\left(\frac{1}{k}-1\right)\sum_{i=1}^{n}z_i(1-kz_i)^{\frac{1}{k}-2}$$

$$\frac{\partial^2 \log_e \ell}{\partial \sigma \partial k} = -\frac{1}{\sigma}\sum_{i=1}^{n}z_i(1-kz_i)^{-1} + \frac{(1-k)}{\sigma}\sum_{i=1}^{n}z_i^2(1-kz_i)^{-2} + \frac{1}{\sigma k^2}\sum_{i=1}^{n}z_i\log_e(1-kz_i)(1-kz_i)^{\frac{1}{k}-1} + \frac{1}{\sigma}\left(\frac{1}{k}-1\right)\sum_{i=1}^{n}z_i^2(1-kz_i)^{\frac{1}{k}-2}$$

sample			
0.806	57.628	1.550	7.057
0.664	1.033	9.098	2.046
0.345	3.532	0.470	0.185
0.001	0.970	0.505	0.435
0.469	0.071	0.033	1.550

parameter	population	moments	Menon	Cohen	MAXLIK
scale	1.6490	1.2300	1.4000	1.3630	1.3610
shape	0.5	0.4300	0.5700	0.5060	0.5051
scale error	-	0.419	0.249	0.286	0.288
shape error	-	0.07	0.07	0.006	0.0051
scale error %	-	-25.4	-15.1	-17.3	-17.3
shape error %	-	-14.0	14.0	1.2	1.0

Table 3.1 Comparison of MAXLIK results with Menon (1963) and Cohen (1965) maximum likelihood estimators

data set 1					
2.58405	-0.50394	-1.71630	0.47867	0.26712	0.87737
0.29929	0.08969	0.42083	-0.25574	-0.06714	0.35253
2.46029	3.05271	0.00297	0.59921	-0.34909	1.30477
1.63881	0.51030	-1.22866	-1.15796	-0.69298	0.94734
0.70814	-1.53493	0.81628	-0.24643	-1.39538	0.38320
data set 2					
2.12546	0.00580	-0.85963	0.55990	0.36670	0.62464
0.18471	0.64917	0.10151	-0.58044	0.06905	0.53800
2.33662	6.05925	0.38144	1.94203	0.31054	1.54564
-0.16174	0.38654	-0.56694	-0.72557	-1.29499	1.90079
-0.06119	-0.84076	3.27829	0.48357	-0.4093	0.10569

Table 3.2 Two datasets of GEV distributed random numbers provided by Coles

quantity	data set 1	data set 2
information matrix	$\begin{bmatrix} -56.678 & 27.716 & 2.1901 \\ & -84.318 & 15.444 \\ sym & & -26.206 \end{bmatrix}$	$\begin{bmatrix} -71.126 & -6.7152 & 28.779 \\ & -61.670 & 15.760 \\ sym & & -45.773 \end{bmatrix}$
covariance matrix	$\begin{bmatrix} 0.02207 & 0.00851 & 0.00686 \\ & 0.01657 & 0.01048 \\ & & 0.04491 \end{bmatrix}$	$\begin{bmatrix} 0.01892 & 0.00107 & 0.01226 \\ & 0.01784 & 0.00681 \\ & & 0.03190 \end{bmatrix}$

Table 3.3 Information and covariance matrices for the Coles data

time series	number of samples	ML mean values for: scale	shape	ML variance for: scale	shape	ML covariance
t1	155	4.06	4.19	0.009	0.004	0.001
20 largest	20	0.70	0.77	0.046	0.022	0.010
t2	102	7.29	8.68	0.010	0.004	0.001
20 largest	20	1.01	0.89	0.072	0.038	0.017
t3	131	2.06	1.89	0.009	0.004	0.001
20 largest	20	0.74	0.88	0.040	0.030	0.012
t4	158	1.76	1.61	0.008	0.004	0.001
20 largest	20	0.49	0.75	0.024	0.016	0.006

Table 3.4 Population and threshold Weibull models fitted to the maxima in the structural response time series. A two parameter Weibull model was fitted using maximum likelihood.

TABLE5.XLS

variable	no. of samples	statistical moments sample (top); weibull (middle); GEV (bottom) mean	variance	skewness	kurtosis	Weibull model shape	scale	location	GEV model shape	scale	location
Hs	5673	2.176	1.589	1.64	4.533	1.595	2.102	0.3	-0.1656	0.8071	1.562
		2.185	1.464	0.967	1.059						
		2.184	1.859	2.675	-16.48						
Uw	5673	8.287	11.57	0.7161	0.9111	2.522	9.117	0.2	0.07796	2.914	6.814
		8.291	11.8	0.3489	-0.1526						
		8.286	11.63	0.735	-0.8373						
Uc	5673	33.7	236	0.4166	-0.03817	2.402	39.09	-0.952	0.146	13.96	27.39
		33.7	236.2	0.404	-0.09537						
		33.66	234.6	0.4511	-0.1651						

Table 3.5 Statistical moments and maximum likelihood parameters for the Weibull and generalised extreme value distributions fitted to the full population of marginal DB1 wind, wave and current data.

variable	monthly maxima model weibull	GEV	population model weibull	GEV
Hs (m)	17.6	32.41	10.04	31.61
Uw (m/s)	33.35	42.54	24.5	29.42
Uc (cm/s)	100.98	97.42	108.23	108.62

Table 3.6 50 year return period estimates for the DB1 data obtained from both monthly maxima and population models.

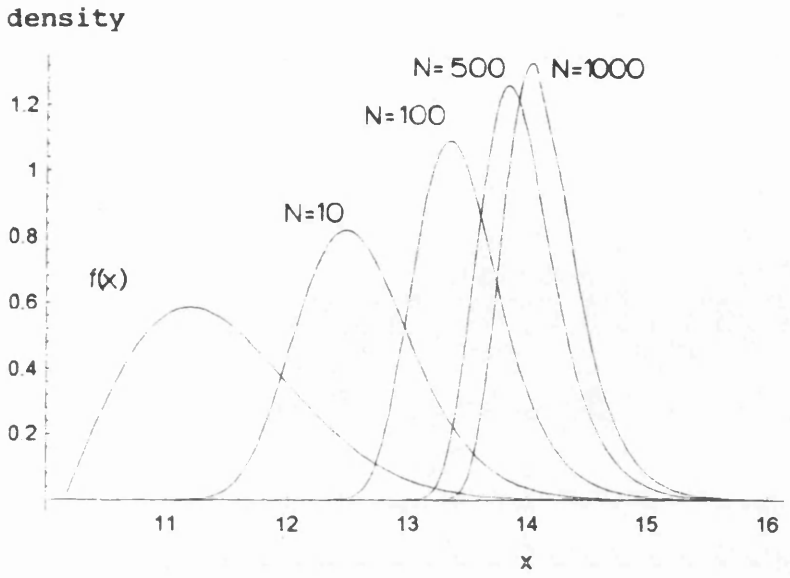


Figure 3.1 Distribution of the largest in n = 10, 100, 500, and 1000 samples with population distribution function $f(x)$.

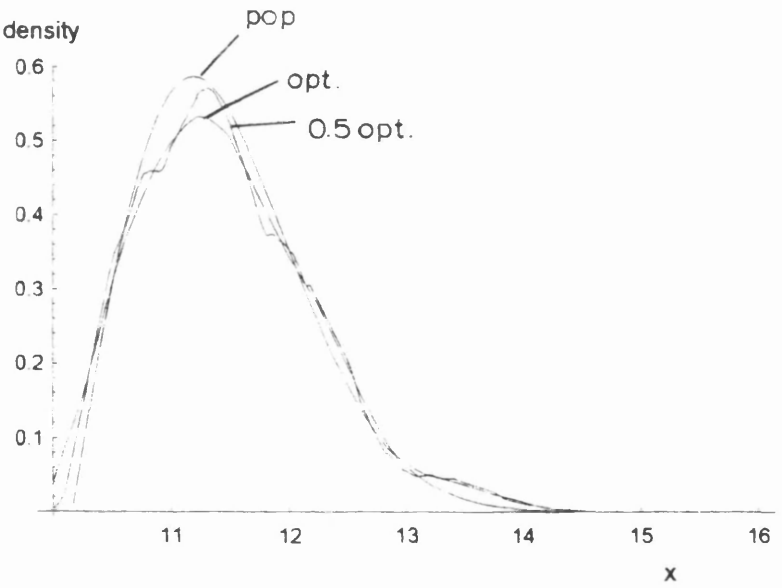


Figure 3.2 Population model and two kernel density estimates: one with the "optimal" window width, and the other with 0.5 times the optimal value.

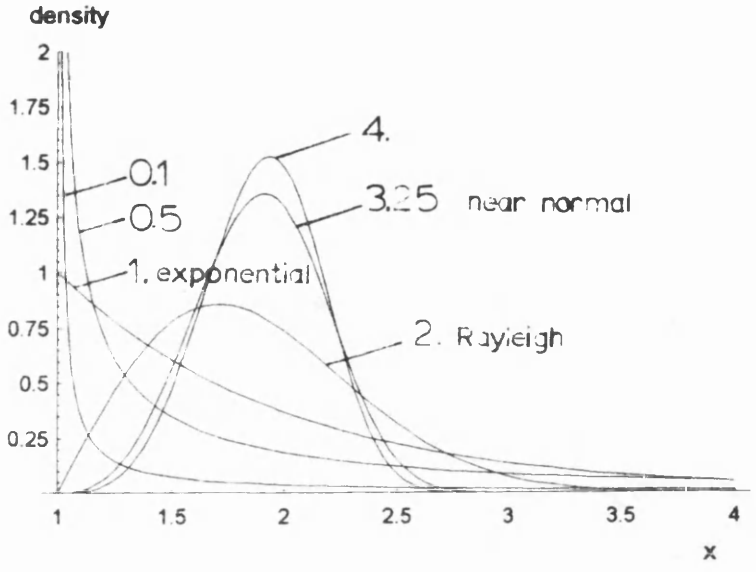


Figure 3.3 Weibull model with: shape parameters 0.1, 0.5, 1.0, 2.0, 3.25,4; scale parameter 1.0; and, location 1.0.

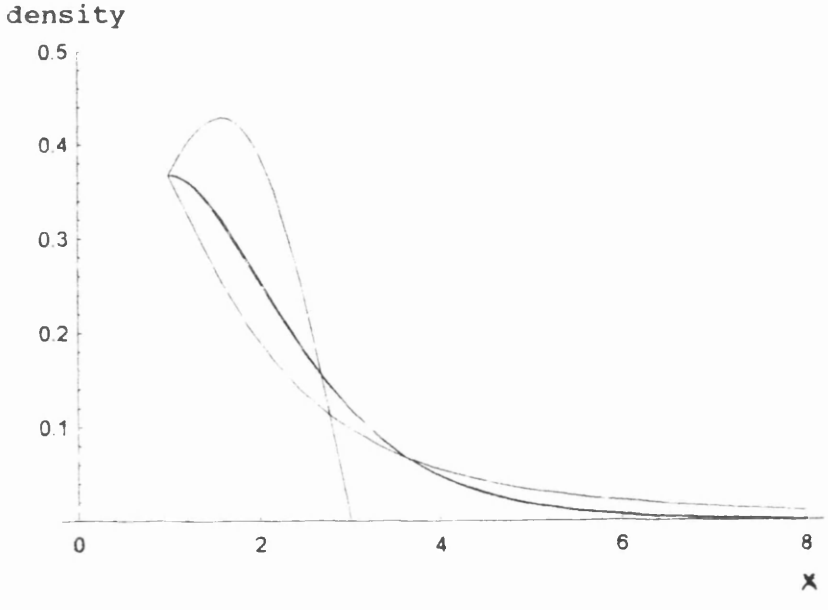


Figure 3.4 Generalised extreme value density functions with: shape parameters -0.5, -0.01, 0.01, 0.5; scale parameter 1.0; and location 1.0.

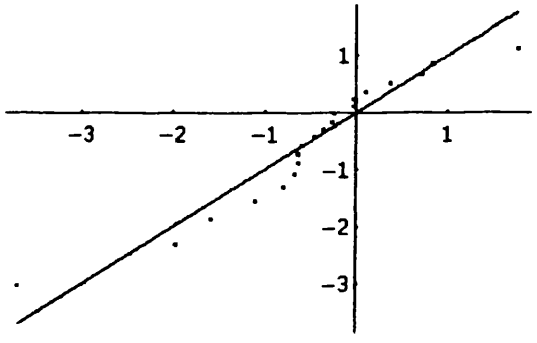


Figure 3.5 Population model and 20 simulated samples with shape parameter 0.5, and scale parameter 1.649 (see Menon (1963)).

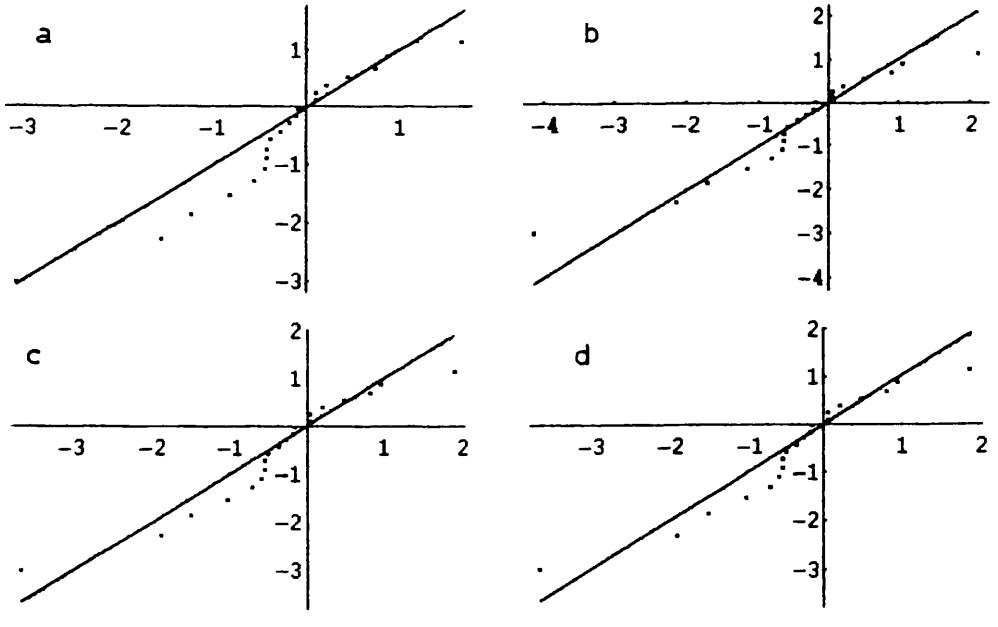


Figure 3.6 Four estimated models for the 20 Menon samples: moments (a), Menon (b), Cohen (c) and MAXLIK (d).

NOTE THE LINEARISATION OF THE WEIBULL MODEL USES THE FORM

$$\log_e\left(\log_e\left(\frac{1}{1-\Pr}\right)\right) = c\log_e(x-\xi) - c\log_e(\alpha); \quad ie \quad y = mx + c$$

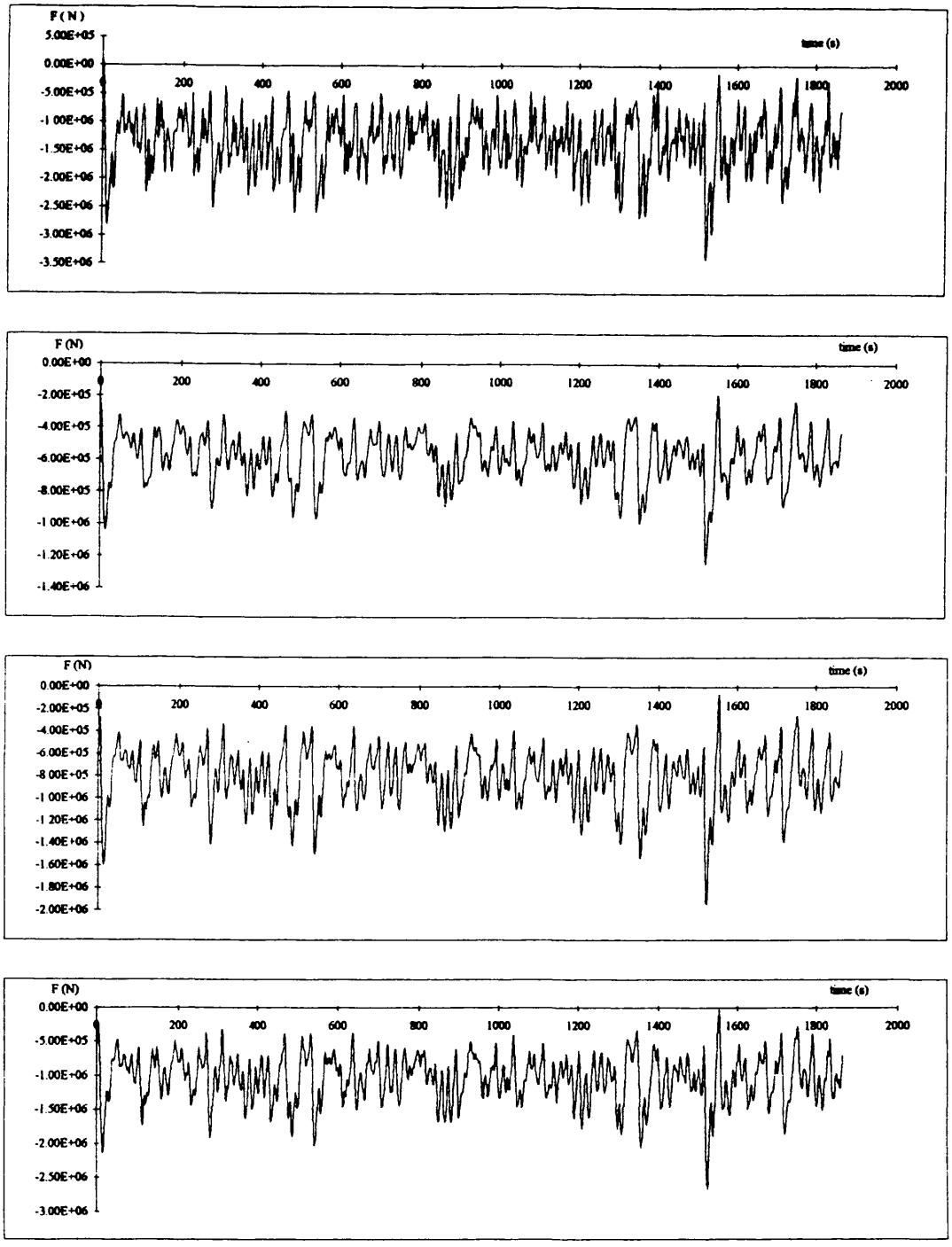


Figure 3.7 Four nonlinear structural member force response time series.

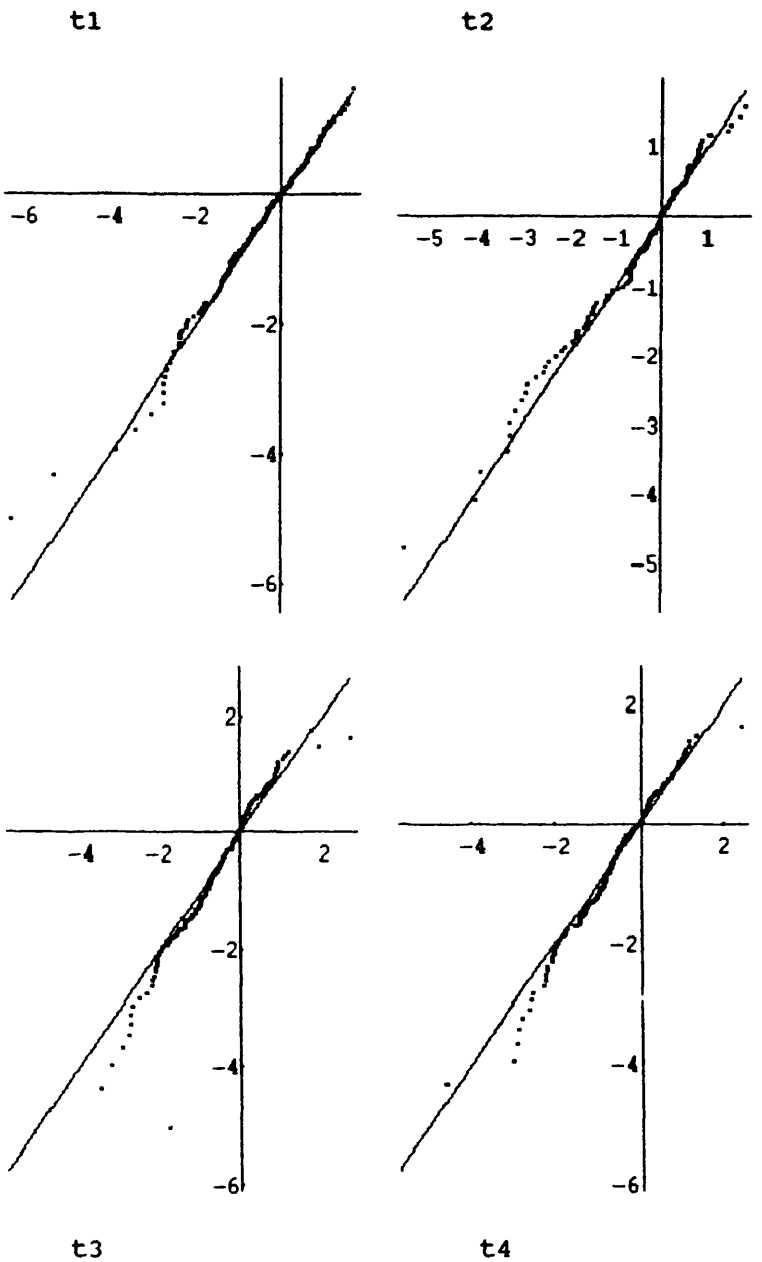


Figure 3.8 Three parameter Weibull distributions fitted to the maxima in four structural response time series.

NOTE THE LINEARISATION OF THE WEIBULL MODEL USES THE FORM

$$\log_e\left(\log_e\left(\frac{1}{1-\Pr}\right)\right) = c\log_e(x-\xi) - c\log_e(\alpha); \quad ie \quad y = mx + c$$

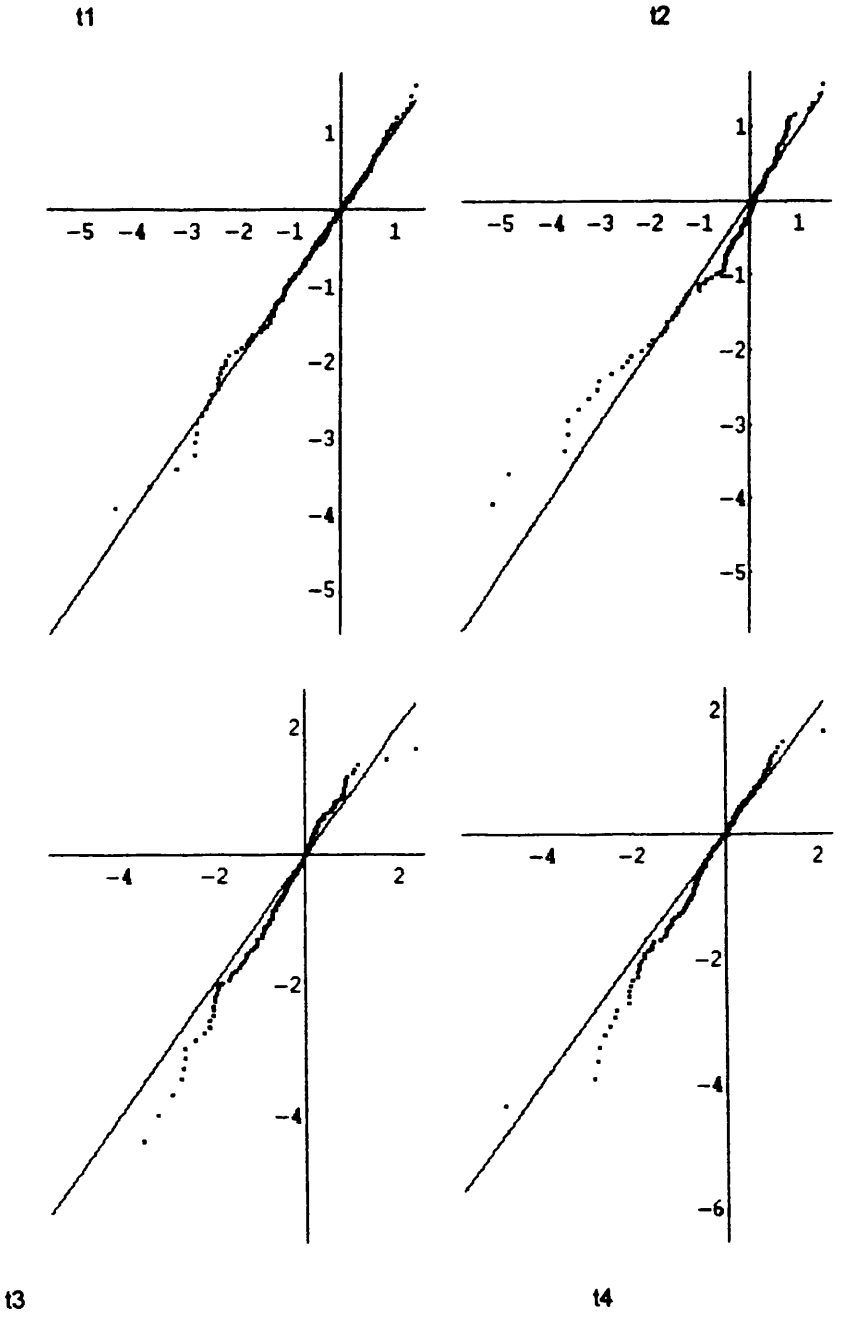


Figure 3.9 Two parameter Weibull distributions fitted to the maxima in four structural response time series.

NOTE THE LINEARISATION OF THE WEIBULL MODEL USES THE FORM

$$\log_e\left(\log_e\left(\frac{1}{1-\Pr}\right)\right) = c\log_e(x-\xi) - c\log_e(\alpha); \quad ie \quad y = mx + c$$

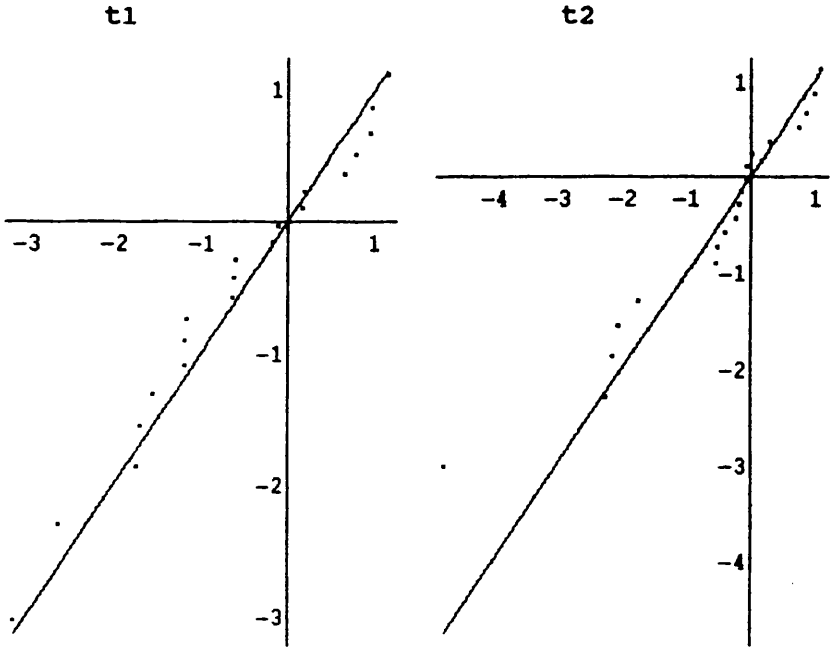


Figure 3.10 Weibull models fitted to the 20 largest observations in two structural response time series: a three parameter model was used for the time series t1, and a two parameter model was used for the time series t2.

NOTE THE LINEARISATION OF THE WEIBULL MODEL USES THE FORM

$$\log_e\left(\log_e\left(\frac{1}{1-\Pr}\right)\right) = c\log_e(x-\xi) - c\log_e(\alpha); \quad ie \quad y = mx + c$$

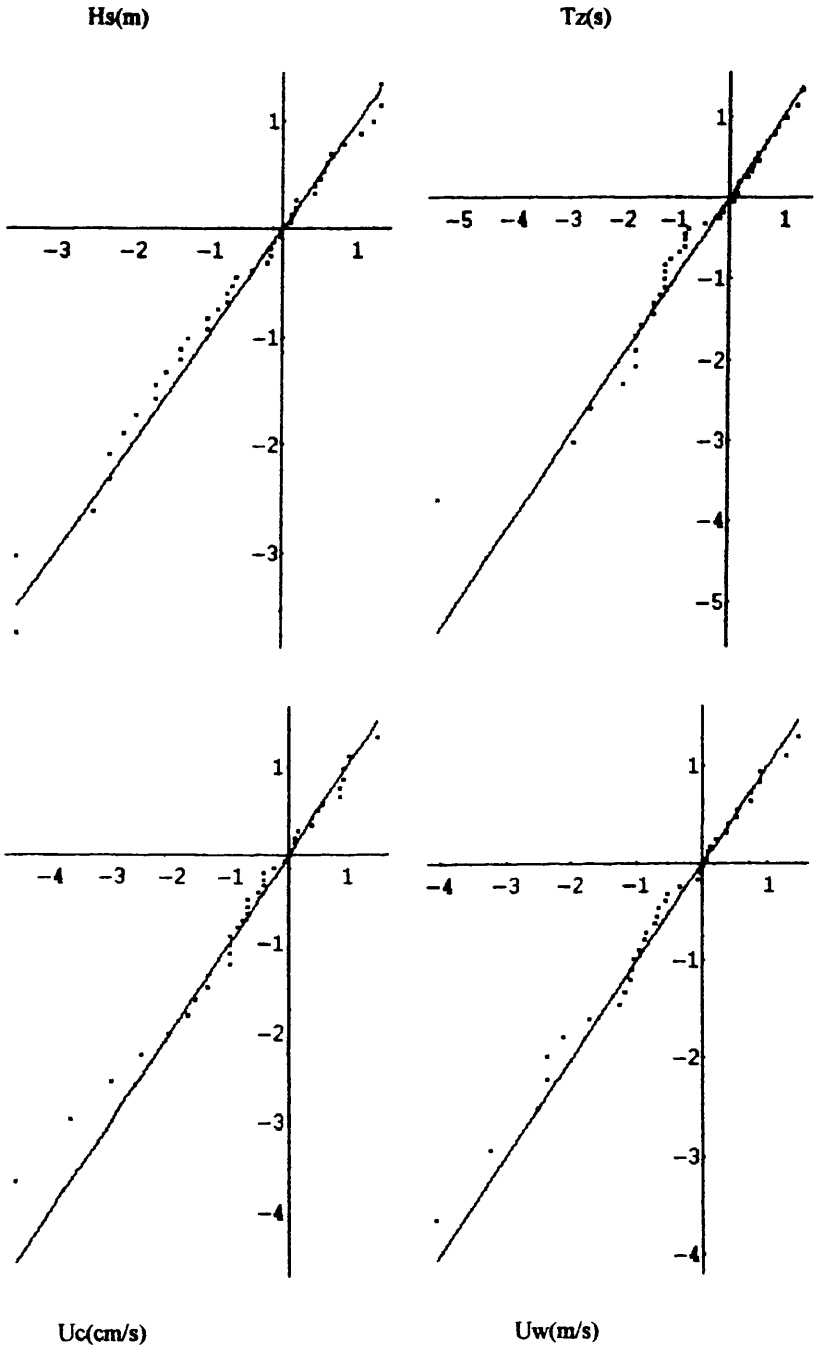


Figure 3.11 Weibull models fitted to the DB1 monthly maxima significant wave height, zero-up-crossing period, current speed, and wind speed using likelihood estimation.

NOTE THE LINEARISATION OF THE WEIBULL MODEL USES THE FORM

$$\log_e\left(\log_e\left(\frac{1}{1-\Pr}\right)\right) = c\log_e(x-\xi) - c\log_e(\alpha); \quad ie \quad y = mx + c$$

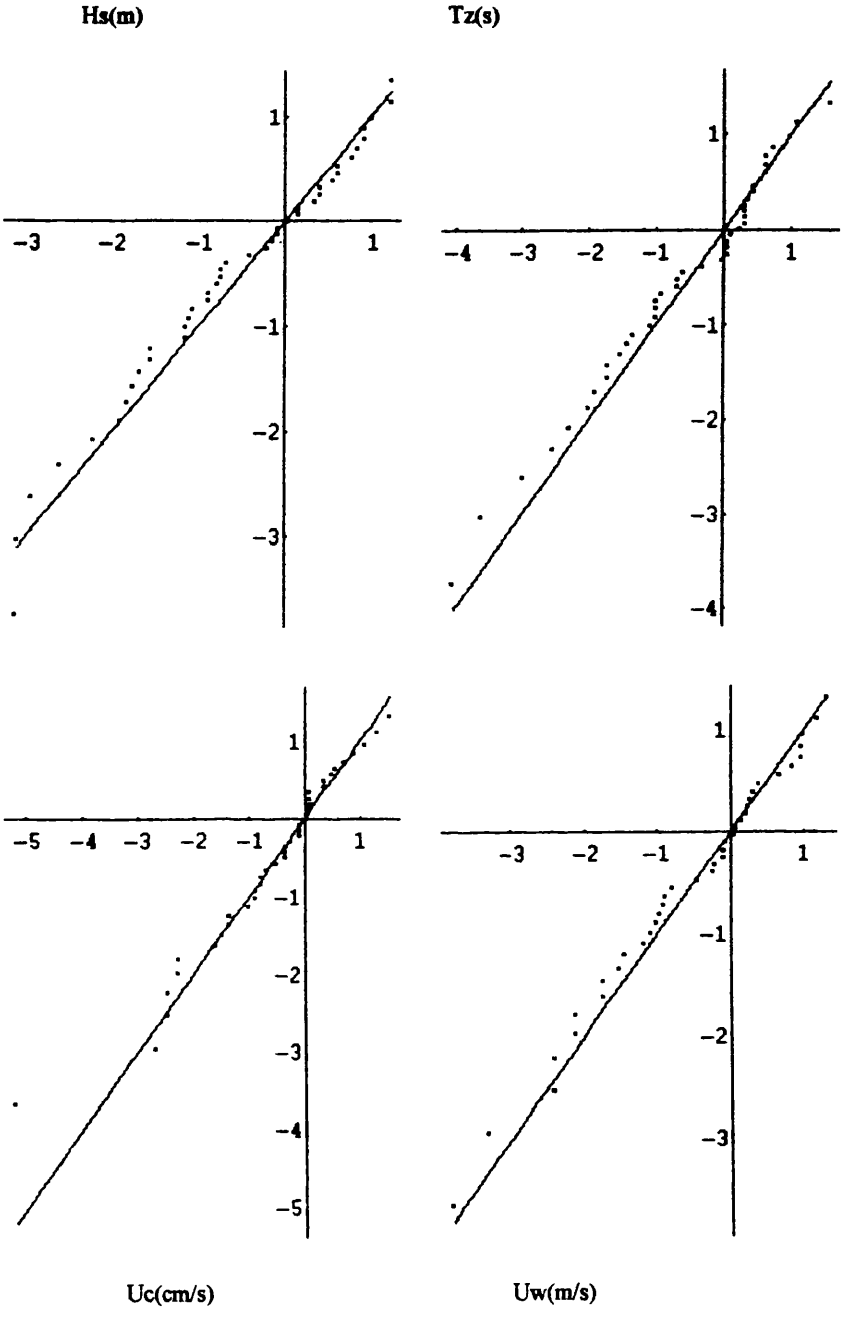


Figure 3.12 GEV models fitted to the DB1 monthly maxima significant wave height, zero-up-crossing period, current speed, and wind speed using likelihood estimation.

NOTE THE LINEARISATION OF THE GEV MODEL USES THE FORM

$$\log_e\left(\log_e(1/\Pr)\right) = \frac{1}{k}\log_e\left(1-k(x-\mu)/\sigma\right), \quad ie \quad y = mx + c$$

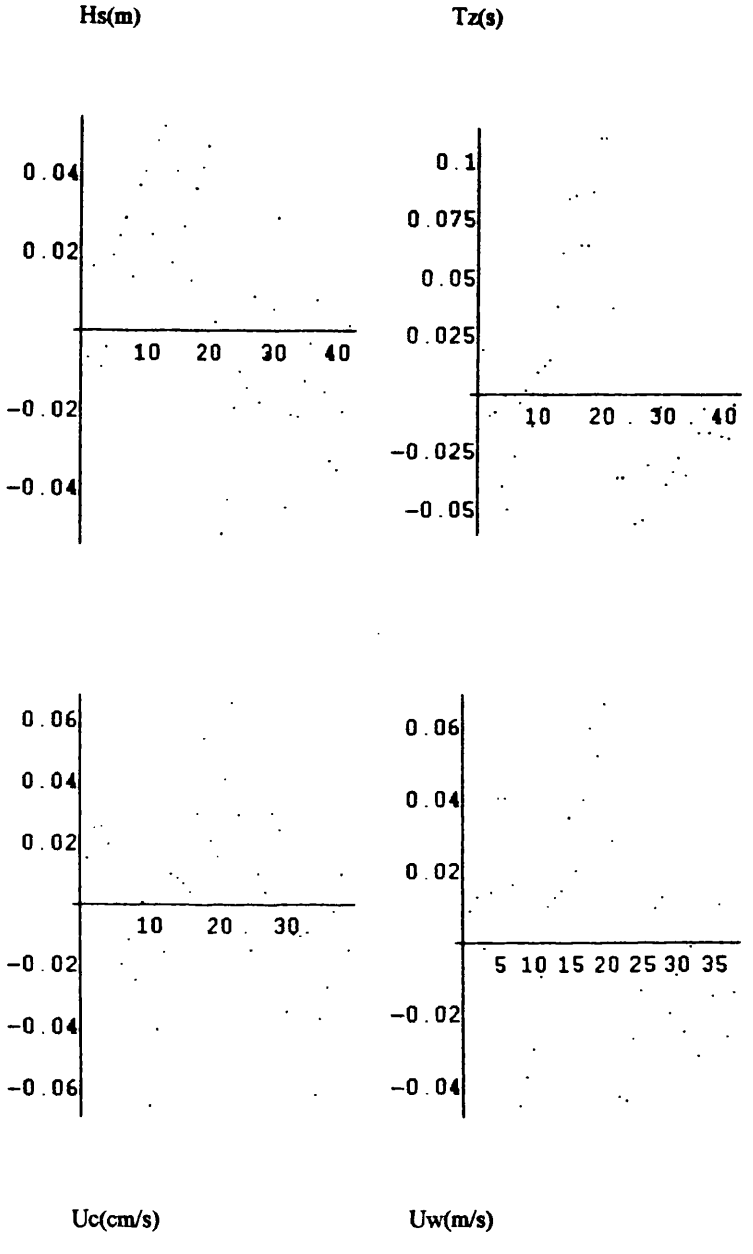


Figure 3.13 Residual plots for the Weibull models fitted to the monthly maxima using likelihood estimation.

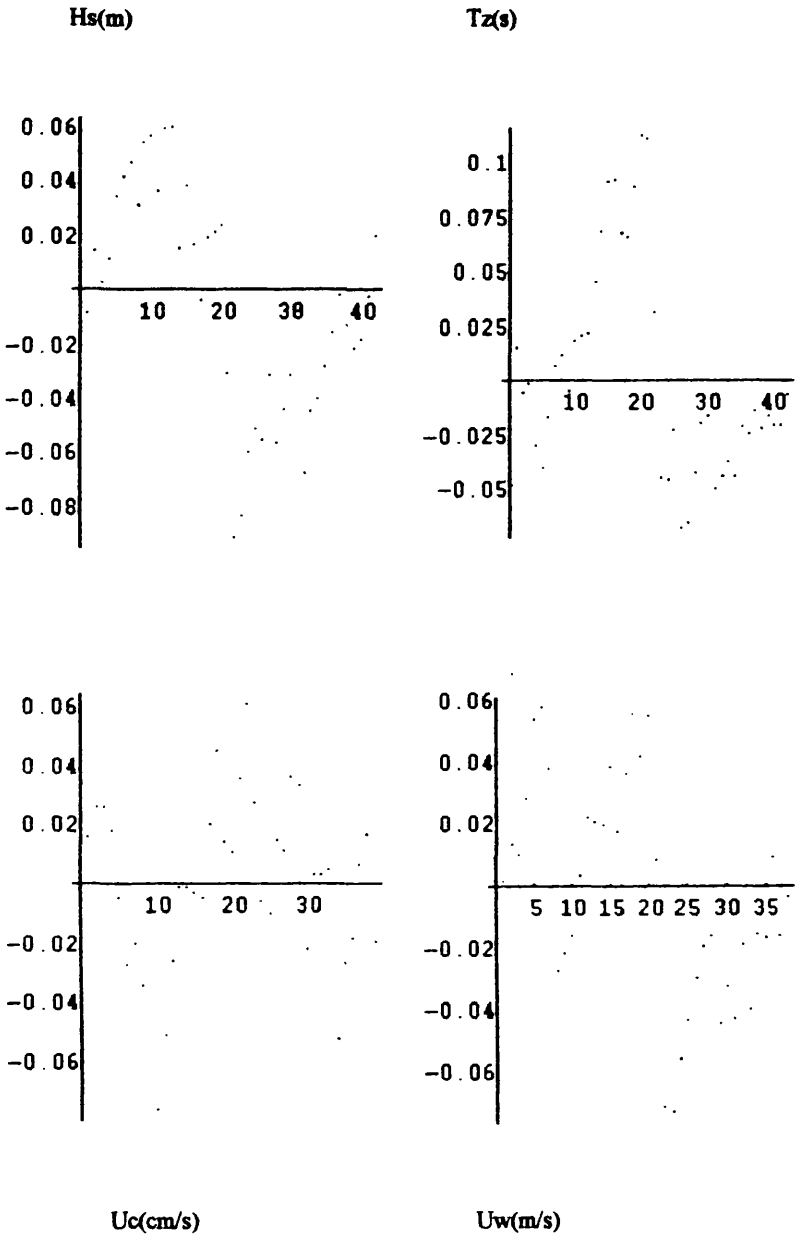


Figure 3.14 Residual plots for the GEV models fitted to the monthly maxima using likelihood estimation.

Chapter 4

MODELLING MULTIVARIATE ENVIRONMENTAL DATA

NOMENCLATURE

1	*vector of 1's:* $\{1,1,...,1\}^T$
A	*matrix of eigenvectors for the principal components*
a_n, b_n	*normalisation coefficients:* $\{a_{1:n},...,a_{p:n}\}, \{b_{1:n},...,b_{p:n}\}$
$A(w)$	*parametric dependence function*
$A^*(w)$	*intrinsic dependence function*
b	*bandwidth*
C	*a constant*
c	*shape parameter*
c_p	*volume of a p - dimensional sphere*
D	*diagonal matrix with* $1\sqrt{s_{ij}}$ *on the leading diagonal*
D_c	*current direction*
D_w	*wind direction*
$E[y_1 \mid y_2]$	*the expected value of* y_1 *given* y_2
$E[.]$	*expected value (the mean)*
$\hat{f}(x)$	*multivariate density estimate*
$F(x,y)$	*cumulative distribution function for x and y*
$F(x)$	*cumulative distribution function*
$H(.)$	*the Heaviside function*
H_s	*significant wave height*
J	*determinant of the transformation matrix*
$K(.)$	*kernel function*
k	*GEV shape parameter*
$min/\ max(a,b)$	*the smallest or largest of a and b*
MAE	*mean absolute error*
MIAE	*mean integrated absolute error*
n	*number of samples*
p	*number of variables*
R	*correlation matrix*
R^p	*Riemann domain of integration*
s_{ij}	*i - j th element of the covariance matrix*
T_z	*zero – up – crossing period*
U_c	*current speed*
U_w	*wind speed*
v	*variance OR extreme value model shape parameter*

$W_{i:n}$	*order statistics for the smallest in a sample*
w	*ratio:* $y / (x + y)$
x	*random vector sample*
X	*random vector*
$\hat{x}$	*modal value of* x
$Z_{i:n}$	*order statistics for the largest in a sample*

α	*scale parameter*
θ	*model parameter*
$\boldsymbol{\theta}$	*vector of observed model parameters*
λ	*shape parameter in the Box & Cox transformation*
μ	*mean value*
$\boldsymbol{\mu}$	*vector of means*
Ξ	*matrix of means*
σ	*standard deviation*
Σ	*covariance matrix*
ξ	*location parameter in the Box & Cox transformation*
ρ	*correlation coefficient*
Φ	*extreme value model dependence parameter*
$\Phi(.)$	*standard normal CDF*

4. INTRODUCTORY REMARKS: MODELLING MULTIVARIATE ENVIRONMENTAL DATA

In Ch. 3 various methods for modelling marginal data using non-parametric, and parametric methods were introduced; with Kernel density estimation, and maximum likelihood estimation selected for special attention. In this chapter, that work is further developed for use in the multivariate context with a view to its ultimate use in a level III reliability calculation.

Multivariate statistical analysis is a broad subject and several approaches can be adopted to solve the same problem. Some of the main techniques which can be used to model a sample are summarised in Fig. 4.1. This flowchart indicates that the approach taken may be determined by the type of analysis to be performed. For example, a fatigue study or a comparative reliability study may solve the problem using multivariate kernel or transformation models which generally give optimal density estimates for all but the extreme values. In particular, we shall see that the transformation approach has the advantage of making the computationally expensive Rosenblatt transformation redundant in a reliability study. On the other hand, if the analysis is concentrating on the extreme events in the design life then either the multivariate parametric models or asymptotic models may be more appropriate.

One problem becomes apparent when using parametric methods, namely, the frequent lack of a suitable model for a given set of data. This problem is well known to statisticians and consequently procedures have been developed to transform observed data into near normally

distributed variates, see Gnanadesikan (1977). This transformation simplifies the semi-analytical structural reliability methods outlined in Ch. 1 because:

- multivariate normal model parameters are useful in all levels of reliability analysis
- Rosenblatt's transformation becomes unnecessary in level III analysis

The major problem with this approach is finding a transformation which is sufficiently flexible to transform the marginal and joint behaviour to a normal model with the correct extremes. Ideally, transformation methods should be used with results from hindcast studies which generally give several decades of simulated metocean data.

By fitting a parametric model to a sample $\boldsymbol{x} \in \boldsymbol{X}$ it is assumed to represent the expected behaviour of the population $\boldsymbol{X}$. In our case, the model is generally a probability density having some functional form dependent on a number of unknown model parameters $\boldsymbol{\theta}$. Fig. 4.1 summarises several estimation procedures which can be used to estimate the parameters, but in this work, Fisher's method of maximum likelihood is adopted since, for a predefined model, it generally provides maximum agreement between observed data and model. The method has the further advantage of generalising to the multivariate case, with the main principles of score statistics and information matrices unchanged from those outlined for the univariate case in Ch. 3.

One problem encountered when modelling 'real' data using parametric methods is the paucity of suitable models. This is because the margins generally have different distributions whereas most multivariate parametric models assume the same form for each margin. In many cases, even the best parametric model may not fit the data well, especially in the extremes, and it is then necessary to assess whether the model is preferential to the mathematically simpler approach of transforming the data to approximate multivariate normality.

This chapter begins in *Section 1* by introducing some of the matrix-based algebra notation used for the multivariate sample and its statistics. This notation is ideal when using symbolic algebra packages like *Mathematica* since operations on matrices are implemented in fast machine code and programming becomes more clear and concise. *Section 2* then shows how multivariate non-parametric kernel methods can be used to estimate the density for a sample. It was initially thought kernel models would be most useful as a graphical tool for comparing parametric models with observed data, but their simplicity, and good performance, suggest they could be of direct use in reliability studies if, like a hindcast dataset, the sample covers a sufficiently long time period. Kernel density estimates for the wave, wind, and current

populations are presented for selected pairs of variables. In *Section 3* , the transformation of non-normal multivariate samples into *near* multivariate normal samples is then examined: first, using univariate and bivariate simulated data; and then with the DB1 data. *Section 4* then shows how the 'best' set of parameters is chosen. *Section 5* then shows how the 'best' model can be transformed to a standard normal format suitable for the reliability calculation.

The initial intention was to use multivariate extreme value theory to model the *componentwise* maxima recorded by the DB1. In this method the margins are first modelled using a generalised extreme value (GEV) distribution and then the parameters of a dependence function are calculated. Unfortunately, there are two problems with this method: first, multivariate extreme value models can only be used for time invariant reliability studies such as those performed on fixed structures; and second, the results obtained from fitting the GEV to the marginal maxima were highly inaccurate for quantiles which correspond to design return periods. The probable cause of the poor behaviour is the limited amount of the data and the use of monthly maxima which are seasonally correlated. Although the method was not pursued - other than fitting the GEV to the margins - a review of the theory is included in *Section 6* and a summary of the essential theory is presented.

4.1 Representing Multivariate Data

Every multivariate data set can be represented as a matrix whose rows refer to each variable, and whose columns refer to each sample. Denoting the multivariate sample by $\boldsymbol{x}$ then

$$\boldsymbol{x}^T = \begin{bmatrix} x_{11} & x_{12} & .... & x_{1p} \\ x_{21} & . & . & x_{2p} \\ . & . & . & . \\ x_{n1} & . & . & x_{np} \end{bmatrix}$$

Several useful summary statistics can be calculated using basic matrix algebra, see for example Kraznowski (1990). For example if $\boldsymbol{1}$ is an $(n \times 1)$ column vector of ones then the column vector of means is given by the matrix dot product

$$\mu^T = \frac{1}{n} \boldsymbol{1}^T . \boldsymbol{x}^T$$

and by writing $\Xi^T = \boldsymbol{1} . \mu^T$, where dot here denotes an outer product the covariance matrix for $\boldsymbol{x}$ is then

$$\Sigma = (1/n)(x - \Xi).(x - \Xi)^T$$

Writing the elements of Σ as $s_{ij}; i = 1,p, j = 1,p$ the correlation matrix is then given by

$$R = D.\Sigma.D$$

where D is a diagonal matrix with $1/\sqrt{s_{jj}}; j = 1,p$ as the leading diagonal.

4.2 MULTIVARIATE KERNEL DENSITY ESTIMATION

Chapter two summarises some graphical tools for exploring univariate statistical data using quantile, kernel density, and residual plots. These non-parametric methods provide estimators which are an essential prerequisite to parametric modelling since they can be used to select the distribution function to be fitted to the data, using the more difficult and expensive parametric methods. For the bivariate case, the simplest graphical method used by engineers is the scatter plot in which the number of occurrences coinciding with a specific range of values, or bin, are recorded. Examples for the DB1 data are given in Ch. 2. In many cases, the number of counts is normalised to simplify estimates of the density corresponding to a specific bin, this often leads to the scatter plot concealing the extreme occurrences.

Whilst scatter plots can be used to estimate the density and cumulative distribution for bivariate random variables, and whilst they provide a general feel for the correlation between the variables, they are an inefficient way of presenting the data because the binning process looses some of the information in the individual samples. More important, the problem is worse in the tails of the distribution where the number of samples (and thus bin count) is small and observations are highly dispersed. A better method of presenting bivariate data, used by Mathiesen & Bitner-Gregersen (1990), is to contour plot the binned data, Fig. 4.2, however the resulting contours are still arbitrary since the bin width is arbitrary and the user has little control over the bias and variance of the density estimate.

Multivariate kernel density estimators are the natural non-parametric choice for estimating the density of a multidimensional sample. The primary use of kernel estimators in this work is to examine the structure of the bivariate pairs of variables (a) before the parametric models are fitted to data using maximum likelihood and (b) after the data has been transformed.

The application of kernel estimators is limited by sample size and the number of dimensions since kernel smoothing uses local averaging of the density, and for a high number of dimensions the neighbourhoods tend to be empty unless the sample is extremely large. Several papers on multivariate kernel estimation have been written, most notably by Epanechnikov (1969) and Rosenblatt (1971). More recently Silvermann (1986) wrote a monograph on the subject, and Scott & Wand (1991) identified the sample sizes required to achieve specified levels of accuracy (defined below). In our case, we are primarily interested in the densities of bivariate samples - although the methods discussed below are generalised to the multidimensional case - for which the sample sizes required to achieve reasonable bivariate density estimates are of the order of hundreds. Therefore typical offshore environmental data samples, which comprise thousands of observations, are sufficiently large for accurate kernel estimation.

A brief outline of the simplest type of kernel estimate is given below and measures of optimality are defined. Although more complex and more accurate kernel methods exist, the simple fixed kernel estimator with an Epanechnikov (or quadratic) kernel is adopted and applied to both simulated bivariate normal data and environmental data recorded by the DB1 data buoy.

4.2.1 MULTIVARIATE KERNEL THEORY

Let $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ be a sample of independent p-dimensional random vectors having unknown probability density $f(\boldsymbol{x})$. Fixed kernel estimators for $f(\boldsymbol{x})$ given a sample $\boldsymbol{x}$ are described by Epanechnikov (1969), where

$$\hat{f}(\boldsymbol{x};b) = \frac{1}{nb^p}\sum_{i=1}^{n} K\left\{\frac{\boldsymbol{x}-\boldsymbol{X}_i}{b}\right\} \qquad [4.1]$$

The window width or bandwith 'b' is a positive deterministic value such that as the sample size $n \to \infty$, then $b \to 0$ and $nb^p \to \infty$. The fixed kernel estimator $K(z)$ is a p-variate density with zero mean vector

$$\int_{R^p} z\,K(z)\,dz = 0 \qquad [4.2]$$

and variance is given by

$$v = \int_{R^p} z^2 K(z)dz \qquad [4.3]$$

Scott & Wand (1991) have examined the statistical properties of this estimator and show how an optimal bandwidth can be defined as that which minimises either the mean absolute error

$$MAE\left\{\hat{f}(x;b)\right\} = E\left[\left|\hat{f}(x;b) - f(x)\right|\right] \qquad [4.4]$$

or the mean integrated error

$$MIAE\left\{\hat{f}(x;b)\right\} = E\left[\int_{R^p}\left|\hat{f}(x;b) - f(x)\right| dx\right] \qquad [4.5]$$

In fact, there is a 'trade -off' between the square bias and the variance which are controlled by the smoothing parameter b. The calculation of optimal window parameters involves the solution of a difficult multidimensional integral to minimise Equations [4.4] or [4.5], which are dependent on the sample size, population density and kernel type. In most cases, closed form solutions for the optimal window are not possible. However, the special case of *unit variance* normal data with a quadratic kernel is solved by Epanechnikov (1969). Defining the optimal window width as b_{opt} then

$$b_{opt} = A(K)n^{-1/(p+4)} \qquad [4.6]$$

where, for the quadratic kernel

$$A(K_e) = \left\{\frac{8p(p+2)(p+4)\left(2\sqrt{\pi}\right)^p}{(2p+1)\,c_p}\right\}^{1/(p+4)} \qquad [4.7]$$

where the coefficients c_p are the volumes of a p-dimensional sphere: $c_1 = 2, c_2 = \pi, c_3 = 4\pi/3$. The optimal window width is therefore inversely proportional to the sample size and converges slowly to zero as $n \to \infty$ at the rate $n^{-1/(p+4)}$. Equations [4.6 & 4.7] are in fact asymptotic approximations for normal data with a quadratic kernel, however, this optimal window width can be used reliably with non-normal data providing the data are not too heavy tailed, as is the case for most pairs of environmental data. In cases where the data are highly non-normal Silverman (1986) warns against the 'blind' use of this optimal width and recommends the data be transformed in the margins for example using $z=log[x]$ to reduce tail length.

An alternative optimal window width has been calculated by Worton (1989) for the case of normal data with a normal kernel, however, the computational penalty associated with calculating the exponential in the kernel outweighs any benefits and consequently for this work

the simple quadratic is used. If the data are lightly asymmetric the use of a radially symmetric kernel, like Epanechnikov's, results in unequal smoothing across each variate. In such cases, when the data are not first transformed the optimal window width h_{opt} should be scaled. One suggestion is to use $h = \sigma h_{opt}$ where

$$\sigma = \sqrt{\frac{1}{p}\sum_{j=1}^{p} s_{jj}} \qquad [4.8]$$

the value σ is then the average of the marginal standard deviations.

4.2.2 QUADRATIC KERNELS

Epanechnikov (1969) proposed a univariate kernel, $k = \frac{3}{4}(1-y^2)\ ;\ |y| \le 1$ with good properties as an estimator. This smoothing function is a simple quadratic defined on the range $-1 \le y \le 1$ such that

$$\int_{-1}^{1} y\,K(y)dy = 0 \quad \& \quad \int_{R^d} K(y)dy = 1$$

This simple univariate kernel points to the use of a multivariate quadratic kernel density for which

$$K(z) = C\{1 - z^T.z\}H\{K(z)\};\quad z^T.z \le 1;\quad i = 1,..,p \qquad [4.9]$$

where T is the transpose of the the vector z

$$z = \frac{x - X_i}{b}$$

$z^T.z$ is the well known quadratic form, Miller (1964), and $H\{.\}$ is the Heaviside function which ensures $K(z) \ge 0$. The form of this kernel for the bivariate case is shown in Fig. 4.3. Substituting this kernel into Equation [4.1] gives the kernel density estimator for the true density

$$\hat{f}(x;b) \;=\; \frac{C}{nb^p}\sum_{i=1}^{n}\{1 - z^T.z\}H\{k(z)\} \qquad [4.10]$$

This kernel density estimator is of the simplest type since the window width is fixed over the sample and equal in each dimension. Other more sophisticated kernels have been suggested, for example, Worton *loc cit* examines the optimal window parameters for adaptive kernels in which the window width is varied over the sample using a pilot estimate of the density. This method gives improved results when the data are highly non-normal since the degree of smoothing is varied according to the local density. However, the method requires longer computational run-times. Another notable method for selecting the optimal window is to use the maximum likelihood estimate; this method again involves further computation and is more suited to studies in which the density estimates are actually used in a reliability calculation. For this work the kernels have been used primarily as a graphical tool for checking the density of the data before and after transformation.

4.3 TRANSFORMATIONS TO MULTIVARIATE NORMALITY

Visual inspection of the scatter plots for the random variables recorded by the DB1 data buoy (Ch. 2) show that in some cases the data are nearly bivariate normal distributed, with the obvious exception that the data are all positive and therefore bounded. In some cases the data can be conveniently transformed into a multivariate normal sample. Box & Cox (1964) present a method for transforming marginal data 'x' into a normal variate 'y' using maximum likelihood and Andrews et al (1971) have extended the method to the multivariate case.

The importance of transforming the data into a multivariate sample is clear when level II and level III reliability methods are examined since each method can directly use the mean and covariance matrix of the transformed random vector, (for example a good account is given by Melchers ,1987) . A further advantage of transforming the data is apparent when fitting parametric models to them since very often a suitable multivariate model does not exist. This is a well-known limitation in the statistical analysis of multi-response data pointed out by Gnanadesikan (1977). As an example, consider the bivariate Weibull models obtained by transformation of the bivariate Gumbel Type I and type II exponential distributions:

type I $$F(x,y) = 1 - e^{-x} - e^{-y} + e^{\{-x-y-\theta xy\}} \qquad x \geq 0; y \geq 0 \qquad [4.11]$$

type II $$F(x,y) = \left(1-e^{-x}\right)\left(1-e^{-y}\right)\left[1+\alpha e^{\{-x-y\}}\right] \qquad x \geq 0; y \geq 0 \qquad [4.12]$$

Setting

$$x = \left\{(v-\xi_1)/\alpha_1\right\}^{c_1} \quad \text{and} \quad y = \left\{(w-\xi_2)/\alpha_2\right\}^{c_2} \qquad [4.13]$$

and then using the Jacobian transformation

$$f(v,w) = J.f(x,y)$$

the resulting distributions $f(v,w)$ are bivariate Weibull. However, the range of correlation that they can model is limited to $-0.4 < \rho < 0$ and $-0.25 < \rho < 0.25$, respectively, see Gumbel (1960). This is too restrictive and if we examine Table 4.1 which shows the vector of means, the covariance matrix, and the correlation coefficient matrix for each pair of random variables recorded by the DB1 we see that these models can only model those values which are either weakly or negatively correlated! (In fact the correlation coefficients for the terms involving direction are meaningless because the directional distributions are multi-modal (Ch. 5)).

Aside from the paucity of multivariate non-normal distributions almost all of the alternatives (many of which are described in Johnson & Kotz (1972)) are defined so that each margin has the same distributional form. Obvious examples are the multivariate exponentials like the Gumbel, and Marshal-Olkin distributions. Multivariate samples of environmental data rarely have the same distributional form on each margin: for example, Bitner-Gregersen & Haver (1991) found the joint distribution significant wave height and zero crossing period had a Weibull marginal distribution for the significant height, and a log-normal distribution for the zero crossing period. It is therefore very unlikely that a single multivariate distribution can be found to model all random variables without first transforming the data. Ultimately, the modelling procedure finally chosen should take into account the merits of each estimator. It may well be the case that a parametric model can be found which fits the data well. However, if this is not the case then the error associated with using a multivariate normal model with transformed variate may be less than, or no worse than, that resulting from imposing an ill-fitting parametric model on the data.

4.3.1 TRANSFORMATION OF THE MARGINS

When data are skewed positively it is often found that the transformations $y = log_e x$ or $y = x^{1/n}$ will produce a symmetric variate y. Both transformations have the effect of lengthening the left tail and shortening the right tail; in particular, as $x \to 0$ then $y \to -\infty$. Box & Cox (1964) consider the transformations

$$y = \begin{cases} (x^{\lambda} - 1)/\lambda & \lambda \neq 0 \\ log_e x & \lambda = 0 \end{cases} \qquad [4.14]$$

This transformation is very flexible and has several important properties which are discussed in both the original paper by Box & Cox (1964) and in Hoaglin et al (1983). The optimal parameter λ can be determined from the data using maximum likelihood in the following way.

Given samples $x_i; i = 1,n$ and the value λ the transformed variate $y_i^{(\lambda)}$, where the superscript (λ) indicates y's dependence on λ, is given by Equation [4.14]. Now, assuming we have transformed correctly to a normal variate with mean μ_λ and standard deviation σ_λ the likelihood is

$$\ell(\mu_\lambda, \sigma_\lambda, \lambda) = (2\pi\sigma_\lambda^2)^{-n/2} exp\left\{-\frac{1}{2\sigma_\lambda^2}\sum_{i=1}^{n}\left(y_i^{(\lambda)} - \mu_\lambda\right)^2\right\}\prod_{i=1}^{n} x_i^{\lambda-1} \qquad [4.15]$$

The maximum likelihood estimates of the mean and standard deviation $\hat{\mu}_\lambda$, $\hat{\sigma}_\lambda$ are given by the sample mean and sample variance respectively

$$\hat{\mu}_\lambda = \frac{1}{n} 1^T . y^T \quad ; \quad \hat{\sigma}_\lambda^2 = \frac{1}{n}(y - \hat{\mu}_\lambda . 1).(y - \hat{\mu}_\lambda . 1)^T \qquad [4.16]$$

Substituting into Equation [4.15] gives the profile likelihood

$$\ell(\lambda) = (2\pi\sigma_\lambda^2)^{-n/2} exp\left\{-\frac{n}{2}\right\}\prod_{i=1}^{n} x_i^{\lambda-1} \qquad [4.17]$$

since

$$\sum_{i=1}^{n}\left(y_i^{(\lambda)} - \mu_\lambda\right)^2 = ns_\lambda^2$$

The maximum likelihood parameter $\hat{\lambda}$ is then found by the one-dimensional maximisation of $\ell(\lambda)$ over the parameter λ. Note, this method also provides a means of checking a sample of data to see if it is already normally distributed since the ML estimate of λ will then be close to one.

A check on this transformation process can be done by: 1, simulating a sample of normal random numbers with known mean and standard deviation; 2, transforming this data using some

known function $y_i = ax_i^b$ to obtain a non-normal sample; and 3, maximising the likelihood to obtain the optimum parameter which transforms the data to near normality. A sample of 1000 normal random numbers was simulated with mean 3.5 and standard deviation 0.4, this data was then transformed using a=2.1 and b=3. This transformation gives a sample which is highly non-normal. The sample likelihood for this data is shown in Fig. 4.4 in which the optimum λ is located near $\lambda = 0.5$. Numerical maximisation of the likelihood gives $\hat{\lambda} = 0.312$ which is close to the actual value of 0.33.

4.3.2 TRANSFORMATION OF JOINT DATA

In the multi-variable case, the data can be transformed to an approximately multivariate normal sample using the above procedure on each margin individually. However, this is an approximation which may result in incorrect modelling of the dependence between the random variables. A more rigourous extension to the Box transformation is given by Andrews et al (1971) who present a transformation process which results in approximate joint normality of a multivariate sample. The advantage of using this transformation is clear when we examine second order reliability analysis methods in which the resistance and loading random variables are transformed into the standard normal u-space, using for example Rosenblatt's transform. These procedures are computationally expensive and can be avoided if the data are modelled as multivariate normal random variables, $N(\mu, \Sigma)$, with known transformation shape and location parameters (λ, ξ). All that is required then is the diagonalisation of the covariance matrix using its orthonormal eigenvectors, Madsen (1986). The new transformed, and now uncorrelated, variables are then given by $z = A^T y$ where A is an orthogonal matrix with columns equal to the orthonormal eigenvectors of Σ.

If $X^T = \{X_{ij}\}^T; \quad i = 1,p, j = 1,n$ is a non-normal multivariate sample, then we seek the transformation which results in a multivariate normal sample Y with mean vector μ and covariance matrix Σ. The procedure for transforming data X into near normal data Y given in Andrews et al (1971) is based on the Box transformation presented for the univariate case above. We shall see later the parameter λ is sensitive to skewness in the data X and that maximum likelihood solutions $\hat{\lambda}_j; j = 1,2$ (found by optimising the bivariate likelihood in Gnanadesikan (1977)) result in a transformed sample with near-zero skewness but variable kurtosis. Non-zero excess kurtosis arises when the fourth moment of the data is not consistent with a jointly normal sample and suggests the largest values in the sample are not adequately transformed for our purposes in which the extremes may be important. To improve the kurtosis Gnanadesikan (1977) suggests including a location parameter in the transformation giving the set of transformations

$$y_{ij}^{(\lambda,\xi)} = \begin{cases} \left(\left[x_{ij}+\xi_j\right]^{\lambda_j}-1\right)/\lambda_j & \lambda_j \neq 0 \\ log_e\left[x_{ij}+\xi_j\right] & \lambda_j = 0 \end{cases} \qquad [4.18]$$

This transformation has two set of parameters for each random variable (λ,ξ) which effectively control skewness and kurtosis, respectively. A procedure for estimating these parameters in the multivariate case is outlined below using a generalisation of the method adopted by Gnanadesikan (1977) - which is for the bivariate case with the one parameter Box transform.

If it is assumed the set of transformations yielding multivariate joint normality is given by Equation [4.18] then the joint density for the random vector x is given by

$$f_X(x) = N(y|\mu,\Sigma,\lambda,\xi).J \qquad [4.19]$$

Where: $N(y)$ is the multivariate normal; J is the jacobian of the transformation; y is the transformed variate ; μ is the vector of means of the transformed random variables; Σ is the covariance matrix of the transformed random variables; and (λ,ξ) are the vectors of transformation parameters.

The general form of the multivariate normal (see Appendix A.1) is given by

$$N(y|\mu,\Sigma) = (2\pi)^{-p/2}|\Sigma|^{-1/2}\, exp\left(-\frac{1}{2}\left(y^{(\lambda,\xi)}-\mu\right)^T \Sigma^{-1}\left(y^{(\lambda,\xi)}-\mu\right)\right) \qquad [4.20]$$

and the Jacobian is the determinant of the matrix of partial derivatives

$$J = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_p} \\ \vdots & & \vdots \\ \frac{\partial y_p}{\partial x_1} & \cdots & \frac{\partial y_p}{\partial x_p} \end{vmatrix} = \prod_{j=1}^{p} \frac{\partial y_j}{\partial x_j}$$

For the transformation in Equation [4.18] the Jacobian is

$$J = \prod_{j=1}^{p} \left(x_j + \xi_j\right)^{\lambda_j - 1} \qquad [4.21]$$

The likelihood for a sample $x_i; i = 1,n$ is then given by

$$\ell(x_1,\ldots,x_n) = \prod_{i=1}^{n} N\left(y_i|\mu,\Sigma,\lambda,\xi\right).J$$

The parameters $\theta = \{\mu,\Sigma,\lambda,\xi\}$ which maximise the sample likelihood are the same as those which maximise the natural logarithm of the likelihood, where

$$log\,\ell = \sum_{i=1}^{n}\left\{-\frac{p}{2}log\{2\pi\} - \frac{1}{2}log|\Sigma| - \frac{1}{2}\left(y_i^{(\lambda,\xi)} - \mu\right)^T \Sigma^{-1}\left(y_i^{(\lambda,\xi)} - \mu\right) + \sum_{j=1}^{p}\left(\lambda_j - 1\right)log\left(x_{ij} + \xi_j\right)\right\}$$

In appendix B it is shown how this simplifies to

$$log\,\ell(\mu,\Sigma,\lambda,\xi) = -\frac{n}{2}log\left|\hat{\Sigma}\right| + \sum_{j=1}^{p}\left(\lambda_j - 1\right)\sum_{i=1}^{n}log\left(x_{ij} + \xi_j\right) \qquad [4.22]$$

Here, the maximum likelihood estimators for the vector of means and the covariance matrix are given by

$$\hat{\mu}^T = \frac{1}{n}\left(1^T . y^{(\lambda,\xi)^T}\right)$$

$$\hat{\Sigma} = \frac{1}{n}\left(y^{(\lambda,\xi)} - \hat{\Xi}\right).\left(y^{(\lambda,\xi)} - \hat{\Xi}\right)^T$$

and the transformation parameters are determined by maximising the log-likelihood over the parameter space. The solutions for the $2p$-unknowns are obtained in the same way as discussed in Ch.3 for the parametric modelling of marginal data using an optimisation program.

Monte Carlo simulation can also be used to check the performance and correct functioning of a multivariate transformation procedure. A sample of 1000 bivariate normal random numbers was generated - using the procedure outlined in appendix C - with *population* mean vector and covariance matrix given by

$$[\mu_1,\mu_2]=[0.0,0.0] \qquad c=\begin{bmatrix}2.0 & 2.2\\ 2.2 & 3.0\end{bmatrix}$$

The *sample* mean vector and covariance matrix was calculated as

$$[\bar{\mu}_1,\bar{\mu}_2]=[-0.0026,-0.0240] \qquad \bar{c}=\begin{bmatrix}2.0324 & 2.2412\\ 2.2412 & 3.028\end{bmatrix}$$

The population density for the data is shown in Fig. 4.5, together with the scatter plot of observed values Fig. 4.6. The density estimate obtained using a quadratic kernel is shown in Fig. 4.7.

The sample x_{ij} was next transformed to sample y_{ij} using the one parameter Box transform on each margin, with $\lambda_1=2$ and $\lambda_2=4$. The sample y_{ij} was highly non-normal with large skewness and kurtosis; the statistics are given in Table 4.2. In the next step, the likelihood given by Equation [4.22] was maximised over the parameter space $(\boldsymbol{\lambda},\boldsymbol{\xi})$ to obtain the optimum parameters which transform the data back to near multivariate normality. The results from the optimisation are also given in Table 4.2 together with the resulting statistics of the transformed sample. The skewness and excess kurtosis are reduced by two orders of magnitude and become near zero as expected for a bivariate sample. The location parameters $\hat{\boldsymbol{\xi}}$ are both zero, as was the case for the original transformation, however the optimum $\hat{\boldsymbol{\lambda}}$ is less close to the expected values of $[0.5,0.25]$ due to the high powers in the transformation.

This simple example indicates the process can transform even highly non-normal data to near multivariate normality with known mean vector, covariance matrix and transformation parameters.

4.3.3 TRANSFORMATION OF THE DB1 DATA

The four years of data recorded by the DB1, summarised in Ch.3, provides a multivariate sample of wind wave and current magnitudes and directions. This data can be used with multivariate inference methods to estimate a model of the joint behaviour suitable for the Monte Carlo reliability analysis described in Ch. 7. The vector of environmental variables examined in this report is taken as

$$\boldsymbol{X}=\{H_s,T_z,U_c,U_w,D_w,D_c\}$$

where the wind and current directions are included in separate brackets because they are multimodal and bounded on $(0,2\pi)$ - note the waves are assumed to be colinear with the wind. This prevents them from being transformed to normal variates and each must be modelled using the Fourier model described in Ch. 5.

Writing the four years of observations as a $(p \times n)$ matrix $\boldsymbol{x}$ we seek the transformation parameters which result in a matrix $\boldsymbol{y}$ of multivariate normal parameters. The maximum likelihood solution for the transformation parameters is that set which maximises the log-likelihood given by Equation [4.22]. The feasibility of solving for the eight parameters $(\boldsymbol{\lambda},\boldsymbol{\xi})$ is improved by first calculating the marginal transformation parameters, and then the parameters for all bivariate pairs, i.e.

$$\{H_s,T_z\}\{H_s,U_w\}\{H_s,U_c\}\{T_z,U_c\}\{T_z,U_w\}\{U_w,U_c\}$$

The values of the transformation parameters obtained from this initial work are then used as the starting points for solving the full eight dimensional problem. This initial analysis also allows the bivariate results to be presented graphically using contour plots of the kernel density estimates.

4.3.4 MARGINAL TRANSFORMATIONS

The sample matrix $\boldsymbol{x}$ for the wind, wave and current magnitudes was used to obtain estimates of the sample mean vector $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{C}$, and correlation matrix $\boldsymbol{R}$. The results are given in Table 4.1 which shows the correlation is most strong for: the significant wave height and zero up-crossing period; and, the wind speed and wave height. The low correlation between the wind speed and the zero up-crossing period is caused by a combination of: the mixing of wind driven and swell seas; and the varying stages of development of the seastates.

One parameter, and two parameter Box & Cox (1964) transformations were used to normalise the data. The statistics for each margin of the matrix $\boldsymbol{x}$ both before and after transformation are given in Table 4.3 and Table 4.4, together with the maximum likelihood shape and location parameters. Plots for the sample likelihoods of the one parameter transformations are given in Fig. 4.8 which shows the maxima are all well defined. The results for the one parameter case show the transformed data have near zero skewness in all cases other than significant wave height, which starts with a skewness of 1.64 and has a skewness of

0.265 after transformation. The results for the excess kurtosis are not as good with the values increasing in two cases.

The two parameter transformation gives improved results: however, it does not confirm Gnanadesikan's claim that the location parameter responds more to kurtosis in the data than the shape parameter. This time the skewness is reduced more than before and the changes of excess kurtosis are similar to those obtained in the one parameter case. Note the results for the mean current speed suggest the un-transformed data are near normal distributed before transformation.

4.3.5 BIVARIATE TRANSFORMATIONS

The values of the transformation parameters for each bivariate pair are given in Table 4.5 together with the skewness and excess kurtosis of each margin. The skewness of the margins for each transformed bivariate pair are all reduced to near zero, with the largest value of 0.11 occurring on the *Hs* margin of the {*Hs*,*Tz*} pair. The excess kurtosis values have also reduced to near zero; with the exception of the current, which started as very near normal before transformation. The effect of transforming each pair can be assessed graphically by plotting the kernel density estimates for the sample both before and after transformation. Figures 4.9 to 4.14 show both the natural logarithm of the kernel density for the data before transformation, and the natural logarithm of the kernel density after transformation. The log of the density has been used to exaggerate the behaviour around the periphery of the density. Contours are plotted at 10 equidistant elevations with intervals determined by the maximum elevation of the modal point.

The kernel density plot for the joint distribution of significant wave height and zero up-crossing period is shown in Fig. 4.9(a). The irregularity of the outer contours, highlighted by taking the log, is caused partly by the density of grid used for the generation of the contour plot and partly by the low numbers of observations in the extremes. The apparent cutoff of the contours at the bottom of each plot is caused by the positioning of the plot frame. The second kernel plot, Fig. 4.9(b), confirms the distribution is near bivariate normal with low correlation between the transformed *Hs* and *Tz*.

The kernel for the joint mean wind speed and significant wave height is shown in Fig. 4.10(a). This set of contours shows the near quadratic dependence of *Hs* on the wind speed, suggesting the sample may be structured, and therefore benefit from the use of a likelihood which reflects this dependence. After transformation, the correlation between *Hs* and *Uw* is strong, compared with the *Hs* and *Tz* data, but the contours are not symmetric about the

principal axes of the data. On the other hand, Table 4.5 shows the skewness and excess kurtosis for this bivariate pair are close to zero on both margins.

Kernel estimates for the remaining pairs are plotted in Fig. 4.11 to 4.14. These all indicate the transformation has reduced the data to near bivariate normal, weakly correlated, variables.

It is important to note that the kernel density plots used in this work are intended to be illustrative and consequently the simplest form of kernel has been used. Scott & Wand (1991) and Worton (1989) have shown both numerically and analytically that improved results can be obtained when adaptive methods are used. Furthermore, since the multivariate kernels are generally symmetric, it is best to first transform the data so that it too is near radially symmetric; this would ensure the degree of smoothing is the same in all directions (dimensions). One way of doing this would be to use Andrews' method and the eigenvalue transformation described previously. The kernels can then be transformed back into the original x-space variables if required.

4.3.6 MULTIVARIATE TRANSFORMATION

The solutions for the ML parameters $\hat{\lambda}, \hat{\xi}$ for the marginal distributions were used as the initial points for the numerical optimisation of the likelihood for the full eight dimensional problem in which the transformation parameters are solved for the vector {*Hs,Tz,Uw,Uc*}. The optimisation was performed using a NAg routine E04JAF in which simple bounds on the parameters can be modelled to allow for the condition $x + \xi > 0$. The solution process for a sample of 6000 observations with four variables was straightforward, and robust, taking some 3000-4000 seconds of CPU on a 486/33 PC. The results are summarised in Table 4.6 which gives: the ML location and shape parameters; the vector of means for the transformed data; the covariance matrix for transformed data; and the sample statistics for each margin. This set of transformation parameters provides a complete probabilistic description for the magnitudes of the jointly occurring wind speed, significant wave height, zero-up-crossing period, and current speed in a format which can be used in a reliability calculation.

4.4 SELECTION OF THE 'BEST' TRANSFORMATION PARAMETERS FOR THE DB1 DATA

The transformation shape and location parameters for the marginal, bivariate, and complete vector cases have been given in Tables 4.3 to 4.6. In most instances the values for each margin are similar for all three cases. This confirms Gnanadesikan's (1977) comment that use of the marginal parameters should result in near multivariate normality. Our main objective is to

model the moderate and extreme values with greatest accuracy and consequently it seems reasonable that we should use the results from all three cases to select those parameters which result in the lowest marginal skewness and excess kurtosis. The 'best' set of transformation parameters are then those given in Table 4.7.

Theoretically, the mixing of marginal bivariate, and multivariate estimates of the location and shape parameters is not unreasonable when the values chosen fall within the α% confidence limits around the ML point obtained from the multivariate optimisation. But, of course, the larger α needs to be for this to be true the more questionable the use of the "mix 'n' match" approach.

4.4.1 BEHAVIOUR OF THE MODEL IN THE EXTREMES

The transformation of the marginal and multivariate data x into our normal variates y is an approximate process. Thus far we have examined the normality of y using both the skewness and kurtosis of each margin; and kernel density plots for each bivariate pair. A further check on the quality of the transformed model must be made to ensure the extremes are correct since they make the biggest contribution to structural failure probabilities, Ch. 1. The approach taken by Mathisen & Bitner-Gregersen (1990) for the bivariate {*hs*,*tz*} pair was to plot contours of the normalised deviations between the proposed parametric model, and the observed data. This gives a good subjective feel for the quality of fit and highlights regions where the model does not agree with the data. On the other hand, the method does not give a qualitative measure for the accuracy of the extreme occurrences.

A more rigourous method for assessing the quality of fit for the largest values would be to compare the marginal extremes, for some return period T_R, obtained from both fitting an asymptotic distribution to say the monthly or annual maxima, and the extremes predicted by the transformed normal model.

The choice of definition for return period is examined by Tucker in Department of Energy (1990b) who suggests that it be defined by:

> *the return period of a stated value of a metocean variable is the average period of time between exceedances of that value.*

This definition of return period is more in keeping with the way metocean data are recorded than the statisticians quantile estimate of return period, Ch. 3. The definition is therefore used

below in the calculation of the return period estimates for marginal wind speeds, wave heights, and current speeds.

After the data have been transformed, using either the Box & Cox (1964) or Andrews et al (1971) method, the marginal distributions are normal with density

$$f(y)=\frac{1}{\sigma_y\sqrt{2\pi}}\,exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}$$

Here, the transformed variate y is given by

$$y=\begin{cases}log(x+\xi) & \lambda=o\\ \left\{(x+\xi)^{\lambda}-1\right\}/\lambda & \lambda\neq o\end{cases}$$

The distribution for the untransformed variate is thus

$$f_x(x)=f_y(y)\left|\frac{dy}{dx}\right|$$

$$f(x)=\frac{(x+\xi)^{\lambda-1}}{\sigma_y\sqrt{2\pi}}\,exp\left\{-\frac{1}{2}\left[\frac{(x+\xi)^{\lambda}-1-\lambda\mu_y}{\lambda\sigma_y}\right]^2\right\} \qquad [4.23]$$

We are essentially interested in how well Equation [4.23] models moderate and extreme values. Defining x_R as the return period value of the metocean parameter we must solve

$$\int_{-\infty}^{x_R} f(x)dx = Pr\{X\le x_R\} \qquad [4.24]$$

Following the discussion above for return period, for the DB1 data we have

$$P_{50}=Pr\{X\le x_R\}=1-\frac{1}{50*365.25*8}=0.999993155$$

for the 50 year survivor probability, and

$$P_{100} = Pr\{X \le x_{R100}\} = 1 - \frac{1}{100 * 365.25 * 8} = 0.9999965777$$

for the 100 year return value. The solution for the unknown value of x_R in Equation [4.24] was done numerically using a root finding algorithm.

A second method has been used to estimate the return period values for the metocean parameters. In Ch. 3 the parameters of both a generalised extreme value model, and a three parameter Weibull model are given. These models correspond to monthly maxima models which have been fitted using maximum likelihood. As has been noted earlier, these models are questionable when so few data are available and no account is taken for the seasonal cycles in the monthly maxima. Nevertheless, we might still expect them to give representative results since they clearly fit the data. The solution for the return period values for these two distributions are given by solving

$$F(x_R) = \frac{1}{12 * T_R} \qquad [4.25]$$

where $F(x_R)$ is either the Weibull or GEV cumulative distribution fitted to the monthly maxima, and T_R is the return period in years.

Each method described above has been used to estimate the 50 year and 100 year return period wind speed, significant wave height, and current speed. Again, with so few data we cannot expect to extrapolate to these return periods without some bias. However, we can check the results for H_s and U_w against the Department of Energy (1990) contour plots, which give $H_{s50} \approx 15.5m$ and $U_{w50} \approx 38.5m/s$.

The wind speed value U_{w50} is the hourly mean wind speed at 10 m above still water level, whereas the DB1 data was taken as the average of two records made at 6.0 and 8.7 m above still water. Using the usual wind speed power laws, this suggests the DB1 values should be factored by ~ 0.955 to allow for the duration of the time average, however this effect cancels the effect of the elevation. The average DB1 wind speeds are therefore taken as equivalent to the hourly mean winds.

The 50 and 100 year return period values for the significant wave height, wind speed, and current speed are given in Table 4.8 for each modelling method. The population estimated wind

speed and significant height compare well with the values suggested by the Department of Energy Guidance notes. The current extremes cannot be compared however, since we have made no allowance for the harmonic tidal components.

In all cases the GEV model gave the worst results and it is difficult to identify why this was the case. Correlation between monthly maxima may contribute to the inaccuracies, but we would then expect the Weibull model to give poor results as well. In fact in two cases the Weibull model agrees closely with the transformed normal model. The irregularities need further investigation.

4.4.2 MOST PROBABLE ZERO-UP-CROSSING PERIOD CONDITIONED ON THE SIGNIFICANT WAVE HEIGHT

The two most important variables in a stochastic response analysis are the significant wave height *Hs* and the wave zero-up-crossing period *Tz*. To correctly predict systems reliabilities it is essential the multivariate normal model accurately predict the most probable *Tz* for the moderate to extreme values of significant wave height. Any bias in the extremes would have a serious effect on a platform's responses.

In normal space the expected value of x_1 conditioned on x_2 is given by

$$E[y_1|y_2] = E[y_1] + s_{12} / s_{22}(y_2 - E[y_2]) \qquad [4.26]$$

where s_{ij} are the elements of the covariance matrix for $\mathbf{y}$ (for more general multivariate results see Appendix A.). For a normal variate the modal and expected values are coincident therefore we can write

$$\hat{y}_1 = E[y_1|y_2]$$

For the Box transformation

$$y = \left\{(x+\xi)^{\lambda} - 1\right\} / \lambda$$

Setting a value of x_2, the value y_2 is then calculated. The most probable value $\hat{y}_1$ given by Equation [4.26] can then be transformed back to *x*-space using the inverse Box transformation

$$\hat{x} = \{\lambda y + 1\}^{1/\lambda} - \xi \qquad [4.27]$$

Using the data in Table 4.7 the distribution for the modal zero-up-crossing period conditioned on some level of significant height can be calculated. For example, the 50 year return period *Hs* is 15.02m, this value is first mapped into y space using the maximum likelihood shape and location parameters

$$y_{Hs} = \{(15.02 - 0.3686)^{0.2703} - 1\} / 0.2703 = 3.9438$$

then the expected value of *Tz* conditioned on *Hs* is calculated in *y*-space

$$E[y_{Tz} | y_{Hs} = 3.9438] = 1.278 + \frac{0.4452}{0.6421}(3.9438 - 0.443) = 3.7053$$

hence the modal zero-up-crossing period is given by

$$\hat{T}_z = (0.4485 * 3.705 + 1)^{2.2296} + 4.5 = 13.3 secs$$

This value can be checked using the limits on wave steepness given by the Department of Energy (1990) who give

$$3.2Hs^{1/2} < Tz < 3.6Hs^{1/2}$$

This gives a range of 12.4 to 14.0 seconds, with central value 13.2. This value is remarkably close to the prediction of the multivariate model modal value of 13.3.

The variation of $\hat{T}_z$ predicted by the multivariate model over the whole range of *Hs* is shown in Fig. 4.15 (top). If this line is superimposed on the scatter plot for *Hs* and *Tz* (Ch. 3) then it is found the model predicts the correct modal period over the complete range of values in the scatter plot.

4.4.3 MOST PROBABLE WIND SPEED CONDITIONED ON THE WAVE HEIGHT

The most probable wind speed for a given seastate significant wave height has also been examined. Fig. 4.15 (bottom) shows the modal wind speed conditioned on the significant wave height together with the upper bound given by the assumption of fully developed seas

$$Hs = 0.022716 U_{10}^{\ 2}$$

In this quadratic model U_{10} is the hourly mean wind speed at 10m elevation. At high wind speeds the model prediction generally falls below the quadratic model because the fetch distances are rarely long enough to develop the seas fully in high winds. Again, when the multivariate model is superimposed on the scatter plot for *Hs* and *Uw* (Ch.3) the model prediction is found to be good.

4.5 TRANSFORMATION TO UNCORRELATED STANDARD NORMAL VECTORS

In Ch. 7 the multivariate normal model for the joint wind, wave, current and structure random variables is transformed into standard normal space U. This enables classical first order reliability methods to be used in the calculation of failure probabilities. To uncorrelate the variables the principal component axes must first be identified, for example using the method in Kraznowski (1988). This is done by first diagonalising the covariance matrix Σ_y of the variables y (which are obtained from the Box transformation on x). The linear transformation $z = A^T y$ is then used to obtain a set of uncorrelated variables Z.. Here A is the matrix containing the eigenvectors of Σ_y. The diagonal covariance matrix (i.e. uncorrelated with $cov_{ij} = o \; for \; i \neq j$) of Z is then given by

$$\Sigma_z = A^T \Sigma_y A \qquad [4.28]$$

Note that the terms on the diagonal of Σ_z are also the eigenvalues of Σ_y. Since the transformation from $y \rightarrow z$ is linear, we have

$$\mu_z = E[z] \;\; = \;\; A^T \mu_y \qquad [4.29]$$

Whilst Z is uncorrelated it must be scaled on each margin to obtain standard normal variates U where

$$u = \Sigma_z^{-1/2}(z - \mu_z) = \Sigma_z^{-1/2} A^T (y - \mu_y) = (A^T \Sigma_y A)^{-1/2} A^T (y - \mu_y) \qquad [4.30]$$

This relationship must be inverted to obtain the format required in Ch. 7, noting $A.A^T = I$ and $\Sigma^{1/2}\Sigma^{-1/2} = I$ we have

$$\begin{aligned} y &= A\Sigma_y^{1/2} u + \mu_y \\ &= A\left(A^T \Sigma_y A\right)^{1/2} u + \mu_y \\ &= Ku + \mu_y \end{aligned} \qquad [4.31]$$

This expression enables us to calculate a y co-ordinate for any u co-ordinate. The transformation matrices and vectors $\Sigma_y, A, \Sigma_z, \mu_y, \mu_z, \Sigma^{1/2}, \Sigma^{-1/2}$ are all given in Table 4.9. These values, together with the Box transformation parameters in Table 4.7 enable us to map between X, Y, and U space as required in the reliability calculations. As an example, Fig. 4.16 (top) shows 300 pairs of $\{H_s, T_z\}$ observations after transformation using the Box & Cox (1964) method. The transformation matrices given in Table 4.9 were used to transform the data into U-space variates, Fig. 4.16 (bottom). The data are now approximately uncorrelated with zero mean values on each margin.

4.6 A REVIEW OF MULTIVARIATE EXTREME VALUE THEORY

The asymptotic theory of extremes was reviewed for this work with the intention of using it in an offshore structure reliability study. Unfortunately, it was not possible to use the method in earnest with the sample of DB1 data for two reasons. First, the GEV models fitted to the monthly maxima were not able to accurately predict extreme values for the 50 and 100 year return period design events. The probable cause was seasonal correlation in the monthly values and the limited sample size of only four years. The second reason is the models are for virtual events which may be physically inadmissible. The basis of multivariate extreme value theory is given in Galambos (1978) so this review concentrates on the more recent developments and applications. A summary of the main bivariate models is given together with appropriate estimation procedures.

The current procedures can be classed as either *intrinsic* (often called non-parametric) or *parametric* estimation for componentwise maxima. By componentwise maxima we mean the largest value in each margin observed during either some time period, or a large number of observations. The implications of this are important in design since the theory is not modelling the simultaneously occurring values. If, for example, the time period is a year then the component maxima for the wind speed, significant wave height, and current speed, may occur during different seastates, days, or even months. The models must therefore be thought of as bounds on the true joint behaviour and the implications need to be examined thoroughly.

The theory and estimation of univariate extremes has received, and still continues to receive (Davison & Smith (1990), Hosking et al (1985), Tawn (1988a)), considerable attention in the statistical and engineering literature. On the other hand, the estimation of multivariate extremes is comparatively recent and many problems are still to be solved. This subject is of potential importance in the design of offshore structures but does seem to require a large sample of annual maxima to be practical. It is therefore suggested the method be considered for use with hindcast data in a future study.

4.6.1 MOTIVATION AND NOTATION

In most practical cases only a small number of extremes can be observed which means the parametric estimation is performed using a very small sample size. A small sample results in large uncertainty in the estimated model parameters, therefore further information is required in the estimation process. This information is provided by the asymptotic theory of extremes which, broadly speaking, tells us that providing suitable normalisation coefficients exist the distribution of the largest in a sample tends to one of the three Fisher-Tippet distributions as

the sample size tends to infinity. This greatly simplifies the estimation process since by choosing the largest from several samples of observations their distribution will be one of the Fisher-Tippet distributions. These three distributions can be generalised using the Jenkinson-von Mises distribution as (Ch. 3)

$$G(x;\mu,\sigma,k)=exp\left\{-\left[1-k\left(\frac{z-\mu}{\sigma}\right)\right]^{1/k}\right\} \qquad [4.32]$$

For the bivariate and multivariate case we seek that set of distributions which characterise the *vector* extremes. However, since no natural ordering exists for dimensions greater than one it is necessary to define the model in terms of the componentwise (or virtual) events.

After Galambos let the random vector $X=\{X_1,.....X_p\}$ and $X_i; i=1,...,n$ be the matrix of n observations. The order statistics of the t^{th} component are written as $X_{t,1:n} \le X_{t,2:n} \le .....\le X_{t,n:n}$. The smallest value of the t^{th} component is denoted as $W_{t:n}=X_{t,1:n}$ and the largest as $Z_{t:n}=X_{t,n:n}$. The problems of interest are the asymptotic joint distributions of $W_{t:n}=\{W_{1:n},W_{2:n},.....,W_{p:n}\}$ and $Z_{t:n}=\{Z_{1:n},Z_{2:n},.....,Z_{p:n}\}$.

The joint population distribution of the random vector X is defined by

$$F(X)=Pr\{X\le x\}=Pr\{X_1\le x_1,.....,X_p\le x_p\}$$

In a similar manner to the theory of univariate extremes we seek vectors a_n and $b_n>0$ such that

$$Pr(Z_n<a_n+b_nz)=G_n(a_n+b_nz)$$

converges to a stable p-dimensional extreme value distribution function $H(z)$.

4.6.2 Properties of Multivariate Distributions

To begin it is necessary to define some simple properties of multivariate distributions. Let X be a random column vector and $x^T=\{x_1,..,x_p\}$ be an observation or point in p-dimensional Euclidean space, then

$$F(X)=Pr\{X\le x\}=Pr\{X_1\le x_1,.....,X_p\le x_p\}$$

Clearly, $F(x)$ is a non-decreasing function of x, and as $x_j \to \infty$, $F(x)$ tends to an $(p-1)^{th}$ dimensional distribution. That is, the distribution of X with X_j removed. The process of removing successive components finally results in the marginal distribution of the remaining variable consequently $F(x)$ defines all marginal distributions $F(x)$. On the other hand the marginal distributions only specify bounds on the multivariate distribution, Galambos (1978).

4.6.3 THE DISTRIBUTION OF VECTOR EXTREMES

Let $F(X)$ be the multivariate distribution function of the independent, identically distributed random vector X. The maxima of the components $X_i; i = 1,...,n$ have the joint distribution function

$$Pr\{X_{1:n} \le x_1,.....,X_{p:n} \le x_p\} = F^n(x)$$

in the limit as $n \to \infty$ we get the asymptotic distribution

$$H(x) = \lim_{n \to \infty} F^n(a_n + b_n x)$$

Here the vectors a_n and b_n are the sequence of the standardising coefficients such that $H_n(a_n + b_n x)$ converges to a non-degenerate distribution function $H(x)$:

$$H_n(a_n + b_n x) \longrightarrow H(x)$$

In fact, Galambos (1978) shows the the vectors a_n and b_n are determined from the margins of $F(X)$ using univariate extreme value theory. Recent research has concentrated on defining the mathematical forms of $H(x)$. For example, the general structure of bivariate extremes was identified by Tiago de Oliveira (1962) and Sibuya (1960). More recently: Pickands (1981), and Tiago de Oliveira (1989b) have identified a number of joint distributions for bivariate extremes based on *non-parametric* estimators for the dependence structure; and Tawn (1990), and Coles (to be published) have identified several bivariate and multivariate distributions based on *parametric* estimators for the dependence structure.

4.6.4 BIVARIATE EXTREME MODELS

Let $\{x_i, y_i\}; i = 1,....,n$ be a sequence of independent indentically distributed pairs drawn from the bivariate random process (X,Y) having distribution $Pr\{X \le x, Y \le y\} = F(x,y)$. Following Tiago de Oliveira, Pickands, and Tawn we must use the notion of componentwise ordering, that is

$$Z_{1:n} = X_{1,n:n}; \quad Z_{2:n} = X_{2,n:n}$$

where

$$Z = \{Z_{1:n}, Z_{2:n}\}$$

One unsolved problem with this approach is that in general some of the virtual observations Z are not physically realisable as vector random observations. An obvious example is the joint distribution of individual wave height and period which is bounded by the breaking wave criterion. Since this problem is unsolved we shall continue to use the component approach and ignore this limitation.

As previously mentioned it is assumed that attraction coefficients $a_{i:n}$ and $b_{i:n}$ can be located for the margins such that

$$\lim_{n \to \infty} F^n(a_{1:n} + b_{1:n}x, a_{2:n} + b_{2:n}y) \to H(x,y)$$

We have seen the bivariate extreme value distribution $H(x)$ uniquely defines the marginal distributions of x and y which we also know will be GEV distributed. It is well known that the Fisher-Tippet distributions are interchangeable through transformation and it is a matter of convenience which form is assumed for the margins of the bivariate extremes model. In this work we follow Tawn, and Pickands who assume the margins are transform into unit exponential survivor functions

$$Pr\{X > x\} = e^{-x}, \quad Pr\{Y > y\} = e^{-y}; \quad (x > 0, y > 0) \qquad [4.33]$$

Pickands (1981) shows that the bivariate extreme value survivor model must have the form

$$G(x,y) = exp\left\{-(x+y)A\left(\frac{y}{x+y}\right)\right\}; \quad (x > 0, y > 0) \qquad [4.34]$$

where the function $A(w)$ is called the dependence function of the pair (x,y). Note this is not the simple function described by Galambos, and Tiago de Oliveira in their earlier works. Assuming we have transformed the marginal data to unit exponentials the bivariate model in Equation [4.34] can be estimated using either intrinsic or parametric methods.

The dependence function $A(w)$ must have several properties for

$$G(x,y) = exp\{-(x+y)A(w)\}$$

to be a bivariate extreme value survivor function:

- (i) $A(0) = A(1) = 1;$
- (ii) $max(w, 1-w) \le A(w) \le 1 \quad (0 \le w \le 1);$
- (iii) $A(w)$ is convex on $(0 \le w \le 1)$
- (iv) $A(w)/(1-w)$ is a non-decreasing function and $A(w)/w$ is a non-increasing function

If we examine the joint survivor function for (x,y) then clearly for the upper bound on $A(w)$ gives

$$G(x,y) = exp\{-(x+y)\} = e^{-x}e^{-y} \qquad [4.35]$$

that is X and Y are independent. The lower bound in condition (ii) corresponds to complete dependence $X = Y$. The correlation between X and Y is given by

$$\rho = \int_0^1 \frac{dw}{A(w)^2} - 1 \qquad [4.36]$$

This can be compared with the usual estimator of correlation

$$\rho^* = \frac{\sum_{i=1}^{n}(x_i - E[x])(y_i - E[y])}{\left[\sum_{i=1}^{n}(x_i - E[x])^2 \sum_{i=1}^{n}(y_i - E[y])^2\right]^{1/2}} \qquad [4.37]$$

where the quantity

$$\sqrt{n}\frac{\rho^* - \rho}{b}$$

is asymptotically normal for a given b. This allows us to test for independence $(\rho = 0; b = 1)$ using the acceptance region

$$\sqrt{n} \leq \Phi^{-1}(1-\alpha)$$

where $\Phi(.)$ is the standard normal distribution evaluated using the error function, see Abramowitz & Stegun (1965). Whilst the properties of $A(w)$ have been examined above no estimators have yet been presented. Generally two approaches are adopted for estimating $A(w)$ namely the intrinsic (or non-parametric) estimators developed by Tiago de Oliveira and Pickands, and the parametric estimators preferred by Tawn. In this review both types of estimator are examined.

4.6.5 INTRINSIC ESTIMATORS OF DEPENDENCE

Several intrinsic estimators for the dependence function $A(w)$ have been proposed. Perhaps the best known are those developed by Pickands (1981)

$$A^*(w) = n\left\{\sum_{i=1}^{n} z_i(w)\right\}^{-1} \qquad [4.38]$$

$$z_i(w) = min\left\{(1-w)^{-1}x_i; w^{-1}y_i\right\}; i = 1,\ldots,n, 0 \leq w \leq 1 \qquad [4.39]$$

and more recently Tiago de Oliveira

$$A^*(w) = 1 - \frac{c(\alpha)}{n}\sum_{i=1}^{n} min\left\{\frac{1-w}{\alpha + x_i}, \frac{w}{\alpha + y_i}\right\} \qquad [4.40]$$

where $\alpha = 1/n$

$$c(\alpha) = \frac{1}{1 + \log n}$$

Given a realisation from a bivariate random process these estimators are very simple to determine and for this reason they are used to explore the structure of a bivariate sample. One simple check on the correctness for the estimated dependence function is to integrate its inverse squared using Equation [4.38], and compare this estimator of the correlation with the sample estimator, Equation [4.37].

4.6.6 PARAMETRIC MODELS OF DEPENDENCE

During the past thirty years a number of parametric dependence functions have been developed, Sibuya (1960), and Gumbel (1960). Parametric estimators for $A(w)$ must satisfy similar conditions to the intrinsic estimators

$$A(0)=A(1)=1;\ \ -1\le A'(0)\le 0;\ \ 0\le A'(1)\le 1;\ \ A''(w)\ge 0;\ \ (0\le w\le 1)$$

Four dependence models are identified in Tawn and have been summarised in Table 4.10, two are mixed models, and two are logistic models. The two mixed models are symmetric about $w=0.5$, making X and Y interchangeable, and the logistic models are either symmetric or asymmetric about $w=0.5$ depending on the choice of the symmetry parameters. As an example we shall examine how to estimate the parameters of the asymmetric logistic model which has three parameters in its dependence function.

4.6.7 SYMMETRIC LOGISTIC DEPENDENCE FUNCTION LIKELIHOOD

In Ch. 3 the joint distribution function $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ for the vector of observed univariate random variables is used to define the likelihood function for the observed data $\ell(\theta;x)$. Given a bivariate sample $\{x_i,y_i\ ;\ i=1,\ldots,n\}$ the joint distribution of the observed value is

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=f(x_1,y_1)\times f(x_2,y_2),\ldots\ldots,\times f(x_n,y_n)$$

which gives the log-likelihood

$$\log_e \ell(\theta;x)=\sum_{i=1}^{n}\log\{f(x_i,y_i;\theta)\}$$

The joint survivor function for the asymmetric logistic model is given by

$$G(x,y)=exp\left\{-(1-\theta)x-(1-\Phi)y-\left(\{x\theta\}^{1/v}+\{y\theta\}^{1/v}\right)^{v}\right\} \qquad [4.41]$$

taking the second partial derivatives gives the joint survivor density

$$g(x,y;\theta)=\frac{\partial^2 G(x,y;\theta)}{\partial x \partial y}$$

The log-likelihood or the sample is then given by

$$log_e\, \ell(\theta;x)=\sum_{i=1}^{n} log\{g(x_i,y_i;\theta)\}$$

The maximum likelihood set of parameters is then found by maximising the likelihood on the bounded feasible region defined by the ranges of the model parameters. In Ch. 3 it is shown that the finite differences method gives acceptable accuracy for the hessian matrix of the log-likelihood and furthermore to avoid calculating the second derivatives analytically the information matrix is also calculated using finite difference estimates of the gradients and curvatures.

Unlike the boundary constraints for the asymmetric mixed model which define a closed region the boundary constraints for the asymmetric logistic model only restricts the search to the region defined by

$$0 \le \theta, \quad \Phi \le 1, \quad 0 \le \nu \le 1$$

This makes solution of the maximum likelihood logistic parameters rather more difficult to obtain.

Complete dependence for the asymmetric logistic corresponds to the $\theta = \Phi = 1$ and $\nu = 0$, complete independence corresponds to $\theta = 0$ or $\Phi = 0$ or $\nu = 1$. Therefore a suitable starting point for locating the ML parameters is $\theta = \Phi = \nu = 0.5$. There are problems with the independence case because the maximum likelihood estimators exhibit NON-REGULAR behaviour at independence. The effect of this is that the expected Fisher information is infinite and cannot be used to assess the variance, covariance matrix for the model parameters. Tawn (1988b) discusses this problem in relation to the logistic model in some depth however we can avoid the problem of non-regular estimation by ensuring we do not use the models with independent random variables for which there is a trivial solution. One method of checking whether X, and Y are in fact only weakly dependent or independent is to examine the non-

parametric estimator $A(w)$. In this way we can easily assess whether Equation [4.41] is a sensible model.

4.7 CONCLUSIONS

Multivariate kernel density methods have been used to examine the bivariate structure of the DB1 environmental data. The results confirm the conclusions in Ch.2 that the data can be modelled as unstructured providing a suitable set of transformations and population model can be found.

The kernel densities used for this study were of the simplest type. Improved estimates can be obtained by using adaptive kernels in which the degree of smoothing is adjusted to the local density.

Univariate and multivariate transformation methods have been examined with a view to mapping the joint wind, wave and current data into a standard normal space. One of the best transformations is the Box & Cox (1964) transformation which includes the log-normal model as a special case. This transformation can be used on each individual margin to reduce a multivariate sample to multivariate normal.

The Box & Cox transformation is powerful enough to transform even highly non-normal data. Furthermore, the shape and location parameters respond strongly to the skewness and kurtosis in the data. This ensures the models are capable of modelling the extremes since these moments are dominated by the tails of the data.

The transformation of the data to a normal model has considerable advantages when used in level III reliability studies. Most important is that it makes the Rosenblatt transformation redundant. Second is that the method is simple to implement and, when used with a good non-linear optimiser, is very efficient.

It is necessary to extend the number of parameters in the multivariate Andrew's transformation to ensure the model is scale invariant and responds to kurtosis. A location and scale parameter have therefore been added to the Box & Cox transformation.

The population model has been assembled from a mixture of ML parameters obtained from the marginal, bivariate,and multivariate datasets. The criterion used for selecting the best set of transformation parameters has been the accuracy with which the resulting population model can predict the marginal 50 year return period values. This approach is reasonable providing the chosen set of parameters lies within the confidence limits obtained from the likelihood

analysis of the full sample of data. In fact the parameters obtained from the marginal, bivariate, and multivariate samples were generally in close agreement.

The margins of the multivariate transformed normal population model have been examined to check the accuracy with which it can predict extreme values. The 50 year return period estimates for the wind speed (33 m/s), current speed (1.0 m/s), and the significant wave height (15 m) were found to be close to those values recommended by the Department of Energy suggesting the model is sufficiently accurate for use in:

operations analysis
long-term fatigue assessments
reliability studies of the extreme events

Checks on the variation of the modal values for zero crossing period conditioned on wave height and wave height conditioned on wind speed indicate that the model can be used to correctly predict values at the 50 year return period levels. This is especially important for the significant height and zero up-crossing period case because compliant structures are sensitive to wave period.

Overall the transformation approach has several advantages over the Rosenblatt transformation. This study shows the population models obtained from the application of likelihood theory are accurate even for the 50 year return period values. Furthermore, the general structure of the fitted model matches the behaviour seen in the scatter plots given in Ch. 2. The only aspect of the model that requires improvement is the forward face of the joint distribution for significant wave height and zero-up-crossing period where the breaking wave condition results in a sharp increase in the density on the forward face of the joint distribution. However, it is likely that this aspect of the model can be improved by using a structured likelihood.

APPENDIX A

A.1 PROPERTIES OF THE MULTIVARIATE NORMAL DISTRIBUTION

In the preceding work, and in the reliability analysis discussed in Ch.7, the multivariate normal distribution is used repeatedly. For example the multivariate normal appears in Section [4.3.2] and the conditional expectations for the transformed normal model are use in Section [4.4.2]. Several useful properties of the normal distribution are presented below in matrix format for the general p-dimesional vector case.

Let $X_p = \{x_1,....,x_p\}$ be a random vector with mean vector A_p and a symmetric, positive definite covariance matrix

$$M_p = E\left[\left(X_p - A_p\right)\left(X_p - A_p\right)^T \right]$$

If X_p is multivariate normal then its probability density function is

$$f_p\left(X_p\right) = \frac{Exp\left[-\tfrac{1}{2}\left(X_p - A_p\right)^T \; M_p^{-1}\left(X_p - A_p\right)\right]}{(2\pi)^{p/2}\sqrt{\left|M_p\right|}} \qquad \text{[A1]}$$

The conditional distribution of $X_{p-r} = \{x_{r+1}, x_{r+2},....,x_p\}$ given $X_r = \{x_1,....,x_r\}$ is then defined by

$$g\left(X_{p-r}/X_r\right) = \frac{f_p\left(X_p\right)}{\int_{-\infty}^{\infty} f_p\left(X_p\right)dX_{p-r}} \qquad \text{[A2]}$$

The conditional density function $g()$ is the $(p\text{-}r)$ dimensional density function of the random variables X_{p-r} with X_r held as constant parameters. If the covariance matrix M_p is partitioned so that the upper left square contains the covariances of X_r and the lower right square contains the covariance of X_{p-r} then the corresponding partitions in the inverse of M_p can be determined using the identity

$$M_p M_p^{-1} = \left[\begin{array}{c|c} M_r & T_r \\ \hline T_r^T & S_{p-r} \end{array}\right]\left[\begin{array}{c|c} Q_r^{-1} & R_r \\ \hline R_r^T & P_{p-r}^{-1} \end{array}\right] = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & I_{p-r} \end{array}\right]$$

Solving for the unknown sub-matrices we get four equations

$$P_{p-r} R_r^T = -T_r^T M_r^{-1}$$

$$Q_r^{-1} = M_r^{-1} + R_r P_{p-r} R_r^T$$

$$|P_{p-r}| = |M_p| M_r^{-1}|$$

$$P_{p-r} = S_{p-r} - T_r^T M_r^{-1} T_r$$

where the bars indicate determinant. The first two are called Shur's identities, and the third is called Jacobi's theorem. Miller (1964) shows the conditional distribution of $g()$ is also multivariate with mean vector B_{p-r} and ($p\text{-}r$*$p\text{-}r$) covariance matrix P_{p-r}

$$g(X_{p-r}/X_r) = \frac{Exp\left[-\frac{1}{2}\left(X_{p-r} - B_{p-r}\right)^T P_{p-r}^{-1}\left(X_{p-r} - B_{p-r}\right)\right]}{(2\pi)^{(p-r)/2}\sqrt{|P_{p-r}|}} \qquad [A3]$$

$$B_{p-r} = A_{p-r} - P_{p-r} R_r^T (X_r - A_r)$$

APPENDIX B

B.1 SIMPLIFICATION OF TRANSFORMED NORMAL LIKELIHOOD

The likelihood for transforming a sample of data to a multivariate normal sample is described in Section [4.3.2]. This simple form is obtained from the general form of the likelihood using the following method.

If the general form for the sample likelihood is given by

$$\ell = \prod_{i=1}^{n} (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}\left(y_i^{(\lambda,\xi)} - \mu\right)^T \Sigma^{-1}\left(y_i^{(\lambda,\xi)} - \mu\right)\right\} \prod_{j=1}^{p} \left(x_{ij} - \xi_j\right)^{\lambda_j - 1}$$

then on taking the log and rearranging we get

$$\log \ell = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}\left\{\left(y_i^{(\lambda,\xi)} - \mu\right)^T \Sigma^{-1}\left(y_i^{(\lambda,\xi)} - \mu\right)\right\} + \sum_{j=1}^{p}\left(\lambda_j - 1\right)\sum_{i=1}^{n}\log\left(x_{ij} - \xi_j\right)$$

By taking

$$y^{(\lambda,\xi)} = y_{ij}^{(\lambda,\xi)}$$

and

$$\Xi^T = 1.\mu^T$$

where $y^{(\lambda,\xi)}$ is now a matrix containing the sample of transformed data, and Ξ^T is a matrix containing repeated copies of the mean vectors, we can rewrite the middle term of the sample likelihood

$$\sum_{i=1}^{n}\left\{\left(y_i^{(\lambda,\xi)} - \mu\right)^T \Sigma^{-1}\left(y_i^{(\lambda,\xi)} - \mu\right)\right\} = tr\left\{\Sigma^{-1}\left(y_i^{(\lambda,\xi)} - \Xi\right)\left(y_i^{(\lambda,\xi)} - \Xi\right)^T\right\}$$

The likelihood can now be written as

$$\log \ell = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}tr\left\{\Sigma^{-1}\left(y^{(\lambda,\xi)} - \Xi\right)\left(y^{(\lambda,\xi)} - \Xi\right)^T\right\} + \sum_{j=1}^{p}\left(\lambda_j - 1\right)\sum_{i=1}^{n}\log\left(x_{ij} - \xi_j\right)$$

However, since

$$\hat{\Xi}^T = 1.\hat{\mu}^T$$

then the sample covariance is

$$\hat{\Sigma} = \frac{1}{n}\left(y^{(\lambda,\xi)} - \hat{\Xi}\right)\left(y^{(\lambda,\xi)} - \hat{\Xi}\right)^T$$

Consequently, the middle term of the likelihood reduces to a simple additive constant which can be ignored for the optimisation of the likelihood. We then have the final form for the likelihood

$$log\,\ell(\mu,\Sigma,\lambda,\xi) = -\frac{n}{2}log\left|\hat{\Sigma}\right| + \sum_{j=1}^{p}(\lambda_j - 1)\sum_{i=1}^{n} log\left(x_{ij} - \xi_j\right)$$

Where, it should be noted the covariance matrix is dependent on the transformation parameters.

APPENDIX C

C1 TRANSFORMATION TO A UNIT HYPERCUBE: ROSENBLATT'S TRANSFORMATION

Most level III reliability calculations performed by engineers involve the transformation of the basic and correlated design variables like wind speed, wave height, material yield strength, etc, into a multivariate standard normal space. The transformation most commonly quoted is one based on the discussion in Rosenblatt (1952) which shows how a general multivariate probability distribution function can be mapped from X -space variates into Z -space variates that are uniformly distributed on the p -dimensional hypercube:

$$z_1 = P\{X_1 \le x_1\} = F(x_1)$$

$$z_2 = P\{X_2 \le x_2 / X_1 = x_1\} = F(x_2/x_1)$$

$$z_p = P\{X_p \le x_p / X_{p-1} = x_{p-1},\ldots\ldots,X_1 = x_1\} = F(x_p/x_{p-1},\ldots\ldots,x_1)$$

These marginal and conditional distribution functions define a transformation vector T which maps the distribution function into space $z = T(x)$. Once the variables are mapped into this new space then it is simple to calculate joint probabilities

$$P\{Z_i \le z_i; i = 1,\ldots\ldots,p\} \quad = \prod_{i=1}^{p} z_i$$

where $0 \le z_i \le 1; i = 1,.....,n$ (that is, the marginal distributions Z_i are uniform on the interval [0,1]). In order to transform from the Z -space to a standard normal space we simply map the variables from the Z-space to U- space using the inverse normal distribution

$$u_i = \Phi^{-1}(z_i) \qquad ; i = 1,.....,p$$

The above transform is widely used for level II and level III reliability calculations which require the location of a design point in standard normal space.

C.2 GENERATING NORMAL RANDOM VECTORS

The special case of the Rosenblatt transformation for a normal model can now be derived from the conditional distribution $g(.)$ in Equation [A3] by writing $X_p = \{x_1,.....,x_p\}$, with $X_{p-r} \equiv \{x_j\}$ and $X_r \equiv \{x_1,.....,x_{r-1}\}$ hence

$$z_j = F_j(x_j / x_{j-1},.....,x_1) = \Phi\left(\frac{x_j - a_j - T_r^T M_r^{-1}(X_r - A_r)}{\sqrt{S_j - T_r^T M_r^{-1} T_r}}\right) \qquad [C1]$$

This equation is a matrix generalisation of the result in Rosenblatt. As an example, consider the bivariate normal distribution with mean vector A_p, and covariance matrix M_p

$$A_2 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad ; \quad M_2 = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

the conditional distribution is then

$$g(X_2 / X_1) = \frac{1}{(2\pi)^{1/2}\sqrt{(c_{11}c_{22} - c_{12}^2)/c_{11}}} Exp\left[-\tfrac{1}{2}\left(\frac{x_2 - a_2 - (c_{12}/c_{11})(x_1 - a_1)}{\sqrt{(c_{22} - c_{21}^2/c_{11})}}\right)^2\right] \qquad [C2]$$

Substituting the covariance terms with the standard deviations and the correlation coefficient gives

$$g(X_2 / X_1) = \frac{1}{(2\pi)^{1/2}\sigma_{22}\sqrt{1-\rho^2}} Exp\left[-\tfrac{1}{2}\left(\frac{x_2 - a_2 - \rho\frac{\sigma_{22}}{\sigma_{11}}(x_1 - a_1)}{\sigma_{22}\sqrt{1-\rho^2}}\right)^2\right] \qquad [C3]$$

If we examine Equation [A3] carefully we can see how it may be used to generate multivariate normal random vectors when the mean vector and covariance matrix are known. The first step is to generate a p-variate uniform random vector $Z = \{z_1,...,z_p\}$, using a pseudo-random number generator. The first uniform random number z_1 is then transformed to a normal variate using

$$x_1 = \Phi^{-1}(z_1)\sqrt{m_{11}} + a_1$$

and then the second z_2 is generated using

$$x_2 = \Phi^{-1}(z_2)\sqrt{m_{22} - m_{12}m_{11}^{-1}m_{21}} + a_2 + m_{12}m_{11}^{-1}(x_1 - a_1)$$

subsequently the j^{th} terms in the vector are given by the matrix form

$$x_j = \Phi^{-1}(z_j)\sqrt{S_j - T_r^T M_r^{-1} T_r} + a_j + T_r^T M_r^{-1}(X_r - A_r) \qquad [C4]$$

By repeating the process a sample of multivariate normal random vectors can be generated with known mean vector and covariance matrix. This method is used to generate the multivariate normal random samples for the kernel density estimation discussed in Section [4.2]. The form of the random number generator given by Equation [C4] is ideal for the *Mathematica* symbolic algebra packages which are written to smooth and contour the wind, wave and current data. However, the efficiency of this multivariate normal random number generator in a Fortran program has not been compared with other methods like for example those given by Rubinstein (1981).

mean vector

Hs	Tz	Uw	Uc	Dw	Dc
2.17552	7.43897	8.2867	33.7114	206.766	166.805

covariance matrix

Hs	Tz	Uw	Uc	Dw	Dc	
1.58877	1.2443	1.95825	-2.04062	16.4086	-5.68063	Hs
1.2443	2.01003	0.557186	-0.783	15.5017	-1.97301	Tz
1.95825	0.557186	11.5705	-4.77697	62.8678	-22.6279	Uw
-2.0406	-0.783	-4.77697	235.809	-7.53868	-155.126	Uc
16.4086	15.5017	62.8678	-7.53868	9852.84	-553.377	Dw
-5.6806	-1.97301	-22.6279	-155.126	-553.377	10140.1	Dc

correlation matrix

Hs	Tz	Uw	Uc	Dw	Dc	
1	0.696297	0.456733	-0.10543	0.131147	-0.044755	Hs
0.6963	1	0.115538	-0.03597	0.110153	-0.01382	Tz
0.45673	0.115538	1	-0.09145	0.186197	-0.066061	Uw
-0.1054	-0.03597	-0.09145	1	-0.00495	-0.100319	Uc
0.13115	0.110153	0.186197	-0.00495	1	-0.055363	Dw
-0.0448	-0.01382	-0.06606	-0.10032	-0.05536	1	Dc

Hs, Tz - significant wave height and zero-up-crossing period from the frequency domain estimates; Uw,Dw ten minute mean wind speed and direction; Uc,Dc 5 minute mean current speed.

Table 4.1 Mean vector, covariance matrix, and correlation matrix for the DB1 wind, wave, and current data.

DATA BEFORE TRANSFORMATION

VARIABLE	MEAN	ADEV	VAR	SDEV	SKEW	KURT	XMIN	XMAX
1	2.051	1.957	9.184	3.03	3.145	13.004	0	25.848
2	31.774	45.279	99	110.832	8.518	102.84	0	1872.412

COVARIANCE MATRIX

9.1746	250.174
250.174	12271.51

DATA AFTER TRANSFORMATION

LAMDA	0.2735	0.1249
ETA	0	0

VARIABLE	MEAN	ADEV	VAR	SDEV	SKEW	KURT
1	-0.013	1.32	2.675	1.635	0.209	-0.239
2	0.777	2.996	13.395	3.66	0.186	-0.397

COVARIANCE MATRIX

2.672	4.2918
4.2918	13.3819

Table 4.2 Transformation of 1000 simulated bivariate normal random numbers. Statistics before and after using Andrews (1971) method.

SUMMARY STATISTICS	sig. wave hgt (m)		zero cross. prd (s)		current speed (cm/s)		wind speed (m/s)	
	x	y	x	y	x	y	x	y
mean value	2.176	0.635	7.439	1.208	33.710	12.760	8.287	3.356
variance	1.589	0.315	2.010	0.004	235.800	19.070	11.570	1.087
skewness	1.640	0.265	0.705	0.010	0.417	-0.056	0.716	-0.005
excess kurtosis	4.533	-0.274	0.381	-0.488	-0.039	-0.221	0.911	0.024
minimum value	0.400	-0.906	4.500	1.025	0.000	-1.573	0.300	-0.936
maximum value	11.000	2.470	14.500	1.403	94.000	26.680	28.550	7.614
standard deviation	1.260	0.561	1.418	0.062	15.360	4.367	3.402	1.043
average deviation	0.948	0.455	1.141	0.051	12.400	3.523	2.684	0.832
maximum likelihood lamda		0.025		-0.548		0.636		0.437
maximum of the likelihood		-2566.		-4069.		-17730.		-9076.

Table 4.3 Summary statistics for the DB1 data before transformation "x" and after transformation "y" using a one parameter Box & Cox method.

SUMMARY STATISTICS	sig. wave hgt (m)		zero cross. prd (s)		current speed (cm/s)		wind speed (m/s)	
	x	y	x	y	x	y	x	y
mean value	2.176	0.622	7.439	1.402	33.710	32.690	8.287	3.400
variance	1.589	0.357	2.010	0.017	235.800	235.500	11.570	1.134
skewness	1.640	0.040	0.705	0.010	0.417	0.417	0.716	0.009
excess kurtosis	4.533	-0.242	0.381	-0.466	-0.039	-0.039	0.911	0.022
minimum value	0.400	-1.006	4.500	1.005	0.000	-1.000	0.300	-0.931
maximum value	11.000	2.633	14.500	1.824	94.000	92.930	28.550	7.780
standard deviation	1.260	0.598	1.418	0.132	15.360	15.340	3.402	1.065
average deviation	0.948	0.483	1.141	0.109	12.400	12.390	2.684	0.850
maximum likelihood lamda		0.079		-0.288		1.000		0.448
maximum likelihood eta		-0.049		-1.224		0.000		0.000
maximum of the likelihood		-2565.		-4064.		-17870.		-9076.

Table 4.4 Summary statistics for the DB1 data before transformation "x" and after transformation "y" using a two parameter Box & Cox method.

variables {1,2}	shape lamda	location eta	skewness	excess kurtosis
tz	0.4821	-4.5	0.03	-0.188
hs	0.2718	-0.3084	0.11	-0.037
uw	0.51	-0.3	0.055	0.069
hs	0.2703	-0.3686	0.011	0.056
uc	0.6329	0	-0.06	-0.22
hs	0.3263	-0.4	0.024	0 .362
uc	0.6323	0	-0.06	-0.219
tz	0.4518	-4.5	-0.017	-0.151
uw	0.48448	-0.3	0.018	0.078
tz	0.4485	-4.5	-0.022	-0.146
uw	0.4392	-0.0105	-0.003	0.025
uc	0.6346	0	-0.057	-0.221

hs: significant wave height (m)
tz: zero up-crossing period (s)
uw: ten minute mean wind speed (m/s)
uc : five minute mean current speed (cm/s)

Table 4.5 Andrews' transformation parameters plus the skewness and excess kurtosis of the transformed margins for bivariate pairs of DB1 significant wave height (m), zero-up-crossing period (s), mean wind speed (m/s) and mean current speed (cm/s).

Transformation parameters:

	hs	tz	uw	uc
shape lamda	0.2799	0.5021	0.4769	0.6352
location eta	-0.2991	-4.5	-0.3	0

parameter	mean	average deviation	variance	standard deviation	skewness	kurtosis
hs	0.5080	0.6170	0.5920	0.7690	0.1380	-0.0360
tz	1.3270	0.6790	0.7000	0.8370	0.0600	-0.2040
uw	3.4200	0.9230	1.3400	1.1580	0.0070	0.0820
uc	12.7500	3.5180	19.0230	4.3620	-0.0560	-0.2210

Covariance matrix:

	hs	tz	uw	uc
hs	0.5915	0.4518	0.4088	-0.3530
tz	0.4518	0.7000	0.0972	-0.1335
uw	0.4088	0.0972	1.3400	-0.4910
uc	-0.3530	-0.1335	-0.4910	19.0198

Determinant of covariance: 0.364627E+01

Table 4.6 Andrews' transformation parameters for the multivariate set of DB1 significant wave height (m), zero up-crossing period (s), wind speed (m/s), and current speed (cm/s).

variables {1,2}	shape lamda	location eta	skewness	excess kurtosis	transformed mean
hs	0.2703	-0.3686	0.011	0.056	0.443
tz	0.4485	-4.5	-0.022	-0.146	1.278
uw	0.448	0	-0.009	0.022	3.4
uc	1	0	0.417	-0.039	32.711

covariance matrix for the transformed data

	hs	tz	uw	uc
hs	0.6421	0.4452	0.3854	-1.2592
tz	0.4452	0.6321	0.0836	-0.3624
uw	0.3854	0.0836	1.0979	-1.5726
uc	-1.2592	-0.3624	-1.5726	235.7674

Table 4.7 Best set of optimisation parameters for the DB1 winds, waves, and currents.

parameter	50 year			100 year			50 year
	transformed normal	weibull	GEV	transformed normal	weibull	GEV	DEn Guidance
hs	15.02	17.61	32.41	15.81	18.88	39.41	15.5
uw	33.02	33.35	42.54	35.6	34.9	42.7	38.5
uc	100.5	100.9	97.4	103.4	102.5	97.98	...

Table 4.8 50 and 100 year return period estimates from the transformation normal, Weibull, and generalised extreme value distributions.

covariance matrix for the Box transformed data Y

0.642141	0.445167	0.392191	-1.259152
0.445167	0.632069	0.085384	-0.362383
0.392191	0.085384	1.135713	-1.598023
-1.259152	-0.362383	-1.598023	235.767429

eigenvalues for the covariance of y

235.786	0.138602	1.45152	0.801576

eigenvectors for the covariance of y

-0.00536893	0.742001	0.553058	0.378857
-0.00155363	-0.627534	0.37419	0.682774
-0.00681952	-0.235868	0.744334	-0.624729
0.999961	0.00140034	0.00862702	-0.00116556

mean vector for Z meanz = Transpose[Ay].meany

32.6822	-1.2299	3.53764	-1.12303

covariance matrix for z

235.786	0	0	0
0	0.138602	0	0
0	0	1.45152	0
0	0	0	0.801576

transformation matrix K

-0.0824416	0.276242	0.66632	0.339194
-0.0238564	-0.233627	0.450821	0.611293
-0.104716	-0.0878121	0.896768	-0.559324
15.3547	0.000521338	0.0103938	-0.00104353

Table 4.9 Vectors and matrices required to transform the y-space variates to u-space variates (or vice versa).

Table 4.10 Bivariate extreme value models for joint survival probabilitites

MODEL NAME	No. of PARAM's	JOINT SURVIVAL FUNCION	DEPENDENCE FUNCTION	CORRELATION COEFFICIENT
GENERAL FORM	-	$G(x,y)=\exp\left\{-(x+y)A\left(\frac{y}{x+y}\right)\right\}$; $x,y>0$	$A(w)=\int_0^1 \max\{(1-w)q,(1-q)w\}dH(q)$ ref. Tawn (1988)	$\rho=\int_0^1 \frac{1}{\{A(w)\}^2}dw-1.0$
MIXED	1	$G(x,y)=\exp\left\{-(x+y)+\frac{\theta xy}{(x+y)}\right\}$	$A(w)=\theta w^2-\theta w+1$; $0\le\theta\le 1$	$\rho=(1-\frac{1}{2}\theta)^{-3/2}\theta^{-1/2}\left\{\sin^{-1}\left(\frac{1}{2}\sqrt{\theta}\right)-\frac{1}{2}\sqrt{\theta}(1-\frac{1}{2}\theta)^{1/2}(1-\frac{1}{2}\theta)\right\}$ $0\le\rho\le 0.4728$; $\theta=0$ independent
LOGISTIC	1	$G(x,y)=\exp\left\{-\left(x^{1/\nu}+y^{1/\nu}\right)^{\nu}\right\}$	$A(w)=\left\{(1-w)^{1/\nu}+w^{1/\nu}\right\}^{\nu}$; $0\le\nu\le 1$	$\rho=\nu\left\{\Gamma(2\nu)\right\}^{-1}\left\{\Gamma(\nu)\right\}^2-1$ $\nu=1$ independent; $\nu=0$ dependent
ASYMMETRIC MIXED	2	$G(x,y)=\exp\left\{-(x+y)+\frac{xy(x(\theta+\Phi)+y(2\Phi+\theta))}{(x+y)^2}\right\}$	$A(w)=\Phi w^3+\theta w^2-(\theta+\Phi)w+1$ $\theta\ge 0$; $\theta+\Phi\le 1$; $\theta+2\Phi\le 1$; $\theta+3\Phi\ge 0$	$\theta=\Phi=0$ independent
ASYMMETRIC LOGISTIC	3	$G(x,y)=\exp\left\{-(1-\theta)x-(1-\Phi)y-\left((x\theta)^{1/\nu}+(y\Phi)^{1/\nu}\right)\right\}$	$A(w)=\left\{\theta^{1/\nu}(1-w)^{1/\nu}+(\Phi w)^{1/\nu}\right\}^{\nu}+(\theta-\Phi)w+1$ $0\le\theta$; $\Phi\le 1$; $0\le\nu\le 1$	$\theta=\Phi=1$ *logistic model* $\theta=0$, or $\Phi=0$, or $r=1$ independent

note dependence refers to complete dependence, and independence refers to complete independence

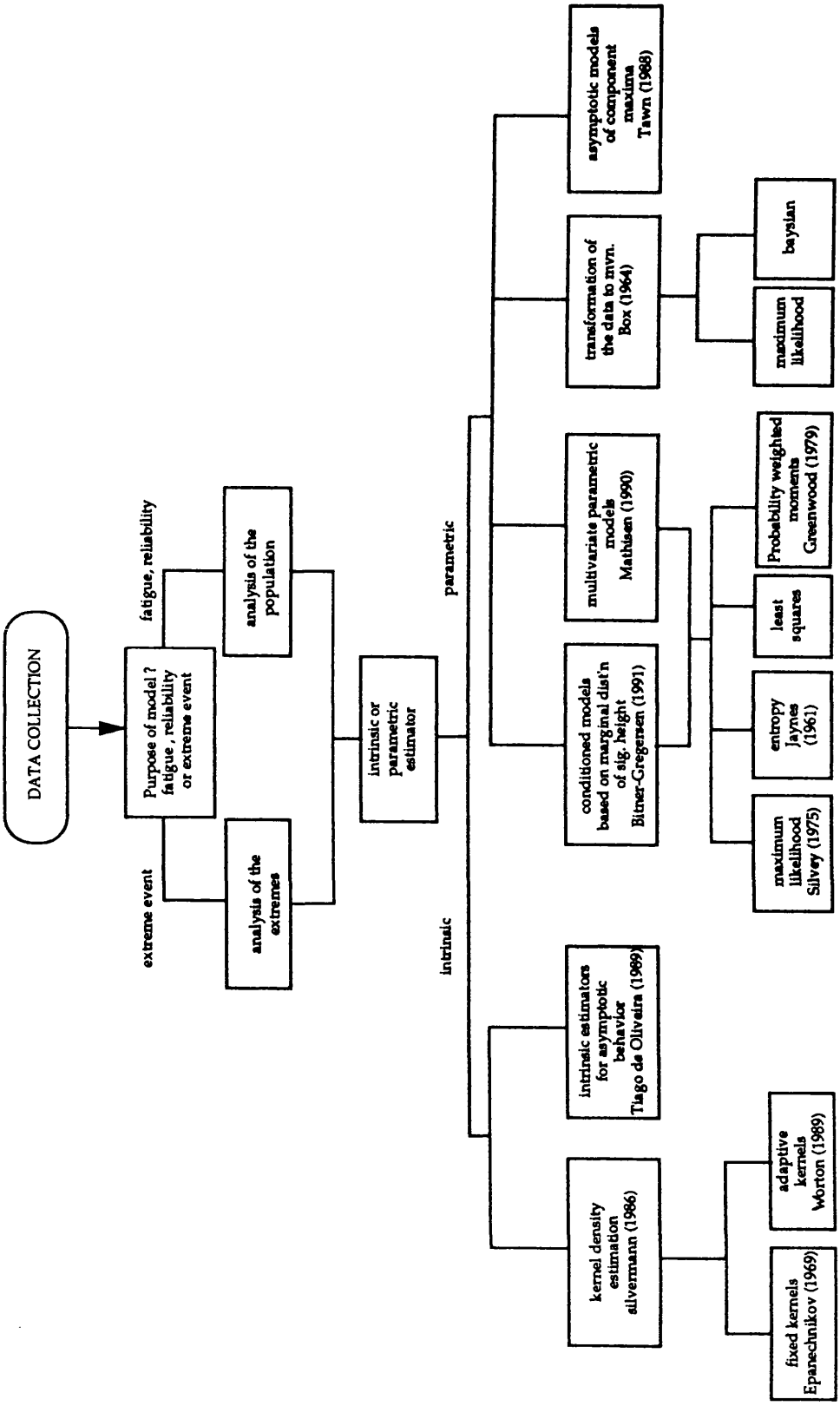


Figure 4.1 Choices for the analysis of environmental data.

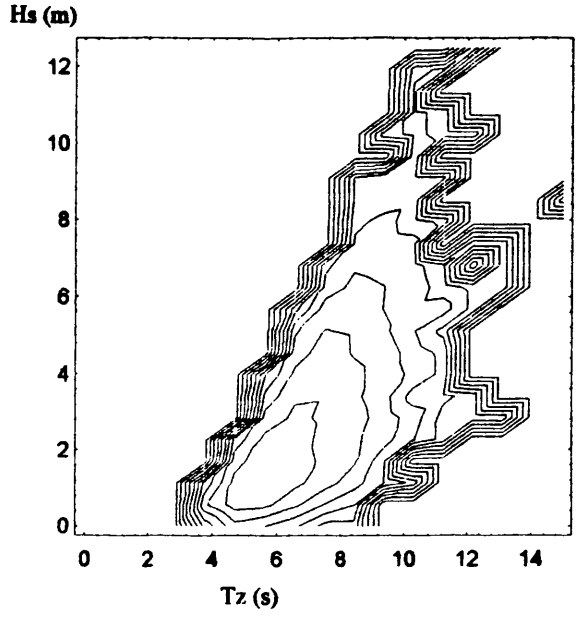


Figure 4.2 Contour plot of the bin-count natural log in the significant wave height (m) and zero-up-crossing period (s) scatter diagram.

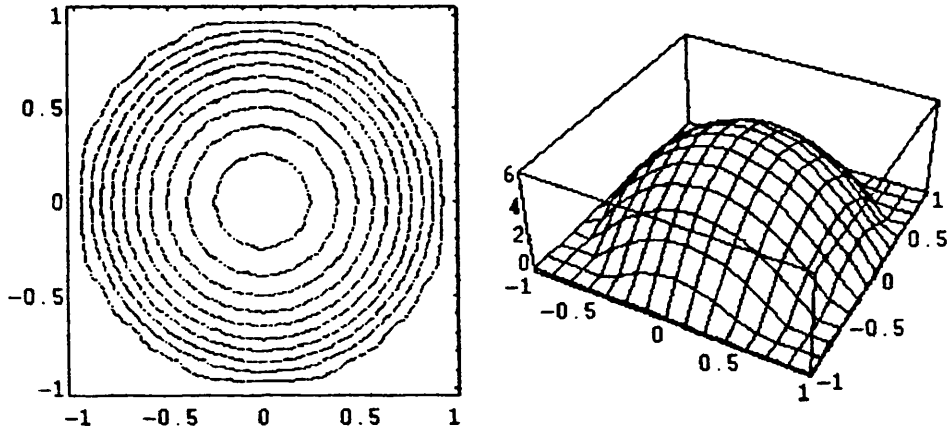


Figure 4.3 Contour and 3d plots of the Epanechnikov (1969) kernel smoothing function.

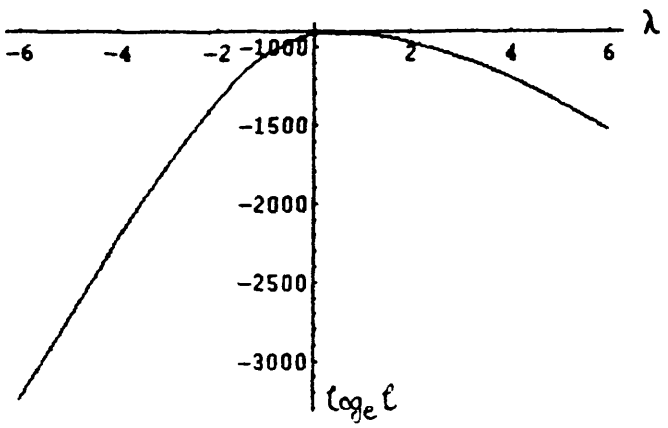
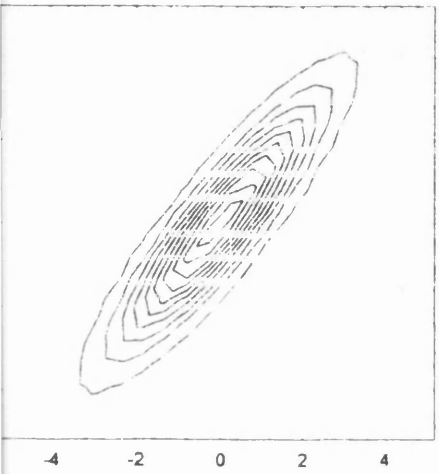


Figure 4.4 Log-likelihood for 1000 transformed normal random numbers.



gure 4.5 Population density with mean={0,0} and
ivariance c={{2.0,2.2},{2.2,3.0}}

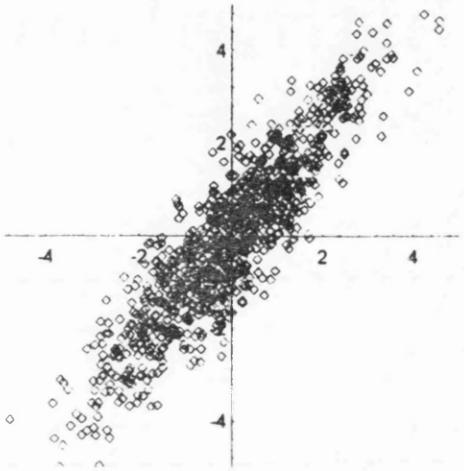


Figure 4.6 Scatter plot for 1000 samples with mean={0,0} and covariance c={{2.0,2.2},{2.2,3.0}}

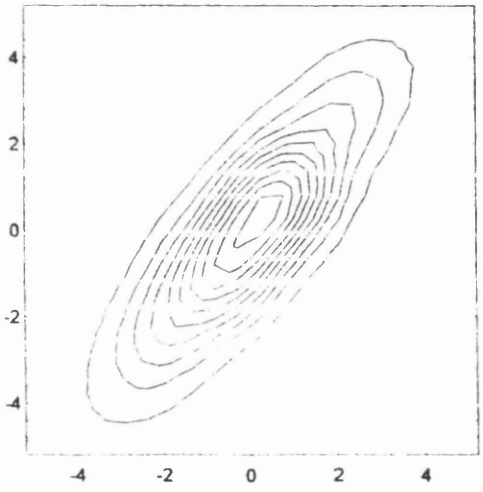


Figure 4.7 Kernel density estimate for 1000 samples with mean = {0,0} and covariance = {{2.0,2.2},{2.2,3.0}}.

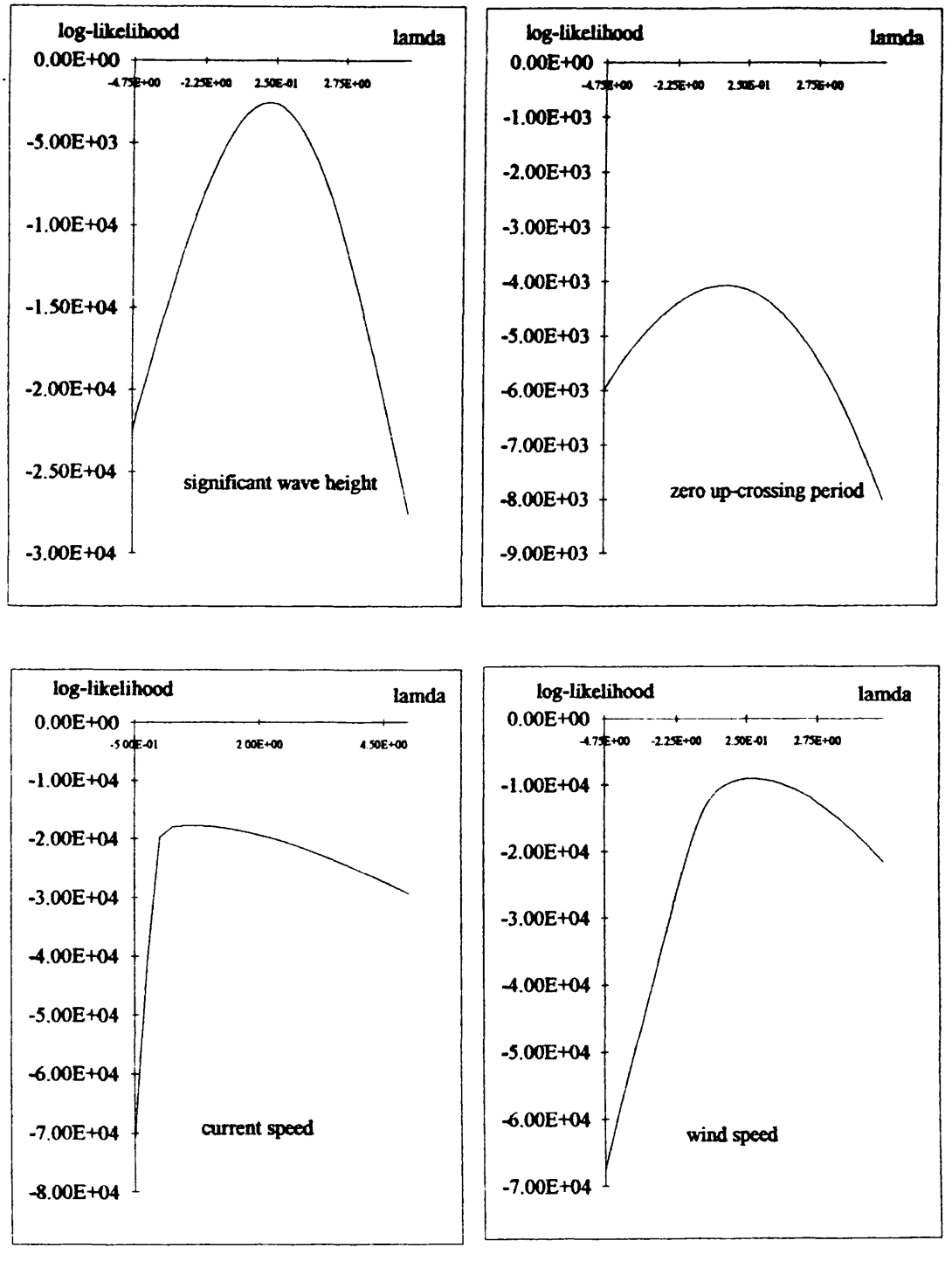


Figure 4.8 Log-likelihood functions for the transformation of the wave, wind and current marginal data, using the Box & Cox method.

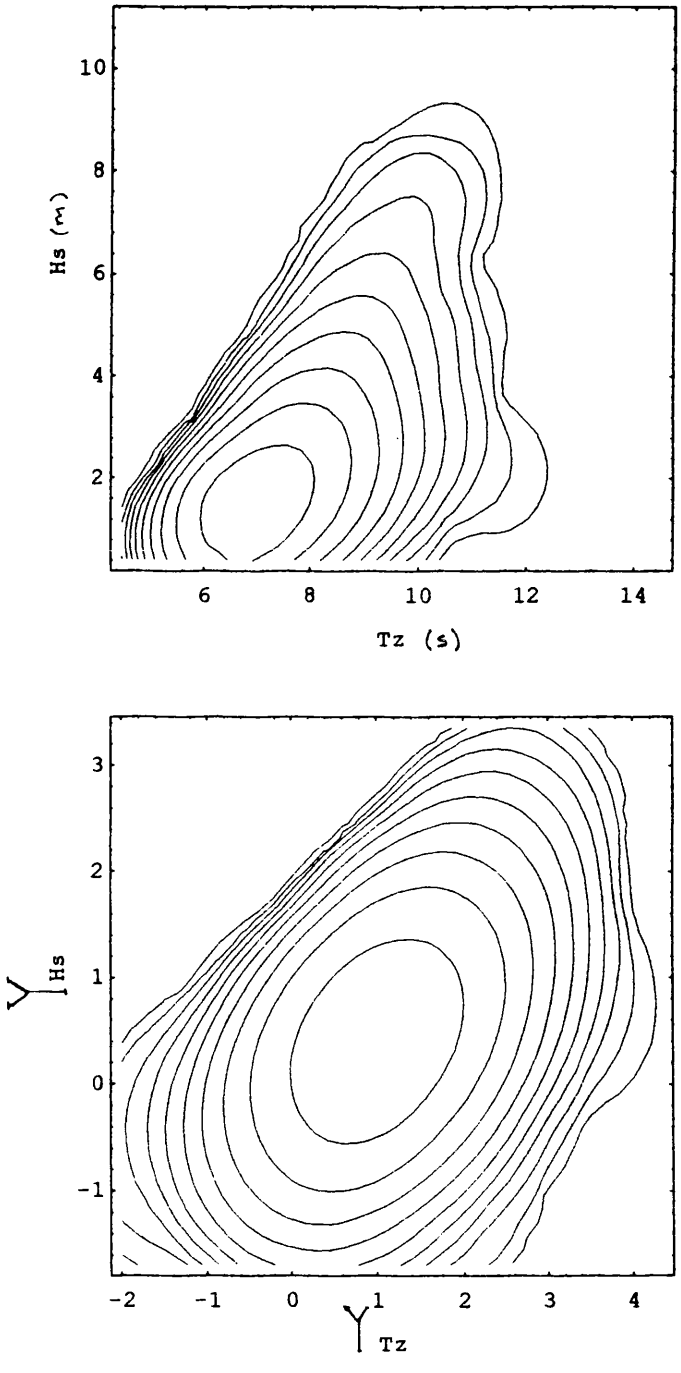


Figure 4.9 Kernel density plots of joint significant wave height (m) and zero-up-crossing period (s); (a) before transformation; (b) after transformation.

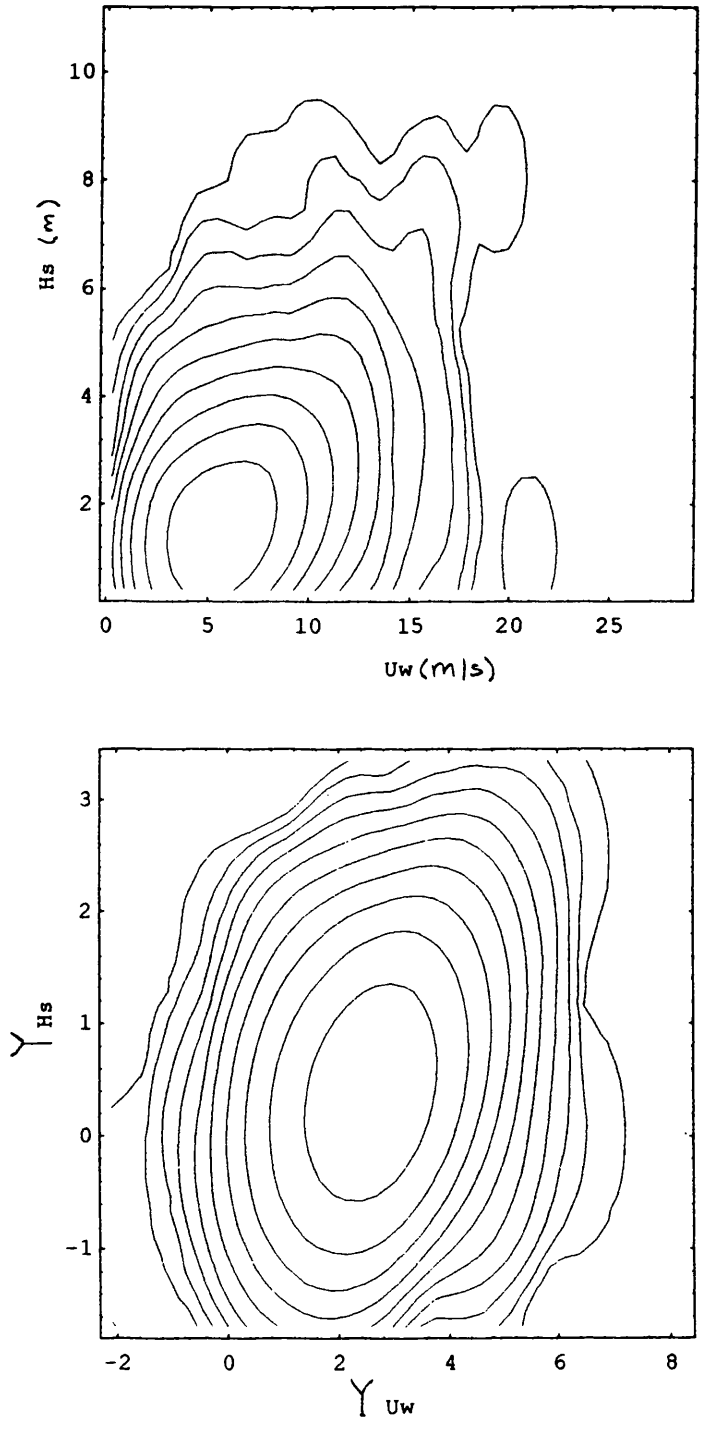


Figure 4.10 Kernel density plots of joint mean wind speed (m/s) and significant wave height (m); (a) before transformation; (b) after transformation.

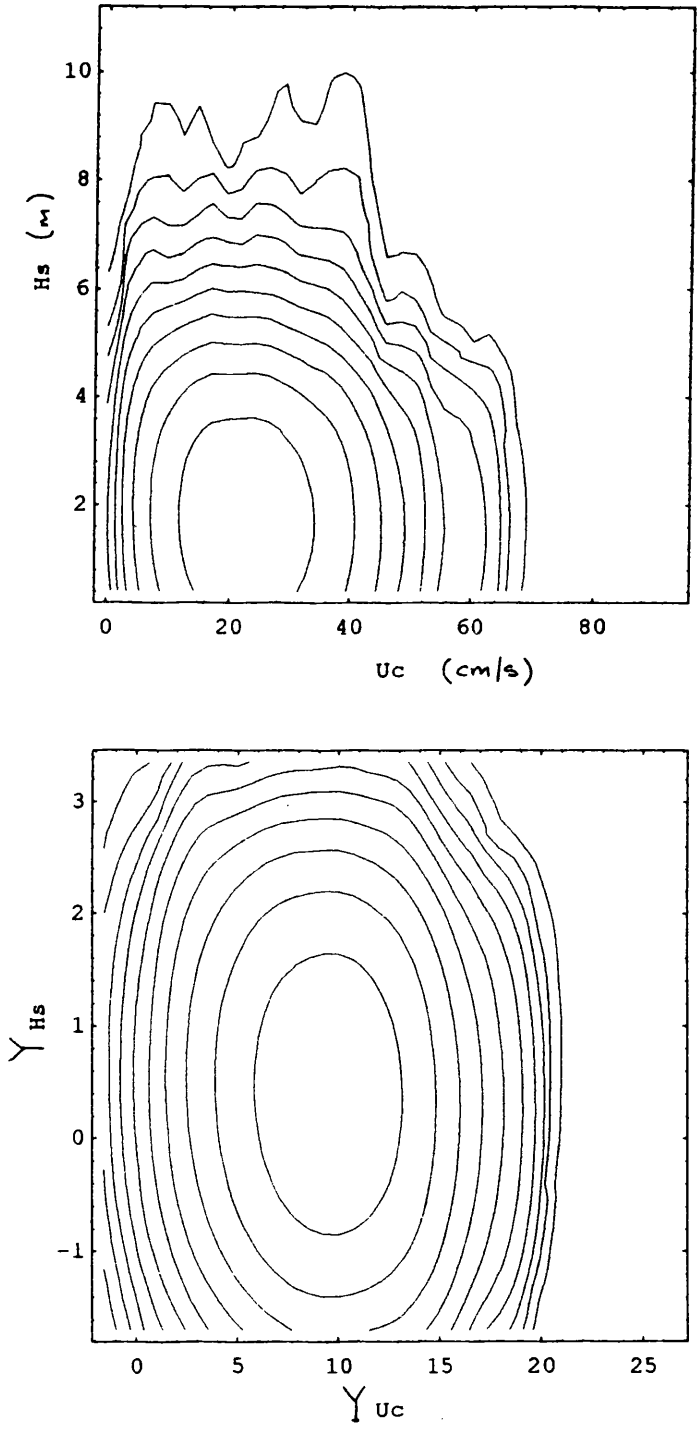


Figure 4.11 Kernel density plots of joint mean current speed (cm/s) and significant wave height (m); (a) before transformation; (b) after transformation.

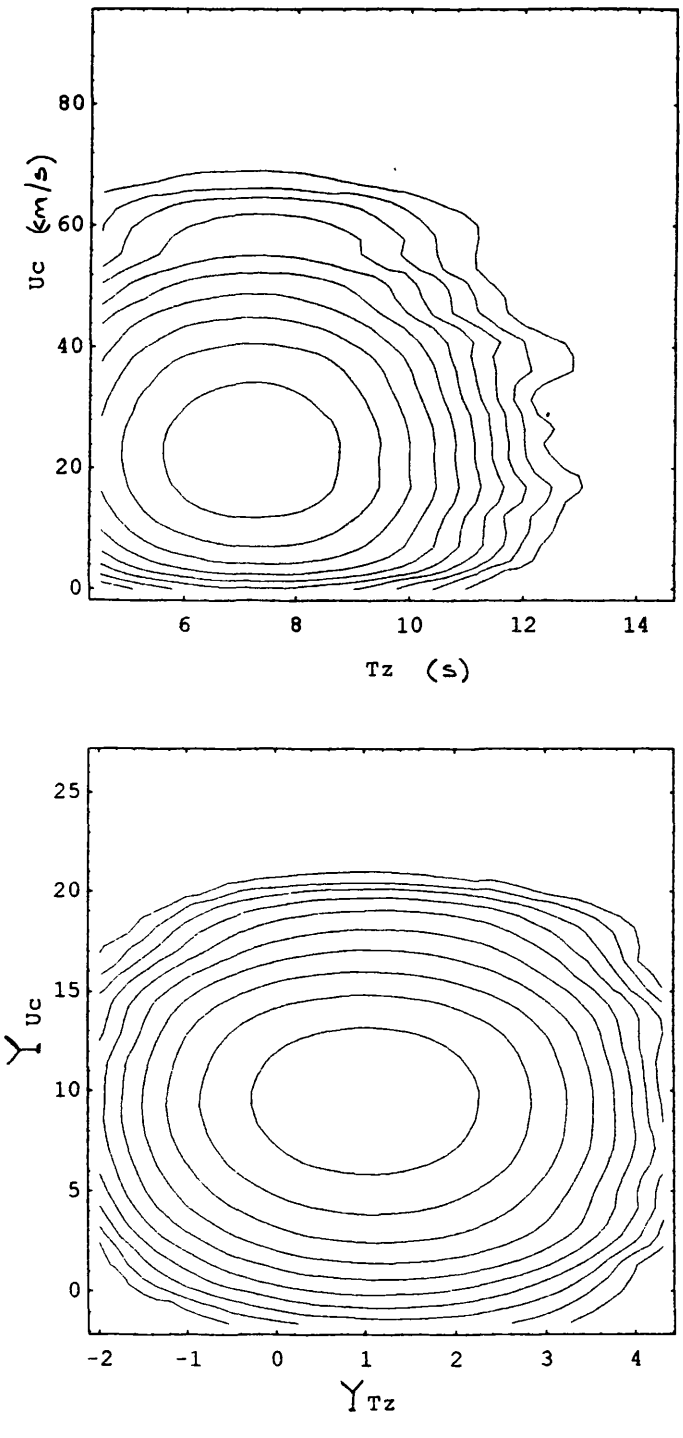


Figure 4.12 Kernel density plots of joint zero-up-crossing period (s) and mean current speed (cm/s); (a) before transformation; (b) after transformation.

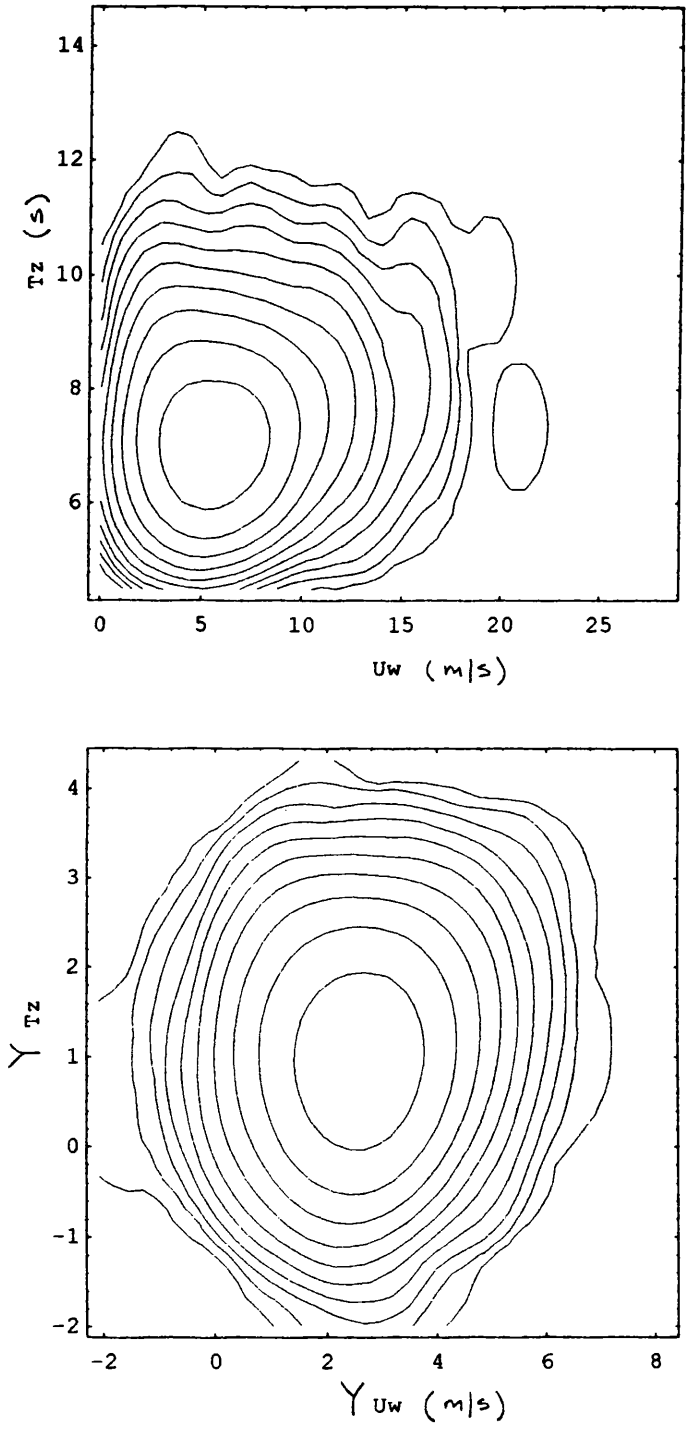


Figure 4.13 Kernel density plots of joint mean wind speed (m/s) and zero-up-crossing period (s); (a) before transformation; (b) after transformation.

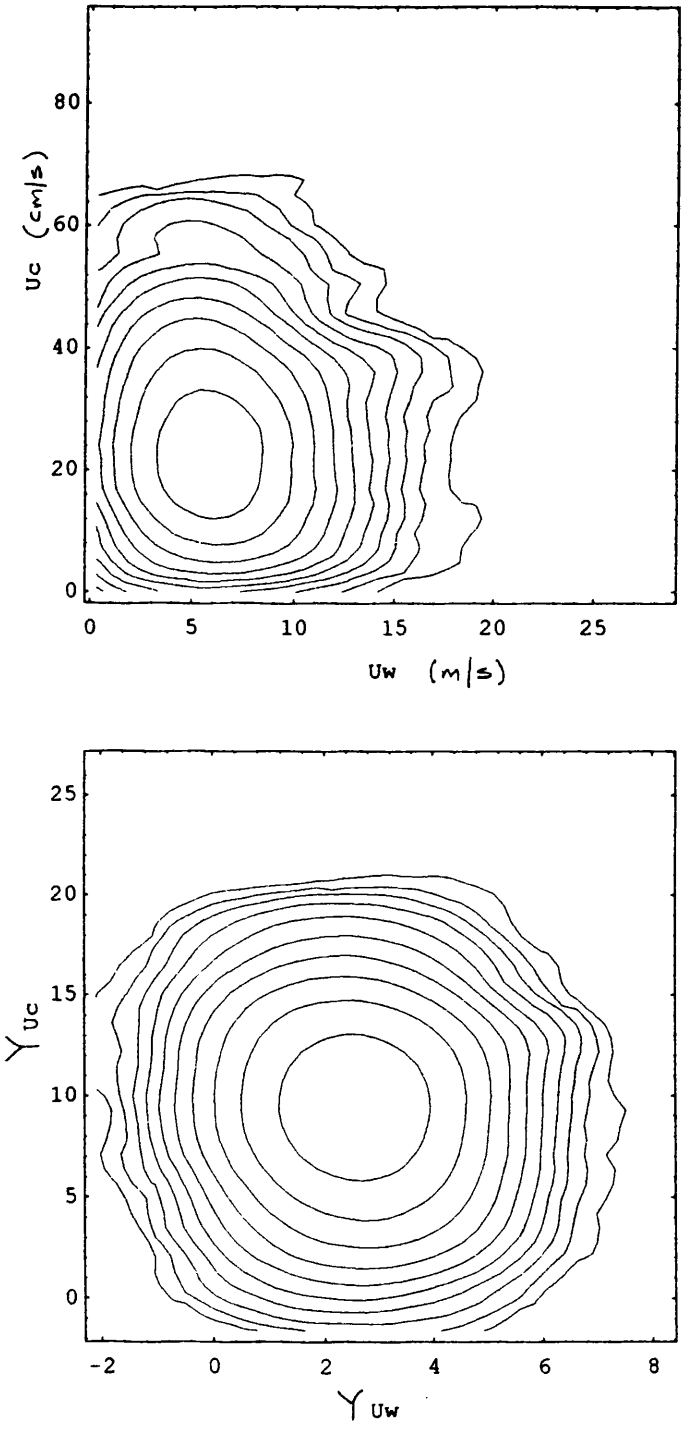
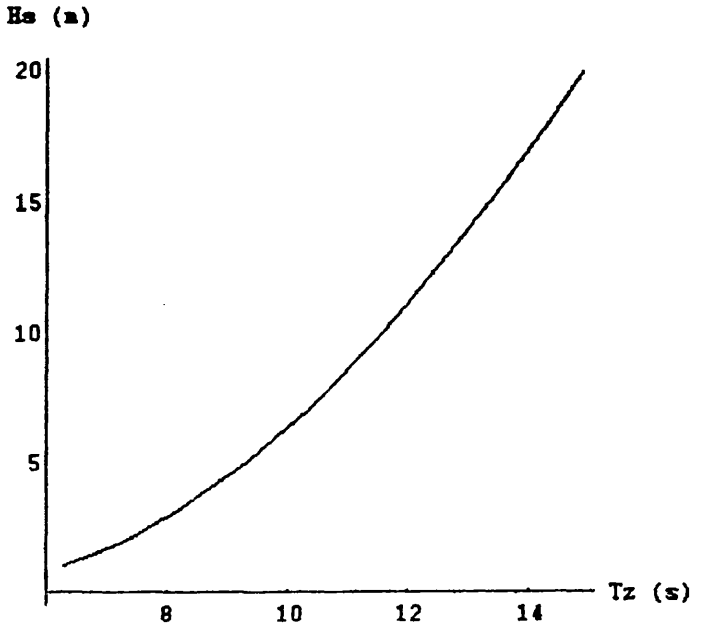


Figure 4.14 Kernel density plots of joint mean wind speed (m/s) and mean current speed (m/s); (a) before transformation; (b) after transformation.



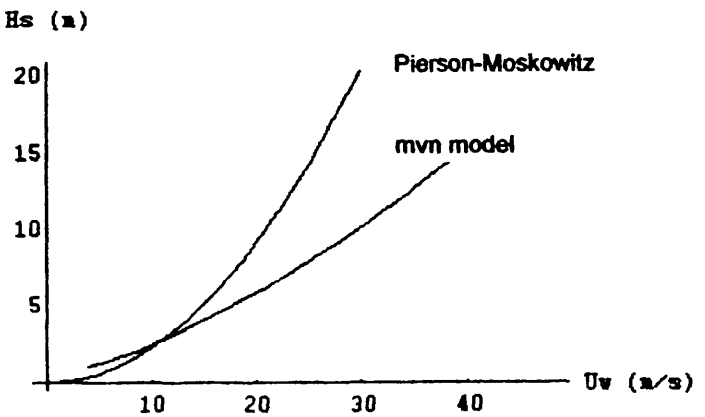


Figure 4.15 Modal zero-up-crossing period (s), and wind speed (m/s) conditioned on the significant wave height (m).

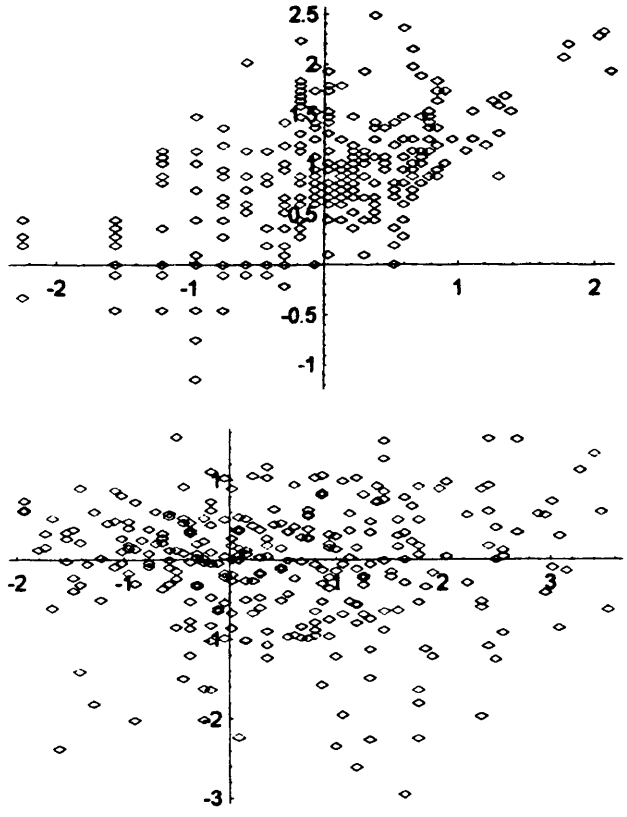


Figure 4.16 300 y-space (top) and corresponding u-space (bottom) variates.

Chapter 5

MODELLING DIRECTIONAL SEAS

NOMENCLATURE

C	*normalisation constant*
$\overline{C}$	*real part of characteristic function*
C_{ij}	*co-spectrum of i^{th} and j^{th} signals*
$D(\theta)$	*directional spreading function*
$D(f,\theta)$	*frequency dependent directional spreading PDF*
E	*expectation operator*
(f_x, f_θ)	*probability distribution function for random variable x*
I_n	*modified Bessel function of order n*
$k(f)$	*frequency dependent wave number*
m_1	*factor on noise component in directional spread*
n	*number of directional observations*
p	*order of trigonometric moment, or cosine power*
$\boldsymbol{P}$	*unit direction vector*
Q_{ij}	*quadrature spectrum of i^{th} and j^{th} signals*
rms	*root mean square*
R	*normalising constant*
$\overline{R}$	*resultant*
s_0	*angular variance*
$s(f)$	*frequency dependent cosine spreading power*
$\overline{S}$	*imaginary part of characteristic function*
$S(f,\theta)$	*directional surface elevation variance spectrum*
$S(f)$	*point surface elevation variance spectrum*
$W(\Phi)$	*weighting function*
α_p	*trigonometric moment of the distribution function*
β_p	*trigonometric moment of the distribution function*
$\overline{\alpha}_p$	*centred moment of the distribution function*
$\overline{\beta}_p$	*centred moment of the distribution function*
Γ	*gamma function*
θ	*Wave propagation direction*
$\overline{\theta}$	*mean wave propagation direction*

$\overline{\theta}_{p0}$	*pth resultant mean wave propagation direction*
μ_o	*Mean direction*
κ	*Concentration parameter*
ϕ	*angle on* $[0,2\pi]$, $\phi = \theta - \overline{\theta}$
Φ	*characteristic function*
f	*frequency in Hz*
σ	*standard deviation*
σ_{0i}	*equivalent linear standard deviation determined from ith angular moments*
γ_1	*circular variance*
γ_2	*circular variance*
ζ	*mean direction of the noise in* $D(f,\theta)$

5. INTRODUCTORY REMARKS: MODELLING DIRECTIONAL SEAS

The extreme responses of fixed, compliant and articulated offshore structures can be affected by the degree of directional wave spreading in storms. It is therefore essential that empirically determined directional distributions accurately represent the expected degree of directional spreading over the full range of frequencies. Calculations of structural response in directional seas generally model the directional spreading of wave energy as deterministic using a frequency dependent uni-modal function. Most often, this function is modelled by the cosine power distribution, originally proposed by Cartwright (1963) and Longuet-Higgins et al (1963a).

Studies of the frequency dependent directional distribution of wave variance using maximum likelihood, Jeffereys (1986), and maximum entropy, Ochoa (1990), all indicate that in real seas the directional distributions contain a considerable amount of detail which may be further complicated by the crossing of wind driven and swell seas. For design this detail must be simplified and in practice no more than the mean direction and the concentration of the directional spread is required. In this respect, the cosine directional distribution is satisfactory. However the Longuet-Higgins approach is sensitive to noise in the co and quadrature spectra, and the occurrence of asymmetric or non-cosine distributed directional spreads. The presence of noise and asymmetry is inevitable in recorded spectra and consequently models based on fitting an exponentiated cosine model using an angular moments approach must be subject to uncertainty and bias.

The effect of noise is examined theoretically by Tucker (1989), who suggests the frequency dependent spreading models developed by Hasselmann and Mitsuyasu may be more a reflection of the buoy response than the true nature of the directional spread, Tucker (1991). To overcome

some of the problems he proposes a method of allowing for imperfect buoy response, however, this does not address the robustness of simply equating the cosine model angular moments to those inferred by the measured spectra. The application of angular statistical methods to the modelling of directional wave spreading results in some improvements to the conventional cosine model estimator. Borgman (1969) proposed the use of circular statistics together with the von Mises directional probability distribution. This model has advantages when compared with the conventional cosine model used by Cartwright (1963), Longuet-Higgins et al (1963a), Mitsuyasu et al (1975), Hasselmann et al (1980) and Ewing & Laing (1987); especially if the distribution can be fitted using weighted estimation techniques.

The need for robust measures of location and spread is noted by Kuik et al (1988) who proposed a set of model free parameters to measure location, circular variance, skewness, and kurtosis. Their estimators are based on the use of those circular statistics which can be determined naturally from pitch-and-roll buoy co- and quadrature spectra using the theory developed by Cartwright and Longuet-Higgins, Appendix A. Whilst these quantities are non-parametric they leave interpretation of the location, skewness, and kurtosis to the designer who will ultimately need to fit some parametric form to the summary statistics.

The problem of fitting parametric models to directional data is addressed by Borgman (1969) who examines Longuet-Higgins et al (1963a) suggestion of using a weighting function to 'smooth' the raw Fourier summation form of the directional distribution. The principle of weighting the distributions assumes that the true directional distribution is $D(\theta)$ and the Fourier summation estimate $D_F(\theta)$ is the weighted distribution

$$D_F(\theta) = \int_0^{2\pi} D(\theta^*) W_1(\theta - \theta^*) d\theta^* \qquad [5.1]$$

where the weight function can be written as

$$W_1(\phi) = \frac{\sin([N+1/2]\phi)}{2\pi \sin(\phi/2)} \qquad [5.2]$$

Unfortunately, this weighting has the undesirable property of broadening the estimated directional distribution and allowing negative estimates for $D(\theta)$. An alternative weight function which removes negative estimates was also proposed with the form: $W_2(\phi) = R\cos^{2N}(\phi/2)$, where R is a normalising constant such that $\int W_2 d\theta^* = 1.0$. However,

this weighting broadens the distribution even more than W_1 and results in a $\cos^4(\phi/2)$ estimate for the directional spread of unidirectional seas.

The obvious way to improve the directional model is to increase the number of Fourier coefficients. This was the motivation for the development of clover-leaf buoys and wave gauge arrays but neither method has become a routine method for data acquisition. The second approach to modelling directional spreading is to assume $D(\theta)$ has some parametric form and then estimate the model parameters using moment estimation. This approach was adopted by: Longuet-Higgins et al (1963a), who used moments to fit a half angle cosine model to the data; and Borgman (1969) who suggests the von Mises and wrapped Normal distributions are appropriate models. Their results were the basis of subsequent experimental work to determine the frequency dependence of the directional spreading in wind driven seas.

In this chapter, the robustness of the parametric modelling is examined in a simple way by simulating double cosine directional distributions and then fitting cosine and von Mises models to them using moment estimation. This demonstrates the sensitivity of the simple circular moments estimator and consequently a weighted estimator based on Huber's W-estimation process (1981) is examined for the parameters of a von Mises distribution. Unfortunately, the new estimator requires some knowledge of the physics of wave generation and consequently the method is only presented for discussion and possible future development.

5.1. EMPIRICAL MODELS OF SPREADING

In order to make use of existing uni-directional wave elevation spectrum models the directional variance spectrum is usually written in the form

$$S_\eta(f,\theta) = S_\eta(f) \times D(f,\theta)$$

where $D(f,\theta)$ is a frequency dependent directional distribution function which must satisfy

$$\int_0^{2\pi} D(f,\theta)d\theta = 1.0$$

Longuet-Higgins et al. (1963a) proposed a half angle cosine spreading function based on observations made using a pitch and roll buoy. Their model is defined on $0,2\pi$ as

$$D(f,\phi) = C\cos^{2s}\left(\frac{1}{2}\phi\right) \; ; \; \phi = \theta - \bar{\theta} \qquad [5.3]$$

where,

$$C = \frac{2^{2s-1}}{\pi} \times \frac{\Gamma^2(s+1)}{\Gamma(2s+1)}$$

The value 's' is the spreading power, and ϕ is the angle measured from the mean wave direction $\overline{\theta}$ which in this work is assumed to correspond to the mean wind direction. Plots showing the form of the cosine spreading function, about the mean wave direction, for a variety of spreading powers (2*s*) between 0 and 20 are given in Fig. 5.1.

Several other models have been proposed as directional distributions, for example Borgman (1969) suggests the wrapped normal and circular von Mises distributions; and Chase et al (1957) recommends the full angle cosine distribution. Whilst each of these models has its own merits, with only the first five Fourier coefficients available to estimate the parameters there is no theoretical reason for favouring any particular model. That said, the von Mises distribution does have some desirable practical properties and a considerable literature to support its use and estimation. Plots showing the form of the von Mises distribution for concentrations between 0 and 20 are shown in Fig. 5.2.

In wind seas the concentration parameter of a model appears to be frequency dependent. For example, Longuet-Higgins et al (1963a) noted the power of the spreading appeared to be related to the dimensionless frequency or (U/c) where U is the wind speed and c is the wave phase speed

$$\tilde{f} = \frac{2\pi f U}{g} = \frac{U}{c}\,;\; c = \frac{g}{2\pi f}$$

This frequency dependence of the spreading power must be modelled accurately because, as shown by Helvacioglu (1990), the responses of articulated and compliant systems can be sensitive to the spreading at the lower frequencies due to their long natural periods. Two frequency dependent spreading models commonly used for structural response calculations are examined below.

5.1.1 MITSUYASU SPREADING FUNCTION

The relationship between resonance angle and the power of the directional spreading was examined by Mitsuyasu et al (1975). Their work was based on surface elevation measurements using clover leaf buoys located at two Japanese sites; one at Hakata Bay in a water depth of 5m; and the other in open seas north of Fukuoka. The two locations both have limited fetch in some directions. They found the normalised form of their spectra for the open sea case corresponded

to the Pierson-Moskowitz form of the spectrum, whilst the results measured at Hakata Bay corresponded to the JONSWAP form. This is consistent with the non-dimensionalised wind fetch for each location and suggests the results should be applicable to the North Sea where designers commonly use both the P-M and JONSWAP spectra. Mitsuyasu finds a relationship between the spreading power s and resonance angle by plotting $s = \frac{1}{2}[s_1 + s_2]$ against U / c and then fitting a suitable model using regression analysis. That is, they take the average of the first and second angular moment estimates of the spreading power giving the result summarised in Table 5.1.

5.1.2 HASSELMANN SPREADING FUNCTION

The directional spreading function adopted by Mitsuyasu was examined by Hasselmann, et al (1980) using data measured during JONSWAP. They conclude the spreading power is not a function of the resonance angle *U/c;* rather it is a function of the normalised frequency *f/fm.* This was substantiated by comparing the correlation coefficients from the linear regression of spreading power onto *U/c* and *f/fm.* Their equations are also reproduced in Table 5.1. Note that the equation for $f/fm > 0$ does not yield the same result at $f/fm = 1.0$ as the equation for $f/fm < 0$. This is overcome by swapping equations at $f=1.05fm$ rather than $f=fm$.

The spreading powers given by Hasselmann's model are consistently lower over the whole frequency range than those given by Mitsuyasu's model. This observation is noted by Hasselmann who points out that their model is strictly only valid for $U/c > 1.0$ whereas Mitsuyasu included data for fully developed seas for which $U/c < 1.0$. Hasselmann suggests the higher values of s may be due to including swell seas - with low U/c - in the Mitsuyasu data. In fact, Hasselmann's model is based on fitting to s_1 to f/fm and not the average of s_1 and s_2. We shall see in the next section that this results in biased estimates of spreading power because of the lack of robustness of the first moment estimator; and that the differences between Hasselmann and Mitsuyasu models are partly due to the poor quality of s_1 as an estimator of s.

5.2. ROBUSTNESS OF THE COSINE POWERS

For this work robustness is defined as insensitivity to small deviations from the modelling assumptions. When fitting parametric models to directional wave data we pre-suppose some functional form for the directional distribution and then use some estimator of the location and concentration parameters. This estimator must be robust to departures of the true directional spread from the idealised models, whilst still being efficient in some sense (efficiency is examined in detail by Huber (1981)). Robust estimates of location and concentration are required if the results of directional wave analysis are to be used as the basis for spectral analysis of offshore structures, since the responses of fixed and compliant systems can be

sensitive to the degree of spreading. For example, when the responses of fixed structures are calculated using directional spectra the in-line forces are decreased by some 5 per cent, Prince-Wright & Percival (1989), to 15 per cent, Haver & Natvig (1991). Conversely, when the responses of articulated systems are calculated using directional spectra the forces can be amplified, Helvacioglu (1990). It is therefore important to ensure directional seas are modelled correctly if unconservative modelling of fixed and articulated structures is to be avoided.

The data recorded by directional wave recorders will always be contaminated by noise and bias which may arise from random error; spectral leakage in the Fourier analysis; or imperfect surface following characteristics. The effect of bias - caused by buoy heave, pitch, and roll response transfer functions not being uncoupled and unity - is examined theoretically by Tucker (1989) who demonstrates how the first angular moments are dependent on the *linear* buoy response transfer functions, and the second angular moments are independent of them. This would suggest that the second moment estimators are more robust and should be used to determine the degree of directional spreading. Ewing & Laing (1987) had previously examined the effects of noise in recorded directional data using numerical simulation and came to the same conclusions as Tucker; whilst also pointing out the first moment estimator has been adopted by most oceanographers which might suggest their models overestimate the degree of directional spreading in real seas.

The effects of noise and bias are important. However, there is a second problem with the Longuet-Higgins estimator for the cosine model parameters and that is the sensitivity of simple moment estimation to deviations from the assumed cosine form for the directional distribution. In real seas it is almost certain that the true directional spread will not be perfectly cosine distributed since wind fields are constantly veering and wind and swell seas may cross each other. The poor robustness of simple moment estimators is well known to statisticians who favour the use of robust measures of mean and concentration when the distribution of errors is known to be non-normal. In this study we do not have discrete observations of direction with error since only the first five Fourier-Stieltjes coefficients are known. These coefficients will contain limited information about the degree of skewness and kurtosis in the recorded data which arises from not only bias and noise but also real asymmetry due to crossing seas. The question is therefore: how robust are the estimates for the cosine model parameters, given that the angular harmonics correspond to real seas and may be corrupted by bias and noise? We can examine this in a simplistic way using simulation.

5.2.1 BIAS IN ESTIMATES OF SPREADING

Assume we have $S(f,\theta)$ for a given frequency and that it can be written in the form $S(f,\theta)=S(f)D(f,\theta)$; where $S(f)$ is the one-dimensional spectrum and $D(f,\theta)$ is the frequency dependent wave directional probability density. By simulating the directional distribution as the mixture of two cosine distributions

$$D(f,\theta)=m_1 cos^{2s}(\theta/2)+(1-m_1)cos^{2p}\left((\theta-\zeta)/2\right) \qquad [5.4]$$

the angular harmonics can be calculated by numerical integration as shown Appendix A. These can be used in Equations [A11 to A14] to estimate the mean directions θ_1,θ_2 and cosine powers s_1,s_2 for the idealised directional distribution. The results of some simulations for a range of directional distributions are shown in Table 5.2. Some examples are plotted in Fig. 5.3 which shows the effect of the second cosine term is to introduce varying degrees of asymmetry when $\zeta \neq 0; m_1 \neq 0$ or flatness when $\zeta = 0; m_1 \neq 0$. In measured spectra such small departures from a perfect cosine model must be commonplace and it is therefore important to note the large bias in the s_1,s_2 values for only small imperfection. The two estimators for mean direction θ_1,θ_2 are generally robust but it should be noted only small degrees of imperfection have been introduced in these simple examples; in cases where wind and swell seas are crossing, the true directional distributions may be bi-modal and then θ_1,θ_2 will be significantly biased.

This simple simulation suggests a more robust approach for the estimation of concentration and location is required if a design model is to be developed for directional wave response analysis. Alternative estimators have been suggested, for example Long & Hasselmann (1979) suggest a variational approach, and Hasselmann et al (1980) examined the use of fitting a double cosine model to the angular harmonics. These models are however probably no more robust than the simple moments estimator, see Kuik et al (1988), Ochoa (1990), and consequently in Section [5.5] the statistical theory for robust 'W-estimates' of directional random variables is discussed.

5.3 CIRCULAR STATISTICS FOR ANGULAR RANDOM VARIABLES

The fundamental difference between a line statistical model and the corresponding circular model, in which the angle is the random variable, is the range of integration. A function f_x is a circular distribution if and only if

$$\int_0^{2\pi} f_\theta(\theta)\, d\theta = 1.0 \;\; ; 0 \le \theta \le 2\pi \quad ; \quad f(\theta) \ge 0$$

In line statistics the model is characterised by its line moments; however, in circular statistics the model is characterised by the p^{th} trigonometric moments about the angular origin

$$\alpha_p = E[Cos\, p\theta] \qquad where \;\; |\alpha_p| \le 1 \tag{5.5}$$

$$\beta_p = E[Sin\, p\theta] \qquad where \;\; |\beta_p| \le 1 \tag{5.6}$$

The expectation operator $E[.]$ is given by

$$E[Z(\theta)] = \int_0^{2\pi} Z(\theta) . f_\theta \, d\theta$$

and the complex characteristic function Φ_p of the random variable θ is defined as

$$\Phi_p = E[e^{ip\theta}] = \alpha_p + i\beta_p \qquad p = 0, \pm 1, \pm 2..... \tag{5.7}$$

To be mathematically exact, a circular distribution is always defined by its trigonometric moments. Use of the characteristic function is more readily understood by thinking of the directional distribution as wrapped around a unit circle in the complex 'z-plane', Fig. 5.4 (a). The characteristic function can then be written as

$$\Phi_p[\theta] = \int_0^{2\pi} \{cos\, p\theta + i\, sin\, p\theta\} D(\theta)\, d\theta = \alpha_p + i\beta_p$$

where the coefficients α_p, β_p can then be thought of as the p^{th} angular moments about the real and complex axes. The characteristic function is similar to the coefficients of a Fourier series . In fact, Mardia shows the numbers $2\pi\Phi_p$ are Fourier-Stieltjes coefficients of $D(\theta)$ and that when the summation

$$\sum_{p=1}^{\infty} \alpha_p^{\,2} + \beta_p^{\,2}$$

is convergent the directional distribution is given by

$$D(\theta)=\frac{1}{2\pi}\sum_{p=-\infty}^{\infty}\Phi_p e^{-ip\theta} \qquad [5.8]$$

Given only the first two angular harmonics Equation [5.8] is not useful as an estimator of $D(\theta)$ however, Mardia (1972) shows how the angular moments can be used to define measures of location, concentration, skewness and kurtosis.

One potential use of Equation [5.8] is to determine the long-term directional distributions of winds, waves and currents using sample estimates of the characteristic function. The process, outlined later, is implemented very simply using observations of direction.

5.3.1 THE MEAN OR LOCATION

A measure of mean, or location for the directional variance spectrum, can be defined as the weighted integral of the unit direction vectors $P_p=[\cos p\theta,\sin p\theta];p=1,2..$; the weighted integral of P_p has resultant $R_p=\left\{\alpha_p{}^2+\beta_p{}^2\right\}^{1/2}$ and direction $\overline{\theta}$ such that $\alpha_p=R_p\cos p\overline{\theta}$, and $\beta_p=R_p\sin p\overline{\theta}$. Recalling Equation [A8] the real and complex components of the first moment can be calculated using the first three Fourier coefficients obtained from the co and quadrature spectra. These values give estimates of the location θ_1,θ_2 and resultant R_1,R_2. Having calculated the mean direction the *centred* circular moments $\overline{\alpha}_p,\overline{\beta}_p$ can be calculated using

$$\overline{\alpha}_p=E\left[\cos\left(p\left(\theta-\overline{\theta}\right)\right)\right]=\alpha_p\cos\left(p\overline{\theta}\right)+\beta_p\sin\left(p\overline{\theta}\right) \qquad [5.9]$$

$$\overline{\beta}_p=E\left[\sin\left(p\left(\theta-\overline{\theta}\right)\right)\right]=-\alpha_p\sin\left(p\overline{\theta}\right)+\beta_p\cos\left(p\overline{\theta}\right) \qquad [5.10]$$

Taking the line moment measures of width, skewness and kurtosis Mardia *op cit* defines circular equivalents with similar properties. His values are essentially those adopted by Kuik et al (1988) who examined the relative performance of the line and circle estimators using Monte Carlo simulations. One useful result which can be used as a quality check on data is the inequality

$$1-\overline{\alpha}_2-\overline{\beta}_2-2\overline{\alpha}_1{}^2\left(1-\overline{\alpha}_2\right)\geq 0 \qquad [5.11]$$

this check was used with the constraints on α_i, β_i to identify gross measurement errors in the DB1 data.

5.3.2 CIRCULAR VARIANCE, SKEWNESS AND KURTOSIS

The dispersion of the directional distribution about the reference vector P', Fig. 5.4 (b), is defined by Mardia (1972) as

$$s_0 = 1 - \int_0^{2\pi} \cos(\theta^* - \bar{\theta}) D(\theta^*) d\theta^* \qquad [5.12]$$

noting that $cos(\theta - \bar{\theta}) = cos\,\theta\, cos\,\bar{\theta} + sin\,\theta\, sin\,\bar{\theta}$ we get

$$s_{01} = 1 - R_1 \; ; \; 0 < s_{01} < 1 \text{ and } s_{02} = 1 - R_2 \; ; \; 0 < s_{02} < 1$$

This measure of dispersion is related to the line equivalent σ_0 using the fact that for symmetric cases $\beta_p = 0$ in which case

$$\Phi_p = \alpha_p = e^{-1/2\sigma^2 p^2}$$

hence,

$$\sigma_{01} = \left\{-2 log_e(1 - s_{01})\right\}^{1/2} \; ; \; 0 < \sigma_{01} < \infty$$

and,

$$\sigma_{02} = \left\{-\tfrac{1}{2} log_e(1 - s_{02})\right\}^{1/2} \; ; \; 0 < \sigma_{02} < \infty \qquad [5.13]$$

Alternatively, both the location and concentration can be defined by the second angular moments. The relative robustness of each method is examined later by fitting the von Mises distribution to Monte Carlo simulated directional distributions.

The asymmetry and peakedness of line variables is generally measured by skewness and kurtosis. Circular analogs of these quantities are defined by Mardia (1972) using the centred second angular moments $\bar{\alpha}_2, \bar{\beta}_2$. He proposes a measure of skewness given by $\gamma_1 = \bar{\beta}_2 / s_0^{3/2}$. For a perfectly symmetric distribution $\bar{\beta}_p$ will be zero and consequently the skewness γ_1 is zero as expected. In the limiting case when the variance s_0 is 1.0 then the data are evenly distributed

on the unit circle and γ_1 has no meaning. In the line case, kurtosis is measured relative to the normal distribution, consequently, Mardia suggests kurtosis be measured by

$$\gamma_2 = \left\{\bar{\alpha}_2 - (1 - s_0)^4\right\} / s_0^2 \qquad [5.14]$$

Kuik et al (1988) points out that this measure is, in a sense, model dependent since it is made to equal zero for the wrapped normal distribution. At first sight this should not be used as justification for rejecting it as a measure of kurtosis, however, the simulation results presented below show that Mardia's model free measure is a better measure of kurtosis, as suggested by Kuik et al (1988), where

$$\gamma_2 = \left\{\bar{\alpha}_2 - 4\bar{\alpha}_1 + 3\right\} / s_0^2 \qquad [5,15]$$

The performance of these measures of width, skewness and kurtosis is examined in Section [5.3.5] using Monte Carlo simulation. Unlike the work of Kuik et al (1988) the simulation is used to examine whether or not these measures can be used to assess likely bias in observed estimates of directional spreading recorded by a floating buoy.

5.3.3 VON MISES DISTRIBUTION

The circular equivalent of the normal distribution proposed by von Mises has a density function $f_\theta(\theta;\mu_0,\kappa)$ given by

$$f_\theta(\theta;\mu_0,\kappa) = \frac{1}{2\pi I_o(\kappa)} exp\left[-\kappa Cos(\theta - \mu_o)\right] \qquad [5.16]$$

$$0 \le \theta \le 2\pi\,;\kappa \ge 0;\, 0 \le \mu_o \le 2\pi$$

where: μ_o is the mean direction; κ is the concentration; and $I_n(x)$ is the modified Bessel function of the first kind and order n

$$I_n(x) = \sum_{k=0}^{\infty} \frac{x^{2k+n}}{2^{2k+n}\, k!\, \Gamma(n+k+1)}$$

Since the argument of the gamma function is a natural number the factorial form can be used where $\Gamma(x+1) = x!$, consequently the zeroeth order Bessel function is given by

$$I_o(x) = \sum_{k=0}^{\infty} \frac{x^{2k}}{2^{2k}(k!)^2} \qquad [5.17]$$

This function is real valued for all real x and takes values greater than zero for all $x \neq 0$. The von Mises distribution is symmetric about the mean and uni-modal. When the concentration about the mean is narrow then κ is large and the distribution tends to the normal distribution $N(x;\mu,\sigma)$, conversely, when the concentration about the mean is wide then κ is small and, in the limit when κ = 0 the distribution tends to the circular uniform distribution. The trigonometric moments of the von Mises distribution are given in closed form by Mardia (1972) as

$$\alpha_p = A_p(\kappa)Cos(p\mu_o) = \frac{I_p(\kappa)Cos(p\mu_o)}{I_o(\kappa)} \qquad [5.18]$$

$$\beta_p = A_p(\kappa)Sin(p\mu_o) = \frac{I_p(\kappa)Sin(p\mu_o)}{I_o(\kappa)} \qquad [5.19]$$

That is, $A_p(\kappa) = I_p(\kappa)/I_0(\kappa)$, where I_p is the p^{th} order modified Bessel function of the first kind. Knowing the first moments of the directional distribution then enables us to solve for the location and concentration.

5.3.4 SOLUTION FOR THE VON MISES PARAMETERS

Characteristics of the directional probability distribution can be estimated using the circular equivalents of the mean, standard deviation, skewness and kurtosis. Kuik et al (1988) show how these parameters can be estimated using the Fourier coefficients derived from the co- and quadrature-spectra inferred from a buoy heave, pitch and roll time series. Recalling Equations [5.18 & 5.19], the parameters in the von Mises distribution can be solved using the four equations in α_p, β_p

$$\alpha_1 = \frac{I_1 Cos\theta_o}{I_o}$$

$$\beta_1 = \frac{I_1 Sin\theta_o}{I_o}$$

$$\alpha_2 = \frac{I_2 Cos2\theta_o}{I_o}$$

$$\beta_2 = \frac{I_2 Sin 2\theta_o}{I_o}$$

Noting $I_p = I_p(\kappa)$, these equations therefore contain two unknowns κ and θ_o. The angle θ_o is most simply solved using either the first moment estimator

$$\frac{\alpha_1}{\beta_1} = \frac{\cos\theta_0}{\sin\theta_0} \quad => \quad \theta_0 = \tan^{-1}\left\{\frac{\beta_1}{\alpha_1}\right\} \qquad [5.20]$$

or the second moment estimator

$$\frac{\alpha_2}{\beta_2} = \frac{Cos 2\theta_0}{\sin 2\theta_0} \quad => \quad \theta_o = \frac{1}{2}\tan^{-1}\left\{\frac{\beta_2}{\alpha_2}\right\} \qquad [5.21]$$

Having found θ_o the first moment estimate of the concentration parameter κ can be solved either by using a numerical algorithm or by using the least squares estimator suggested by Lenth (1981)

$$\frac{1}{\kappa_1} \approx 2(1-\overline{R}_1) + (1-\overline{R}_1)^2.\{0.48794 - 0.82905\overline{R}_1 - 1.3915\overline{R}_1^2\} \qquad [5.22]$$

this simple estimator has an absolute error of less than 0.005 per cent for resultants $\overline{R}_1 \geq 0.12$ which will be the case for nearly all directional wave data in which there is a predominant wave direction. The second moment estimator for the concentration can be obtained by noting

$$I_2(\kappa) = I_0(\kappa) - 2I_1(\kappa)/\kappa \qquad [5.23]$$

and,

$$R_2 = \sqrt{\alpha_2^2 + \beta_2^2} = I_2(\kappa)/I_0(\kappa) \qquad [5.24]$$

Then

$$\begin{aligned} R_2 &= 1 - 2I_1(\kappa)/\{I_0(\kappa)\kappa\} \\ &= 1 - 2R_1/\kappa \\ \kappa_2 &= -2R_1/(R_2 - 1) \end{aligned} \qquad [5.25]$$

The second moment estimate is therefore calculated using both the first and second moment resultants R_1, R_2; it is therefore necessary to examine the behaviour of this estimator using simulation since we have seen the first moment resultant gives highly biased estimates for the

cosine model spreading power s_1 when only small amounts of noise are introduced into the directional distribution.

5.3.5 ROBUSTNESS OF MARDIA'S DESCRIPTIVE PARAMETERS

Before applying the modelling methods discussed above to the DB1 data it is instructive to examine the robustness of the cosine and von Mises model parameter estimates. A sample of randomly simulated directional distributions was generated using the double cosine mixture Equation [5.4], with parameters given by

$$\begin{aligned} s_i &= 1.0 + 20U(0,1) \\ p_i &= 1.0 + 4.0U(0,1) \\ m_i &= 0.4U(0,1) \\ \zeta_i &= (\pi / 2)\,U(0,1) \end{aligned} \qquad [5.26]$$

Here $U(0,1)$ is the uniform random number on the interval (0,1). Each parameter used a separate random number to ensure independence. The real and imaginary components of the characteristic function (Equation [5.7]) were calculated by numerical integration of Equation [5.4] and the solutions for the cosine and von Mises model parameters were obtained by using Equations [A11 to A14] and Equations [5.20 to 5.25]. The range of simulated distributions given by the double cosine mixture is of course arbitrary, nonetheless, if $m_i \cos^{2p_i}(\theta / 2)$ is regarded as the true peak, and $(1-m_i)\cos^{2p_i}\left(\left[\theta-\zeta_i\right]/2\right)$ is regarded as noise, then tentative estimates of the bias in estimated directional parameters may be made using the results from this simulation.

The angular harmonics $\{\alpha_i,\beta_i;i=1,2\}$ are plotted in pairs in Fig. 5.5 which shows that only α_1,α_2 are strongly correlated. In fact Mardia (1972) shows that $\alpha_2 \approx 2\alpha_1^2 - 1$. The correlation increases as α_1,α_2 increases, that is, correlation increases as the directional variance decreases. Other pairing show little or no correlation, on the other hand the bounds $\alpha_i^2 + \beta_i^2 \le 1$ can be seen in the scatter plots. One important property is the correlation between resultants R_1 and R_2 shown in Fig. 5.6 for the 100 simulated distributions. As the resultants increase so does the correlation, which suggests for narrow beam seas the buoy response can be checked by comparing the first and second moment resultants.

Mardia's measure of skewness is based on scaling the second imaginary component of the characteristic function by the variance obtained from the first resultant s_0. The bias in cosine

power s_2 is plotted against both $\overline{\beta}_2$ and skewness, $\overline{\beta}_2 / s_0^{3/2}$, in Fig. 5.7. The bias has been defined as

$$b = (s_2 - \hat{s}_2) / s_2 \qquad [5.27]$$

Strong correlation does not exist in either case, however, a least squares fit to $\overline{\beta}_2 / s_0^{3/2}$ *vs* bias gives

$$b_1 = 0.2|\gamma_1| + 0.22 \qquad [5.28]$$

The bias of cosine power s_2 is also affected by kurtosis γ_2. Various estimators have been proposed for kurtosis, for example one suggestion is $\overline{\alpha}_2$ alone , and Mardia (1972) suggests standardising with respect to the wrapped normal giving

$$\gamma_2 = \left\{\overline{\alpha}_2 - (1 - V_0)^4\right\} / s_0^2 \qquad [5.29]$$

More recently, Kuik et al (1988) recommend a model free kurtosis

$$\gamma_2 = \left\{\overline{\alpha}_2 - 4\overline{\alpha}_1 + 3\right\} / s_0^2 \qquad [5.30]$$

The correlation between these estimators and bias in cosine power s_2 is shown in Fig. 5.8. The simple $\overline{\alpha}_2$ estimator (b) and excess kurtosis (c) shows little correlation with s_2 bias, only the model free kurtosis (d) shows any real correlation and consequently it has been adopted for this work. Using least squares regression it was found an estimator of bias in s_2 due to kurtosis is given by

$$b_2 = 0.061\gamma_2 + 0.07 \qquad [5.31]$$

If we can assume skewness and kurtosis are bivariate Normal the expected total bias of b_T is given by $E[b_T] = E[b_1] + E[b_2]$ where $b \leq 1$. The effect of using Equations [5.28 & 5.31] as corrections can be seen in Fig. 5.9 which shows that for $s \leq 10$ the correlation between 'true' spread power and the corrected second moment estimate is good, but, for $s > 10$ the method results in some significant overestimates of spreading power.

5.4. SPREADING RECORDED BY DB1 IN STORM SEAS.

Previous studies of wave directional spreading have concentrated on establishing a model which is consistent with the physics of ocean wave generation. For example, Ewing &

Laing (1987) examine the spectra of seas near full development with a view to checking Komen's assertion that spreading is most narrow at some frequency greater than the spectral peak. Such studies are aimed at developing a unified model of directional spreading which can be used by oceanographers and designers in the same way that Pierson-Moskowitz and JONSWAP spectra are used as frequency dependence models. At present, the most accurate models of directional spreading are probably those due to Hasselmann et al (1980) and Mitsuyasu et al (1978). However, both models are based on data recorded by pitch and roll buoys in relatively low seastates and both show the spreading is sensitive to the frequency ratio f / f_m.

The importance of directional spreading on structural response is discussed in detail in Ch. 6 in which it is shown that fixed and compliant system response can either be attenuated by up to 15% if spreading is allowed for; or amplified by some 10% in some of the modes of motion of an articulated structure like a tower-tanker mooring. However, there remains considerable scepticism amongst designers about the degree of spreading in extreme seas which effectively dominate structural design (ISSC 1991, Committee I.2). For this reason it was decided this chapter should examine the spreading in the most extreme seas recorded by the DB1 buoy.

In what follows no attempt was made to ensure the environmental conditions correspond to a stationary wind field and seas near full development. Instead, the spread in each extreme sea was examined unconditionally in order to identify whether, or not, the spread of extreme seas was more narrow than predicted by Hasselmann or Mitsuyasu models.

The DB1 buoy recorded heave acceleration, pitch and roll for 20 minutes every three hours at 1.2 second intervals, giving a Nyqvist frequency of 0.4167Hz. As mentioned previously, a faster sampling rate would have been desirable, especially in low seastates, however, for the extreme seas examined in this chapter the minimum zero crossing period is 10 seconds which suggests aliassing should be less of a problem. The co- and quadrature spectra recorded by the buoy were corrected to correspond to slope time series in compass north-east axes. The auto and cross spectra were formed from averages of 11 non-overlapping sections, each of 100 second duration, resulting in spectra estimates with a resolution of 0.01Hz and having 22 degrees of freedom. This sampling scheme was later recognised as responsible for leakage in the spectra which may contribute to some of the scatter found in the directional estimates, Tucker (1991).

The frequency domain analysis results reported in Ch. 3 were used to identify the dates and times when seastates exceeded 6.0m significant height. This identified 100 samples of which 68 datafiles had satisfactory quality control flags. The year, month, day, time, significant wave height and zero up-crossing period are given for the largest 100 in Table 5.3.

The cosine model parameters were calculated for each of the 68 seastates using the method proposed by Lonquet-Higgins et al (1963a) and Cartwright (1963), Appendix A, and the von Mises model parameters were calculated using the procedures outlined in Section [5.3.4]. A check on the results was made by comparing them with Fig. 3 from Ewing & Laing (1987).

The results for each of the 68 storms were first examined visually to determine the overall quality of the estimates. As shown previously the effect of noise and bias due to 'non-cosine' form can be assessed using

A the check ratio $R = (2\pi f)^2 / g.\left[c_{11} / (c_{22} + c_{33})\right]^{1/2}$

B the skewness and kurtosis

C the difference between first and second angular resultants in narrow seas

D Mardia's inequalities

The check ratios for the largest seastates are plotted in Fig. 5.10 which shows that values at or near the spectral peak $(f / f_m \approx 1.0)$ are close to 1.0. Below the peak the ratio k(theory)/k(buoy) is generally less than 1.0 and above the peak the check ratio is generally > *1.0*. This suggests values of power s or concentration κ are likely to be unreliable above and below the spectral peak.

When skewness is examined no obvious correlation or trend with (f / f_m) exists with values ranging from *0* to + 4; which is similar to the values obtained from the simulations described in Section [5.3.5]. Whether this is evidence of asymmetry in buoy response, or the wave elevation variance is not clear. Ewing (private communication) suggests that in fact the true directional distribution will be very nearly symmetric in which case the buoy response is probably responsible for the asymmetry. This is supported by the fact that the sign of the skewness changes over the range of frequencies in most spectra, if turning wind fields were responsible for asymmetry we would expect the signs to be consistent over the energetic parts of the spectrum.

Kurtosis values are weakly correlated with the concentration, as suggested by the simulation results. However, in many cases negative values were given at several frequencies in a spectrum. The likely cause of negative values is bias in $\overline{\alpha}_1$, due to noise. This suggests that $\overline{\alpha}_2$ may well be a more practical measure of kurtosis providing the seas are narrow beam and buoy response is a reasonable measure of surface displacement and slope.

In Section [5.3.5] it was found that correlation between resultants R_1 and R_2 increased with decreasing angular variance. The resultants at the spectral peak for each of the 68 storm spectra are shown in Fig. 5.11, where the continuous line, suggested by Tucker (1991), is given by solving

$$R_1/(1-R_1)=\left\{1+3R_2+\left(1+14R_2+R_2^2\right)^{1/2}\right\}/2(1-R_2) \qquad [5.33]$$

Note that for $R_1 < 0.5$ only negative solutions are obtained for R_2 using Equation [A14]. In fact the plot of s_1,s_2 against R_1,R_2 shown in Fig. 5.12 shows the minimum possible value of s_2 is 1.0. This explains the behaviour shown in Fig. 4 of Hasselmann et al (1980) which shows poor correlation between s_1 and s_2 for low s_1, together with a bound on s_1.

5.4.1 ANGULAR VARIANCE

The variation of angular variance σ_{01} with resonance angle or frequency ratio (f/f_m) is generally accepted as a real physical phenomenon. Unfortunately, the presence of noise and bias in the DB1 data makes it difficult to categorically confirm that the spread is narrowest at the spectral peak. Nonetheless, in all 68 cases the spreading at or near the spectral peak was found to be more narrow than in the fore and aft tails of the spectrum, this is shown in Fig. 5.13 in which, after allowing for noise, the spread is clearly narrowest at the peak. This is reflected in the second moment estimates of cosine spread power s_2 shown in Fig. 5.14 in which the values at the peak range from $3 \rightarrow 37$.

We have seen how large values of skewness γ_1 and kurtosis γ_2 can reduce the concentration parameter, consequently, it is worth looking at some examples where low powers were estimated. Three examples are shown in Fig. 15,16 & 17 which correspond to 13th February 1979 1800hrs, 28th March 1980 1500hrs and 10th December 1979, 0000hrs. In each case the frequency axis has been scaled to (f/f_m), where f_m is the frequency corresponding to the largest observed c_{11} value the spectrum. The directional spectrum recorded on 13th February 1979, Fig. 5.15, shows some variability away from the main peak and could be due to a mixture of wind driven and swell seas. The directional variance s_0 shows considerable variation over the full range of frequencies with a definite minimum near the peak of the spectrum. Neither skewness, kurtosis or check ratio suggest that the spread parameters s_2 and κ_2 will be excessively biased, and the first moment estimator of mean direction $\overline{\theta}_1$ is stable over most of the spectrum. If this is a pure wind driven spectrum, and not a mixture of swell and wind seas, then this seems to be a case where storm seas are not narrowly spread. It is interesting to note, however, that the frequency dependence of angular variance is relatively weak compared to Hasselmann or Mitsuyasu models.

In the second example, Fig. 5.16, recorded 28th March 1980 at 1500hrs, the spectrum is more narrow band with the lowest angular variance at one frequency interval past the spectral peak. Again, the skewness, kurtosis and check ratio do not suggest the results are doubtful and this is another case where spreading is not narrow beam.

The final example, Fig. 5.17, recorded 10th December 1979 at 0000hrs, is a case where the seas are very narrow beam at the spectral peak, with large skewness and kurtosis and check ratio of 1.006. This is clearly a case of near extreme seas which are practically uni-directional at the most energetic parts of the spectrum, and shows clearly that adoption of Hasselmann or Mitsuyasu models will in some cases result in unrealistic attenuation of wave forces on offshore structures.

5.4.2 PARAMETRIC MODELS OF SPREADING PARAMETERS

Hasselmann and Mitsuyasu propose a parametric form for their observed spreading powers based on least squares fitting to s_1 values (Hasselmann) and $\frac{1}{2}(s_1 + s_2)$ values (Mitsuyasu). Since recent work by Ewing, Tucker and this author demonstrates second moment estimators are more robust we propose a parameterisation based on s_2 using the model suggested by Hasselmann et al (1980)

$$\left.\begin{aligned} s &= \tilde{s}_m \{f / f_m\}^{\tilde{\lambda}} \\ \kappa &= \tilde{\kappa}_m \{f / f_m\}^{\tilde{\beta}} \end{aligned}\right\} \quad f < f_m$$

$$\left.\begin{aligned} s &= s_m \{f / f_m\}^{\lambda} \\ \kappa &= \kappa_m \{f / f_m\}^{\beta} \end{aligned}\right\} \quad f \geq f_m \qquad [5.34]$$

The s_m parameter effectively controls the maximum spread which is assumed to occur at the spectral peak, and the A parameter controls the 'peakyness' of the model. The use of two parameterisations results in a discontinuity at $f = f_m$ which Hasselmann resolved by switching equations at $f = 1.05 f_m$. Unlike the work of Hasselmann no attempt is made to further parameterise 'A' using the resonance angle U / c_m, furthermore, by simply modelling the 68 largest seastates recorded by the DB1 we have mixed seas at different stages of development. This can be seen in three examples discussed earlier.

To fit Equations [5.34] the data was subdivided into two groups: group I corresponds to data with $0.2 < f / f_m < 1.0$ and group II corresponds to data with $1.0 \leq f / f_m \leq 3.0$. Both groups are plotted on double log scales in Fig. 5.18 for s_2 and κ_2; truncation of the upper and

lower tails of the spectrum was chosen to be consistent with Fig. 3 of Hasselmann. The results of a least squares fit to the data gives

$$\left.\begin{aligned} s &= 9.1\{f/f_m\}^{0.84} \\ \kappa &= 4.9\{f/f_m\}^{0.89} \end{aligned}\right\} \quad f < f_m$$

$$\left.\begin{aligned} s &= 7.3\{f/f_m\}^{-0.92} \\ \kappa &= 4.7\{f/f_m\}^{-1.05} \end{aligned}\right\} \quad f \geq f_m \qquad [5.35]$$

Comparing these models with Table 5.1 shows that the peak values at $f = f_m$ are similar to Hasselmann's results with *7.3* instead of *9.77* and *9.1* instead of *6.97*. On the other hand, the decay of spreading powers $A, \tilde{A}$ is slower than Hasselmann's on both sides of the peak, Fig. 5.19. At first sight it may seem contradictory that the peaks agree so closely when in fact Equations [5.35] use s_2 when Hasselmann uses s_1. However, the simulations in Section [5.3.5] demonstrate that as angular beam width reduces the correlation between resultants, and therefore spreading powers, increases. Either side of the spectral peak the two resultants reduce and so too does their correlation giving rise to large differences between s_1 and s_2. This is probably the reason why our exponents, $A, \tilde{A}$, are lower than those found by Hasselmann.

5.4.3 WEIBULL MODEL OF SPREAD AT THE SPECTRAL PEAK

These models are mean value fits with considerable spread either side of this mean. For design purposes we need both upper and lower bound models for the spread and given the uncertainty in the spreading estimates away from the spectral peak (caused by non-unity check ratios) it is sensible to adopt a probabilistic description of the spreading which can be used in conventional offshore structure reliability analysis.

A 3-parameter Weibull model was fitted by maximum likelihood to the 68 s_2 and κ_2 values observed at the spectral peak. The result, shown in Fig. 5.20, shows a lower bound of 1.2 for both s_2 and κ_2 and a long upper tail, corresponding to near unidirectional seas. The Weibull model parameters are:

model	shape	scale	location	mean
cosine	1.15	17.46	1.6	17.68
von Mises	1.25	8.12	1.28	8.73

Note the mean values demonstrate the regression fit to the data is biased downward at the spectral peak and therefore the model given by Equations [5.35] are only appropriate for

fatigue calculations and should not be used in design event simulation. It is recommended that the Weibull models be used for this purpose.

5.6. ROBUSTNESS OF VON MISES PARAMETER ESTIMATES

The discussion in Section [5.4] examines the classical approach for maximum likelihood estimation of the parameters of a von Mises distribution. In Section [5.3.5] the robustness of Longuet-Higgins estimator for the cosine model spreading powers was examined and the results confirm Kuik et al (1988) assertion that the procedure is not suited to the routine analysis of directional wave data. This will also be the case for the simple von Mises estimator described above since it also relies on equating circular moments. A robust estimator is therefore required which is not sensitive to the presence of either non-colinear wind and wave seas, or asymmetric directional spreads due to bias, noise or the veering of wind systems.

Robust estimation of directional data is examined by Lenth (1981) for the case when discrete observations have been taken. Unfortunately there is no direct extension to this method for this work in which only the first two angular moments are available through co and quadrature spectra. However, the *W*-estimator sub-class of Huber's *M*-Estimation (1981) principal leads to the development of a weighted density approach for estimating the directional parameters for a chosen distribution. The concept of a weighted distribution was formalised by Rao (1965) who recognised that real observations of random variables are very often modified in some way, for example, by truncation, corruption, and missing values.

Weighted estimation is defined by Patil et al (1977) as follows. Consider a natural process generating a random variable X with *pdf* $f(x;\theta)$ where $\theta \in \Omega$ is the parameter space. When measuring a random sample of observations on X we have to use a method of selection which gives the same chance of including in the sample any observation produced by the true process. However, in practice our measurements are not perfect and the relative chances of inclusion of two observations x and y, say, are $w(x)$ and $w(y)$ where $w(.)$ is some non-negative valued function. Then the recorded value X_w has the pdf

$$f^w(x;\theta) = w(x)f(x;\theta) / \omega \qquad [5.36]$$

where

$$\omega = E[w(x)] = \int w(x)f(x;\theta)\, dx \qquad [5.37]$$

For our purposes $f(x;\theta)$ is the directional distribution and $w(x)$ is some weighting function chosen to 'correct' the measured data which has been affected by imperfect buoy response or veering winds. Unlike the classical *M*-Estimation process in which discrete observations on direction are available we only have the real and imaginary components of the first two angular moments. Consequently, if the form of $f(x;\theta)$ and $w(x)$ are parameterised in some way then the solution for the weighted estimates of the model's location and scale parameters is obtained from solving

$$\alpha_p = E\left[w(\theta-\mu;\kappa).\cos p\theta\right] = \frac{1}{\omega}\int_0^{2\pi} w(\theta-\mu;\kappa).\cos p\theta.f(\theta;\mu,\kappa)d\theta \qquad [5.38]$$

$$\beta_p = E\left[w(\theta-\mu;\kappa).\sin p\theta\right] = \frac{1}{\omega}\int_0^{2\pi} w(\theta-\mu;\kappa).\sin p\theta.f(\theta;\mu,\kappa) \qquad [5.39]$$

where,

$$\omega = E\left[w(\theta-\mu;\kappa)\right] = \int_0^{2\pi} w(\theta-\mu;\kappa).f(\theta;\mu,\kappa)d\theta \qquad [5.40]$$

The density $f(\theta;\mu,\kappa)$ might be the von Mises probability density, and the weight function $w(\theta-\mu;\kappa)$ might be some function based on known characteristics of either instrumental or buoy response.

5.6 LONG-TERM DIRECTIONAL PROBABILITY DISTRIBUTIONS

Two methods can be used to determine the directional distribution of winds and waves given a sample of observations $\theta_i; i = 1,n$. The simplest method is to divide the range $[0,2\pi]$ into class intervals and then count the frequencies in each class, much in the same way as for a line histogram. Historically, this approach has been modified to produce rose and circular histogram estimators which are useful graphical aids, but can be difficult to use in practical analysis of offshore structures. The second approach uses the characteristic function defined in Section [5.3]. This method is appealing because it enables us to use some of the statistical theorems outlined in Mardia (1972) for univariate and bivariate observations of directional data and, in addition, provides a simple parametric form for calculating the directional density in design or analysis calculations.

Line histogrammes of circular data are probably the most simple density estimators to interpret and construct. The range $[0,2\pi]$ is simply divided into a number of class intervals and the frequency corresponding to each class is counted. Three examples are shown both in Fig. 5.21 and Fig. 5.22 which show estimates of the directional distribution of currents and winds recorded by the DB1. The first example has class intervals 1 degree wide and shows a noisy sinusoidal trend. In places spikes occur which exceed the range of the plot; these are probably due to instrumental or recording error and could be removed from the data by inspection and replaced by the average of the estimates either side. The second example, with bin width of 10 degrees, has several minor peaks which do not show up clearly in the first example, and the third example, with bin width of 20 degrees, shows a smoother variation.

The second method for estimating the directional probability density uses the characteristic function. Given a sample of discrete observations of directional data $\theta_i; i = 1,n$, sample estimates of the real and imaginary components of the characteristic function are given by

$$\overline{C}_p = \frac{1}{n}\sum_i Cos\, p\theta_i \qquad \overline{S}_p = \frac{1}{n}\sum_i Sin\, p\theta_i \qquad [5.35]$$

Where

$$\overline{R}_p = \left(\overline{C}_p^{\,2} + \overline{S}_p^{\,2}\right)^{1/2} \quad \overline{C}_p = \overline{R}_p Cos\,\overline{\theta}_{po} \quad \overline{S}_p = \overline{R}_p Sin\,\overline{\theta}_{po} \qquad [5.36]$$

In the limit as $n \to \infty$ then $\overline{C}_p \to \alpha_p$ and $\overline{S}_p \to \beta_p$. An estimate of the directional distribution is given by the characteristic function using

$$D(\theta) = \frac{1}{2\pi}\sum_{p=-\infty}^{\infty} \phi_p e^{-ip\theta}$$

from which , if $\sum \left|\Phi_p\right|^2$ is convergent as $p \to \infty$, the directional density is given by

$$D(\theta) = \frac{1}{2\pi}\left\{1 + 2\sum_{p=1}^{\infty}\left[\overline{C}_p cos(p\theta) + \overline{S}_p sin(p\theta)\right]\right\} \qquad [5.37]$$

Further, the cumulative directional distribution is simply

$$F(\theta) = \frac{1}{2\pi}\left\{\theta + 2\sum_{p=1}^{\infty}\left[\overline{C}_p sin(p\theta) + \overline{S}_p(1 - cos\, p\theta)/p\right]\right\} \qquad [5.38]$$

If the directional distribution is known to be unimodal then the resultants $\overline{R}_p$ and the locations $\overline{\theta}_{po}$ can be solved using a sample set $\{\theta_i\}$. Use of Equation [5.37] requires sample estimates of (α_p, β_p) which are formed by summation over the weighted harmonics of the directional distribution. To begin with 1000 summations were performed resulting in the current and wind directional distributions shown in Fig. 5.23 and Fig. 5.24. In both cases the result is noisy and the general trend is difficult to identify. The effect of increasing the number of (α_p, β_p) terms was examined by increasing the number of summations to 50000, however, this required considerable computing effort and resulted in a more noisy estimator with large bias, this estimator is therefore not consistent in the sense defined in Ch. 3, Section [3.1]. It was also noticed that $\sum|\Phi_p|^2$ did not seem to be converging as $p \to \infty$, this should be examined in more detail in a future study.

A further problem with using a large number of terms in the summation is the difficulty of using the result. In fact, what is required is a smoothed estimate, however, we shall see below that it is better to simply take only the first few terms from Equation [5.38] since this results in a model which is simple enough to use in design calculations.

5.6.1 SMOOTHING DIRECTIONAL DISTRIBUTIONS

The principles of smoothing have been discussed in the context of kernel density estimation for line random variables in Ch.3. In this case, where we have noisy histogram and characteristic function estimators, the simplest approach is to smooth the data $f(x)$ subjectively by convolution with a kernel weighting function $k(x)$

$$\tilde{f}(x) = \int_{-\infty}^{\infty} f(y)k(y-x)dy$$

For large samples, this convolution is performed most quickly by using the fact that convolution in the space domain is equivalent to multiplication in the frequency domain: that is

$$\tilde{f}(x) = \mathcal{F}^{-1}\left[\mathcal{F}\left[f_i(y)\right]\mathcal{F}\left[k_i(y-x)\right]\right]$$

where, $\mathcal{F}[\ ]$ is the Fourier transform, and $\mathcal{F}^{-1}[\ ]$ is the inverse Fourier transform. Several kernel types exist; two simple examples are the square kernel, and the exponential kernel. Both were tried and the best results were obtained using an exponential kernel. The smoothed

histogram and characteristic function estimators are shown for the currents and winds in Fig. 5.25 and Fig. 5.26. Both types of estimator agree well with the exception of the sharp discontinuities in the current estimators (which are due to the spikes observed in the 1 degree interval histogram).

5.6.2 DESIGN MODELS OF DIRECTIONALITY

The two estimators discussed so far have not been suitable for design or analysis being too detailed. A simple parametric model is given by summing only the first few terms of Equation [5.38], for example, Fig. 5.27 shows the current and wind directional distributions obtained using only the first ten terms in the summation. In both cases the main and secondary peaks evident in the smoothed results shown in Fig. 5.25 and Fig. 5.26 can be seen in these simple estimators. Of course fewer terms in the summation can be used and some experimentation is necessary to obtain an acceptable fit. The values of the first ten real and imaginary components of Φ_p are given for the currents and winds in Table 5.4.

5.7 CONCLUSIONS

Measures of location, spread, skewness and kurtosis have been taken from the work of Mardia. These non-parametric estimators can be used to assess: the performance of a buoy's surface following characteristics; and the robustness of the simple moments estimator developed by Cartwright.

The robustness of equating the first and second angular moments of directional wave data has been examined by simulating double cosine mixtures for the directional distributions. The results have then been used to develop an estimator for the probable bias in the estimated spread given the amount of circular skewness and kurtosis.

Solutions for the location and concentration parameters of a von Mises distribution have been derived. The solution uses the real and imaginary components of the characteristic function which are obtained directly from the co- and quadrature spectra for the buoy heave, pitch and roll time series. The advantages of the von Mises distribution are the 'common sense' nature of the concentration parameter and the considerable literature which supports the various statistical aspects of estimating its parameters.

The largest significant wave heights observed by the DB1 buoy have been used to select those co- and quadrature spectra which were recorded in extreme seas (Hs > 6m). The intention being to test whether spreading reduced as the seastate increased. After screening the data 63 samples were assumed to be uncorrupted and composed of wind driven seas. These data were then used to develop a regression model for the frequency dependent - second moment estimate - cosine model spreading powers and von Mises model concentration parameters.

The models are largely in agreement with the Hasselmann study results - despite their use of the average of the first and second moment estimates for the cosine spreading power. The major differences are either side of the spectral peak where this study predicts a slower increase in the directional width. There did not seem to be any correlation between the significant height and the angular variance.

For design it is suggested that the spreading power be taken as frequency independent and that the results at the spectral peak be used for the whole range of f/fm.

The results at the spectral peak show considerable scatter which has been modelled by fitting a Weibull distribution to the observed values. The result has a lower bound of 1.2, a

modal value of 4.6 and a mean of 17.7. It is suggested the effect of this variability be examined at a later date by including the uncertainty in a reliability analysis.

A method of weighted estimation has been presented in which the moments inferred from the co- and quadrature spectra are assumed to be from a weighted directional distribution. This method was tested but with limited success. The major problem seemed to be the large variability in the skewness values between discrete frequencies in each set of co - and quadrature spectra. A further problem was that the physical nature of the weighting (which is creating the skewness) was unknown. If the general form of the weighting can be identified then it is likely the method could be used successfully. In particular, it would be possible to develop a method in which the fitting is done for the full set of data across all frequencies rather than on a frequency-by-frequency basis, as for the present case.

The use of circular statistical methods is equally valid for describing the long-term directional distributions of the winds, waves and currents. A method based on the use of Fourier-Stieltjes series is presented which enables multimodal directional distributions to be described using only a few terms from the characteristic function. This method compares well with conventional circular histogrammes and has the added advantage of being a continuous function on the circle.

APPENDIX A - CALCULATION OF THE DIRECTIONAL SPECTRUM

The calculation of a directional spectrum, for example using the heave, pitch and roll motions of the DB1 buoy, is based on the assumption that the sea surface topography can be represented as the summation of an infinite number of cosine wave components, linearly superposed over a range of directions i and frequencies j. This assumption ignores the non-linear effects like wave breaking and shallow water harmonics but, nonetheless, provides a model which correlates well enough with real seas for engineering applications. The double summation is based on the two dimensional cosine wave equation

$$\eta(x,t) = a_j \cos\left(k_j x - 2\pi f_j t + \psi_j\right) \qquad [A1]$$

where k_j is the wave number, f_j is the frequency in Hz, and ψ_j is a phase. Equation [A1] may be transformed into three dimensions, Fig. 1, giving

$$\eta(X,Y,t) = a_{ij} \cos\left(k_j\left[X\cos\theta_i + Y\sin\theta_i\right] - 2\pi f_j t + \psi_{ij}\right) \qquad [A2]$$

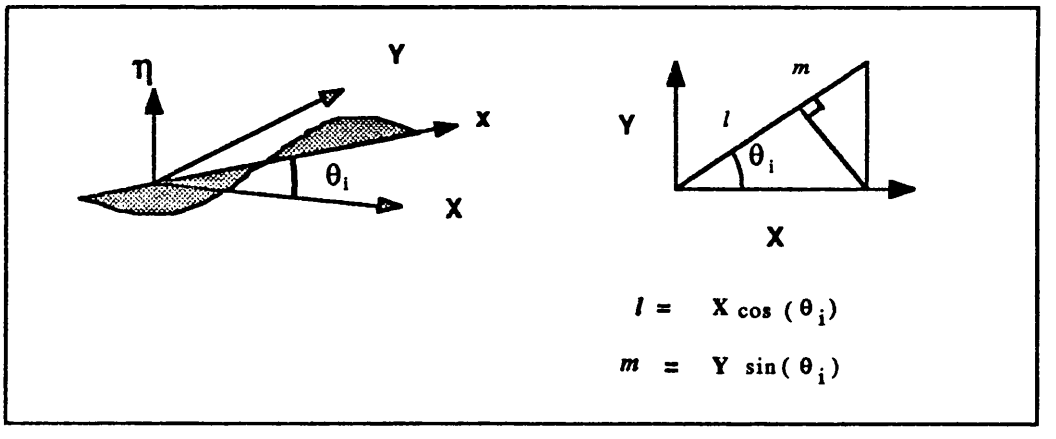


Figure 1 2_D Wave transformation

Assuming a real sea surface can be modelled as a finite number of linearly superposed waves with the form of Equation [A1], then

$$\eta(X,Y,t) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} \cos\left(k_j\left[X\cos\theta_i + Y\sin\theta_i\right] - 2\pi f_j t + \psi_{ij}\right) \qquad [A3]$$

Equation [A3] represents the variation of the sea surface in both space and time, however, the recordings made with a pitch and roll buoy are the time histories of the surface elevation and slope derivatives of this equation at some fixed point in space (X',Y'), i.e.

$$\eta(X',Y',t);\frac{\partial\eta}{\partial X};\frac{\partial\eta}{\partial Y} \qquad [A4]$$

The uni-directional variance spectrum $S(f)$ may be determined using one of the standard procedures - like the fast Fourier transform or maximum likelihood method - and only requires the heave response history. For this simple case, Equation [A3] reduces to

$$\eta(t)=\sum_{j=1}^{m} a_j \cos\left(2\pi f_j t+\psi_j\right) \qquad [A5]$$

Note in this equation, the amplitudes a_j and the phase angle ψ_j do not represent physical reality since at each frequency 'j' the amplitude and phase are the result of waves from all directions 'i' between 0 to 2π radians. To determine the directional variance spectrum $S(f,\theta)$ both the heave and two slope derivatives are required. The procedure for calculating the directional spectrum uses the co and quadrature spectral density matrices for the heave,pitch and roll time series, for examples see Hasselmann (1980) or Tucker (1991). If subscript 1 denotes vertical acceleration, and subscripts 2 and 3 denote the tilts in the north and east directions then there are six independent non-zero spectra

$$\begin{bmatrix} C_{11} & 0 & 0 \\ & C_{22} & C_{23} \\ & & C_{33} \end{bmatrix} \text{ and } \begin{bmatrix} 0 & Q_{12} & Q_{13} \\ & 0 & 0 \\ & & 0 \end{bmatrix}$$

These co and quadrature spectra are given by Longuet-Higgins as

$$
\begin{aligned}
C_{11}(f) &= \int_0^{2\pi} S(f,\theta)\, d\theta \\
C_{22}(f) &= \int_0^{2\pi} k^2 \cos^2(\theta)\, S(f,\theta)\, d\theta \\
C_{33}(f) &= \int_0^{2\pi} k^2 \sin^2(\theta)\, S(f,\theta)\, d\theta \\
Q_{12}(f) &= \int_0^{2\pi} k\cos(\theta)\, S(f,\theta)\, d\theta \\
Q_{13}(f) &= \int_0^{2\pi} k\sin(\theta)\, S(f,\theta)\, d\theta \\
C_{23}(f) &= \int_0^{2\pi} k^2 \sin(\theta)\cos(\theta)\, S(f,\theta)\, d\theta
\end{aligned}
\qquad [A6]
$$

where k is the wave number, and the unknown quantity is the directional spectrum $S(f,\theta)$. To solve for the directional spectrum a Fourier approximation is assumed, where

$$
a_n + ib_n = \frac{1}{\pi}\int_0^{2\pi} e^{in\theta} S(f,\theta) d\theta \qquad n = 0,1,2. \qquad [A7]
$$

setting n=0,12 and equating the Fourier terms to the relevant co and quad spectra gives solutions for a_n, b_n. Ewing & Laing (1987) point out the convenience of the normalised angular harmonics $A_n = a_n / a_0$, $B_n = b_n / a_0$ which are then given as

$$
\begin{aligned}
A_1 &= Q_{12} / \left[C_{11}(C_{22} + C_{33})\right]^{1/2} \\
B_1 &= Q_{13} / \left[C_{11}(C_{22} + C_{33})\right]^{1/2} \\
A_2 &= (C_{22} - C_{33}) / (C_{22} + C_{33}) \\
B_2 &= 2C_{23} / (C_{22} + C_{33})
\end{aligned}
\qquad [A8]
$$

The wave number can be obtained from both

$$
k^2(f) = (C_{22} + C_{33}) / C_{11} \qquad [A9]
$$

and $k = (2\pi f)^2 / g$; the ratio of the two estimators provides a simple check ratio which is used to assess the effect of noise in the signals. Longuet-Higgins et al (1963a) and Cartwright (1963) proposed the directional distribution of variance $S(f,\theta)$ be written as the product $S(f,\theta) = D(f,\theta)S(f)$; where the directional spread $D(f,\theta)$ is uni-modal at each frequency. They suggested a suitable functional form for $D(f,\theta)$ is given by the exponentiated cosine

$$D(f,\theta) = C\cos^{2s}\left(\tfrac{1}{2}\left(\theta - \bar{\theta}\right)\right) \quad ; \quad -\pi < \theta < \pi \qquad \text{[A10]}$$

in which the cosine power s is usually frequency dependent $s = s(f)$. By equating the angular moments of $D(f,\theta)$ to the angular harmonics from the Fourier coefficients, Equation [A8], two estimates for the cosine power and the mean direction can be calculated.

(i) First moment estimator

$$\theta_1 = tan^{-1}\left\{\frac{B_1}{A_1}\right\} \qquad \text{[A11]}$$

$$s_1 = c_1 / (1 - c_1) \qquad where \qquad c_1^2 = A_1^2 + B_1^2 \qquad \text{[A12]}$$

(ii) Second moment estimator

$$\theta_2 = \frac{1}{2} tan^{-1}\left\{\frac{B_2}{A_2}\right\} \qquad \text{[A13]}$$

$$s_2 = \left\{1 + 3c_2 + \left(1 + 14c_2 + c_2^2\right)^{1/2}\right\} / \left\{2(1 - c_2)\right\} \qquad \text{[A14]}$$

$$where \qquad c_2^2 = A_2^2 + B_2^2$$

The robustness of these estimators is discussed in Section [5.5]. It should be noted the normalised angular harmonics can be written as

$$A_1(f) = \int_0^{2\pi} Cos(\theta)\, D(f,\theta) d\theta$$

$$B_1(f) = \int_0^{2\pi} Sin(\theta)\, D(f,\theta) d\theta$$

$$A_2(f) = \int_0^{2\pi} Cos(2\theta)\, D(f,\theta) d\theta$$

$$B_2(f) = \int_0^{2\pi} Sin(2\theta)\, D(f,\theta) d\theta \qquad \text{[A15]}$$

Written in this form it will be clear these are equivalent to the real and complex components of the *characteristic function* of angular statistics.

In order to calculate more terms in the Fourier approximation additional C and Q terms are required. Mitsuyasu et al (1975) used a clover-leaf articulated buoy to measure heave, slope, and curvature, however, measurement of the curvature requires more complex instrumentation, and equipment has yet to be developed to the level of reliability achieved with the more common pitch and roll buoy.

	$f < f_m$	$f \geq f_m$
Mitsuyasu $s = 11.5\tilde{f}^{a}\tilde{f}_m^{b}$	$a = 5$ $b = -7.5$	$b = -2.5$ $a = 0.0$
Hasselmann $s = s_m\{f / f_m\}^{\mu}$	$s_m = 6.97\{\pm 0.83\}$ $\mu = 4.06\{\pm 0.22\}$	$9.77\{\pm 0.43\}$ $(-2.33\{\pm 0.06\}) - (1.45\{\pm 0.45\})\left[\frac{U}{c_m} - 1.17\right]$

where: f_m is the modal frequency ; $\tilde{f}$ is the dimensionless frequency; c is the wave phase speed; and U is the wind speed.

Non-dimensional fetch: $$\tilde{F} = \frac{gF}{U^2}$$

Frictional velocity: $$U_{10} = U = 25u_*$$

Mitsuyasu fetch relation: $$\hat{f}_m = 1.00\hat{F}^{-0.33}$$

where: $$\hat{F} = \frac{gF}{u_*^2}$$

and:

$$f_m = \frac{g}{u_*^{0.34}.F^{0.33}}$$

Table 5.1 Parametric form for the cosine model spreading power from Hasselman *et al* (1980), and Mitsuyasu *et al* (1975).

SIMULATED DOUBLE COSINE MODEL					FIRST AND SECOND MOMENT COSINE MODELS				- VON MISES	BETA DENSITY		
NUMB	TRUE POWER	NOISE POWER	NOISE ZETA	NOISE 1-m	EST 1 THETA	EST 2 THETA	EST 1 S1	EST 2 S2	VONM K	R	T	MEAN
1	14.27	1.66	43.41	0.19	10.98	2.24	3.59	5.4	2.67	2.09	12.43	10.98
2	8.63	3.35	13.16	0.13	1.94	1.23	6.82	7.08	4.32	14.15	34.96	1.94
3	19.28	3.61	27.93	0.18	6.99	3.94	7.44	8.64	4.63	3.77	22.61	6.99
4	4.25	3.18	16.95	0.2	2.97	2.42	3.89	3.93	2.83	9.96	20.94	2.97
5	5.89	3.45	13.33	0.14	1.88	1.39	5.24	5.32	3.52	12.72	27.61	1.88
6	13.03	1.54	25.01	0.15	5.22	1.09	4.6	6.17	3.19	4.28	19.73	5.22
7	14.29	2.54	12.27	0.2	3.14	1.51	6.34	7.3	4.07	9.17	30.61	3.14
8	7.51	3.44	18.72	0.04	0.93	0.59	6.94	7.05	4.38	16.29	35.96	0.93
9	3.81	1.23	5.33	0.18	0.89	0.16	2.81	3.05	2.26	7.9	16.43	0.89
10	2.42	3.4	17.13	0.17	2.37	3.04	2.48	2.49	2.08	7.13	13.95	2.37
11	7.13	1.4	38.2	0.2	7.99	1.33	3.19	4.25	2.46	3.99	15.46	7.99
12	3.11	1.66	8.16	0.05	0.45	0.2	2.96	2.99	2.34	8.1	16.4	0.45
13	19.12	0.51	5.25	0.1	0.76	359.79	3.01	5.89	2.37	11.05	26.7	0.76
14	7.66	1.33	18.41	0.08	1.89	0.29	4.98	5.78	3.39	9.79	27.06	1.89
15	17.01	0.32	33.12	0.16	5.6	358.01	1.58	4.29	1.56	3.24	12.83	5.6
16	6.04	1.41	30.58	0.02	0.65	0.11	5.51	5.74	3.66	13.25	29.44	0.65

Table 5.2. Comparison of cosine, von Mises and beta parametric estimates for 16 simulated directional distributions

no.		year	month	date	time	Hs	To	no.		year	month	date	time	Hs	To
1	.	78	dec	11	18	8.4	11.5	51	.	80	dec	15	12	7.7	11.8
2	.		dec	11	21	8.3	12.1	52	.		dec	18	9	8.0	12.0
3	x		dec	12	12	8.3	10.5	53	.		dec	18	12	8.0	12.0
4	x		dec	12	15	9.7	11.4	54	.		dec	18	15	7.8	12.6
5	x		dec	12	21	12.7	11.3	55	x		dec	20	9	9.0	11.9
6	x		dec	13	3	10.8	12.2	56	.		dec	20	12	8.1	11.6
7	x		dec	13	6	10.5	12.0	57	.		dec	20	15	9.3	12.1
8	x		dec	13	9	9.0	12.0	58	x		dec	20	18	8.0	11.2
9	x.		dec	13	12	9.6	11.4	59	.	81	jan	17	9	7.8	10.9
10	.		dec	13	15	9.4	11.4	60	.		mar	1	9	7.8	11.4
11	.		dec	13	18	9.2	11.2	61	.		mar	28	0	10.5	13.0
12	.		dec	13	21	9.2	12.1	62	.		oct	27	9	7.6	14.0
13	x		dec	14	0	7.8	11.1	63	.		dec	14	21	7.6	11.9
14	.	79	feb	13	6	7.7	14.0	64	.			26	12	8.2	11.1
15	.		feb	13	9	7.5	13.7								
16	.		feb	13	18	7.5	10.9	65	.	78	sep	29	12	6.1	11.4
17	.		mar	4	6	8.0	14.6	66	.		sep	29	15	6.0	10.5
18	.	79	dec	10	21	8.6	10.9	67	.		dec	2	3	6.2	10.4
19	.		dec	10	0	8.1	11.2	68	.		dec	3	10	6.6	12.0
20	.		dec	13	0	8.0	12.9	69	.		dec	3	15	7.4	11.4
21	.		dec	13	3	7.7	11.8	70	.		dec	6	6	6.1	9.3
22	x		dec	14	21	9.9	12.1	71	.		dec	6	9	6.0	9.6
23	x		dec	15	0	11.3	12.4	72	.		dec	7	0	6.8	9.6
24	x		dec	15	3	11.7	12.8	73	.		dec	7	3	6.0	9.1
25	x		dec	15	6	10.1	12.2	74	.		dec	7	6	6.6	9.4
26	x		dec	15	9	10.1	12.3	75	.		dec	7	12	6.1	9.7
27	x		dec	15	12	10.0	12.1	76	x		dec	7	18	6.3	10.3
28	x		dec	15	15	8.0	11.4	77	x		dec	7	21	6.6	11.4
29	x		dec	19	0	7.5	10.9	78	.		dec	8	0	6.1	10.6
30	x		dec	27	6	7.5	10.2	79	.		dec	8	3	6.2	10.3
31	.	80	jan	21	6	8.6	10.8	80	.		dec	8	6	6.2	10.3
32	.		jan	21	9	10.1	11.6	81	.		dec	8	18	6.2	11.0
33	.		jan	21	12	9.2	11.4	82	x		dec	9	0	6.6	10.6
34	.		jan	21	15	8.2	11.5	83	.		dec	9	3	6.2	10.6
35	.		jan	21	18	7.7	11.0	84	.		dec	9	6	6.3	11.0
36	.		jan	31	9	8.6	11.2	85	.		dec	9	9	6.4	10.7
37	.		feb	5	3	10.0	13.3	86	.		dec	9	12	6.2	10.9
38	.		feb	5	9	8.6	11.9	87	.		dec	10	9	6.3	9.9
39	.		feb	5	12	7.8	11.7	88	.		dec	11	9	6.5	10.1
40	x		mar	7	3	9.2	11.5	89	.		dec	11	12	6.5	10.4
41	x		mar	7	6	10.9	12.9	90	.		dec	11	15	7.1	11.4
42	x		mar	7	9	11.5	11.8	91	.		dec	12	0	6.5	10.6
43	x		mar	7	12	12.2	13.6	92	.		dec	12	3	6.1	9.5
44	x		mar	7	15	11.3	13.1	93	.		dec	12	6	6.9	10.1
45	x		mar	7	18	11.0	12.8	94	.		dec	12	9	6.5	9.7
46	x		mar	7	21	8.9	11.6	95	.		dec	14	3	7.0	11.0
47	x		mar	8	0	8.6	11.2	96	.		dec	14	6	6.7	10.9
48	.		mar	28	15	8.5	10.7	97	.		dec	14	9	6.3	11.0
49	x		mar	28	18	7.7	10.6	98	.		dec	14	12	6.8	10.9
50	.		mar	28	21	7.9	11.1	99	.		dec	14	15	7.1	10.0

Table 5.3 Date and time series for seastates with Hs>6.0m.

Fourier Stieltjes term	currents alpha	beta	winds alpha	beta
1	-0.05138	0.04149	0.019	-0.25684
2	0.08418	0.17532	0.03298	0.0031
3	-0.04266	-0.03284	-0.06006	-0.04333
4	0.03241	0.00804	-0.00542	0.02079
5	-0.00844	-0.00605	0.00175	0.03763
6	-0.00081	0.02436	-0.02273	0.00191
7	-0.00907	0.00315	-0.03654	-0.00586
8	-0.00388	-0.02905	-0.02288	0.00329
9	-0.00709	-0.00669	-0.03387	0.02291
10	0.0061	0.00768	-0.01781	0.01174

Table 5.4 Real and imaginary components of the characteristic function for the DB1 currents and waves.

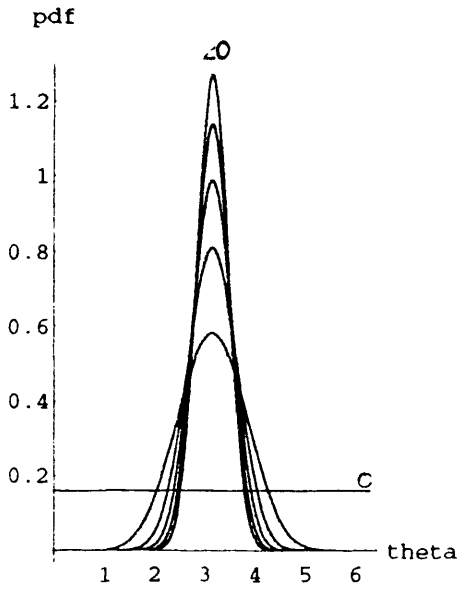


Figure 5.1 Cosine spreading model, s= 0,4,8,12,20.

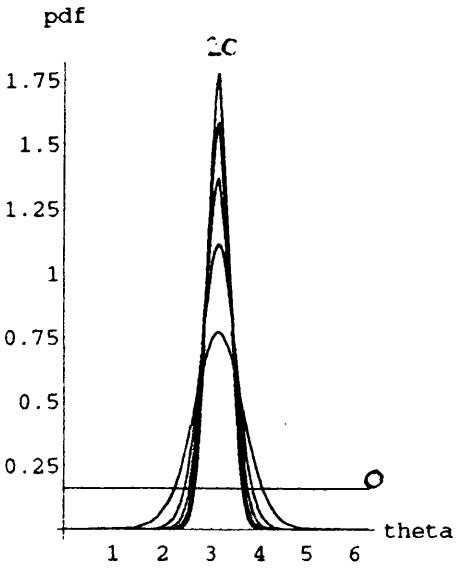


Figure 5.2 von Mises spreading model, k= 0,4,8,12,20.

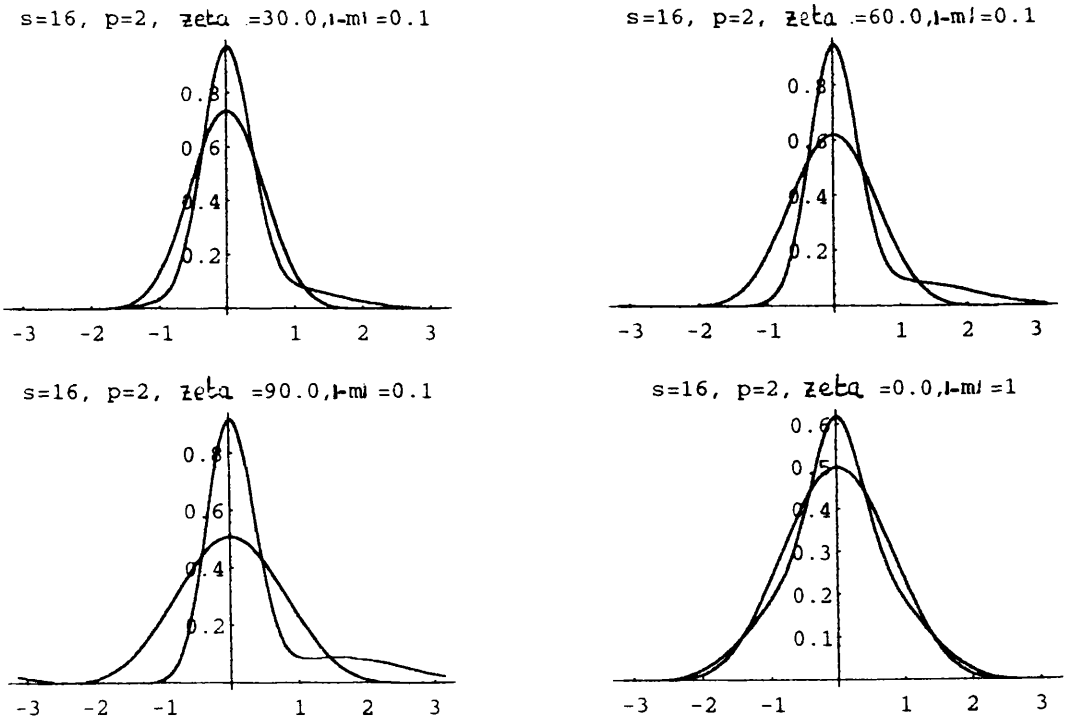


Figure 5.3 Four directional distributions simulated from a double cosine mixture and the corresponding cosine, moment estimated models.

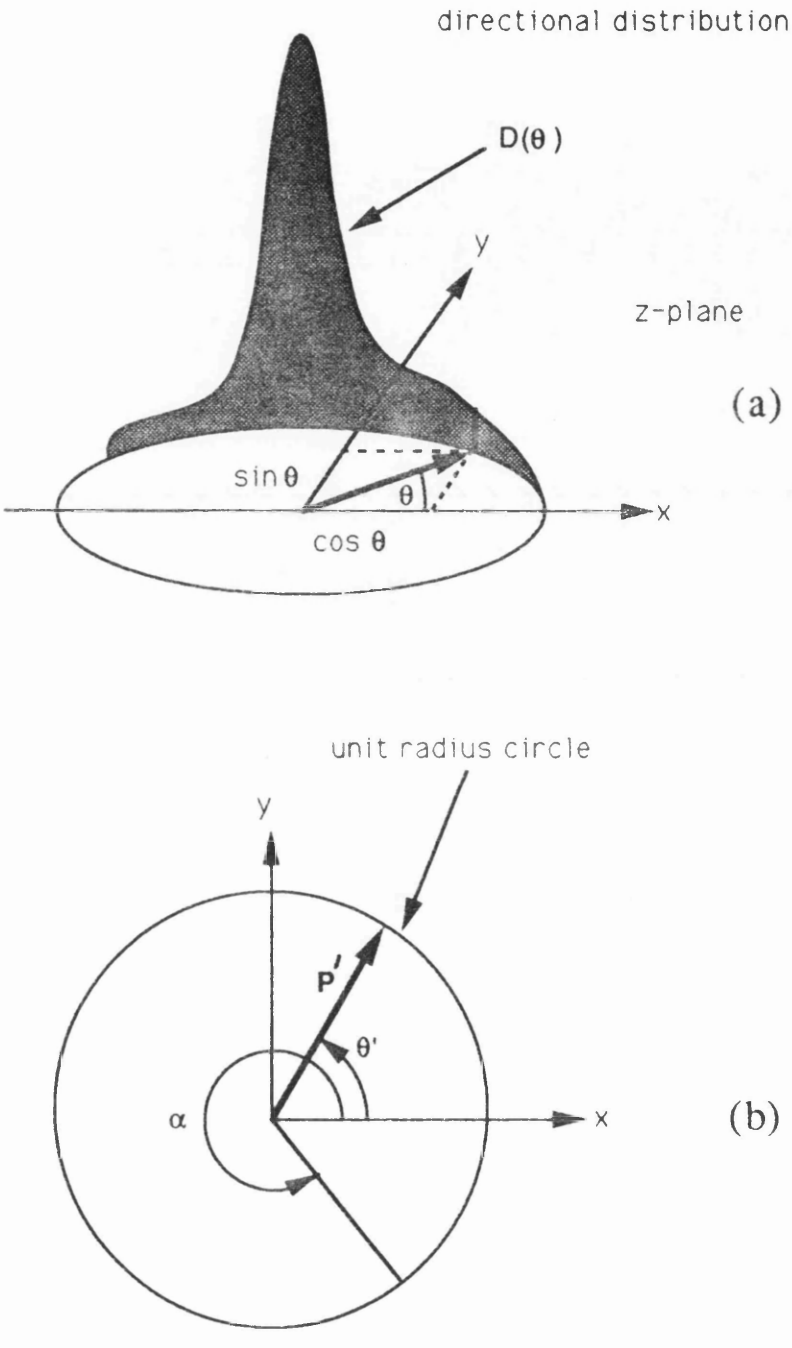


Figure 5.4 The characteristic function for angular moments.

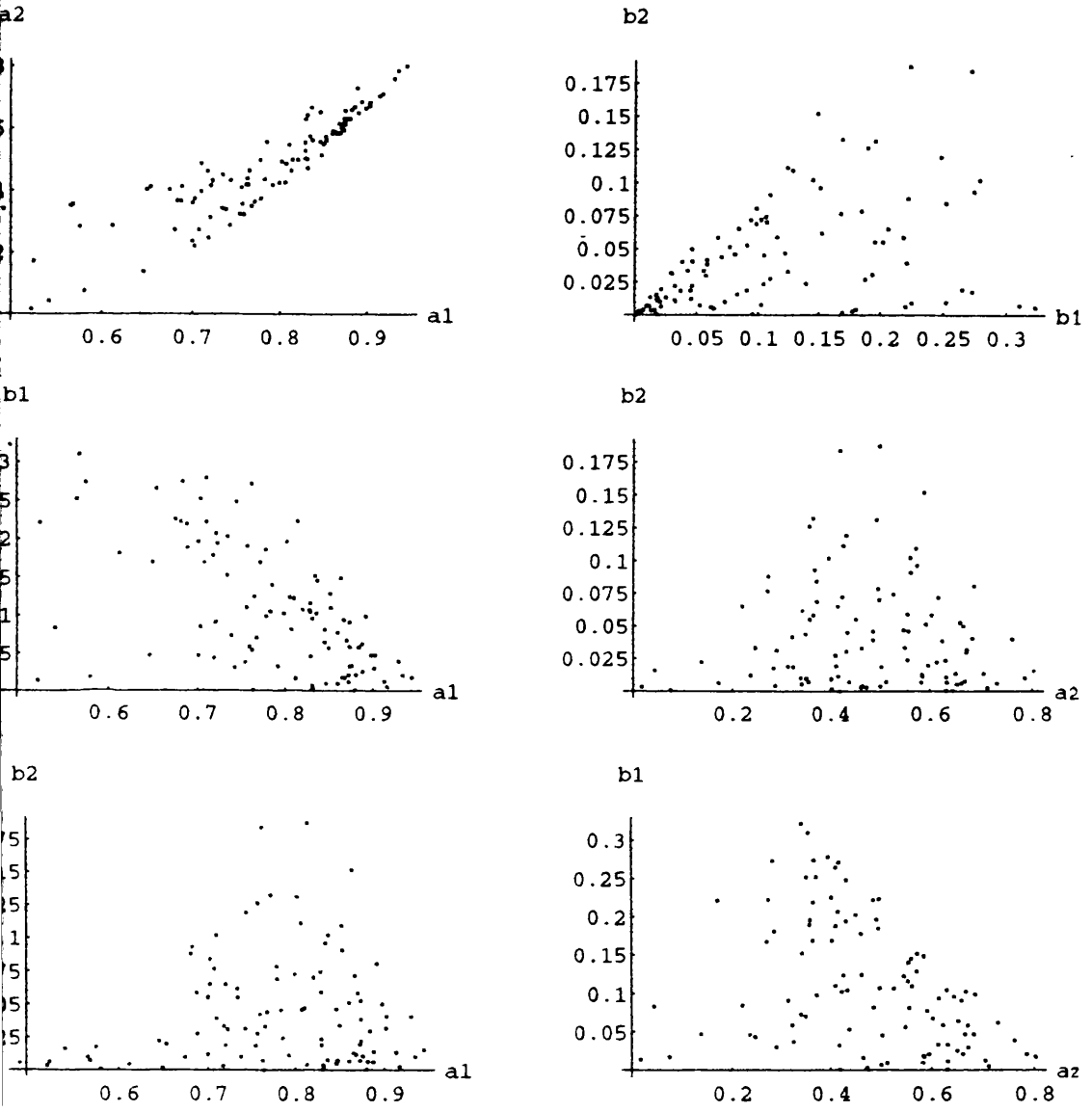


Figure 5.5 Results for 100 simulated directional distributions with non-zero skewness. alpha 1 vs alpha 2 (a); beta 1 vs beta 2 (b); alpha 1 vs beta 1 (c); alpha 2 vs beta 2 (d); alpha 1 vs beta 2 (e); alpha 2 vs beta 1 (f).

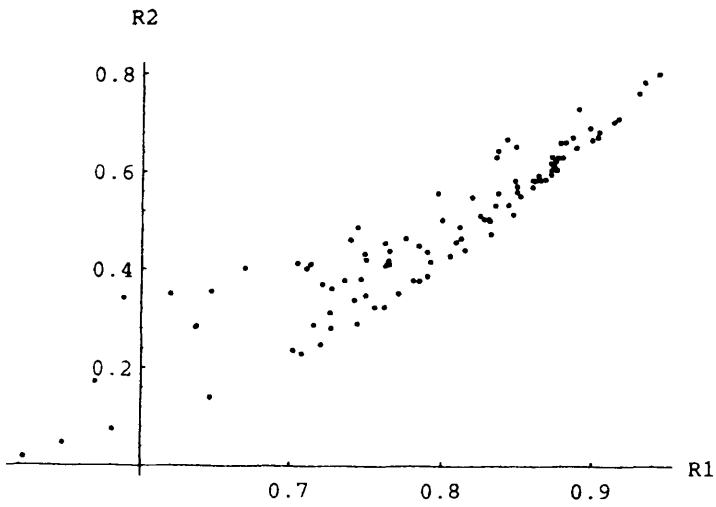


Figure 5.6 Resultants R1 vs R2 for 100 simulated directional distributions with NON-zero skewness.

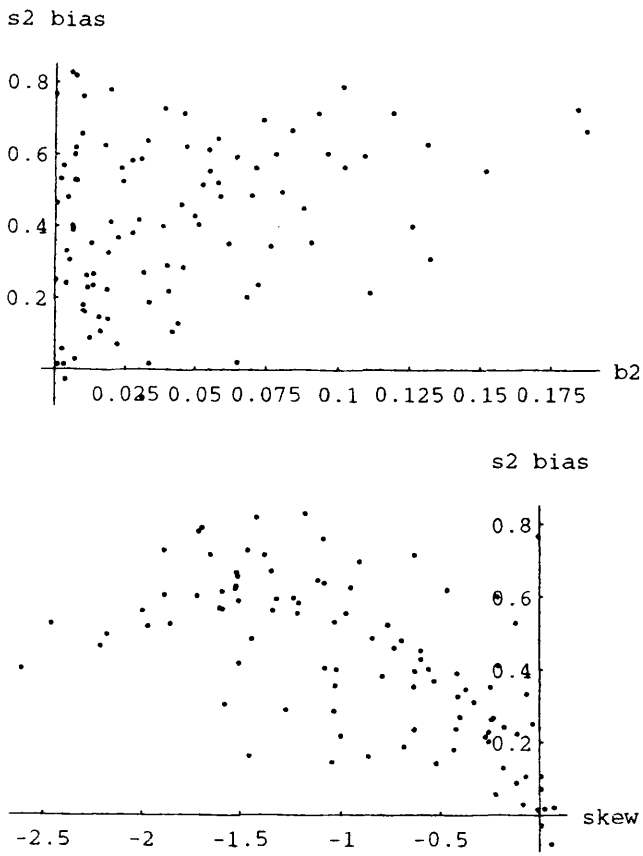


Figure 5.7 Results for 100 simulated directional distributions with non-zero skewness. Bias in cosine power vs: second angular moment (a); Mardia's measure of skewness (b).

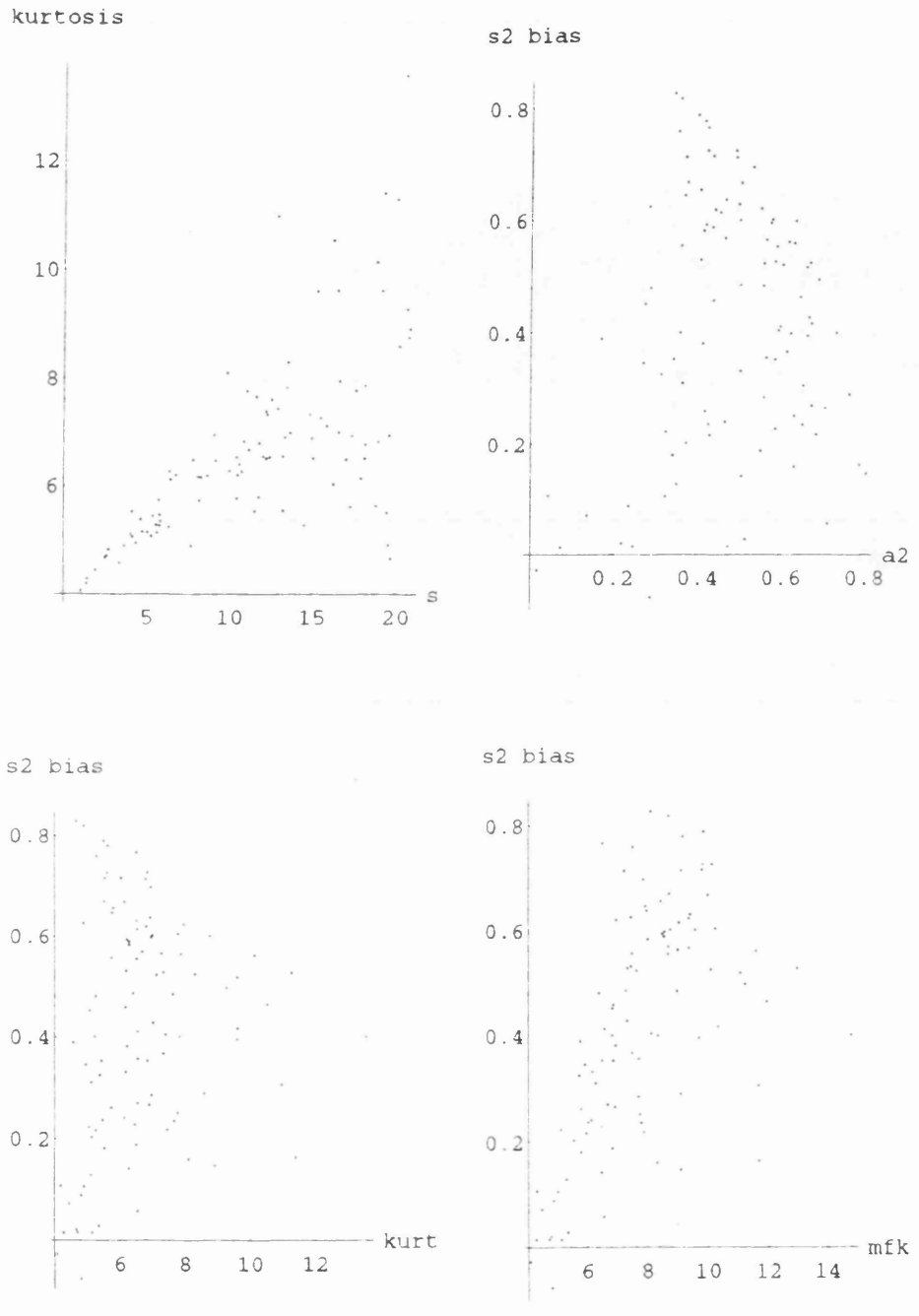


Figure 5.8 Results for 100 simulated directional distributions with non-zero skewness. Excess kurtosis vs cosine power (a). Bias in cosine power vs: second angular moment (b); excess kurtosis (c); model free kurtosis (d).

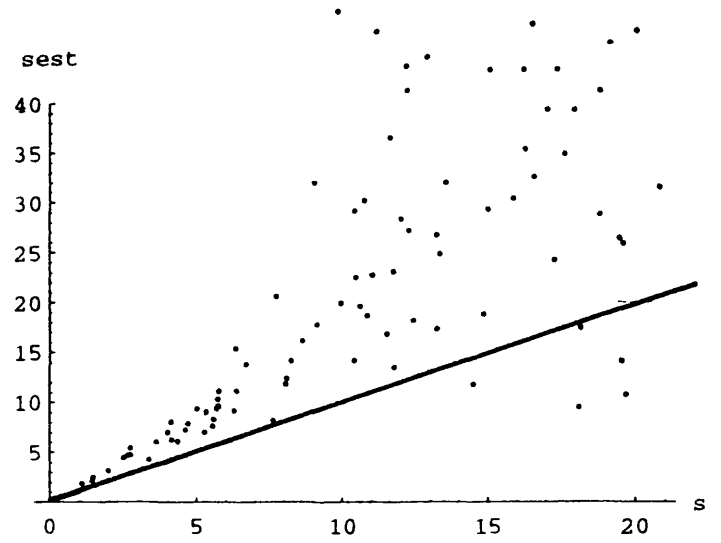


Figure 5.9 Results for 100 simulated directional distributions showing true spreading power 's' plotted against bias corrected estimate

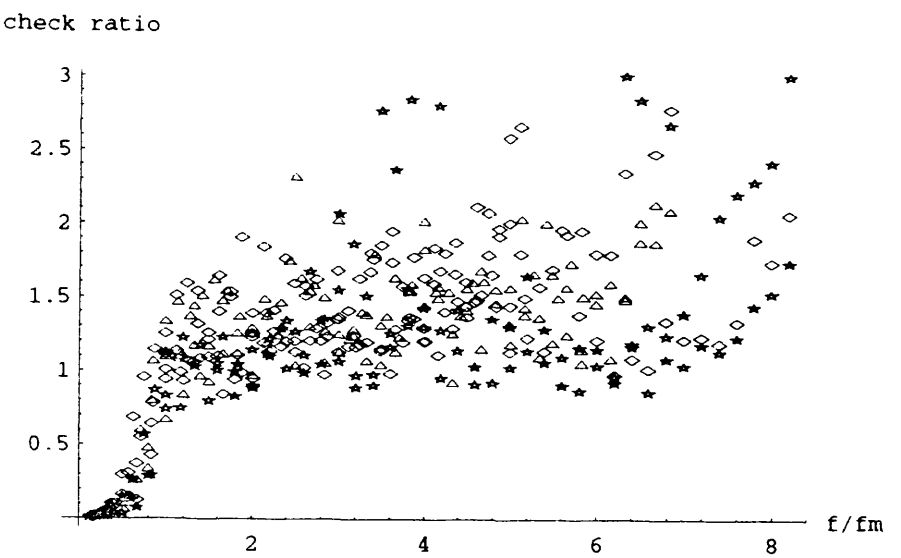


Figure 5.10 Check ratios for the largest ten seastates recorded by the DB1

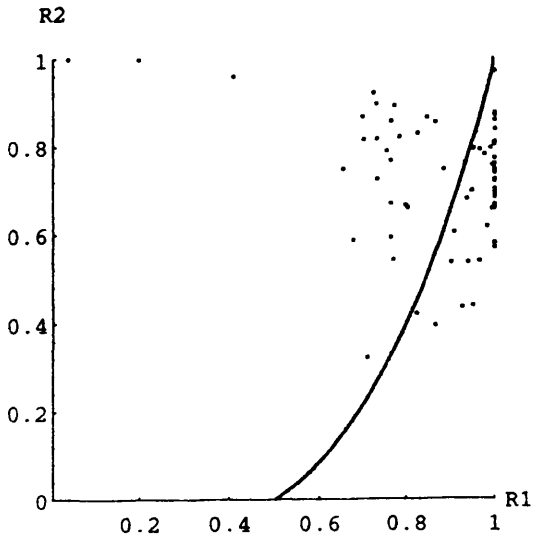


Figure 5.11 Resultant lengths for the 68 most extreme seastates recorded by the DB1 data buoy. Note values which equal 1.0 are those for which an alpha and beta moment is >1.0.

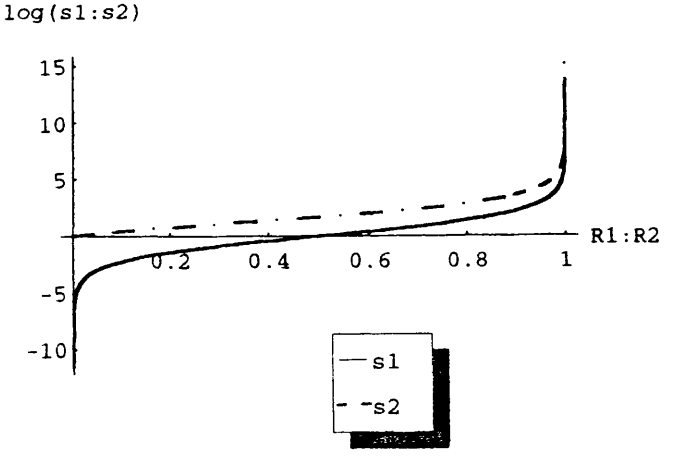


Figure 5.12 Variation of natural log of spreading power with resultants R1 and R2.

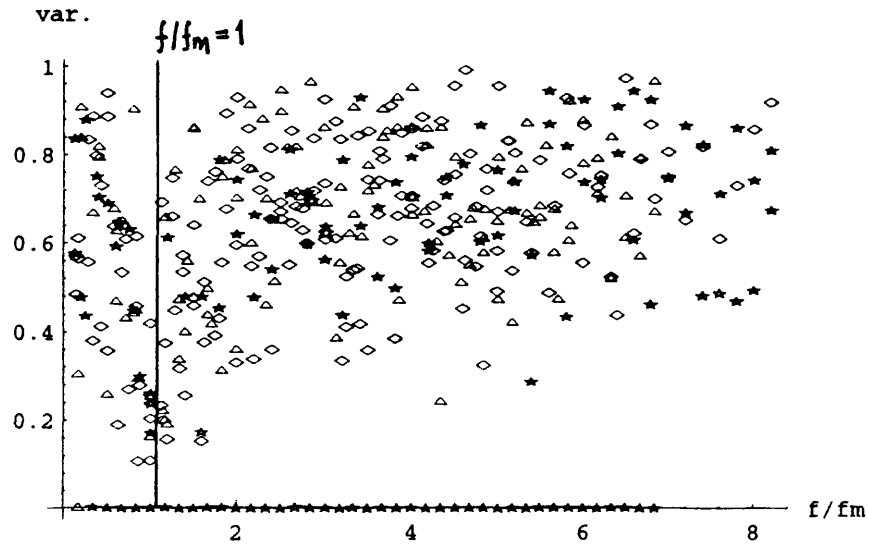


Figure 5.13 Angular variance vs frequency ratio f/fm for the ten largest storms recorded by the DB1. Note the check ratios are not screened for this plot.

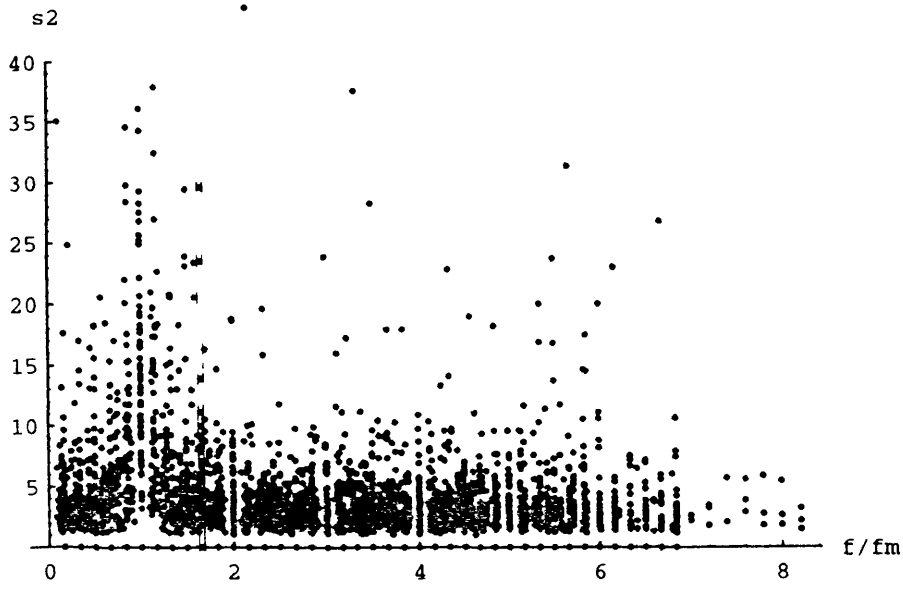


Figure 5.14 Second moment estimate of spreading power s2 vs frequency ratio for the most extreme seas.

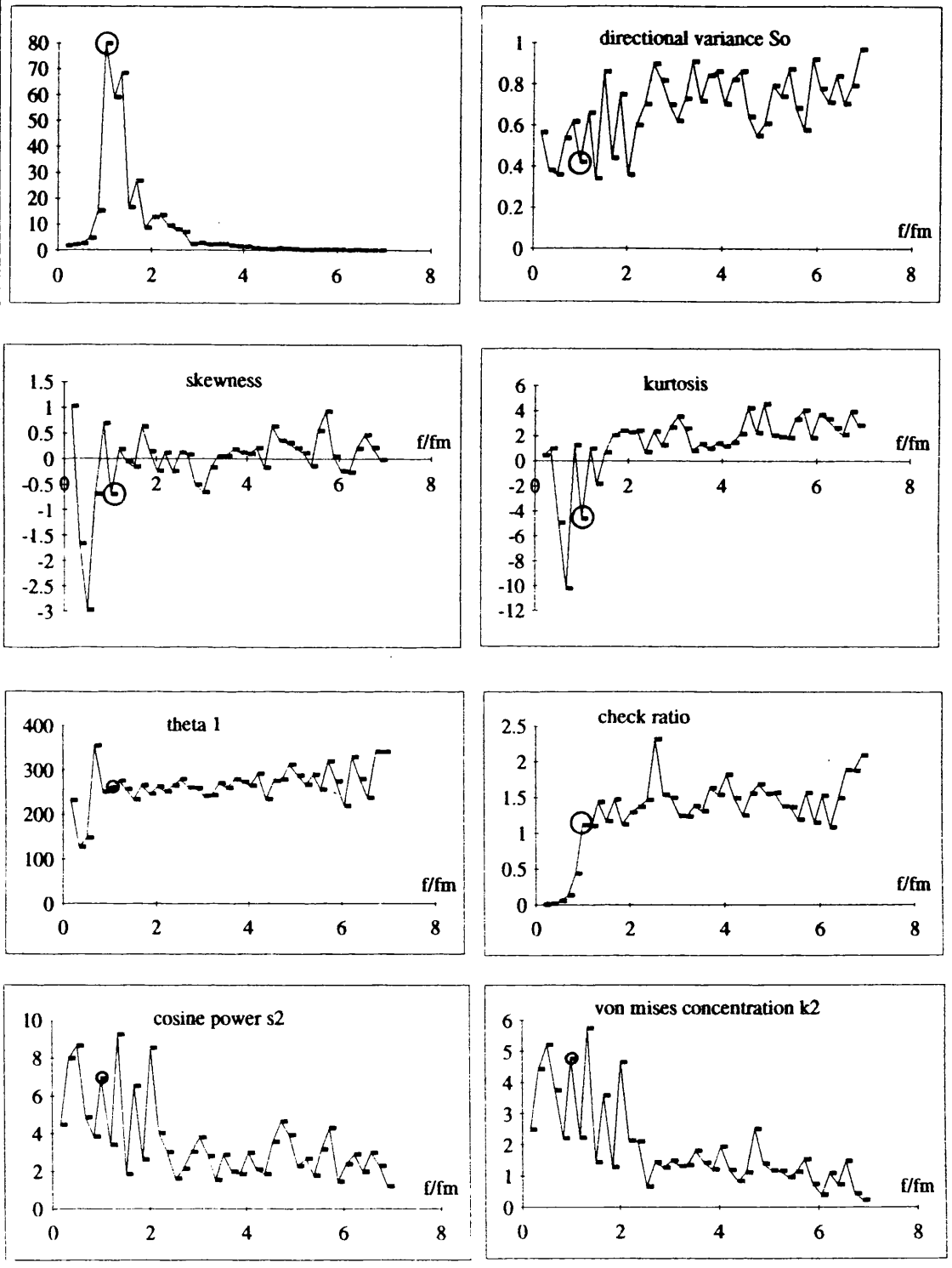


Figure 5.15 Directional spectrum parameters measured by the DB1 buoy at 18.00hrs, 13 Feb 1979. Significant wave height 7.8m and zero crossing period 10.1 sec.

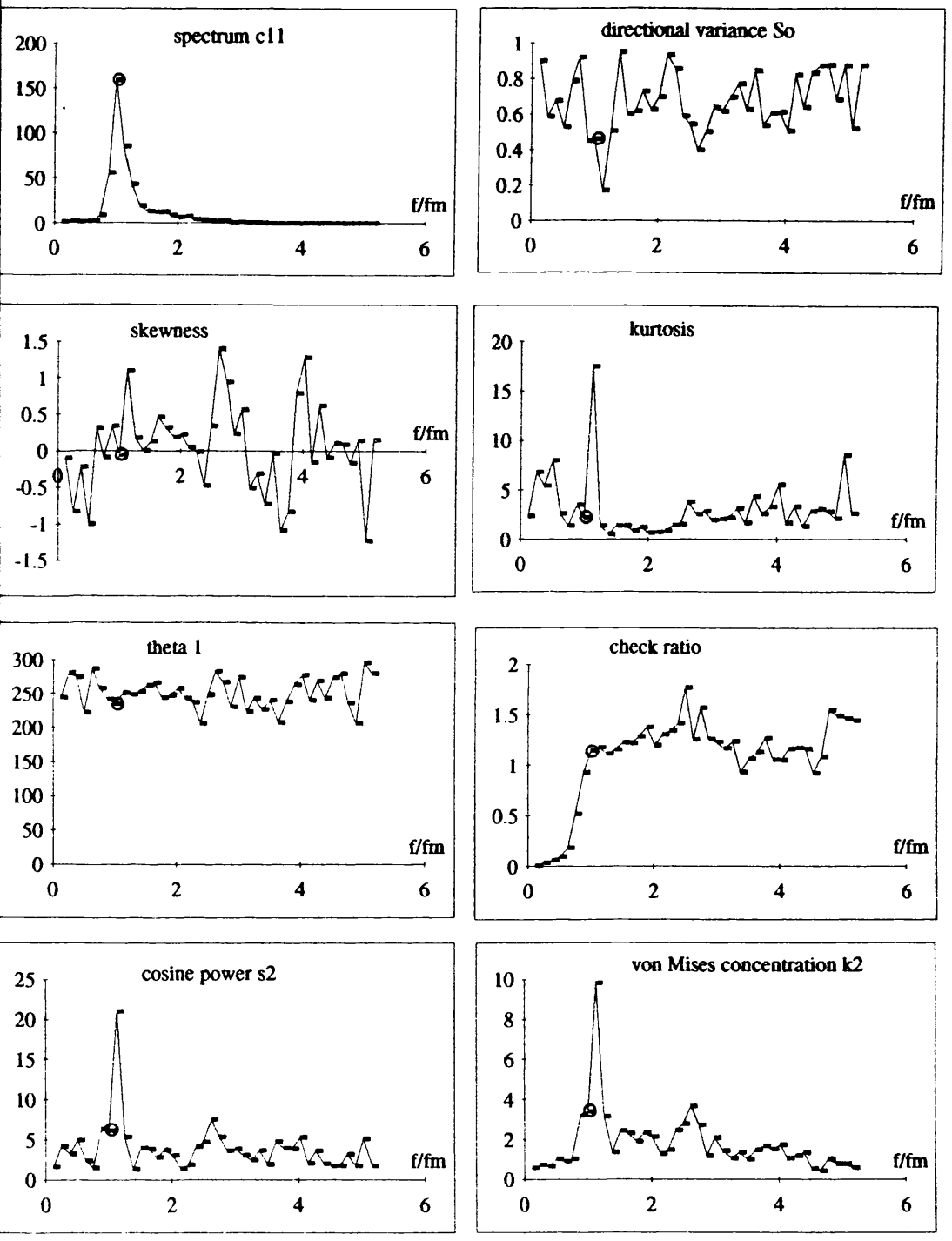


Figure 5.16 Directional spectrum parameters measured by the DB1 buoy at 15.00hrs, 28 march 1980. Significant wave height 8.5m and zero crossing period 11.2 sec.

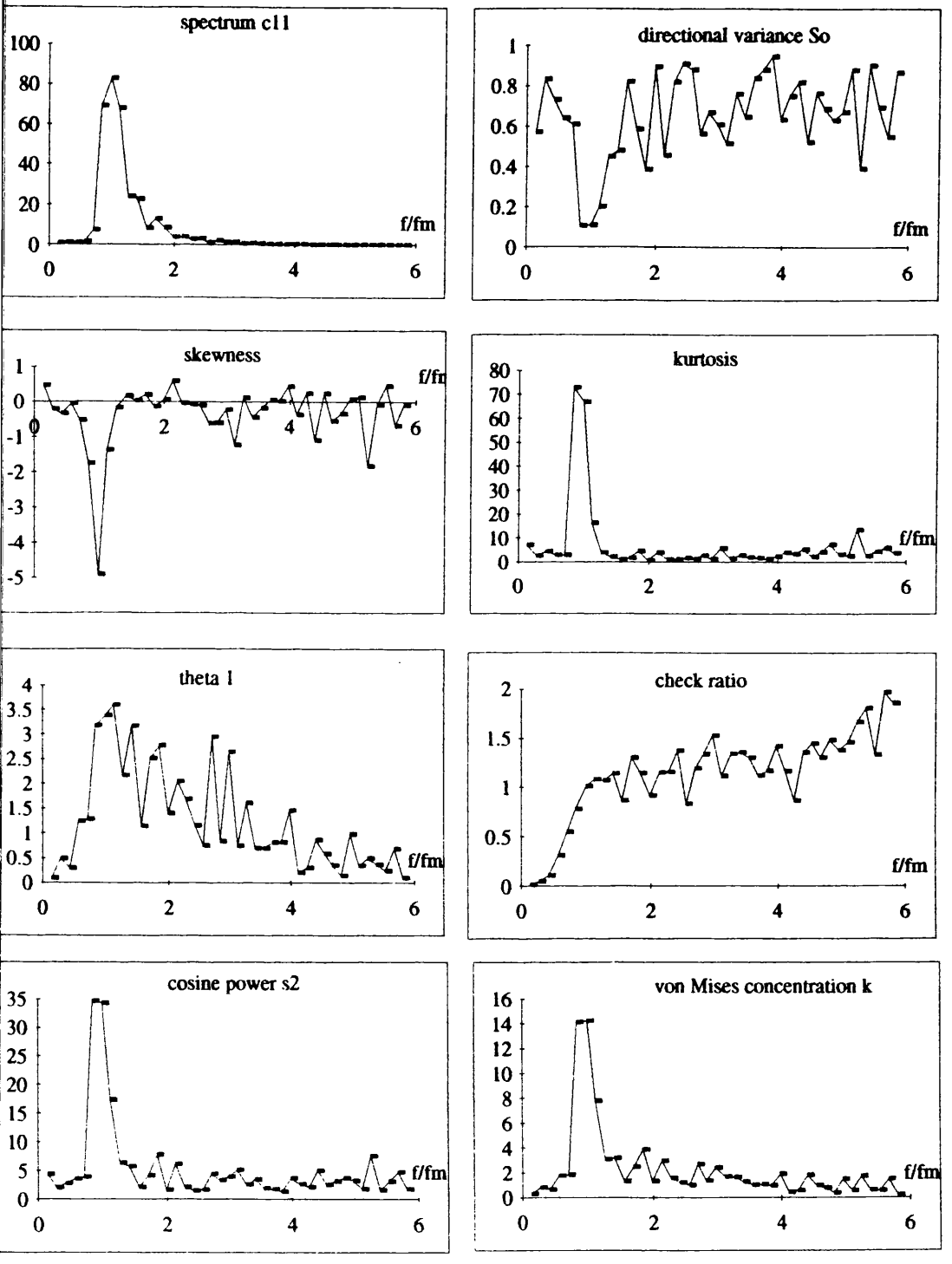


Figure 5.17 Directional spectrum parameters measured by the DB1 at 0000hrs, 10 dec. 1979. Significant wave height 8.1m and zero crossing period 11.2 sec.

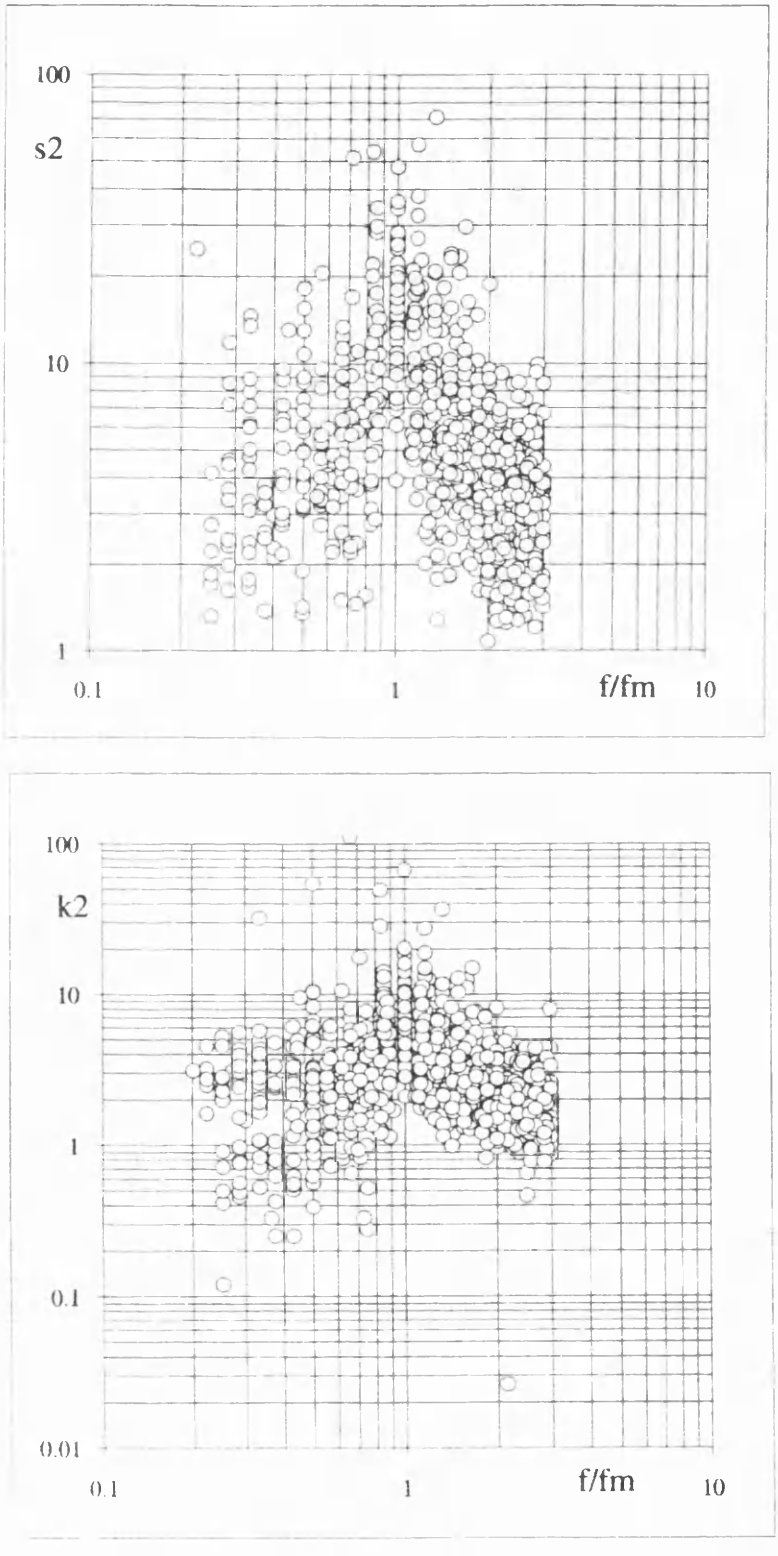


Figure 5.18 Second moment spread parameters s2 and k2 for the extreme seas recorded by the DB1. (compare with Fig. 3 of Hasselmann et al, 1980)

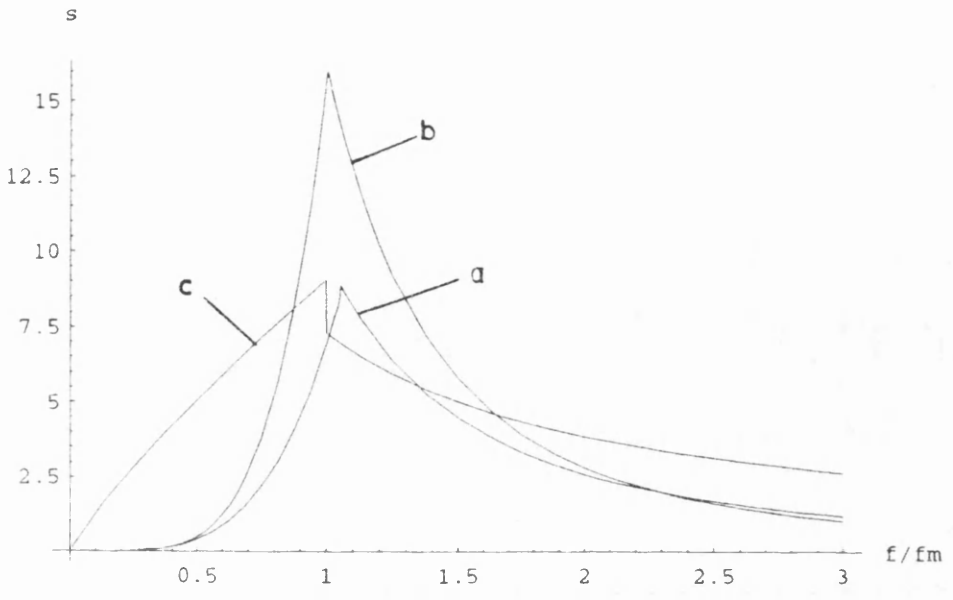


Figure 5.19 Cosine model spreading powers: (a) Hasselmann et al (1980); (b) Mitsuyasu et al (1975); (c) DB1 - this work.

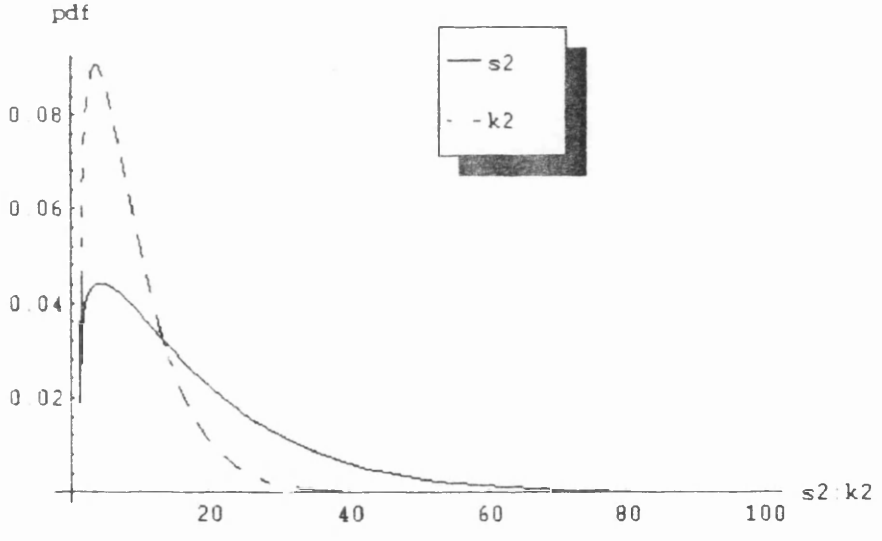


Figure 5.20 Density functions for the cosine and von Mises spread parameters at the spectral peak f = fm. Cosine model parameters are: shape = 1.15, scale = 17.46, and location 1.263. von Mises model parameters are: shape = 1.25, scale = 8.12, and location 1.28.

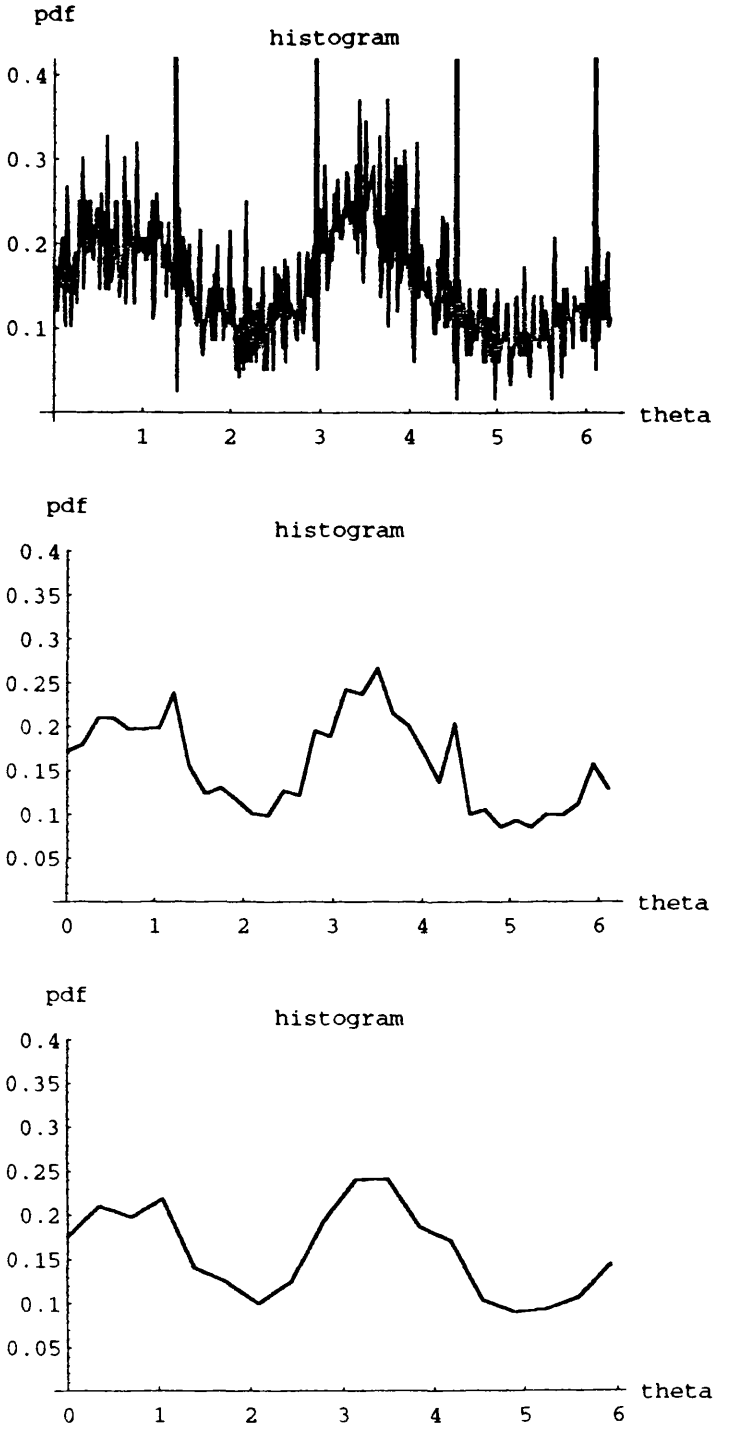


Figure 5.21 Three histogram estimates (360:36:18 bins) of the directional distribution of currents measured by the DB1 in its four years operating in the South Western approaches.

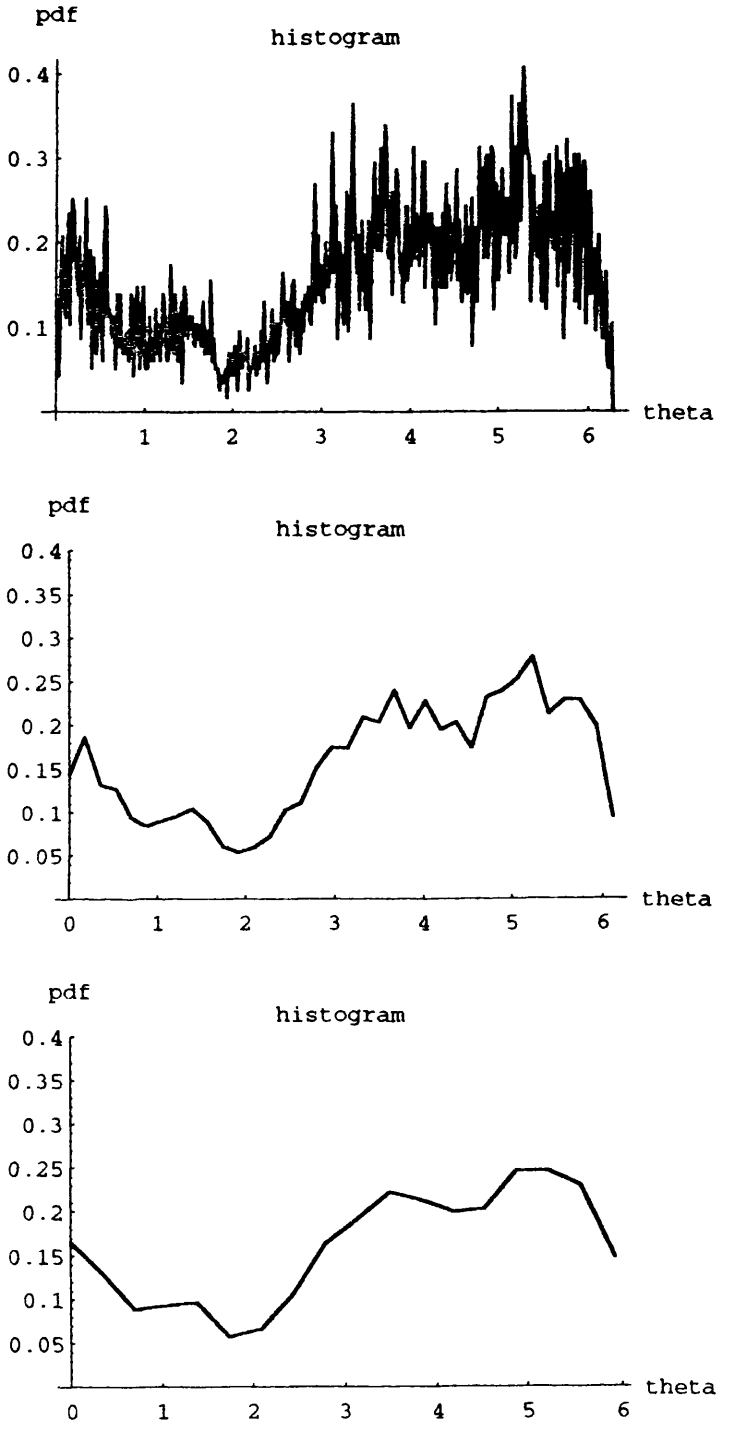


Figure 5.22 Three histogram estimates (360:36:18 bins) of the directional distribution of winds measured by the DB1 in its four years operating in the South Western approaches.

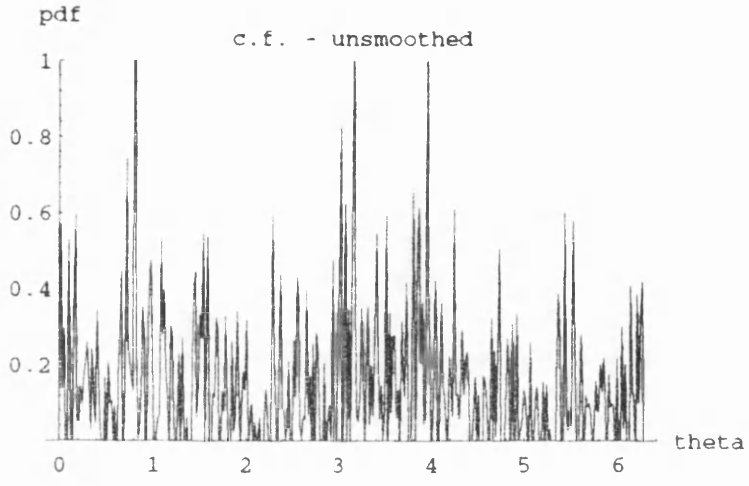


Figure 5.23 Raw characteristic function estimate of the directional density for currents recorded by the DB1.

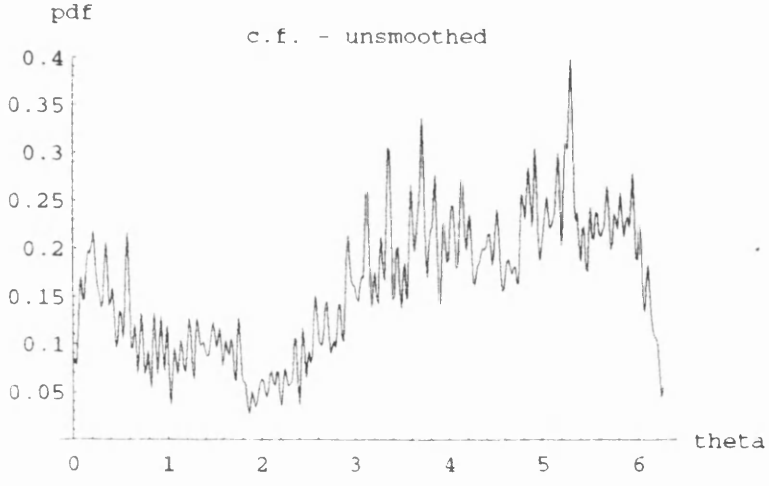


Figure 5.24 Raw characteristic function estimate of the directional density for winds recorded by the DB1.

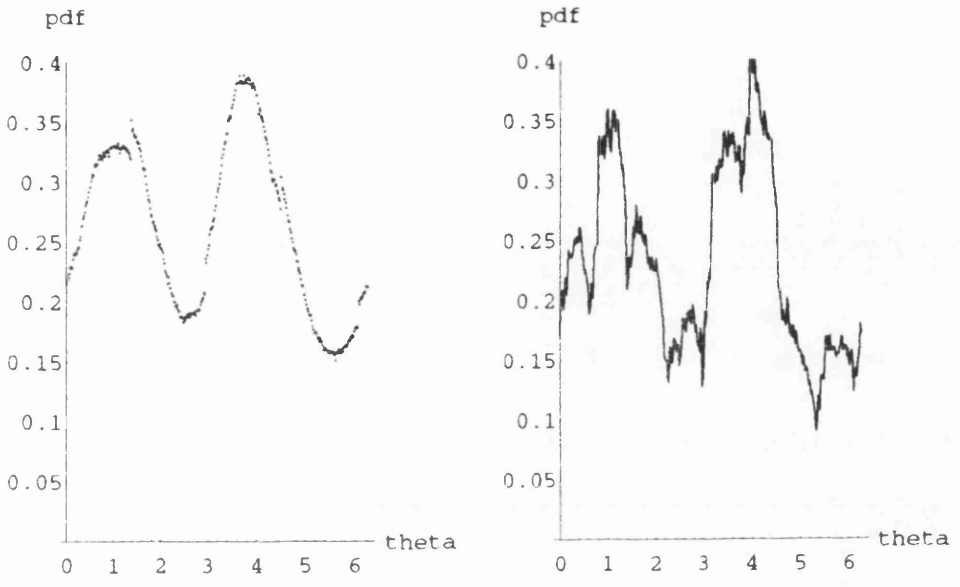


Figure 5.25 Smoothed density estimates of current direction recorded by the DB1: (left) histogram estimate with 360 class intervals; (right) smoothed characteristic function estimate evaluated at 360 angles. In both cases an exponential kernel was used.

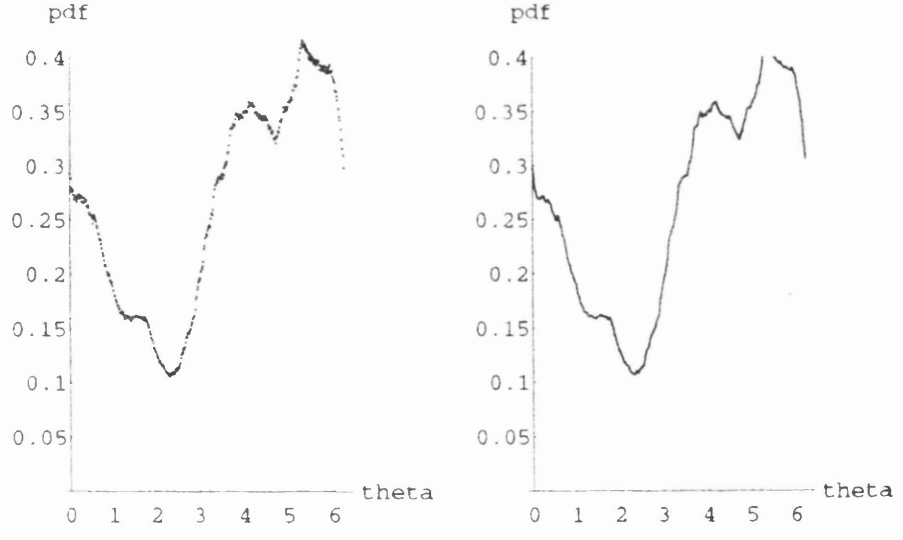


Figure 5.26 Smoothed density estimates of wind direction recorded by the DB1: (left) histogram estimate with 360 class intervals; (right) smoothed characteristic function estimate evaluated at 360 angles. in both cases an exponential kernel was used.

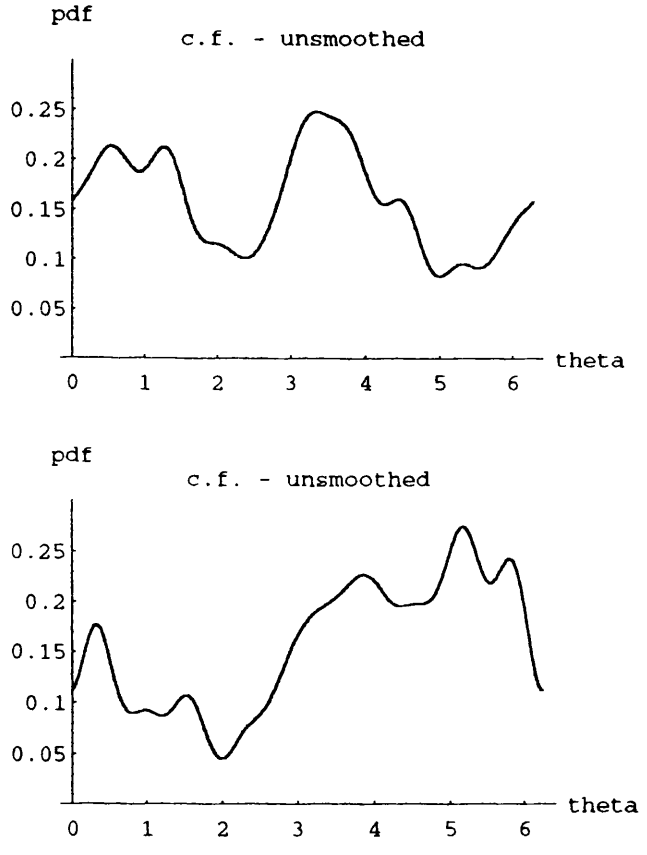


Figure 5.27 Characteristic function estimates of the directional density of currents (top) and winds (bottom) recorded by the DB1.

Chapter 6

LOADING AND RESPONSE MODEL OF A TENSION LEG PLATFORM

NOMENCLATURE

A	*exposed area for wind force calculations*
A_t	*cross – sectional area for the tendons*
B	*spacing between column centrelines*
C	*shape coefficient for drag force calculation*
C_i	*polynomial coefficients for Havelock approximation*
c_{10}	*surface drag coefficient*
C_a	*added mass coefficient*
C_d	*drag coefficient in waves*
$\overline{C}_d$	*drag coefficient in steady flow*
C_{dl}	*linearised drag force*
C_f	*spectrum modification factor*
C_m	*inertia coefficient*
C_{mk}	*vector of added inertia coefficients*
d	*damping ratio*
D	*characteristic dimension*
D	*vector of characteristic dimensions*
$D(\theta; f)$	*directional spreading function*
dl_k	*elemental lengths*
E	*Young's modulus*
f	*frequency in hertz*
f_*	*non - dimensional frequency*
f_0	*apparent frequency*
F	*vector of Morison forces*
F_c	*steady current force*
F_{DF}	*vector of diffraction forces*
F_{DR}	*vector of viscous drag forces*
F_{FK}	*vector of Froude - Krylov forces*
$\overline{F}_1$	*mean viscous drag drift force*
$\overline{F}_2$	*mean wave elevation drift force*
$\overline{F}_3$	*mean velocity head drift force*
$\overline{F}$	*total drift force*
F_r	*resultant steady force*
F_{sd}	*steady slow drift force*
$F_w(t)$	*time varying wind force*
F_w	*steady wind force*
G_i	*ith limit state function*
g	*acceleration due to gravity*
H	*wave height*
$H(f; \theta)$	*response transfer function*
$J_n(.)$	*Bessel function of the first kind order n*
k	*wave number*
k_{ij}	*element of the hydrostatic stiffness matrix*
k_c	*Keulegan - Carpenter number: $u_0 T / D$*

$\boldsymbol{K}_H$	*hydrostatic stiffness matrix*
$\boldsymbol{K}_T$	*tendon stiffness matrix*
l	*wave length*
L_T	*length of the tendons*
$\boldsymbol{M}$	*structure mass matrix*
$\boldsymbol{M}_A$	*matrix of added masses*
$\boldsymbol{M}_{FK}$	*matrix of Froude - Krylov added inertias*
$\vec{n}$	*vector of unit normals*
p	*pressure in the fluid*
$Q(w)$	*dynamic amplification factor*
$\boldsymbol{R}$	*relative displacements between the structure and the fluid*
$S_{wf}(f)$	*spectral density of the wind force*
s	*surface area*
$\boldsymbol{S}_k$	*vector of cross - sectional areas*
$S_R(f)$	*response spectral density*
$S_\eta(f)$	*surface elevation spectral density*
$S_w(f)$	*spectral density for wind speed*
T	*wave period*
T_0	*mean force in a tendon*
(u,v,w)	*wave particle velocities*
u_0	*peak horizontal velocity*
u_*	*shear velocity*
z_i	*ith degree of freedom in the structure reference axes*
U_c	*mean current speed*
U_w	*mean wind speed*
U_z	*mean wind speed at elevation z*
$v(t)$	*time varying component of wind speed*
V	*submerged volume*
X_0	*non - dimensional distance*
$Y_n(.)$	*Bessel function of the second kind order n*
$\boldsymbol{Z}$	*velocity vector for the platform*
Ψ	*wave phase angle*
δT	*change in tendon tension due to dynamic motions*
ΔT	*steady change in tendon tension*
Φ_0	*undisturbed Airy wave potential*
γ	*JONSWAP peak enhancement parameter*
η_a	*wave amplitude*
$\boldsymbol{\lambda}$	*vector of eigenvalues*
μ	*frequency rad / sec*
θ	*direction of wave propagation*
ρ	*mass density of water*
ρ_a	*mass density of air*
σ_0	*mean water level tendon stress*
τ	*JONSWAP shape parameter*
ω	*frequency in rad / sec*
ω_p	*spectral peak frequency in rad / sec*
η_a	*wave amplitude*

6. INTRODUCTORY REMARKS: LOADING AND RESPONSE MODEL OF A TENSION LEG PLATFORM

A deep water tension leg platform has been chosen to illustrate the relative sensitivity of an offshore structure's response predictions to the use of joint probabilities and directional wave spreading. This type of structure is suitable for several reasons: first, its long natural periods in surge, sway, yaw make it sensitive to wave directional spreading; second, the effects of currents and winds are important since they contribute a significant amount to the static and dynamic offsets of the platform; third, the limit states for the tendons are simple functions of the platform motions; and finally, the capital cost of these platforms is sufficiently large to reward serious analysis of the platform responses.

This chapter describes a mathematical model of the loading on a tension leg platform which has been designed for a Monte Carlo simulation in which the long-term responses are examined using probabilistic methods. The multivariate, environmental density used in the simulation is based on the work described in ch's 3,4, and 5. This data was recorded in the conventional way using time-averaged wave, wind, and current measurements recorded at regular intervals over a four year period. The multivariate probability models obtained from statistical inference on this data are therefore random vector models for the time-averaged quantities. This conforms to the requirements of a discrete, point-in-time reliability calculation in which the long-term variation of the environment is modelled as a discrete vector random process; and the within seastate behaviour is modelled using a stochastic process model for the waves and constant values for the winds and currents, Bjerager (1988). The details of the probabilistic analysis are described in more detail in Ch. 7.

Section 2 of this chapter defines the geometry of the platform together with the so called limit state functions for the tendons and riser system; which are essentially constraint equations which define the boundaries between the safe and unsafe responses of the platform as functions of the system parameters. *Section* 3 describes the model of the wind, first order wave, and currents forces. This takes into account viscous, potential, and motion induced forces on the submerged hull, tendons and risers. *Section* 4 then shows how the non-linear lateral stiffness of the tendons has been modelled for the calculation of the platform offsets and tendon tension changes. These arise in presence of steady forces due winds, currents and waves.

Section 5 shows how the dynamical behaviour has been modelled as a linear six degree of freedom system in which the coupling between modes has been included to ensure the tendon tensions are correctly modelled. This model is used to calculate the first order wave force, motion, and limitstate responses in both long-crested and short-crested seas using the methods described in *Section* 6. The spreading models used for the analysis (which are defined in Ch. 5) are compared to assess the importance of the angular variance of the wave spread.

Section 7 presents the simplified calculation method used for the mean and slowly varying second order wave drift forces on the TLP hull. The results of this simple method are then compared with those obtained from more complex analysis using 3-D source distribution method. Finally, *Section* 8 examines the force, motion, and limitstate responses of the TLP for a variety of wind, wave, and current magnitudes and directions.

For this work, only a small number of design limit states are examined since the purpose is to assess the differences between: reliability estimates assuming simultaneous occurrence of load process extremes in unidirectional seas (the design wave method), and reliability estimates using joint probability models of the load process with directionally spread seas.

6.1. TENSION LEG PLATFORM MODEL

The tension leg platform (TLP) model used in this study is based on published details of a test structure originally proposed by Tan & de Boom (1981), and later chosen to assess some of the commercially available compliant systems analysis packages by Eatock-Taylor & Jeffereys (1986). The geometric model reported by Tan & de Boom (1981) is copied in this work to enable comparison of: (i) our first order wave force results with the work of Eatock-Taylor & Jeffereys (1986); and (ii) our second order force results with the results in de Boom et al (1984). By choosing this structure we can ensure the responses are correctly calculated in the Monte Carlo analysis discussed in chapter 7.

The structural and geometric details for the TLP tendons given in Table 6.1, and illustrated in Fig. 6.1, have been designed for this project using the known vertical stiffness and information on the dimensions and material strengths given in Burns (1983), and Woo-Sun et al (1991).

Several methods have been developed for calculating the hydrodynamic loads on compliant systems. Denis & Heaf (1979) used both linear diffraction-radiation theory and non-linear time domain solutions to calculate the responses of a TLP. Tagaki et al (1985) carried out an extensive study of the responses of a semi-submersible using 34 computer programs from 28 organisations. The theories used by these programs were either Morison based or source distribution based. His results demonstrate that for surge and sway forces the computer programs, including those based on Morison's approach, all correlated well with their experimental results. However, agreement was poor in the other modes of motion. More recently Incecik et al (1987) examined the loading on a semi-submersible and an articulated tower using both 2D-source distribution and Morison theory. The results confirm that the Morison approach provides accuracy equivalent to the 2D-source method for surge and sway given carefully chosen drag and inertia coefficients. Since this work examines the effects of joint probabilities on both TLP motion and tether force system limit states during the platform lifetime a simple Morison based loading model has been chosen. The primary advantage of this method is that responses can be calculated quickly in the frequency domain, by using a simple linearisation of the drag term. Patel& Witz (1991) shows this approach is reasonable and can produce results which agree quite well with time domain solutions.

In what follows, the simplified loading and response model of the TLP is discussed. The response vector of platform displacements for the platform are then used to define the system limit states for the tethers.

6.1.1 STRUCTURE LIMIT STATES

The structure limit states are a mathematical definition of the design criteria required for safe operation of the structure. For TLP's the major design criteria are outlined in the American Petroleum Institute's API:RP2T (1987) regulations. In a detailed design several limit states must all be examined to ensure that operational, environmental, stability and strength design criteria are all adhered to. Only the most simple limit states which do not require frame or finite element analysis are considered in this work. These include limit states associated with the platform motions like: lateral offsets; platform setdown; and the air gap between the underside of the deck and the most extreme waves. The most simple structural limit states considered are those for the tendons: for example, the ultimate tensile strength, and the requirement that the tendons do not become slack.

Using the platform reference system shown in Fig. 6.1 the limit states chosen for the tendon tensions are similar to those given by Woo-Sun et al (1991), with the addition of the effects of static and dynamic offset forces which change the tendon tensions. Two simple limit state functions are considered for each tendon group at the corners of the tension leg platform. The first examines the margin between the pretension in the tethers and the compression which arises due to heaving, pitching and rolling. The second examines the margin between the ultimate tensile capacity of the tendon groups and the maximum tensile forces induced by heaving, pitching and rolling. These two limit states can be expressed as eight equations

tendon 1

$$G_1 = T_0 - \{WT - k_t[z_3 + Bz_4 + Bz_5] - \delta T(z_1, z_2, z_6)\}$$
$$G_2 = A_t\sigma_y - \{A_t\sigma_0 + k_t[z_3 + Bz_4 + Bz_5] + \delta T(z_1, z_2, z_6)\}$$

tendon 2

$$G_3 = T_0 - \{WT - k_t[z_3 - Bz_4 + Bz_5] - \delta T(z_1, z_2, z_6)\}$$
$$G_4 = A_t\sigma_y - \{A_t\sigma_0 + k_t[z_3 - Bz_4 + Bz_5] + \delta T(z_1, z_2, z_6)\}$$

tendon 3

$$G_5 = T_0 - \{WT - k_t[z_3 - Bz_4 - Bz_5] - \delta T(z_1, z_2, z_6)\}$$
$$G_6 = A_t\sigma_y - \{A_t\sigma_0 + k_t[z_3 - Bz_4 - Bz_5] + \delta T(z_1, z_2, z_6)\}$$

tendon 4

$$G_7 = T_0 - \{WT - k_t[z_3 + Bz_4 - Bz_5] - \delta T(z_1, z_2, z_6)\}$$
$$G_8 = A_t\sigma_y - \{A_t\sigma_0 + k_t[z_3 + Bz_4 - Bz_5] + \delta T(z_1, z_2, z_6)\} \qquad [6.1]$$

Where: A_t is the cross sectional area of tendons at each corner; T_0 is the initial tendon tension in stil water; WT is the effective tendon weight; σ_y is the yield strength of the tendon material; σ_0 is the initial tendon stress; B is half the distance between column centrelines; and k_t is the axial stiffness, and z_i are the coordinates of the tendon group centre.

These eight limit states correspond to: exceedance of the ultimate strength in tension, and the requirement that no tendon group goes into compression at the seabed level. This gives two limit states per corner which must be evaluated for each stationary seastate.

6.2. FORCE MODELLING

The three primary sources of environmental loads on a TLP are wind, wave and current induced. In Ch. 7 the wind and current forces have both been modelled as random variables which are constant throughout the duration of a seastate: though in fact both will fluctuate randomly and could be modelled as stochastic processes. The combination of the wind, and current forces with the wave first and second order forces is then simplified.

The wave forces are more complex with low wave and high frequency components. Only the low *difference-frequency* and *wave frequency* components have been modelled in this work since they are the primary source of loading on the submerged parts of the platform. Although the low frequency lateral forces are small when compared to the first order wave force they can induce large meandering surge and sway oscillations because of the low damping at the surge, sway and yaw natural frequencies, see Faltinsen & Demirbilek (1989). High frequency wave loads arise from second order *sum-frequency* forces and have been identified as the cause of tendon leg 'ringing' in which high frequency axial vibrations are set-up at the natural frequency of the tendons.

The slowly varying drift motions are important because they set-up large axial forces in the tendons upon which the first order wave frequency forces are superposed. The combination of the first and second order force components is simplified in this work by using Turkstra's rule in which the square root of the sum of the squares is taken for the combined load process. This is effectively an assumption of independent gaussian loads which is not true in practice because the largest wave will occur when the wave envelope is largest. This suggests the first and second order maxima will be strongly correlated, albeit with a difference in phase. This problem has been examined by Naess (1989) and needs further research.

The linear first order wave loading on a TLP is given by the linear superposition of: the Froude-Krylov pressure forces; the acceleration (or diffraction) forces; and the linearised viscous drag forces. Each force contribution has been modelled for this TLP study and a brief description of the way each was calculated follows.

6.2.1 WIND FORCES

The conventional method for calculating wind forces on an offshore structure uses the approximation

$$F_w(t) = \tfrac{1}{2}\rho_a C A\left(U_w + v(t)\right)\left|\left(U_w + v(t)\right)\right| \qquad [6.2]$$

Where: ρ_a is the mass density of air; C is the shape coefficient (DnV Table B.5); A is the projected area normal to the wind direction; U_w is the mean wind speed averaged over one hour; and $v(t)$ is a time varying wind speed about the mean U_w. The wind force expression can be splitplit into a steady force $\overline{F}_w$ and a gust force $F^*_w(t)$

$$F_w(t) = \overline{F}_w + F^*_w(t)$$

Where the first term is given by

$$\overline{F}_w = \tfrac{1}{2}\rho_a C A U_w^2$$

and the second term is approximated by a linear term plus a quadratic "error" term which is assumed to be small

$$F^*_w(t) = \tfrac{1}{2}\rho_a C A U_w v(t) + E\left(v(t)^2\right) \qquad [6.3]$$

The error in the gust force approximation is small if the fluctuating component is small compared to the mean wind speed. The primary purpose of this approximation is to simplify the calculation of the wind force using frequency domain methods, Barltrop & Adams (1991), however, the results discussed below suggest the approximation in extreme winds needs to be validated using time domain analysis since the significant amplitude (based on spectrum m_0) of the gusting component is generally some 20-30% of the mean wind speed. The exposed area of the total structure for this work has been estimated using the Hutton TLP as a typical structural configuration giving the mean forces shown in Table 6.2.

The wind forces and platform responses can be modelled by a linear system. since the wind gust process has a near-normal probability distribution, ESDU (1974). The most appropriate spectral form for gust speeds over the open ocean is that suggested by Eidsvik (1985), and later adopted by Ochi (1988). Ochi *loc cit* defines the non-dimensional wind spectrum as

$$S_w(f) = \frac{u_*^2 S_w(f_*)}{f} \qquad [6.4]$$

$$f_* = zf / U_z$$

$$u_* = U_{10}\sqrt{c_{10}}$$

$$c_{10} = 6.7e^{-5}U_{10} + 6.2e^{-4}$$

$$U_z = U_{10} + 2.5u_* ln(z/10)$$

Here: $S_w(f)$ is the gust velocity spectrum; u_* is the shear velocity; f_* is the non-dimensional frequency; f is the frequency in Hz; c_{10} is the surface drag coefficient; and U_z is the elevation dependent wind speed. Note the wind speeds recorded by the DB1 were measured by cup anemometers located at 6.0m and 8.7m above sea level; this is effectively close enough to the international reference height of 10.0m used in this and the JONSWAP spectrum.

The spectrum of the wind gust force is given by

$$S_{wf}(\omega) = \frac{S_w(f)}{2\pi}\left\{\frac{F^*_w}{v}\right\}^2$$

where the 2π is included to rescale the spectrum from hertz to radians. The lateral displacement response of the platform can then be calculated using the solution for a single degree-of-freedom system:

$$S_R(\omega) = S_{wf}(\omega).(Q(\omega).L_T/T_0)^2$$

Here, the dynamic amplification factor $Q(\omega)$ is a function of the natural frequency in surge ω_n, and the damping ratio d

$$Q(\omega) = \left\{\left[1-\left(\frac{\omega}{\omega_n}\right)^2\right]^2 + \left[2\left(\frac{\omega}{\omega_n}\right)^2 d\right]^2\right\}^{-1/2}$$

Some results for the wind gust statistics, forces, and platform responses are given in Table 6.2 (a) for a range of wind speeds from 10 to 50 m/s. The significant amplitude of the winds speed increases as the wind speed increases, whereas the zero up-crossing period decreases, Fig. 6.1(a). Contrary to the assumption made earlier the gusting component of the wind speed is not insignificant which suggests for extreme winds the non-linear drag term should be modelled, either by time domain simulation or linearised frequency domain methods.

6.2.2 CURRENT FORCES

The current forces are calculated using a modification of the Morison wave force equation with the extra force due to the current included as a third term

$$F_D = A_I C_m \frac{\partial u}{dt} + A_D\left\{(C_D + \overline{C}_D)u|u|\right\}$$

Here, the drag coefficients C_D, $\overline{C}_D$ are the wave and current drag coefficients, respectively, which for this study have been taken as constant values of: 1.2 for the columns tendons and risers; and 2.0 for the pontoons. Modelled in this conventional way the current force is calculated as though the waves were not present whereas, in practice, there is interaction which not only shifts the frequency distribution in random seas but also alters the particle kinematics and creates a drift force term. In design calculations this interaction is difficult to model and usually ignored.

For simplicity, the current profile is modelled using a modification of the Department of Energy recommended profile, Fig. 6.2, with a current speed of U_c for the columns and pontoons, and $0.85U_c$ for the tendons and risers. The drag forces are assumed to be fully correlated over the whole structure so that the component forces can be summed to obtain the total force on the TLP. Examples of the total forces on the TLP for a range of currents are given in Table 6.2. Note the wind and current forces have the same order of magnitude for the extreme storm cases.

6.2.3 CURRENT MODIFICATION OF SPECTRA

The interaction of random waves and currents alters the spectral form by a significant amount and should therefore be allowed for when combining wave and current process. The spectral parameters recorded by the DB1 (summarised in chapter 2) were recorded without correction for current speed. This means the significant wave height and zero crossing period, in effect, correspond to the *corrected* values discussed below.

In the simplest case, the currents and waves are co-linear resulting in a change in the apparent wave frequency, that is, a single regular wave of frequency f will be transformed by a current giving an apparent frequency f_0 where $f_0 = f + kU_c / 2\pi$. Here a positive U_c is taken as being in the direction of the wave. The dispersion relation then becomes $(2\pi f_0 - kU_c)^2 = gk\tanh(kd)$ which in deep water can be simplified (see Huang et al (1972)) to give

$$k = \frac{4(2\pi f_0)^2 / g}{\left\{1+\sqrt{1+8U_c\pi f_0 / g}\right\}^2} \qquad [6.5]$$

When the waves are not co-linear with the current the interaction is more complex due to refraction, so for most work only the component resolved into the direction of the waves is taken as influencing the apparent wave frequency .

The current modification of wave elevation spectral form was studied by Huang et al (1972) who proposed a modification factor C_f such that $S^*(f) = C_f.S(f)$. For deep water

$$C_f = \frac{4}{\{1+8U_c\pi f/g\}^{1/2}\left\{1+(1+8U_c\pi f/g)^{1/2}\right\}^2} \qquad [6.6]$$

The effect of current on the apparent spectrum is shown in Fig. 6.3 for a range of current speeds from -1.0 to +1.0 m/s. Negative current increases the area under the spectrum and positive current decreases it. The influence of current on the spectrum is therefore most important when the current is in the opposite direction to the waves.

6.2.4 EFFECT OF CURRENT ON DRAG AND INERTIA LOADING

First order hydrodynamic fluid loading has been modelled by the Morison equation using frequency independent values of drag C_d and added mass coefficients C_a. The fluid loading on a TLP is dominated by Froude-Krylov and added mass forces so the effect of current on the added mass coefficients C_a, can be significant. For example, Sarpkaya et al (1984) found that C_m values $(= 1 + C_a \quad for\ a\ cylinder)$ in the critical range of Keulegan-Carpenter number K_c are increased significantly when currents are introduced. The uncertainty is further compounded by the large variability in observed measurements and experimental conditions. For this work, the statistical and modelling uncertainty for the added mass coefficients has been consolidated into a single uncertainty with the mean values of C_a determined using Det norske Veritas rules (1981) and the distribution of C_a taken as Normal with coefficient of variation equivalent to observed cov's for virtual mass coefficients C_m.

6.2.5 FROUDE-KRYLOV FORCES

The undisturbed, incident wave potential is used to calculate the Froude-Krylov forces F_p. These are given by the integrated hydrodynamic pressure on each wetted structural element

$$F_{FK} = \int_s p\vec{n}\, ds$$

Here: F_p is the vector of forces in structure x, y and z directions; n is the vector of unit normals; s is the element of surface area; and p is the pressure determined by $-\rho\partial\Phi/\partial t$. In this work, a simple linear Airy wave potential Φ_0 has been used where

$$\Phi_0 = (igH/2\omega)\frac{\cosh(k[y+d])}{\cosh(kd)} e^{i(kx\cos\theta + ky\sin\theta - \omega t - \psi)} \quad [6.7]$$

Here: i is the complex number $\sqrt{-1}$; g is acceleration due to gravity; H is the wave height; ω is the angular frequency; k is the wave number $2\pi/l$; d is the water depth; θ is the direction in which the wave propagates; t is time; Ψ is the phase angle ; and (x,y) are the coordinates in the (X,Y)-plane. This potential gives the the velocity vector (u,v,w), in which $u = \partial\Phi_0/\partial x$, $v = \partial\Phi_0/\partial y$, and $w = \partial\Phi_0/\partial z$.

When the section sizes are small compared with the incident wave length the forces for each element can be estimated using

$$F_{FK} = M_{FK}\ddot{\eta}$$

Where $\ddot{\eta}$ is the vector of accelerations, and the matrix M_{FK} is the (6 x 6) Krylov 'added inertia' matrix determined by

$$\sum_k \rho S_k dl_k$$

Here: S_k is the cross sectional area in a plane parallel to the flow; and dl_k is the kth elemental length. The summation is taken over the whole structure and the fluid particle accelerations $\ddot{\eta}$ are determined at the centre of the member cross-section parallel to the flow. This approach is only acceptable when the wave length is greater than approximately five times the maximum dimension of the structural components. For the TLP examined in this study the column diameter is 16.88m which suggest the wave length must be greater than 84m; that is the period must be greater than 7.3 seconds. In consequence, the Krylov forces on the columns must be calculated analytically using closed form solutions for the integrated pressure using for example the results given by Chakrabarti (1987) for the first order wave force

$$F_x = \rho V \frac{2J_1(kD/2)}{kD/2}\frac{\sinh[kl/2]}{kl/2}\dot{u}_0 \quad [6.8]$$

Here: J_1 is the Bessel function of the first kind and order 1; k is the wave number; D is the column diameter; V is the volume; and l is the length of the cylinder.

6.2.6 DIFFRACTION FORCES

The presence of the structure disturbs the incident wave potential creating diffraction forces. These forces can either be calculated using source distribution methods or by simple Morison approximation in which, for small sections, the diffraction forces are estimated as

$$F_{DF} = M_A\left(\ddot{\eta} - \ddot{Z}\right)$$

The matrix M_A is the (6 x 6) added mass matrix for the submerged structure, and the vector $\ddot{Z}$ is the vector of platform acceleration responses. The added mass coefficients have been calculated for the columns and pontoons using the Det norske Veritas (1981) Appendix B recommended values.

The sum of the Froude-Krylov and the diffraction forces $F_{FK} + F_{DF}$ can be written as

$$-M_A\left(\ddot{Z}\right) + \ddot{\eta}\left(M_A + M_{FK}\right)$$

For circular cylinders the second term is approximated in Morison's equation as

$$\ddot{\eta} \sum_k \rho\, Cm_k\, S_k dl_k$$

Here, Cm_k is a vector of inertia coefficients in the x, y, z directions.

6.2.7 DRAG LOADING

Viscous effects and flow separation result in drag forces which are a non-linear function of wave amplitude. Resolving the drag forces into x, y and z directions we get

$$F_{DR} = \frac{1}{2}\rho \sum C_d . H\left(\dot{\eta} - \dot{Z}\right)\left|\left(\dot{\eta} - \dot{Z}\right)\right|$$

Where: H is the vector of element widths normal to the flow; C_d is the vector of drag coefficients in the x, y, z directions; and $\left(\dot{\eta} - \dot{Z}\right)$ is the relative velocity at the centre of the member.

Solution of the general equation of motion responses is complicated by the presence of non-linear forcing and it is necessary to linearise the drag term. The most common method assumes the linearised drag force dissipates the same energy per wave cycle as the non-linear drag with

the result that the non-linear term $C_d\left(\dot{\eta}-\dot{Z}\right)\left|\left(\dot{\eta}-\dot{Z}\right)\right|$ can be replaced by the linear 'equivalent' $C_{dl}\left(\dot{\eta}-\dot{Z}\right)$. Here the linearised drag coefficient is given by

$$C_{dl}=\frac{8\omega}{3\pi}C_d(\eta-Z)$$

which is a function of the platform motions.

6.2.8 MOTION INDUCED FORCES

For fixed structures the velocities and accelerations used in Morison's equation are given by the water particle kinematics, however, for compliant systems the relative velocities and acceleration must be used

$$\dot{R}=\left(\dot{\eta}-\dot{Z}\right) \quad and \quad \ddot{R}=\left(\ddot{\eta}-\ddot{Z}\right)$$

Where, R is the relative displacement vector, $R=\left\{R_x,R_y,R_z,\theta_x,\theta_y,\theta_z\right\}$ and dot denotes differentiation with respect to time. At the start of a motion analysis the vector R is unknown and consequently so is C_{dl}. An iterative scheme is then needed in which the linearised drag force is updated by the results of the previous iteration.

6.3. TETHER TENSION MODEL

The response of the tendons is complicated by the large displacements induced by the wind, wave steady drift, and current forces which change the stiffness characteristics of the system. In this work, the changes in tendon mean tension induced by the steady offset forces are not included in the first order response calculation since this would require that the general equation of motion be solved for each Monte Carlo or FORM iteration. Consequently, in the generation of the motion and force transfer functions, the tethers are idealised as weightless, perfectly linear elastic, with constant lateral and axial stiffness.

The non-linear change of tendon tension has been modelled in the calculation of the mean tension used for the limit state functions discussed in Section [6.2]. Denoting the wind, current, and steady drift forces as F_w, F_c , and F_{sd} then the vector resultant F_r is simply $\sqrt{F_w^{\,2}+F_c^{\,2}+F_{sd}^{\,2}}$. This force creates a change in the tendon mean tension ΔT at each corner group and in turn stretches the tendons by an amount Δl, inducing strains $\varepsilon=\Delta l/L_t$. The force

diagram for a group of tendons is shown in Fig. 6.4 from which we can see vertical equilibrium gives

$$T_0 \cos\theta + \Delta T \cos\theta = T_0 + \tfrac{1}{4}k_{33}L_T\left(1 - \cos\theta - \frac{\Delta T}{AE}\cos\theta\right)$$

where: θ is the inclination of the tendons; T_0 is the initial tension in the group; $\frac{1}{4}k_{33}$ is the heave hydrostatic stiffness at one corner; A is the cross sectional area of the tendons; E is youngs modulus for the steel; and the stretch is given by $\Delta l = \Delta T L_T / AE$. On algebraic manipulation we get

$$\Delta T = \frac{(T_0 + \frac{1}{4}k_{33}L_T)(\sec\theta - 1)}{[1 + \frac{1}{4}k_{33}L_T / AE]} \qquad [6.9]$$

Taking moments about the base gives

$$\tfrac{1}{4}F_r \cos\theta - T_0 \sin\theta - \tfrac{1}{4}k_{33}L_T\{1 - (1 + \Delta T / AE)\cos\theta\}\sin\theta = 0 \qquad [6.10]$$

The two unknowns θ, and ΔT are solved by substituting for ΔT in Equation [6.10] and then root finding with a bisection algorithm. One point worth noting is the small effect of the extension of the tendons. This supports the use of less stiff materials such as parafil and kevlar for deep water tension legs.

6.4. PLATFORM DYNAMICS

The motion responses have been calculated using the structure reference system shown in Fig. 6.1. Assuming the platform is modelled as a rigid mass with six degrees of freedom the general equation of its motion can be written as

$$M\ddot{Z} + C\dot{Z} + KZ = F(t) \qquad [6.11]$$

Here: $M = (M_s + M_A)$ is the (6x6) structural plus added hydrodynamic mass matrix; Z is the vector of unknown platform displacements; $C = (C_v + C_R)$is the (6x6) matrix of viscous and radiation damping coefficients; $K = (K_T + K_H)$is the (6x6) matrix of tether and hydrostatic stiffnesses; F is the vector of Morison wave forces. The evaluation of each term is given explicitly in Patel & Witz (1991) and will be discussed briefly in what follows.

6.4.1 THE MASS MATRIX

The structure mass matrix is diagonal when the structure axes coincide with the principal axes. Consequently, $m_{11} = m_{22} = m_{33}$ = the total structure mass (54.5 x 106 kg); and the pitch, roll, and yaw moments of inertia are those given in Table 6.1.

$$M = \begin{bmatrix} m_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & m_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & m_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & m_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & m_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & m_{66} \end{bmatrix}$$

The added mass matrix is given by summation of the volume element added masses for the submerged parts of the structure. Patel *op cit* gives the added mass matrix terms for a non-elongated body and these have been used for this work.

6.4.2 THE FLUID DAMPING MATRIX

Fluid damping results from viscous drag effects. The terms of the 6x6 damping matrix can be derived element by element and then summed to give the total fluid damping. Unlike the added mass the damping matrix is asymmetric being proportional to the non-linear velocity squared term in Morison's equation. Again Patel gives expressions for element damping terms and they have been adopted for this study.

6.4.3 TENDON STIFFNESS MATRIX

The idealised tether stiffness matrix is calculated by assuming the tethers are weightless, linear elastic and tensioned by a constant amount. These assumptions allow us to ignore the non-linear tether dynamics and change in tension as the platform oscillates. Taking the axial elastic stiffness per tether as AE/L_T and the equilibrium position tension as T_0 the restoring stiffness terms are those given by Patel op cit.

6.4.4 HYDROSTATIC STIFFNESS

The hydrostatic restoring forces arise in the heave, pitch and roll degrees of freedom. Writing $a33_i$ for the water plane area for the ith surface piercing element and ∇ as the vessel displacement volume we get

$$k_{33} = \rho g \sum a33_i$$
$$k_{43} = \rho g \sum y_w a33_i$$
$$k_{53} = -\rho g \sum x_w a33_i$$
$$k_{54} = \rho g \sum x_w y_w a33_i$$
$$k_{44} = \rho g \nabla (GM_p)$$
$$k_{55} = \rho g \nabla (GM_r)$$

where: GM_p and GM_r are the pitch and roll metacentric heights; and x_w, y_w are the x and y coordinates for the centriods of the cut waterplane areas. All other terms are zero.

6.4.5 MOTION NATURAL FREQUENCIES

The natural frequency for each mode of motion of the TLP can be calculated from the mass, and stiffness matrices for the system. Referring to Equation [6.11] the equation of dynamic equilibrium for free, undamped vibrations of the platform is given by

$$M\ddot{Z} + KZ = 0$$

Substituting the solution $Z = Be^{i\omega t}$ then gives

$$-\omega^2 MZ + KZ = 0$$

which can be identified as a classical eigenvalue problem by rearranging to give

$$(M^{-1}K - \lambda I)Z = 0 \qquad [6.12]$$

Here: I is the identity matrix; and the eigenvalues λ correspond to the square of the platform natural frequencies ω_n. Solution for the eigenvalues of a non-symmetric matrix is not trivial and consequently a set of library routines is used, Press et al (1989). The procedure consists of first balancing the matrix, then reducing it to a Hessenberg form, and finally extracting the eigenvalues using the Housholder QR algorithm. The results are the six natural frequencies

1surge	97.07 sec	2 sway	97.07 sec
3 heave	1.706 sec	4 roll	1.766 sec
5 pitch	1.766 sec	6 yaw	83.52 sec

These values are consistent with the results obtained by assuming the modes are all uncoupled and then using the single degree of freedom solution $T_i = 2\pi\sqrt{k_i / m_i}$.

6.5 STRUCTURAL RESPONSE IN IRREGULAR WAVES

The eight limit state functions outlined above require estimates of the platform displacements for each realisable event defined by the joint density of environmental parameters. Both frequency and time domain methods can be used to solve for motion responses in irregular waves. Time domain methods have the advantage of accurate modelling of free surface, viscous and stiffness non-linearities, however, the method is computationally too expensive to be used in Monte Carlo simulation. A frequency domain method has therefore been adopted for this work.

If the response of the TLP is linearly related to wave amplitude over all frequencies then classical response analysis, Chakrabarti (1987), gives

$$S_R(f) = S_\eta(f).[H(f)]^2$$

Where: $S_\eta(f)$ is the frequency distribution of surface elevation variance in m^2s; $H(f)$is the response amplitude operator in response/m; and $S_R(f)$ is the response spectrum with units response^2 second . The response of a linear system with gaussian forcing is also gaussian and several useful statistical properties of the response can be calculated from the moments of the response spectra using the results obtained by Rice (1944/45). An alternative approach based on the so-called out-crossing approach developed by Veneziano et al (1977) has been suggested for vector process models which could be of future use in this work.

In multidirectional seas the response transfer function for each mode of motion and limit state can be calculated from the directional spectrum $S_\eta(f,\theta)$ and the directional transfer function $H(f,\theta)$

$$S_{Ri}(f) = \int_0^{2\pi} S_\eta(f,\theta).H_i(f,\theta)^2 d\theta$$

The directional wave spectrum $S_\eta(f,\theta)$ is normally written as the product of the point spectrum and a directional spreading function

$$S_\eta(f,\theta) = S_\eta(f).D(\theta,f)$$

The response spectrum can then be evaluated at discrete frequencies f_i using transfer functions calculated at the same frequencies and directions $\theta_j; j = 1,nd$

$$S_R(f_i) = \sum_{j=1}^{nd} S_\eta(f_i).D(\theta_j;f_i).H(f_i,\theta_j)^2 \Delta\theta \qquad [6.13]$$

Models for the frequency dependent directional probability distribution $D(\theta_j;f_i)$ are discussed in detail in chapter 5. The evaluation of Equation [6.13] therefore requires calculation of the linear first-order transfer function for each mode of motion and limit state margin G_i covering the range of frequencies for which significant energy is present in the wave surface elevation spectrum. A piece-wise linear estimate of the safety margin transfer functions $H_{G_i}(f,\theta)$ can be obtained by calculating the responses induced by a unit amplitude wave with range of frequencies $f_i; i = 1,...,nf$, and directions $\theta_i; i = 1,nd$.

6.5.1 JONSWAP SPECTRUM

The variance spectral density for fetch limited seas was examined during the JONSWAP project, Hasselmann et al (1976b). The average shape of their fetch limited seas was found to have form given by a modified Pierson-Moskowitz (1964) spectrum

$$S(\omega) = \frac{\alpha g^2}{\omega^5} \quad exp\left\{-1.25\left[\frac{\omega}{\omega_p}\right]^{-4}\right\} \quad \gamma^{exp\left\{\frac{-(\omega-\omega_p)^2}{2\tau^2\omega_p^2}\right\}} \qquad [6.14]$$

$$\omega_p = 2\pi * 3.5\left(\frac{g}{U_w}\right)[X_0]^{-0.33}$$

where: γ is the peak enhancement factor ; ω is the wave frequency; ω_p is the spectral peak frequency; and τ are the shape factors. The parameter alpha is a function of the fetch

$$\alpha = 0.076\,(X_0)^{-0.22}$$

Here, X_0 is the non-dimensionalised distance

$$X_0 = \frac{gX}{U^2_w}$$

where U_w is the wind speed at an elevation of +10.0m, and X is the fetch distance. When the fetch is unknown Phillip's constant is usually used ie α =0.0081.

When the peak enhancement factor equals 1.0 the spectrum reduces to the form of a Pierson-Moskowitz spectrum. This enables us to determine the fetch - for a given wind speed - at which the seas become fully developed using the expression for the spectral peak frequency.

The parameters of the JONSWAP spectrum were determined by fitting the parametric model to 121 observed spectra using least squares. Hasselmann et al (1976b) reported some estimates of the statistical variability of $\{\gamma, \alpha, \tau\}$ using some 333 spectra observed by a number of researches; the results given below are useful for reliability studies.

parameter	mean	S.D.(%)	regr. coef.
γ	2.65	44	0.32
τ_a	0.85	76	-0.32
τ_b	0.097	47	-0.16
α	0.0109	38.6	0.87

The form of the JONSWAP spectrum given previously is not suited to engineering applications and has been modified by a number of researchers. Goda (1985) proposed a form which is a function of the significant wave height, i.e

$$S(\omega) = \frac{\alpha^* Hs^2}{\omega^5} \omega^4_p \quad exp\left\{-1.25\left[\frac{\omega}{\omega_p}\right]^{-4}\right\} \quad \gamma^{exp\left\{\frac{-(\omega-\omega_p)^2}{2\tau^2\omega^2_p}\right\}} \qquad [6.15]$$

The value of alpha is a function of the peak enhancement factor

$$\alpha^* = \frac{0.0624}{\left\{0.230 + 0.0336\gamma - \frac{0.185}{(1.9+\gamma)}\right\}}$$

$$\tau_a = 0.07 \qquad \omega \le \omega_p \qquad \tau_b = 0.09 \qquad \omega > \omega_p$$

Regression analysis of measured data has shown

$$H_s = (0.11661 + 0.01581\gamma - 0.00065\gamma^2)T_0^2 \qquad [6.16]$$

where the spectral peak period is given by

$$T_0 = \frac{2\pi}{\omega_p} = (1.49 - 0.102\gamma + 0.0142\gamma^2 - 0.00079\gamma^3)T_z \qquad [6.17]$$

6.5.2 FIRST ORDER WAVE TRANSFER FUNCTIONS

The first order wave drag and inertia forces have been calculated for a range of frequencies and directions using the Morison method described previously. The results, which are illustrated graphically for forty frequencies and five directions in Fig. 6.5 and 6.6, compare very favourably with the diffraction results reported by Eatok-Taylor & Jeffereys (1986). The only major discrepancies occur at high frequencies where the structure dimensions are large when compared to the wave length and diffraction effects cannot be ignored. For this work this is not important for two reasons: first, only the extreme seastates which have longer period waves contribute to the failure probabilities of interest in the reliability analysis; and second, we are only interested in the difference frequency second order forces for the calculation of mean and slowly varying drift effects.

The force and motion response functions shown in Fig. 6.5 and 6.6 demonstrate that only the surge, sway, roll, and pitch responses are sensitive to the direction of the waves. The heave is little affected by the approach angle.

6.5.3 RESPONSES IN DIRECTIONALLY SPREAD SEAS

The directional spreading of wave elevation variance is examined in Ch. 5 in which the Hasselmann, Mitsuyasu, and Prince-Wright directional distributions are defined. These have all been included in the TLP program to enable comparison of the TLP response in seas with different directionally spread seas. An example of the JONSWAP spectrum and Hasselmann spreading powers is shown in Fig. 6.7(a) for a seastate with 10m significant wave height, and 12 second zero-crossing period. Note the mismatch of the peaks is the result of swapping equations at f/fm = 1.05 (Ch.5). The shape of the directional spectrum is shown in Fig. 6.7(b) for a range of frequencies and directions.

A 3-D plot of the directional surge force and motion transfer functions is shown in Fig. 6.8, and the surge, heave and pitch force and response spectra are shown in Fig. 6.9 for a seastate with Hasselmann spreading function. In all three modes, the force spectra in long crested seas are higher than in short crested seas. This is also true for the surge and heave motions but not for the pitch motion. This is due to the motions being largest in pitch for a quartering sea case which can be seen in the motion transfer functions, Fig. 6.6. The ratios *long crested surge / short crested* surge are given for a range of JONSWAP shape parameter and three spreading models in Table 6.3. The results are also plotted in Fig. 6.10 which shows more clearly that, in all cases, as the seas become more fully developed the long crested seas model overestimates the motions and forces by up to a maximum of 20% when the Mitsuyasu and Hasselmann models are used. However, the Prince-Wright model indicates the overestimate is only 10% for the forces and 16% for the motions. The differences are due to the slower decay of the Prince-Wright spreading powers away from the spectral peak. Table 6.4 shows in greater detail the ratio of the force, motion, and limitstates for a Hasselmann sea.

The analysis of the DB1 directional data in Ch. 5 demonstrated the large variability of the spreading powers at a given frequency ratio f/fm. This variability is due to a combination of the effects of: randomness; noise in the recorded heave pitch and roll time series; and sampling error. The relative importance of this variability can be modelled in a reliability analysis by including the angular variance as a random variable.

6.6. SLOWLY VARYING AND MEAN MOTIONS DUE TO WAVE DRIFT

Fixed and floating offshore structures in regular and irregular waves experience second order forces with small mean value and frequencies which are harmonics of the first order wave force. There are several reasons for these forces and the contribution of each cause to the total effect depends on the type and dimensions of the structure. Large structures like ships or barges are dominated by the potential drift forces which are linearly related to the wave amplitude squared - and therefore simple to calculate in irregular seas. Smaller structures like semi-submersibles and tension leg platforms experience both viscous and drift forces, the latter being proportional to the cube of the wave elevation and therefore difficult to include in spectral analysis. In this work the potential drift forces are modelled using a simplified method developed by Chakrabarti (1984).

The calculation of accurate drift forces on complex geometries generally requires the use of 3D-source distribution methods, see Chan (1990), however for this work a simple approximation is used in which the MacCamy & Fuchs (1954), and Chakrabarti (1984) closed form solutions for seabed mounted, surface piercing structures, are used (despite the fact the

columns do not reach the seabed). This simplification is justified because the majority of the drift force is located near the wave free-surface zone, away from the pontoons, risers, and tethers.

The mean *viscous* drift forces on a cylinder in the direction of each axis, $\overline{F_1}$, can be approximated by averaging the Morison drag force over the whole cylinder for one wave cycle

$$\overline{F_1} = \tfrac{1}{2}\rho C_d D \int_0^\eta \vec{n}.|\nabla\Phi|^2 \, dl \qquad [6.18]$$

Here: C_D are the drag coefficients; D is the cylinder diameter (radius r); n is the vector of direction cosines ; $|\nabla\Phi|^2$ is the first order velocity vector squared; and *dl* is the elemental length of the cylinder. On substituting the deep water Airy wave potential in to Equation [6.18], and then integrating, Chakrabarti (1984) gives the simple result

$$\overline{F_1} = \frac{\rho g C_D D k H^3}{12\pi} \qquad [6.19]$$

Note this force is proportional to the wave amplitude cubed, unlike the potential forces outlined below which are proportional to the wave amplitude squared.

The mean *potential* drift force on a cylinder in regular seas with frequency ω is used to calculate the potential drift force transfer functions. These are the basis for Pinkster's (1974) approximate method for calculating drift forces in narrow-band irregular seas.

The mean *wave elevation* drift force $\overline{F_2}$ in a regular wave train is found by averaging the integral of the hydrodynamic pressures on the wetted area between the still water level and the instantaneous water level, over one whole wave cycle. For a vertical cylinder we have

$$\overline{F_2} = \overline{-\tfrac{1}{2}\rho \int_S \vec{n}.\partial\Phi / \partial t.ds} \qquad [6.20]$$

Here: $-\rho\partial\Phi / \partial t$ is the hydrodynamic pressure; and *ds* is the elemental surface area.

The *velocity head* drift force $\overline{F_3}$ is found by averaging the integral of the velocity head pressure over the full length of the submerged cylinder, for one wave cycle

$$\overline{F_3} = \overline{\tfrac{1}{2}\rho \iint_s \vec{n}|\nabla\Phi|^2 .ds} \qquad [6.21]$$

The deep water (depth > 0.5 wave length) closed form solutions for the integrals Equations [6.20] and [6.21] were determined by Havelock (1940), MacCamy & Fuchs (1954), and Chakrabarti (1984) who express the sum of the wave elevation and velocity head forces as

$$\overline{F_2} + \overline{F_3} = \frac{\rho g H^2 r}{\pi^2 (kr)^3} / \sum_{n=0}^{\infty} \left[1 - \frac{n(n+1)}{(kr)^2} \right]^2 \frac{1}{A_n(kr) A_{n+1}(kr)} \qquad [6.22]$$

Here: $A_n(x) = J_n'(x)^2 + Y_n'(x)^2$, where $J_n(x)$ is the Bessel function of the first kind, order n; $Y_n(x)$ is the Bessel function of the second kind, order n; and prime denotes the derivative with respect to x. The p^{th} derivatives of the Bessel functions can be expressed in terms of the basic Bessel function using the results in Abramowitz & Stegun (1965)

$$\ell_n^{(p)} = \frac{1}{2^p} \left\{ \ell_{n-p} - \binom{p}{1} \ell_{n-p+2} + \binom{p}{2} \ell_{n-p+4} - \ldots\ldots\ldots + (-1)^p \ell_{n+p} \right\} \qquad p = 0,1,2,.... \qquad [6.23]$$

Here ℓ is J,or Y , or any linear combination of the functions, and (p) denotes the p^{th} derivative. For our case, where $p = 1$, we have

$$\ell_n^{(1)}(z) = \frac{1}{2} \{ \ell_{n-1}(z) - \ell_{n+1}(z) \}$$

in which it should be noted that $\ell_{-n} = (-1)^n \ell_n$. This simple form enables us to calculate $A_n(x)$for all $x = kr$. Plots showing the Bessel functions and their derivatives are shown in Fig. 6.11.

The wave potential drift forces calculated from Havelock's closed form solution correspond to the forces on an isolated cylinder in an ideal fluid with unidirectional waves. The wave elevation and velocity head components of the drift force on a cylinder with radius equal to the TLP column radius are shown in Fig. 6.12 for a range of *kr*. Here we can see the velocity head force is negative and the wave elevator force is positive, these sum to give a positive wave drift force which is nearly constant for values of kr > 5.0.

A procedure for calculating both the mean drift force $\overline{F}$ and the drift force $S_F(\mu)$ and response spectra $S_R(\omega)$ in irregular seas is presented by Faltinsen & Demirbilek (1989). This method is, strictly speaking, only valid for narrow-band seas and uses the force results described above for regular seas with frequency ω and amplitude η_a. The mean drift is given as

$$\overline{F} = 2\int_0^{\infty} S_\eta(\omega')\left\{\frac{F(\omega')}{\eta_a^{\,2}}\right\}d\omega' \qquad [6.24]$$

where, $S_\eta(\omega)$ is the spectral density of the wave surface process, and the transfer function $F(\omega')$ is given by summing the viscous and potential drift forces. The spectrum of the slowly varying drift excitation force $S_{\bar{F}}(\omega)$ is then calculated using Pinkster's (1974) method in which it is assumed the drift forces are a quadratic function of the wave amplitude

$$S_{\bar{F}}(\mu) = 8\int_0^{\infty} S_\eta(\omega')S_\eta(\omega'+\mu)\left\{\frac{F(\omega'+\mu/2)}{\eta_a^{\,2}}\right\}^2 d\omega' \qquad [6.25]$$

This force spectrum is transferred to a response spectrum for the lateral motions of the platform by idealising it as a single degree of freedom system with damping d and lateral stiffness T / L_T, giving

$$S_R(\omega) = S_{\bar{F}}(\omega).\left(Q(\omega).L_T / T_0\right)^2$$

Here, the dynamic amplification factor $Q(\omega)$ is a function of the natural frequency in surge ω_n, and the damping ratio d

$$Q(\omega) = \left\{\left[1-\left(\frac{\omega}{\omega_n}\right)^2\right]^2 + \left[2\left(\frac{\omega}{\omega_n}\right)^2 d\right]^2\right\}^{-1/2}$$

This model for the wave drift forces is only accurate for low significant wave heights because of the amplitude cubed term in the viscous drift force equation. As the wave heights increase the viscous forces dominate the drift loading on the TLP columns. The effect of this simplification is to underestimate the total forces due to the viscous drift. This problem has been examined by Kato & Kinoshita (1990) who suggests the second order forces be formulated as a Volterra functional series and solved using Wiener's filter theory. This approach will be

more accurate but computationally too expensive for the Monte Carlo analysis. It is suggested this be the subject of future research

6.6.1 COMPUTED WAVE DRIFT FORCES: COMPARISON WITH 3-D SOURCE RESULTS

The drift force model developed for this work is a simple approximation which does not allow for wave force cancellation, spreading, or interference effects. The results compare favourably with Tan & de Boom (1981) in which a 3-D source distribution method is used to generate the drift transfer functions. A comparison of our results, with the head seas drift force spectra, as given in Tan *loc cit*, is shown in Fig. 6.13.

As an example, the mean and slowly varying wave drift forces on the TLP have been calculated for a JONSWAP sea with Hs=10.0 metres, Tz=12.0 seconds, and γ=3.3. The intermediate results in the calculation of the drift statistics are illustrated in Fig. 6.14 which shows: the seastate spectrum; the viscous forces; the potential drift force; the drift force spectrum; the single degree of freedom magnification factor; and finally, the drift motion response spectrum. The drift motion response spectrum is narrow with a peak located at the surge natural period of the TLP (97sec.) . The mean wave drift force for this case is 3.16e+5 (N), which is very small when compared with the most probable, maximum (mpm) first order wave surge force of 7.56e+8 (N). The slowly varying responses are larger with a three hour mpm force of 2.03e+6 N and a three hour mpm motion amplitude of 20.4m. This motion is significant when compared with the first order wave mpm amplitude of 53.0m.

6.7. COMBINED WIND, WAVE AND CURRENT MOTIONS AND FORCES

The motions of the TLP and the forces in the tendons consist of three components: a mean drift response; a first order wave frequency response; and a second order slowly varying drift response, Fig. 6.15. The mean value of each force acting individually has been examined previously. In the conventional design approach it is usually assumed that each force acts colinearly and that they attain their maxima simultaneously. This is conservative for most structures and it is interesting to examine the effect of varying the directions and magnitudes of the winds, waves and currents.

Tables 6.6, 6.7,and 6.8 compare the platform responses for seastates with: 12,14 and 16 metre significant height; zero crossing periods ranging from 11 to 14 seconds; a constant wind speed of 50m/s; and a constant current speed of 2.0m/s. These tables show the platform first order wave frequency motions and tendon responses are sensitive to both the significant wave height and the zero crossing period. The increases in the motions and the limit state stresses are nearly linear with increasing significant height and zero-crossing period. However, the second order

mean and slowly varying wave responses increase with increasing height, but decrease with increasing period.

The change in motion, force, and limit state variance with height and period demonstrates the importance of correctly modelling the joint distribution for the significant wave height and zero crossing period. For example, if we compare the motion and limit state responses in a seastate with *Hs=12* metres, and *Tz=14* seconds with the responses in a seastate with *Hs=16* metres, and *Tz=11* seconds then we find they are almost equal! The period sensitivity of this sort of structure - due to its' long natural periods in surge sway and yaw - can only be understood using probabilistic methods since the contribution of the lower, longer period, seastates to the expected population of extreme response events during the platform life cannot be ignored. The use of a design wave approach, in which a return period wave is combined with a most probable or associated wave period,. could in this case be unconservative The significance of the occurrence of long period, moderate significant height, seastates must of course take into account the likelihood of their occurrence and this lends support to the use of kernel density estimates for the joint distribution of the heights and periods, rather than crude scatter plot estimates.

6.8 CONCLUSIONS

A loading and linear response model of a tension leg platform is presented which is suitable for reliability calculations. Environmental wind and wave forces are modelled as stochastic and current forces are modelled as steady throughout the seastate. The responses of the system are calculated in the frequency domain by assuming the system is linear, and the combined load effect is modelled using a sum-of-squares approach.

A simple Morison-based method was used to calculate the first order wave forces on the TLP columns and pontoons. The force and response transfer functions compare very favourably with the results from several source-sink distribution programs which in themselves gave widely differing results. The only major differences occur at high frequencies in which case the structure dimensions are large when compared with the wave length.

The simple MacCamy & Fuchs (1954)and Chakrabarti (1984) closed form solutions for the second order potential wave drift forces compare reasonably well with the results from the study by Tan & de Boom (1981). However, in neither their case, or ours, is the cubic non-linearity of the drag effect properly allowed for and this is probably significant for the extreme wave case. Since this effect is important it is suggested the effects be quantified by comparing results with time domain solutions.

The sensitivity studies for the wind,wave and current forces indicate the dominating force is from the first order wave loads. The slowly varying wind and drift forces contributing the next largest component, and the steady current forces are small in comparison.

An Ochi wind spectrum was chosen because it is supposed to model spectra over the open ocean. However, the results from this spectrum seem questionable because the significant amplitude is generally some 20~30% of the mean wind speed. If the gusts are narrow-band Gaussian then we would expect the 3-hour most probable maximum amplitude to be some 40~60% of the mean wind speed. It is suggested therefore some time series analysis is performed on data recorded during extreme storms to check this result.

APPENDIX A - POLYNOMIAL APPROXIMATION FOR THE POTENTIAL DRIFT FORCES

The CPU time required for the calculation of the potential drift forces was considerably larger than was practical for the Monte Carlo analysis and consequently a nth order polynomial approximation was determined by fitting the form

$$\overline{F}_2 + \overline{F}_3 = \sum_{i=1}^{n} C_i (kr)^i \qquad i = ,1,20$$

to several exact values of the force calculated using Havelock's closed form solution. The result, summarised in Fig. 6.16 together with the coefficients of the polynomial, gives accurate approximates for all kr>4.5. For values greater than this the Havelock asymptotic approximation has been used. Note the coefficients are defined in double precision to ensure the correct values are calculated for the larger (kr>1.0) values of ka.

COLUMNS AND PONTOONS:	
spacing between column centres	86.25 m
column radius	8.44 m
pontoon width	7.50 m
pontoon height	10.50 m
draft	35.00 m
displacement	54.5×10^6 kg
total mass	40.5×10^6 kg
Longitudinal metacentric height	6.0 m
Transverse metacentric height	6.0 m
Roll moment of inertia	82.37×10^9 kg m^2
Pitch moment of inertia	82.37×19^9 kg m^2
Yaw moment of inertia	98.07×10^9 kg m^2
Vertical position of C of G above keel	38.0 m
TETHER PROPERTIES:	
Vertical stiffness of combined tethers	0.813×10^6 kN/m
Total tether pre-tension	14.0×10^6 kg
no. per corner	4 off
outside diameter	1.0m
inside diameter	0.9338m
wall thickness	0.033088m
cross-sectional area	0.10051m^2
Youngs modulus	2.098×10^{11} N/m^2
mass per unit length	789 kg/m
submerged weight per unit length	0 kg/m
axial stiffness	5.081×10^7 N/m
length	415m
mean ultimate strength	620.0 N/mm^2
s.d. of ultimate strength	37.5 N/mm^2
Weibull parameters of yield strength	
shape	1.8
scale	92 N/mm^2
lower bound	488 N/mm^2

Table 6.1 Structural data for the TLP, from Tan & de Boom (1981).

currents (m/s)	force (N)	wind (m/s)	force (N)
0.5	1.49e6	10	1.44e5
1.0	5.95e6	30	1.30e6
1.5	1.34e7	50	3.60e7

Table 6.2 Wind and current forces on the tension leg platform

MEAN WIND SPEED (m/s)		10	20	30	40	50
ZERO CROSSING PERIOD (s)		48.46	43.53	40.74	38.81	37.42
SIGNIFICANT WIND SPEED (m/s)		1.794	4.46	7.557	10.94	14.55
THREE HOUR MPM LARGEST SPEED (m/s)		2.95	7.405	12.62	18.35	24.49
SPECTRAL MOMENTS	M0	0.8047	4.972	14.28	29.9	52.94
	M1	7.58E-02	0.5411	1.704	3.828	7.143
	M2	1.35E-02	0.1036	0.3395	0.7836	1.493
ZERO CROSSING PERIOD OF MOTIONS (s)		108.5	106.4	104.8	103.5	102.3
SIGNIFICANT WIND MOTION AMPLITUDE (m)		0.8633	4.224	10.66	20.49	33.83
THREE HOUR MPM LARGEST MOTION (m)		1.309	6.42	16.23	31.23	51.63
SPECTRAL MOMENTS	M0	0.1863	4.461	28.42	104.9	286.1
	M1	1.02E-02	0.2512	1.632	6.129	16.95
	M2	6.24E-04	1.56E-02	0.1021	0.3869	1.079
ZERO CROSSING PERIOD OF FORCES (s)		48.46	43.53	40.74	38.81	37.42
SIGNIFICANT WIND FORCE AMPLITUDE (N)		4.30E+04	2.14E+05	5.44E+05	1.05E+06	1.75E+06
THREE HOUR MPM LARGEST FORCE (N)		7.07E+04	3.55E+05	9.08E+05	1.76E+06	2.94E+06
SPECTRAL MOMENTS	M0	4.63E+08	1.14E+10	7.39E+10	2.75E+11	7.61E+11
	M1	4.36E+07	1.25E+09	8.82E+09	3.52E+10	1.03E+11
	M2	7.78E+06	2.38E+08	1.76E+09	7.21E+09	2.15E+10

Table 6.2 (a) Statistics of wind speed, wind force and platform motion for a range of mean wind speeds and the Ochi (1988) spectrum

jonswap gamma	JONSWAP wave spectrum with Hs=10.0m and Tz=12.0sec. ratios long crested/short crested					
	hassemann		mitsuyasu		prince-wright	
	surge force	surge motion	surge force	surge motion	surge force	surge motion
1	1.181	1.198	1.176	1.155	1.107	1.161
2	1.164	1.173	1.15	1.12	1.098	1.154
3	1.15	1.154	1.139	1.098	1.092	1.151
4	1.14	1.14	1.131	1.085	1.081	1.142
5	1.132	1.13	1.124	1.075	1.08	1.144
6	1.126	1.129	1.12	1.067	1.077	1.14
7	1.121	1.116	1.118	1.06	1.071	1.138

Table 6.3 Ratio of the TLP surge responses in long crested seas and short crested seas.

hasselmann

MOTIONS	surge	sway	heave	roll	pitch	yaw
VARIANCE OF MOTIONS :	1.150	0.000	1.024	0.000	0.776	0.000
ZERO CROSSING PERIODS :	1.008	0.863	0.986	1.237	1.033	1.632
SIGNIFICANT AMPLITUDE :	1.150	0.000	1.024	0.000	0.776	0.000
THREE HOUR MPM LARGEST:	1.149	0.000	1.025	0.000	0.775	0.000
SPECTRAL MOMENTS M0 :	1.322	0.000	1.049	0.000	0.603	0.000
M1 :	1.309	0.000	1.065	0.000	0.583	0.000
M2 :	1.300	0.000	1.080	0.000	0.564	0.000

FORCES	surge	sway	heave	roll	pitch	yaw
VARIANCE OF FORCES :	1.147	0.000	1.024	0.000	1.067	0.000
ZERO CROSSING PERIODS :	0.955	0.000	0.985	1.035	0.974	1.668
SIGNIFICANT AMPLITUDE :	1.147	0.000	1.024	0.000	1.067	0.000
THREE HOUR MPM LARGEST:	1.151	0.000	1.025	0.000	1.069	0.000
SPECTRAL MOMENTS M0 :	1.316	0.000	1.048	0.000	1.138	0.000
M1 :	1.358	0.000	1.064	0.000	1.155	0.000
M2 :	1.444	0.000	1.079	0.000	1.201	0.000

LIMIT STATES	1	2	3	4	5	6	7	8
VARIANCE OF LIMIT STATES :	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001
ZERO CROSSING PERIODS :	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SIGNIFICANT AMPLITUDE :	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001
THREE HOUR MPM LARGEST :	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001
SPECTRAL MOMENTS M0 :	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002
M1 :	1.003	1.003	1.003	1.003	1.003	1.003	1.003	1.003
M2 :	1.003	1.003	1.003	1.003	1.003	1.003	1.003	1.003

Table 6.4 Ratios of response statistics for long-crested and short crested seas with Hasselmann spreading function and Hs=10.0m, Tz=12.0sec.

	current direction (degrees) 0	45	90	135
SEA DRIFT FORCE FX (N)	3.16E+05	3.16E+05	3.16E+05	3.16E+05
SEA DRIFT FORCE FY (N)	0.00E+00	0.00E+00	0.00E+00	0.00E+00
CURRENT FORCE FX (N)	2.38E+07	1.54E+07	2.76E-08	-1.54E+07
CURRENT FORCE FY (N)	0.00E+00	1.54E+07	2.38E+07	1.54E+07
WIND FORCE FX (N)	3.60E+06	3.60E+06	3.60E+06	3.60E+06
WIND FORCE FY (N)	0.00E+00	0.00E+00	0.00E+00	0.00E+00
RESULTANT FORCE FR (N)	2.77E+07	2.48E+07	2.41E+07	1.93E+07
WIND,CURRENT, DRIFT OFFSET DIR.	0	38.573	80.646	126.727
INCLINATION OF TENDONS (deg)	8.75	8.091	7.959	6.772
INCREASE IN TENDON TENSION (N)	1.14E+07	9.73E+06	9.42E+06	6.80E+06
STRETCH OF THE TENDONS (m)	0.056	0.048	0.046	0.033
STATIC OFFSET DISTANCE (m)	63.129	58.407	57.461	48.939
STATIC OFFSET SETDOWN (m)	4.83	4.131	3.997	2.896

Table 6.5 Effect of varying the current direction on the mean wind, wave and current drift forces and responses.

zero crossing period (s) ----->		11sec	12sec	13sec	14sec
response	unit				
SEA DRIFT FORCE FX	N	5.99E+05	4.55E+05	3.54E+05	2.78E+05
SEA DRIFT FORCE FY	N	0.00E+00	0.00E+00	0.00E+00	0.00E+00
CURRENT FORCE FX	N	2.38E+07	2.38E+07	2.38E+07	2.38E+07
CURRENT FORCE FY	N	0.00E+00	0.00E+00	0.00E+00	0.00E+00
WIND FORCE FX	N	3.60E+06	3.60E+06	3.60E+06	3.60E+06
WIND FORCE FY	N	0.00E+00	0.00E+00	0.00E+00	0.00E+00
RESULTANT FORCE FR	N	2.80E+07	2.79E+07	2.78E+07	2.77E+07
WIND,CURRENT, DRIFT OFFSET DIR.	D	0	0	0	0
INCLINATION OF TENDONS	D	8.838	8.794	8.794	8.75
STRETCH OF THE TENDONS	m	0.057	0.056	0.056	0.055
STATIC OFFSET DISTANCE	m	63.758	63.444	63.444	63.129
STATIC OFFSET SETDOWN	m	4.927	4.878	4.878	4.83
INCREASE IN TENDON STRESSES	N/mm2	28.614	28.328	28.328	28.043
INCREASE IN TENDON TENSION	N	2876.027	2847.217	2847.217	2818.556
VARIANCE OF DRIFT MOTIONS	m**2	7.54	5.67	4.36	3.40
ZERO CROSSING PERIOD OF MOTIONS	S	101.80	101.70	101.60	101.60
SIGNIFICANT DRIFT MOTION AMPLITUDE	m	15.08	11.35	8.72	6.81
THREE HOUR MPM LARGEST MOTION	m	23.02	17.33	13.32	10.40
VARIANCE OF DRIFT FORCES	m**2	9.59E+05	7.66E+05	6.25E+05	5.13E+05
ZERO CROSSING PERIOD OF FORCES	S	7.45E+00	7.03E+00	6.68E+00	6.39E+00
SIGNIFICANT DRIFT FORCE AMPLITUDE	m	1.92E+06	1.53E+06	1.25E+06	1.03E+06
THREE HOUR MPM LARGEST FORCE	N	3.66E+06	2.93E+06	2.40E+06	1.98E+06
MOTION SPECTRUM					
VARIANCE OF MOTIONS	m**2	1.18	1.42	1.61	1.75
ZERO CROSSING PERIODS	S	14.20	15.36	16.39	17.30
SIGNIFICANT AMPLITUDE	m	2.36	2.84	3.23	3.57
THREE HOUR MPM LARGEST	m	4.30	5.14	5.82	6.31
SPECTRAL MOMENTS M0		1.39	2.02	2.61	3.09
M1		0.60	0.81	0.99	1.11
M2		0.27	0.34	0.38	0.48
FORCE SPECTRUM					
VARIANCE OF FORCES	N**2	2.20E+07	2.29E+07	2.26E+07	2.19E+07
ZERO CROSSING PERIODS	S	1.07E+01	1.17E+01	1.26E+01	1.34E+01
SIGNIFICANT AMPLITUDE	N	4.53E+07	4.58E+07	4.52E+07	4.38E+07
THREE HOUR MPM LARGEST	N	8.43E+07	8.47E+07	8.31E+07	8.01E+07
SPECTRAL MOMENTS M0		5.14E+14	5.26E+14	5.12E+14	4.80E+14
M1		2.86E+14	2.65E+14	2.39E+14	2.10E+14
M2		1.76E+14	1.50E+14	1.25E+14	1.04E+14
LIMIT STATE SPECTRUM					
VARIANCE OF LIMIT STATES	(N/mm^2)^2	2.49	2.30	2.48	2.90
ZERO CROSSING PERIODS	S	12.00	12.60	14.10	15.70
SIGNIFICANT AMPLITUDE	N/mm^2	4.99	4.61	4.96	5.80
THREE HOUR MPM LARGEST	N/mm^2	9.20	8.47	9.04	10.40
SPECTRAL MOMENTS M0		6.22	5.32	6.15	8.43
M1		8.21	2.56	2.64	3.24
M2		1.70	1.30	1.22	1.34

Table 6.6 Variation of TLP surge response statistics with zero-crossing period for Hs=12.0m

zero crossing period (s) ---->		11sec	12sec	13sec	14sec
response	unit				
SEA DRIFT FORCE FX	N	8.15E+05	6.20E+05	4.82E+05	3.79E+05
SEA DRIFT FORCE FY	N	0.00E+00	0.00E+00	0.00E+00	0.00E+00
CURRENT FORCE FX	N	2.38E+07	2.38E+07	2.38E+07	2.38E+07
CURRENT FORCE FY	N	0.00E+00	0.00E+00	0.00E+00	0.00E+00
WIND FORCE FX	N	3.60E+06	3.60E+06	3.60E+06	3.60E+06
WIND FORCE FY	N	0.00E+00	0.00E+00	0.00E+00	0.00E+00
RESULTANT FORCE FR	N	2.82E+07	2.80E+07	2.79E+07	2.78E+07
WIND,CURRENT, DRIFT OFFSET DIR.	D	0	0	0	0
INCLINATION OF TENDONS	D	8.882	8.838	8.794	8.794
STRETCH OF THE TENDONS	m	0.057	0.057	0.056	0.056
STATIC OFFSET DISTANCE	m	64.073	63.758	63.444	63.444
STATIC OFFSET SETDOWN	m	4.976	4.927	4.878	4.878
INCREASE IN TENDON STRESSES	N/mm2	28.902	28.614	28.328	28.328
INCREASE IN TENDON TENSION	N	2904.987	2876.027	2847.217	2847.217
VARIANCE OF DRIFT MOTIONS	m**2	10.26	7.72	5.94	4.63
ZERO CROSSING PERIOD OF MOTIONS	S	101.80	101.70	101.60	101.60
SIGNIFICANT DRIFT MOTION AMPLITUDE	m	20.52	15.44	11.87	9.27
THREE HOUR MPM LARGEST MOTION	m	31.34	23.58	18.13	14.15
VARIANCE OF DRIFT FORCES	m**2	1.31E+06	1.04E+06	8.51E+05	6.98E+05
ZERO CROSSING PERIOD OF FORCES	S	7.45E+00	7.03E+00	6.68E+00	6.39E+00
SIGNIFICANT DRIFT FORCE AMPLITUDE	m	2.61E+06	2.08E+06	1.70E+06	1.40E+06
THREE HOUR MPM LARGEST FORCE	N	4.98E+06	3.99E+06	3.27E+06	2.69E+06
MOTION SPECTRUM					
VARIANCE OF MOTIONS	m**2	1.37	1.65	1.88	7.05
ZERO CROSSING PERIODS	S	14.02	15.30	16.39	17.30
SIGNIFICANT AMPLITUDE	m	2.75	3.30	3.77	4.10
THREE HOUR MPM LARGEST	m	5.01	6.00	6.79	7.36
SPECTRAL MOMENTS M0		1.89	2.75	3.55	4.21
M1		0.82	1.10	1.34	1.51
M2		0.37	0.45	0.52	0.55
FORCE SPECTRUM					
VARIANCE OF FORCES	N**2	2.64E+07	2.67E+07	2.64E+07	2.56E+08
ZERO CROSSING PERIODS	S	1.07E+01	1.17E+01	1.26E+01	1.34E+01
SIGNIFICANT AMPLITUDE	N	5.29E+07	5.35E+00	5.28E+07	5.11E+08
THREE HOUR MPM LARGEST	N	9.84E+08	9.88E+07	9.69E+07	9.35E+08
SPECTRAL MOMENTS M0		7.00E+14	7.15E+14	6.97E+14	6.54E+14
M1		3.90E+14	3.61E+14	3.25E+14	2.86E+14
M2		2.40E+14	2.04E+14	1.71E+14	1.42E+14
LIMIT STATE SPECTRUM					
VARIANCE OF LIMIT STATES	(N/mm^2)^2	2.91	2.69	2.89	3.38
ZERO CROSSING PERIODS	S	12.00	12.60	14.00	15.70
SIGNIFICANT AMPLITUDE	N/mm^2	5.82	5.38	5.79	6.77
THREE HOUR MPM LARGEST	N/mm^2	10.74	9.88	10.50	12.24
SPECTRAL MOMENTS M0		8.47	7.24	8.30	11.48
M1		4.37	3.48	3.59	4.41
M2		2.32	1.77	1.66	1.82

Table 6.7 Variation of TLP surge response statistics with zero-crossing period for Hs=14.0m.

zero crossing period (s) ----->		11sec	12sec	13sec	14sec
response	unit				
SEA DRIFT FORCE FX	N	1.06E+06	8.09E+05	6.30E+05	4.95E+05
SEA DRIFT FORCE FY	N	0.00E+00	0.00E+00	0.00E+00	0.00E+00
CURRENT FORCE FX	N	2.38E+07	2.38E+07	2.38E+07	2.38E+07
CURRENT FORCE FY	N	0.00E+00	0.00E+00	0.00E+00	0.00E+00
WIND FORCE FX	N	3.60E+06	3.60E+06	3.60E+06	3.60E+06
WIND FORCE FY	N	0.00E+00	0.00E+00	0.00E+00	0.00E+00
RESULTANT FORCE FR	N	2.85E+07	2.82E+07	2.80E+07	2.79E+07
WIND,CURRENT, DRIFT OFFSET DIR.	D	0	0	0	0
INCLINATION OF TENDONS	D	8.925	8.882	8.838	8.794
STRETCH OF THE TENDONS	m	0.058	0.057	0.057	0.056
STATIC OFFSET DISTANCE	m	64.387	64.073	63.758	63.444
STATIC OFFSET SETDOWN	m	5.025	4.976	4.927	4.878
INCREASE IN TENDON STRESSES	N/mm2	29.192	28.902	28.614	28.328
INCREASE IN TENDON TENSION	N	2934.096	2904.987	2876.027	2847.217
VARIANCE OF DRIFT MOTIONS	m**2	13.40	10.08	7.75	6.05
ZERO CROSSING PERIOD OF MOTIONS	S	101.80	101.70	101.60	101.60
SIGNIFICANT DRIFT MOTION AMPLITUDE	m	26.80	20.17	15.51	12.10
THREE HOUR MPM LARGEST MOTION	m	40.93	30.80	23.68	18.49
VARIANCE OF DRIFT FORCES	m**2	1.71E+06	1.36E+06	1.11E+06	9.11E+05
ZERO CROSSING PERIOD OF FORCES	S	7.45E+00	7.03E+00	6.68E+00	6.39E+00
SIGNIFICANT DRIFT FORCE AMPLITUDE	m	3.41E+06	2.72E+06	2.22E+06	1.82E+06
THREE HOUR MPM LARGEST FORCE	N	6.51E+06	5.21E+06	4.27E+06	3.51E+06
MOTION SPECTRUM					
VARIANCE OF MOTIONS	m**2	1.57	1.84	2.15	2.34
ZERO CROSSING PERIODS	S	14.02	15.30	16.30	17.30
SIGNIFICANT AMPLITUDE	m	3.14	3.79	4.31	4.69
THREE HOUR MPM LARGEST	m	5.72	6.86	7.76	8.41
SPECTRAL MOMENTS M0		2.47	3.59	4.64	5.50
M1		1.07	1.44	1.75	1.97
M2		0.48	0.60	0.68	0.72
FORCE SPECTRUM					
VARIANCE OF FORCES	N**2	3.02E+07	3.05E+07	3.01E+07	2.90E+07
ZERO CROSSING PERIODS	S	10.70	11.70	12.68	13.46
SIGNIFICANT AMPLITUDE	N	6.05E+07	6.48E+07	6.03E+07	5.80E+07
THREE HOUR MPM LARGEST	N	1.12E+08	1.13E+08	1.11E+08	1.07E+08
SPECTRAL MOMENTS M0		9.15E+14	9.35E+14	9.10E+14	8.54E+14
M1		5.09E+14	4.72E+14	4.25E+14	3.74E+14
M2		3.14E+14	2.66E+14	2.23E+14	1.86E+14
LIMIT STATE SPECTRUM					
VARIANCE OF LIMIT STATES	(N/mm^2)^2	3.31	3.07	3.31	3.87
ZERO CROSSING PERIODS	S	12.00	12.69	14.08	15.70
SIGNIFICANT AMPLITUDE	N/mm^2	6.65	6.15	6.61	7.73
THREE HOUR MPM LARGEST	N/mm^2	12.20	11.30	12.02	13.90
SPECTRAL MOMENTS M0		11.07	9.46	10.95	14.90
M1		5.70	4.55	4.71	5.76
M2		3.03	2.32	2.17	2.38

Table 6.8 Variation of TLP surge response statistics with zero-crossing period for Hs=16.0m.

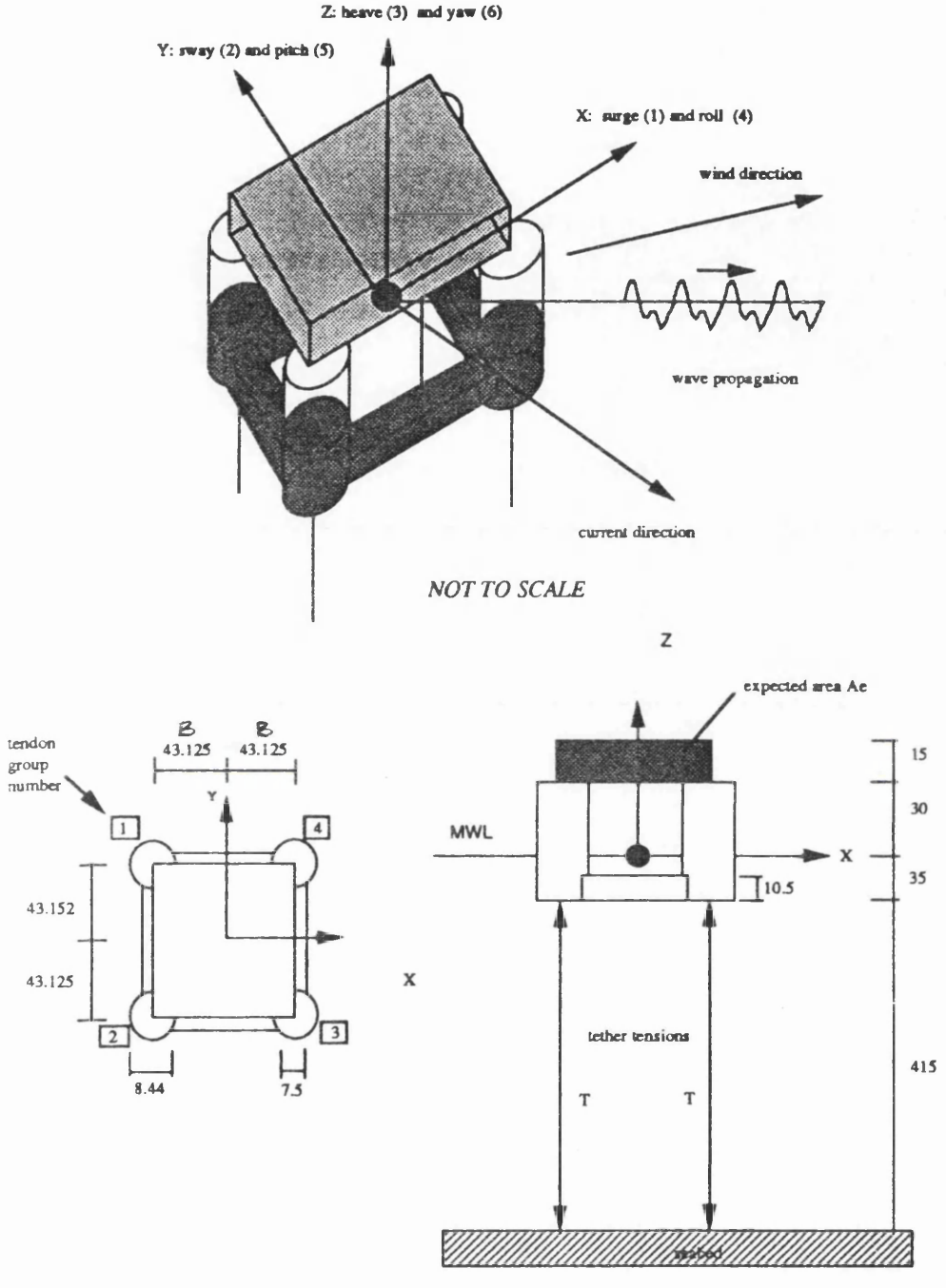


Figure 6.1 General arrangement of the tension leg platform and the load / structure reference axes. All dimensions are in metres.

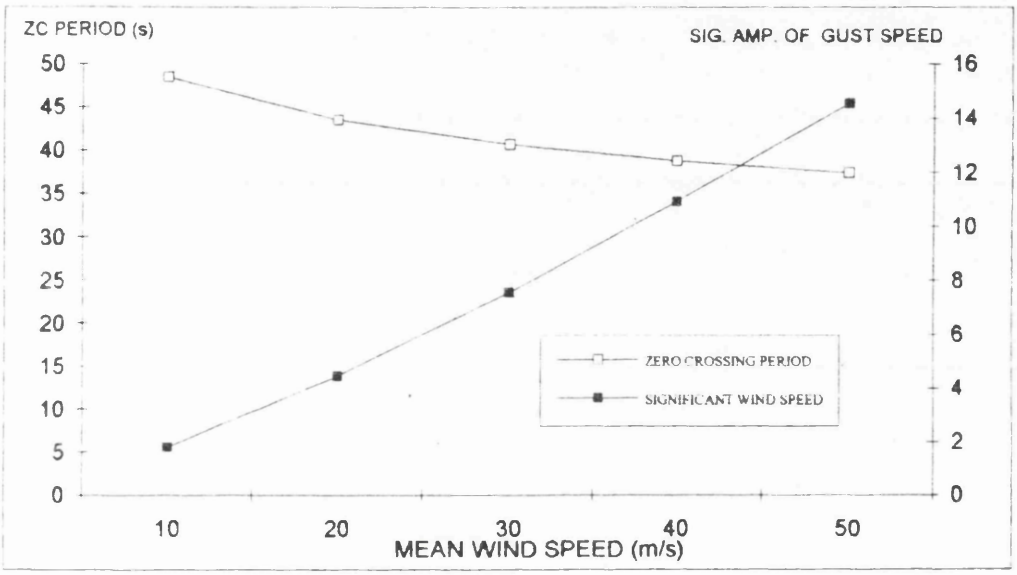


Figure 6.1(a) Significant amplitude and zero crossing periods of the wind gust velocity obtained from the Ochi spectrum for a range of wind speeds.

current	normalised depth
1.3	1
0.85	0.95
0.85	0.9
0.85	0.8
0.85	0.5
0.61	0.05
0.54	0.02
0.49	0.01

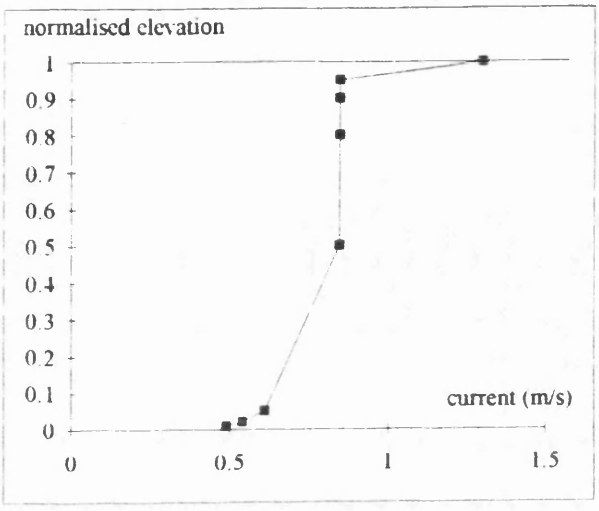


Figure 6.2 Department of Energy guidance notes current profile.

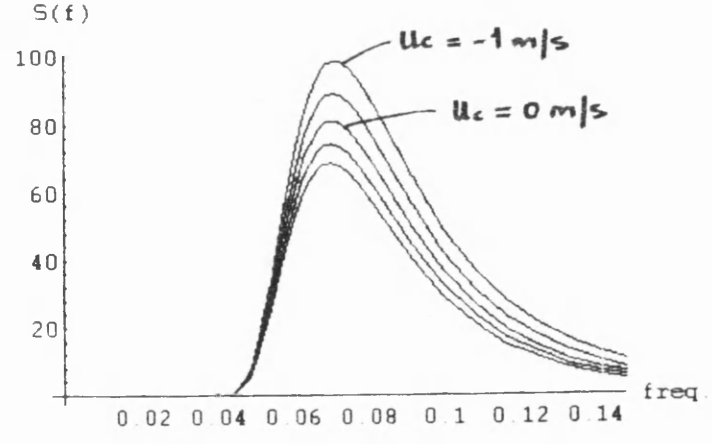


Figure 6.3 The effect of current on modifying a PM spectrum. The upper curve is with -1.0 m/s current and the ones below are for currents -0.5, 0.0, 0.5, and 1.0.

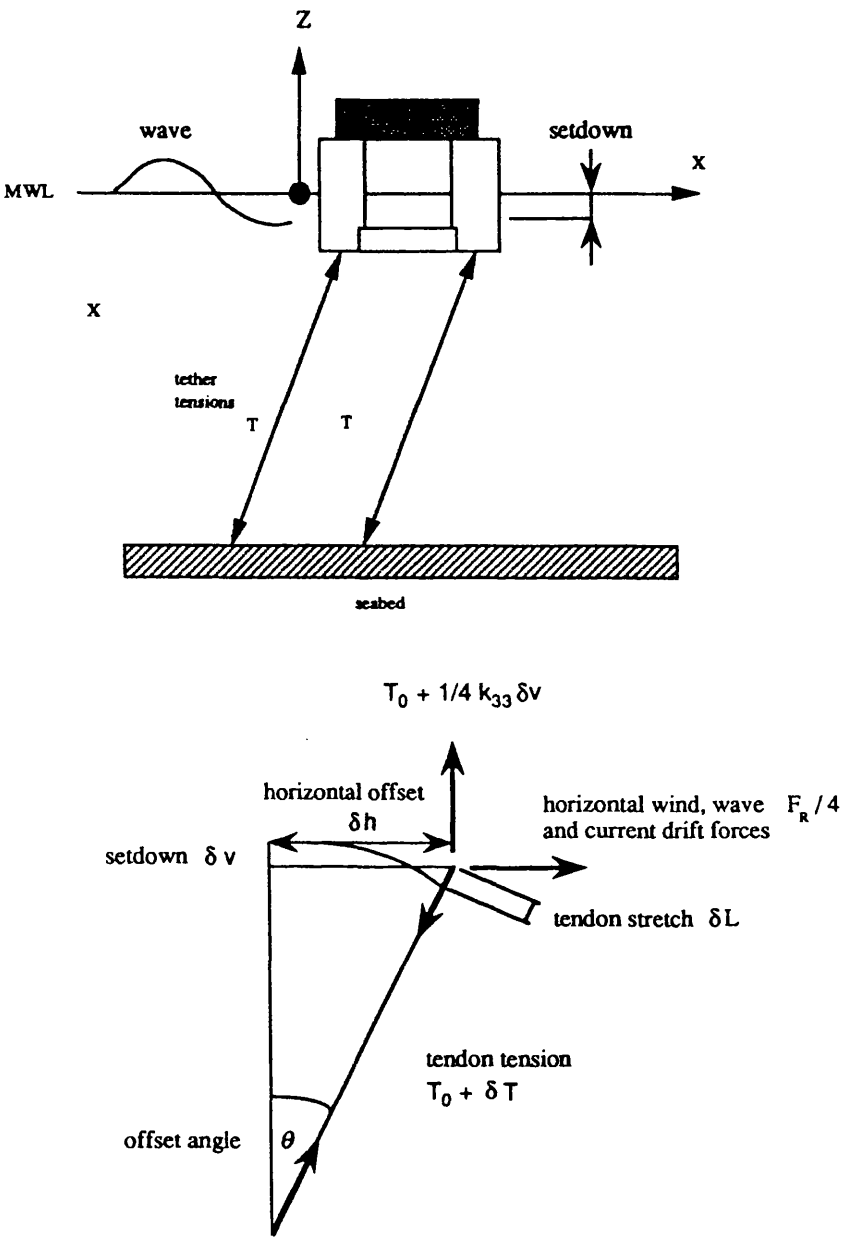
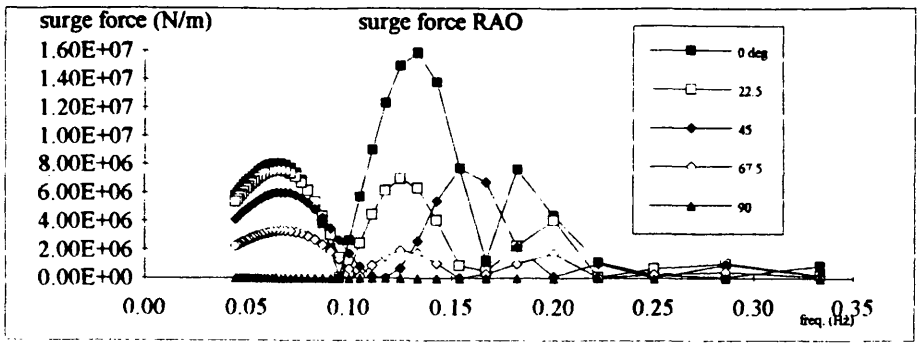
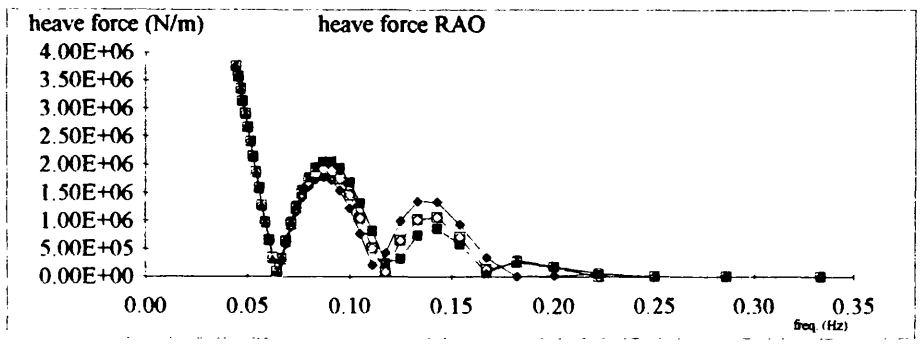
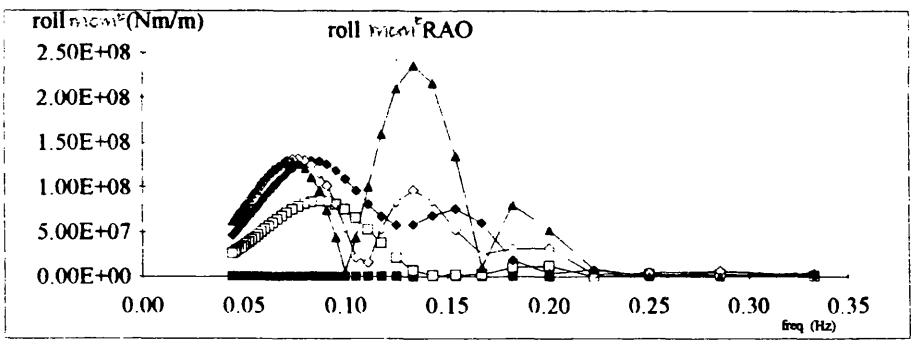


Figure 6.4 Tendon force variation due to the TLP offsets.







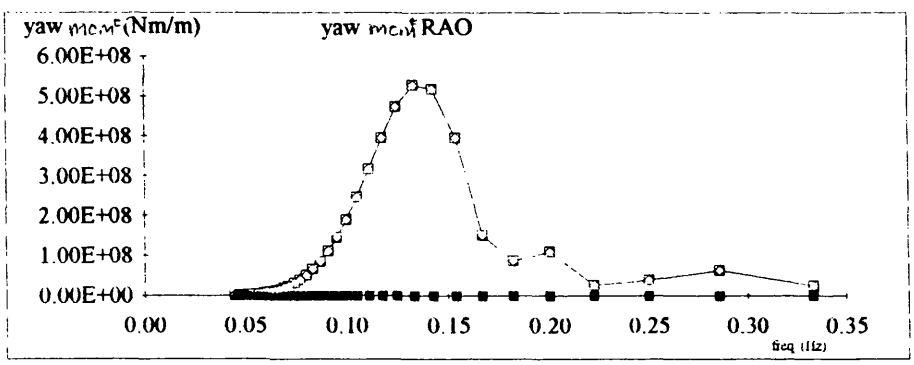
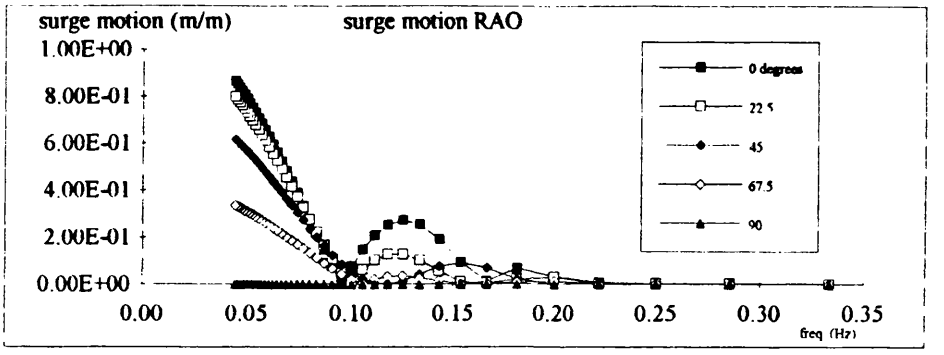
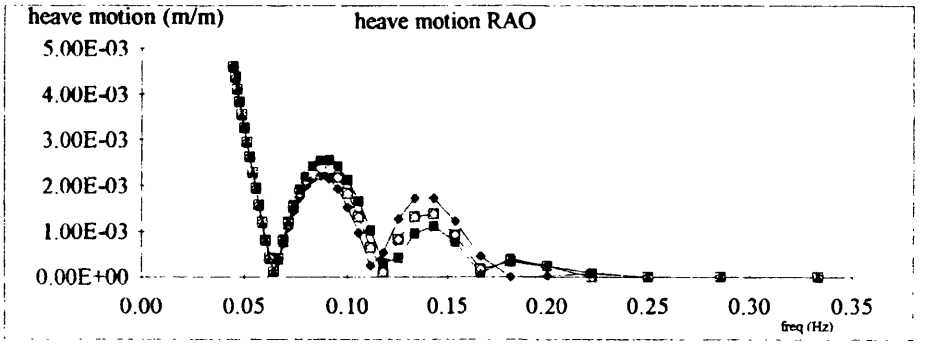
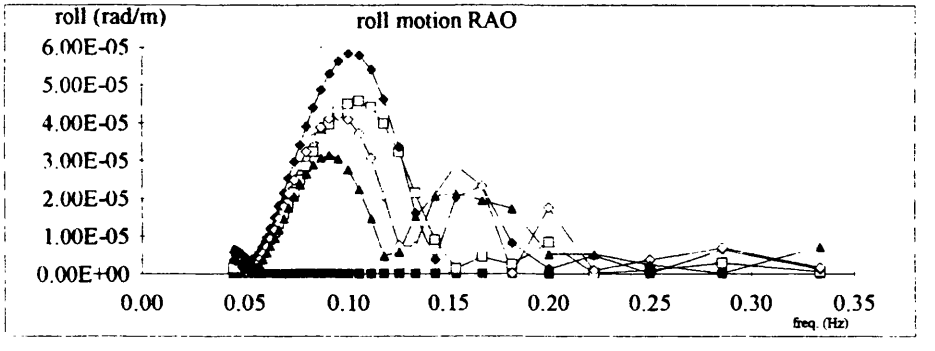


Figure 6.5 Force response amplitude transfer functions for the TLP for a range of wave headings from 0 to 90 degrees.







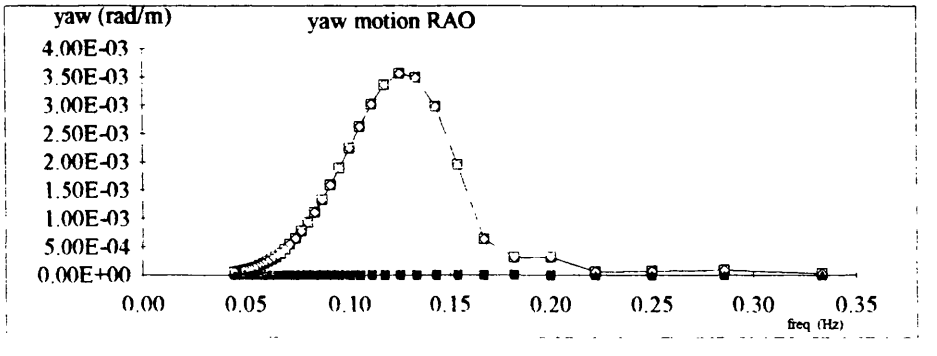
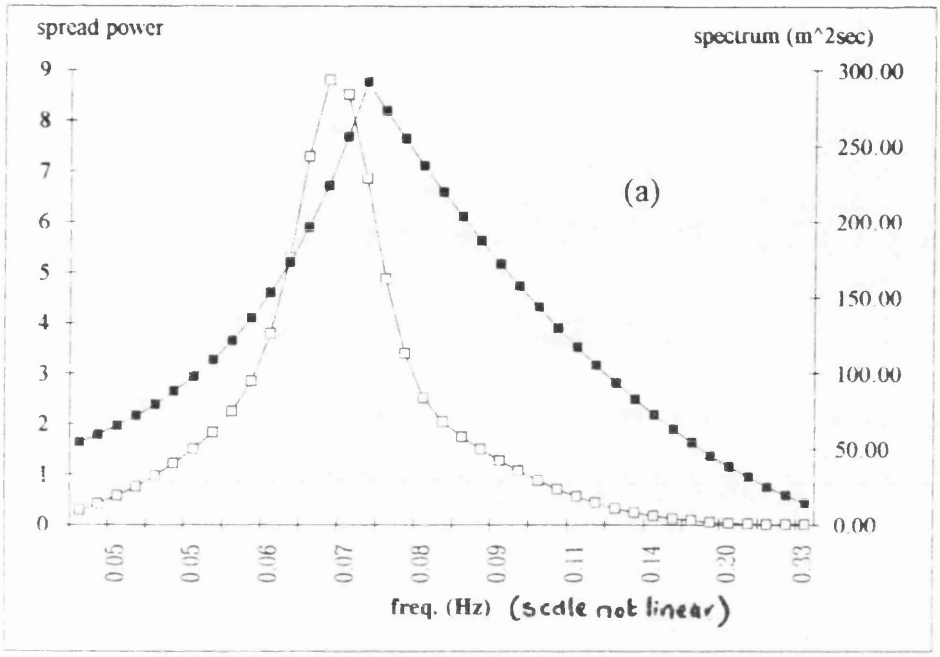


Figure 6.6 Motion response amplitude transfer functions for the TLP for a range of wave headings from 0 to 90 degrees



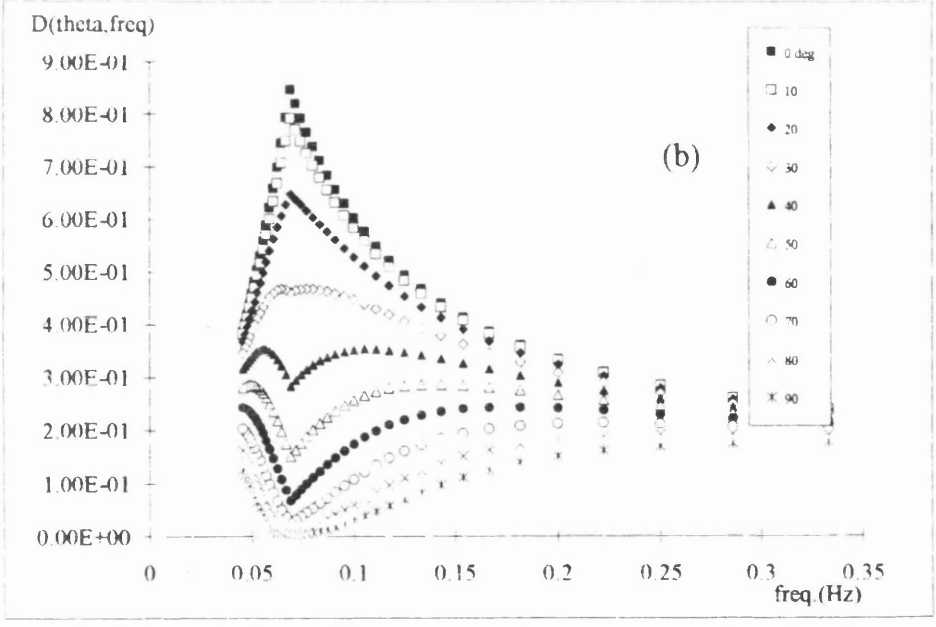


Figure 6.7 (a) Example spectrum and spreading powers for a range of frequencies.
(b) directional distribution for a range of frequencies and directions.

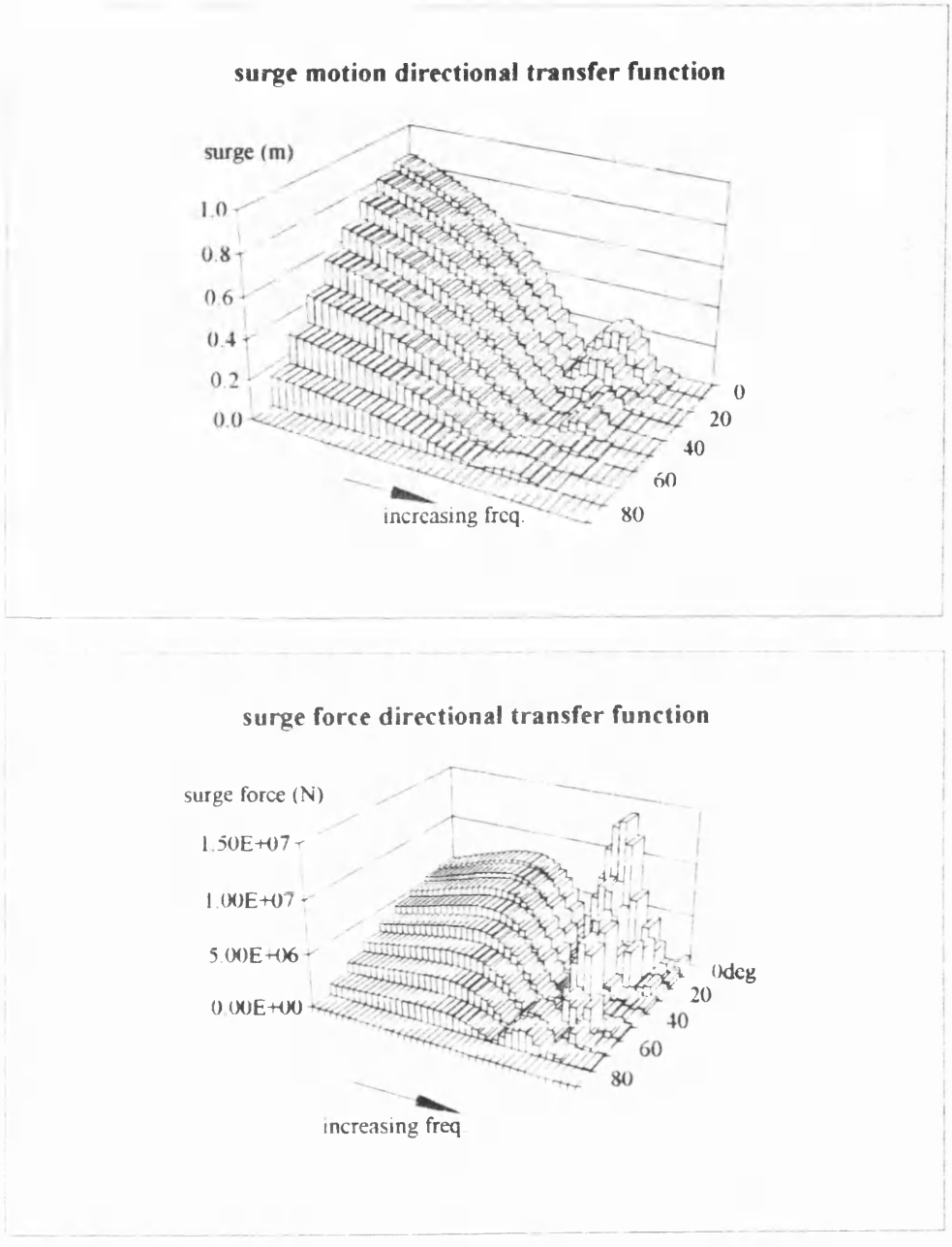


Figure 6.8 Directional surge force and motion transfer functions for the TLP.

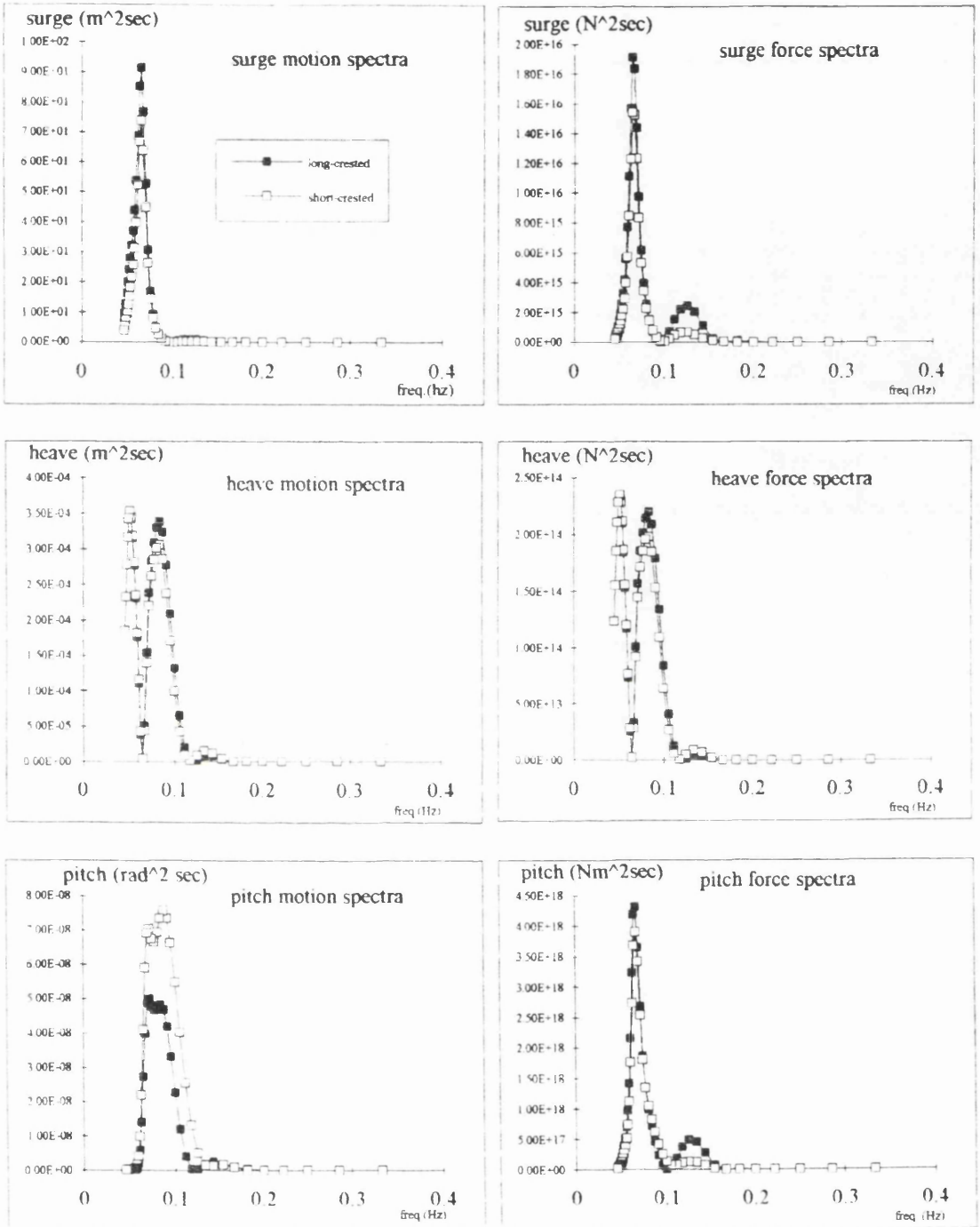
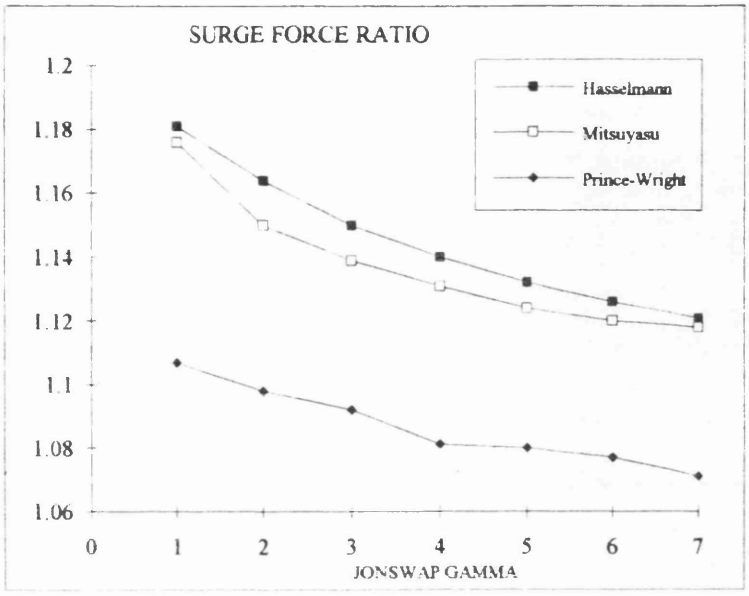


Figure 6.9 TLP response and force spectra for both uni and omni-directional seas with an Hs = 10.0m, Tz = 12.0 and Hasselmann directional spreading function.



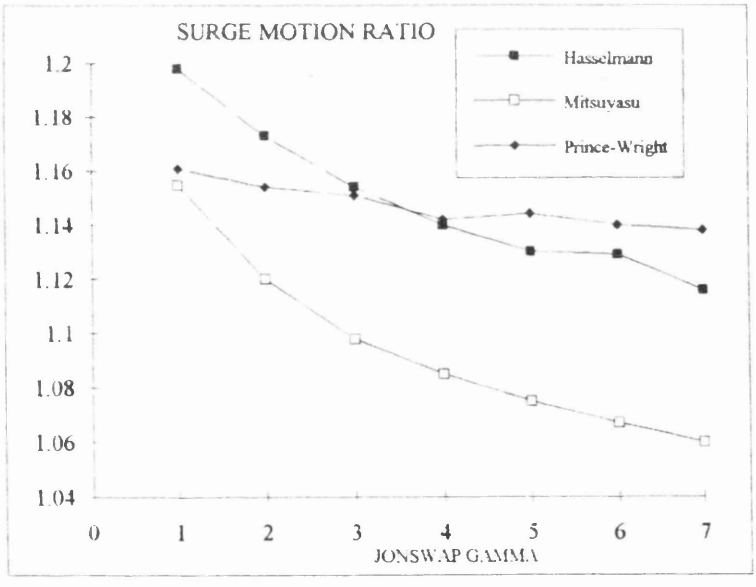


Figure 6.10 Ratio of TLP surge responses in long crested and short crested head seas.
(JONSWAP spectrum, Hs = 10.0 m, Tz = 12.0 sec.)

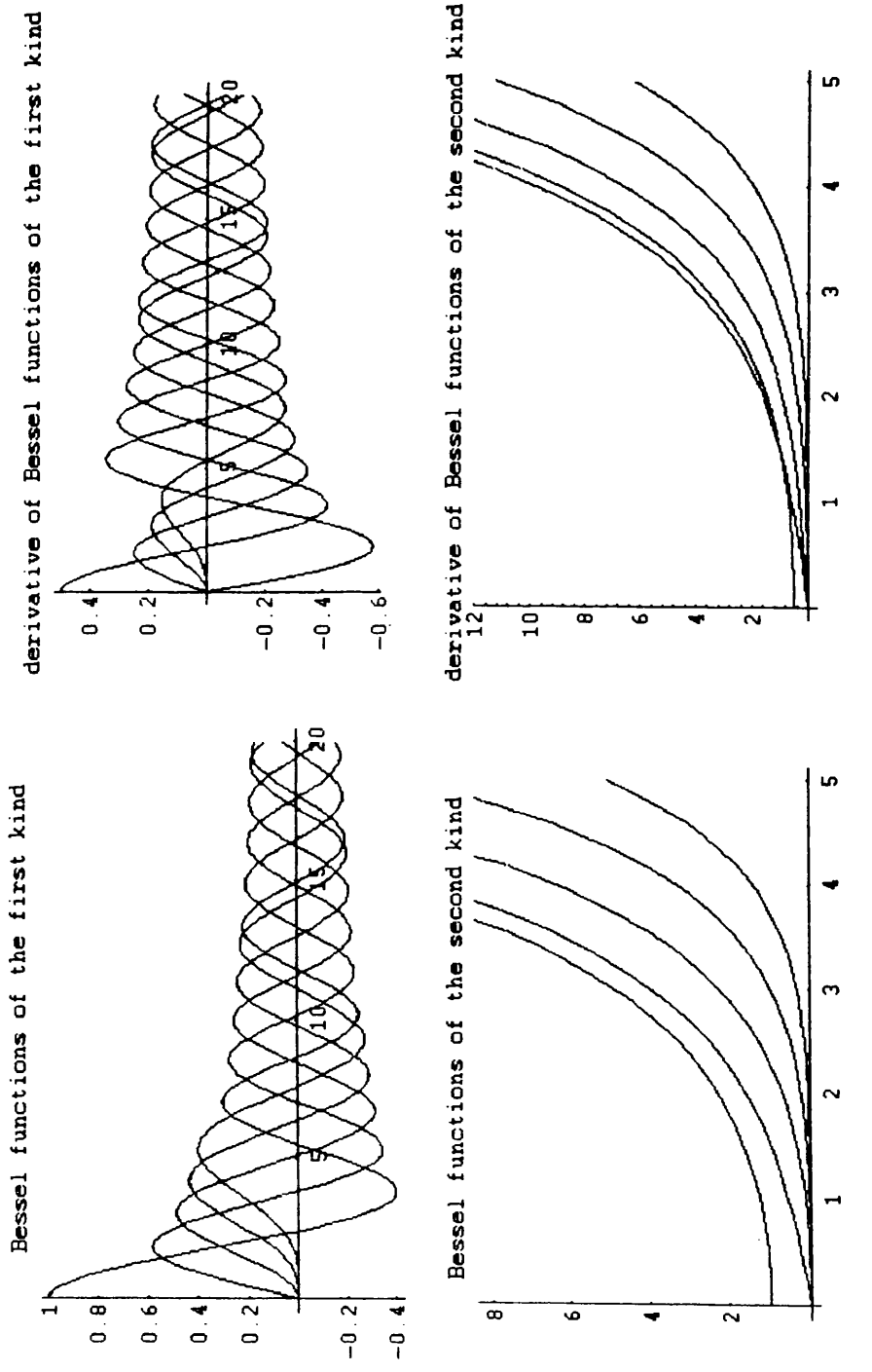


Figure 6.11 Bessel functions of the first and second kind and their derivatives for orders 0 to 4.

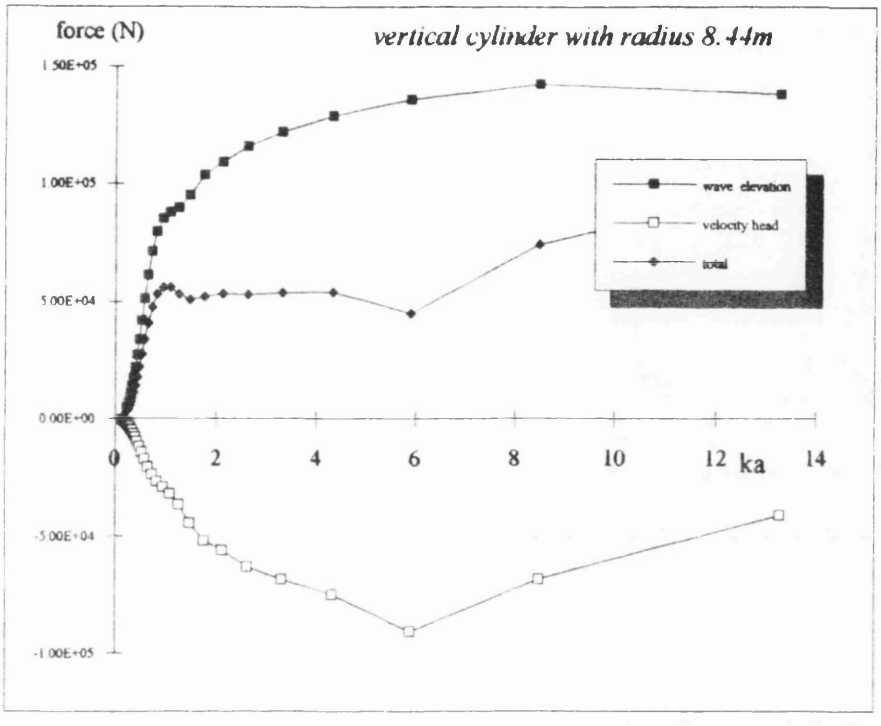


Figure 6.12 Potential drift force on a column of the TLP given by Havelock's closed form solution for the incident wave and diffracted potential.

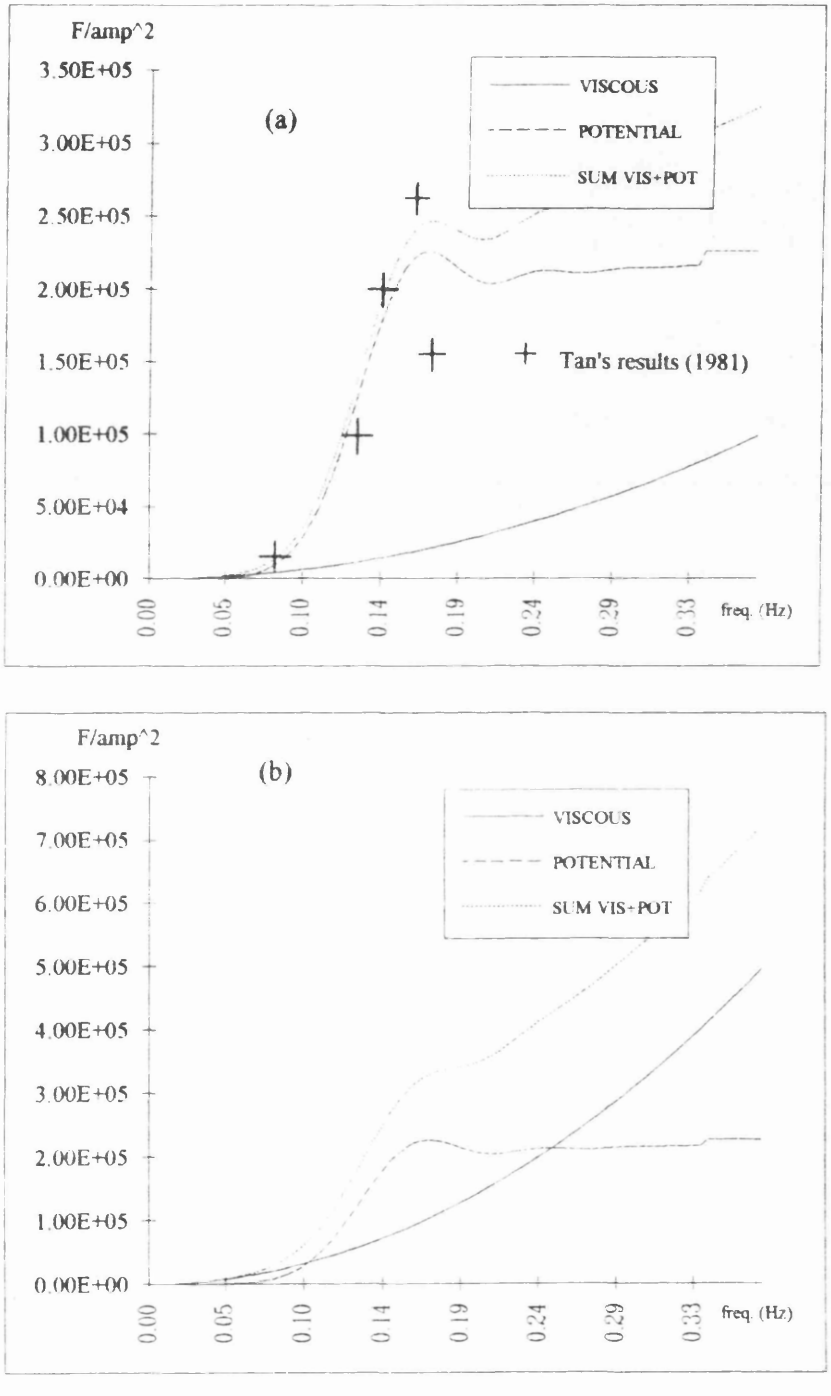


Figure 6.13 Comparison of normalised forces on the TLP: (a) transfer functions were generated using a unit amplitude wave; (b) transfer functions generated using wave amplitude equal to the significant amplitude of the seastate.

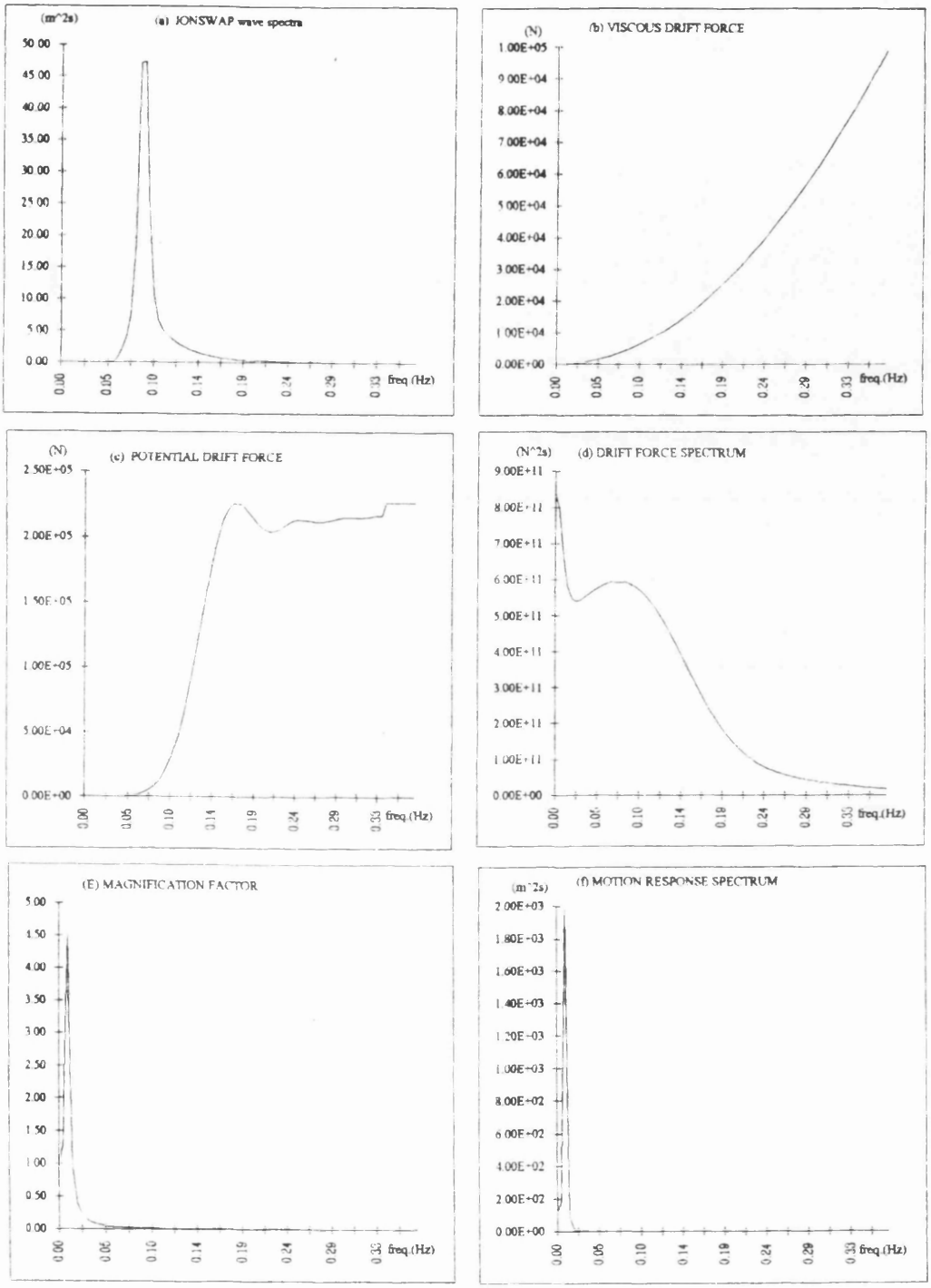


Figure 6.14 Results from calculation of the second order wave drift forces on the TLP. (a) JONSWAP wave elevation variance spectrum with Hs=10.0m, Tz=12.0s, gamma=3.3; (b) viscous drift forces; (c) potential drift forces; (d) drift force spectrum; (e) single degree of freedom system magnification factor; (f) drift response spectrum.

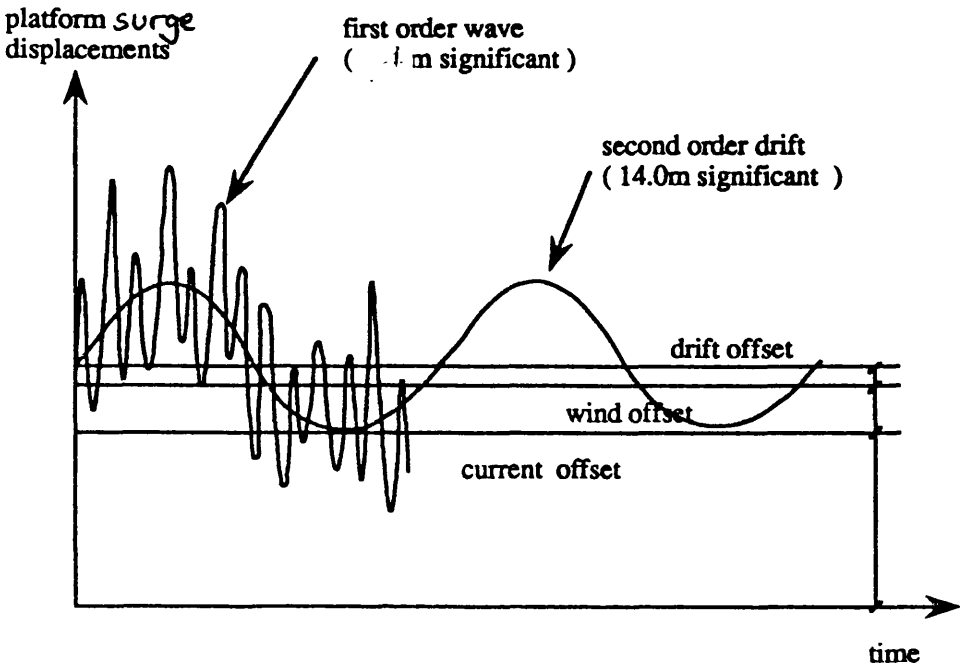
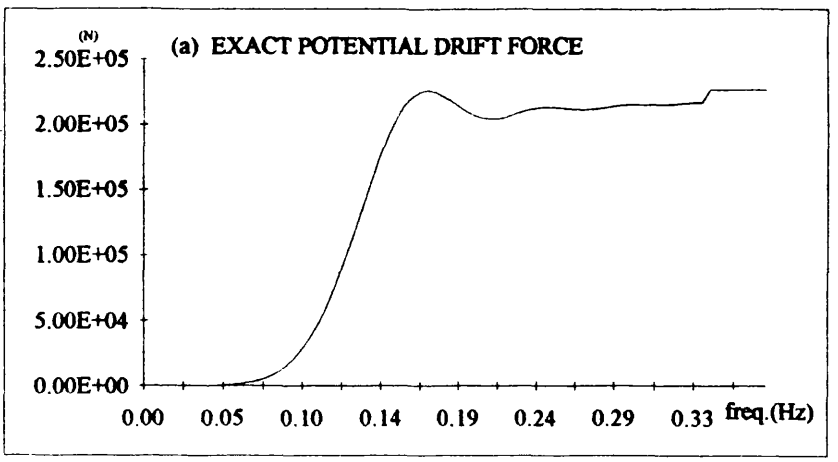
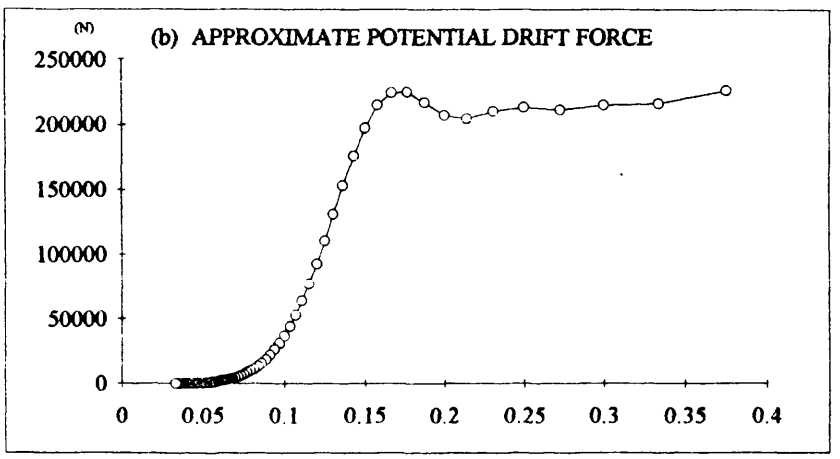


Figure 6.15 Platform offsets due to winds, waves, and currents.





standardised force =
-0.0011977359711557 65*x - 0.0195425307685574*x**2 +
- 1.656984611743997*x**3 - 9.18374474281434*x**4 +
- 45.66562370545643*x**5 - 134.4366911641065*x**6 +
- 243.4742889529963*x**7 - 298.3043856709336*x**8 +
- 264.1103792591544*x**9 - 176.11587981946*x**10 +
- 90.6548614002557*x**11 - 36.48534635955184*x**12 +
- 11.52441097793968*x**13 - 2.845719309265964*x**14 +
- 0.5429987974379911*x**15 -
- 0.07840355593997139*x**16 +
- 0.0082736751859680 5*x**17 -
- 0.0006013728404993089*x**18 +
- 0.0000268883119520 3449*x**19 -
- 5.57029105777337e-7*x**20

where x = wave number * cylinder radius, and the standardisation is
F/(rho * g * h^2 a)

Figure 6.16 Comparison of exact solution and polynomial approximation for Havelock's potential steady drift forces on a vertical cylinder: (a) is the exact; (b) is the polynomial approximation.

Chapter 7

RELIABILITY ANALYSIS OF A TLP

NOMENCLATURE

A	*matrix of eigenvectors for principal components see Section 4.5*
A	*cross - sectional area of a tendon*
a	*amplitude*
${}_{W}c_{w}$	*Weibull model shape parameter*
dx	$= dx_1 \ldots dx_p$
${}_{C}D_{c}$	*current direction*
${}_{W}D_{w}$	*wind direction*
E	*Young's modulus*
$E[.]$	*expected value*
$f(x)$	*joint probability density for the loads and strength*
F	*failure domain*
$F(x)$	*cumulative distribution function:* $Pr(X \leq x)$
f	*frequency in hertz*
$G(x)$	*limit state function in the basic variable space*
$G(u)$	*limit state function in the standard normal space*
H_s	*significant wave height*
$h(.)$	*auxillary limit state function*
k	*number of limit states*
$L(x_L)$	*loading: function of random variables* x_L
M	*number of current directions*
m_i	*ith spectral moment*
N	*number of Monte Carlo samples OR numberof directions.*
$N(u;0,I)$	*multivariate standard normal density*
N	*number of waves in a seastate*
p	*number of random variables in the analysis*
P_f^*	*target failure probability*
P_F	*failure probability for any one three hour period*
P_{FL}	*long – term failure probability*
$P_f(x)$	*short - term failure probability conditioned on the seastate*
$R(x_R)$	*resistance: function of random variables* x_R
S	*safe domain*
$S(f)$	*spectral density*
T_s	*sample vector*

$T^{-1}(.)$	*inverse mapping function from* $\boldsymbol{U} \rightarrow \boldsymbol{X}$
T_D	*design life of the structure*
$\boldsymbol{u}$	*standard normal vector*
U_c	*number of samples*
U_w	*random variable*
$v[.]$	*variance*
W_{ij}	*directional weighting function for winds and currents*
$\forall x$	*all x defined by some domain*
σ_y	*yield stress*
δ	*geometric imperfections*
$\Phi(.)$	*standard normal CDF*
$\Phi^{-1}(.)$	*inverse standard normal CDF*
$\varphi(.)$	*standard normal PDF*
λ	*Box & Cox model shape parameter*
ξ	*Box & Cox model location parameter OR standardised variate*
ξ_T	*standardised threshold level for tendon tension*
ξ_C	*standardised threshold level for tendon compression*
ξ_w	*Weibull model location parameter*
α_w	*Weibull model scale parameter*
α	*Box & Cox model scale parameter OR regularity factor*
β	*radius of balloon that just touches the failure surface in the transformed normal space*
Σ_z	*covariance matrix for z variate: see Section 4.5*
Σ_L	*covariance matrix for loading random variables in y - space*
Σ_R	*covariance matrix for strength random variables in y - space*
μ_y	*mean vector for y - space variate: see Section 4.5*
μ_y	*mean value for y*
μ	*mean vector for all y - space random variables*
μ_L	*mean vector for loading variables in y - space*
μ_R	*mean vector for resistance variables in y - space*
$\delta T(t;x)$	*time - varying change in tendon tension conditioned on event x*
$\overline{\delta T}(x)$	*steady change in tendon tension conditioned on event x*

7. INTRODUCTORY REMARKS: RELIABILITY ANALYSIS OF A TLP

The large number of loading and strength random variables used in the design of an offshore structure makes it difficult to use reliability methods in a routine way. The two major problems are: the paucity of suitable probabilistic models; and the work required to implement a generalised algorithm for modelling correlated random variables (using for example the Rosenblatt method outlined in Appendix C: Ch. 4). This chapter outlines a totally new methodology for calculating structural reliability using conventional time-variant reliability methods. The basis of the approach developed out of the work in Ch. 4 in which multivariate modelling, using maximum likelihood, was applied to the DB1 data.

Structural reliability calculations are usually formulated in terms of loading and resistance random variables. The loading on an offshore structure is a function of many variables like significant wave height, wind speed etc., and the resistance is a function of the material properties, strength model and so on. Each random variable can be defined either probabilistically or deterministically in the reliability analysis. In this chapter we are primarily concerned with modelling the environmental loading parameters and use is made of the model developed from the DB1 data. Nonetheless, it is important to realise that the modelling methodology is general and so two examples of how a resistance model can be developed using the new method are included. This demonstrates how the available strength information, for examples see Smith et al (1987), can be used to formulate the required models.

Models for environmental variables are difficult to synthesize because failure is generally caused by one or a combination of correlated extremes in the wave, wind or current process. Unlike the strength models, which are primarily independent descriptive models, we require

predictive models for the environment which are accurate at very high quantile levels. The problem is further compounded by the site dependence of the environmental variables.

The reliability analysis discussed below uses a model which has eight random variables, seven of which describe the environment, and one for the strength. The variables are:

significant wave height	Hs
wave zero-up-crossing period	Tz
wind speed	Uw
wind direction	Dw
current speed	Uc
current direction	Dc
spectral shape parameter	γ
tendon material yield strength	σ_y

Variables like tide, storm surge, pretension, tether geometry, material characteristics, fluid loading coefficients, wave theory and many others could be added for a more complete analysis. However, this would increase the computational run time and make it more difficult to interpret the results. Lotsberg (1991) includes the effect of variables not modelled by using a normally distributed 'response uncertainty' with a coefficient of variation of 8 per cent. However, the merits of this are not clear and the COV seems rather low.

7.1 GENERAL OUTLINE OF RELIABILITY THEORY

This chapter reports on the results from a reliability analysis of the tension leg platform described in Ch. 6. The joint probabilities model is based on the work described in Ch.2, 3, 4 & 5 in which statistical inference was used to model both the wind, seastate, and current magnitudes and directions. A First Order Reliability (FOR) analysis methodology as been adopted for the calculation of the failure probabilities to illustrate the real benefit of a transformed normal multivariate model for the environmental random variables (see Ch.4). However, a brief description of a Monte Carlo analysis methodology is given below since it provides the most tangible introduction to the calculation of structural reliability.

The procedure for calculating structural reliability can be most readily understood by considering what is actually happening to a real structure, which is one from the total population of 'equivalent' structures. If we accept the discretisation of the long-term variation of the environment into independent events of finite duration, during which the winds, waves, and current are treated as stationary stochastic processes then we can examine the behaviour of

the structure for every seastate it encounters during its design life. The total failure probability P_F during any one event encountered by the structure is then given by

$$P_F = \int \ldots \int_{\forall x} P_f(x) f(x) dx \qquad [7.1]$$

in which x are the random variables governing the behaviour of the system, and $\forall x$ denotes all x. This is most easily interpreted in its discrete form as the sum of all seastate failure probabilities $P_f(x_i)$ weighted by the probability of occurrence of the seastate $f(x_i)\Delta x$. The process is illustrated in Fig. 7.1. In this form, it is apparent there are two kernel problems. The first is evaluation of a suitable joint density function for both the loading random variables $\{H_s, T_z, U_w \ldots . etc\}$ and the resistance $\{\sigma_y, E, \delta, \ldots . etc\}$. The second is calculation of the within seastate failure probability $P_f(x)$ for the event x. The central tenet of this thesis is to define a methodology for estimating $f(x)$ using measured data. The problem has been solved using the main principles of classical multivariate, and directional analysis, as described in Ch. 4 and 5.

Solution of the integral Equation [7.1] is simple in principle, however for a TLP the calculation of $P_f(x)$ is a time consuming process since it requires the structural responses be solved for each event x used in the solution. Furthermore, the calculation becomes more time consuming as the number of variables increases. The most direct way to solve the integral would be to use crude Monte Carlo methods to sample the joint density $f(x)$ a large number of times and then use the 'hit-and-miss' approach to estimate the integrand, Rubinstein (1981). This is not a feasible approach for problems involving large numbers of dimensions since the number of points required for a specific level of accuracy increases as N^p, where N is the number of sampling points required for each of the p variables.

In the crude Monte Carlo method described above the loading and strength random variables are sampled a large number of times: a separate response analysis is then performed for each sampled vector x_i to create a sample of response 'observations' which can be used to establish the statistics of the long term responses. In the past, the method was generally regarded as computationally too expensive for all but the simplest problems, Thoft-Christensen & Baker (1982). However, this author shares Shinozuka's (1989) opinion that the more sophisticated Monte Carlo methods, Rubinstein (1981), are a valuable aid in reliability analysis and that use of adaptive Monte Carlo methods on modern computers provides a realistic design tool. This opinion is shared by Ang et al (1992) who describe a method (which

uses the multivariate kernel theory described in Ch. 4) to improve the "importance sampling " Monte Carlo technique.

7.1.1 THE TIME INVARIANT METHOD

A less expensive estimate of structural reliability can be obtained using the so called *time invariant* Level III methods, Madsen et al (1986), in which the joint distribution of the loading and strength parameters must be known. The generic approach requires the definition of limit state functions which specify the margins between the demand on a component of the structure and its capacity. Clearly, when demand exceeds capacity the system is judged to be in the failure set F, rather than the safe set S. Denoting all random variables as $\boldsymbol{X}$ then the resistance random variables are $\boldsymbol{X}_R \in \boldsymbol{X}$ and the loading random variables are $\boldsymbol{X}_L \in \boldsymbol{X}$. The margin, or limit state function, can then be defined as

$$G(\boldsymbol{x}) = R(\boldsymbol{x}_R) - L(\boldsymbol{x}_L) \qquad [7.2]$$

When capacity exceeds demand $G_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0$ and the system is in the safe domain, in shorthand $\{G_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0\} \in S$ and the complementary state is then $\{G_{\boldsymbol{X}}(\boldsymbol{x}) < 0\} \in F$. This is illustrated graphically in Fig. 7.2 (top) which shows a system governed by two limit state functions, G_1 and G_2, which are functions of the random load x_1 and resistance variable x_2. The notional contours of the joint density function of $f(\boldsymbol{x})$ are also shown. The solution of the failure probability can now be rewritten as a p-dimensional integral over the failure domain, Melchers (1991)

$$P_F = \int \ldots \int_F f(\boldsymbol{x}) d\boldsymbol{x} \qquad [7.3]$$

where the domain of integration in X-space F is

$$\bigcup_{i=1}^{k} G_i(\boldsymbol{x}) < 0$$

Here, k is the number of limit states, $d\boldsymbol{x} = dx_1.dx_2.....dx_p$, and the failure domain is defined by the union of the limit states G_i. The evaluation of this integral using multidimensional quadratures or Simpson's method is only feasible for low numbers of random variables ($p < 4$). Consequently, a procedure has been developed to approximate P_F called FORM analysis (first order reliability method). In this method, the random variables $\boldsymbol{X}$ must be transformed into uncorrelated normal variables in $\boldsymbol{U}$-space using for example Rosenblatt's method, Appendix C, Ch. 4.

The mapping of $X \rightarrow U$ results in the new integral

$$P_F = \int \ldots \int_F N(u:0,I)du \qquad [7.4]$$

where the domain of integration in U-space is

$$\bigcup_{i=1}^{k} G_i(u) < 0$$

and $N(u;0,I)$ is the p-dimensional standard normal distribution illustrated in Fig. 7.2 (bottom). The approximation of this integral is remarkably simple. If we place a spherical balloon at the origin of the u-space and then 'blow it up' at some point it will just touch one or more failure surfaces. The radius of this balloon is usually denoted by β, and the point at which it touches the failure surface is called the design point (which is also the most likely failure point u^*). The failure probability is then estimated using an important property of the multivariate standard normal, that is the volume contained within the p-dimensional hypersphere (or circle in the case shown in Fig. 7.2) is given very simply by $\Phi^{-1}(\beta)$, where Φ is the univariate standard normal cumulative distribution. The whole problem of calculating failure probabilities using this approach is then reduced to a multidimensional minimisation of

$$\beta^2 = u^T.u \qquad [7.5]$$

subject to the constraint $G(u) = 0$. This method is an approximation since the failure function is assumed to be a hyperplane in the u-space, which is not the case. However, Madsen et al (1986) found the results for $P_F = \Phi^{-1}(\beta)$ are generally within a factor of 2 to 5 times the actual value, which is sufficiently accurate for comparative reliability studies like this one.

Checks can be made on the condition of the failure surface at the design point using the fundamentals of unconstrained extremum theory. The first condition to be satisfied is the gradients must all be zero

$$\frac{\partial \beta}{\partial u_1} = \frac{\partial \beta}{\partial u_2} = \ldots\ldots\ldots \frac{\partial \beta}{\partial u_p} = 0.$$

The second condition is the hessian matrix - the symmetric matrix of second derivatives - must be positive definite at u^*. Both conditions are checked internally by the NAg routine used in

this work, however, the hessian matrix is also useful for assessing the curvature of the failure surface at the design point; this gives an indication of the accuracy of the tangent hyperplane assumption which is implicit in the use of $\Phi^{-1}(\beta)$ as an approximation for the survival probability.

7.1.2. A TIME VARIANT SOLUTION

The approximate integration method above assumed the loading was time invariant so that it could be modelled as a random vector process. This method is therefore appropriate for an offshore platform if it is not dynamically sensitive since a time-invariant extreme event approach Equation [7.3] can then be used in which the dominant load process (the waves) is modelled by an extreme value distribution, Turner & Baker (1988). Compliant systems cannot be modelled using this approach because the stochastic nature of the winds and waves, within seastate, must be modelled using the methods given in Ch. 6. This is reflected in the integral Equation [7.1] which is often referred to as a nested reliability integral because the solution for P_f within a seastate can also be obtained using a FOR method. In this work, however, the combination of the wind and wave stochastic process has been simplified by using a simple linear system model for the responses and a 'sum of squares' load combination. This allows one to use the Rice (1944) distribution for the response maxima of each limit state during some event x.

The second level of the integral in Equation [7.1] can also be solved by FOR methods if the integral is first recast in a time invariant format. The method, due to Wen & Chen (1987), requires the unbounded integral be rewritten as a bounded integral with auxiliary limit state function h. This is done by introducing an auxiliary standard normal random variable u_{p+1} such that

$$P_F = \int \ldots \int_{\forall x} P_f(x) f(x) dx = \int \ldots \int_{h(u,u_{p+1})<0} f(u).f(u_{p+1}) du du_{p+1} \qquad [7.6]$$

Here, $f(u)$ is the joint density in U-space, $f(u_{p+1})$ is the standard normal density, and the auxiliary limit state function h is defined as

$$h(u,u_{p+1}) = u_{p+1} - \Phi^{-1}\left[P_f\left(T^{-1}(u)\right)\right] \qquad [7.7]$$

The term Φ^{-1} is the inverse cumulative standard normal, and $P_f\left(T^{-1}(u)\right)$ is the conditional failure probability for the event $x = T^{-1}(u)$. In this case, $T(x)$ is the Box & Cox (1964) transformation followed by the eigenvector transformation, as discussed in Ch. 4. The rhs of Equation [7.6] is now in a form which allows the use of the time invariant integration method discussed above to approximate the long term failure probability. Again, the solution is obtained by minimisation of the Euclidian distance (see Melchers 1987)

$$\beta^2 = u_{p+1}{}^2 + u^T.u \quad subject\ to \quad h\left(u, u_{p+1}\right) = 0 \qquad [7.8]$$

where the objective function is simply

$$\beta^2 = u_{p+1}{}^2 + \sum_{j=1}^{p} u_j{}^2 \qquad [7.9]$$

The constraint - which is a function of the auxiliary variable u_{p+1} and the conditional failure probability $P_f\left(T^{-1}(u)\right)$ - can be used to reduce the problem to an *unconstrained minimisation* giving

$$\beta^2 = \left\{\Phi^{-1}\left[P_f\left(T^{-1}(u)\right)\right]\right\}^2 + \sum_{j=1}^{p} u_j{}^2 \qquad [7.10]$$

The minimisation of β is very simple if a suitable optimisation program is available. For this work the NAg routine E04VDF was used successfully although CPU times were considerable on a 486 PC.

7.1.3 BOUNDS ON THE INVERSE BOX MODEL

Written in its general form the objective function for the distance β appears to be unbounded in u-space, however, in practice the Box transformation introduces a set of indirect, model dependent, bounds in the y-space. Inverting the Box transformation we get

$$x_i = \begin{cases} \left(1 + \lambda_i y_i\right)^{1/\lambda_i} - \xi_i & \lambda_i \neq 0 \\ e^{y_i} - \xi_i & \lambda_i = 0 \end{cases}$$

To obtain a real solution for x_i we must satisfy two sets of inequalities for each variable y_i:

$$y_i > -1/\lambda_i; \quad \lambda_i > 0$$
$$y_i < -1/\lambda_i; \quad \lambda_i < 0$$

These bounds on the vector $\boldsymbol{y}$, in the multivariate normal space, are mapped into $\boldsymbol{u}$-space through the linear transformation

$$\boldsymbol{u} = \Sigma_z^{-1/2} A^T \left(\boldsymbol{y} - \mu_y \right) \qquad [7.11]$$

see Ch. 4, Section [4.5]. This results in a set of simple bounds on $\boldsymbol{u}$ with the sign of the inequality determined by the sign of each λ_i. Of course, if we had used a different transformation then the bounds would not be the same.

7.2 THE LONG-TERM JOINT DENSITY

The dominant load process for fixed offshore structures is normally wave induced and the majority of reliability assessments of offshore platforms use an environmental model of the load process in which the joint density $f(\boldsymbol{x})$ is given by the marginal extreme value distribution of the seastate significant wave height $f(H_s)$, and a series of conditional distributions $f(x_i|H_s), f(x_{i-1}|H_s, x_i)$..... etc. This approach is often called a storm based approach since an extreme value model is usually used for $f(H_s)$. The method therefore assumes an extreme in one of the other processes combined with a moderate significant wave height will not result in an out-crossing of the safe domain S. This is not the case for tension leg platforms in which the winds and tidal elevation play an important role in the platform responses.

The storm based method allows us to use a time independent approach owing to the use of an extreme value distribution. At this point the motivation for using an extreme value distribution should be noted. The integral in Equation [7.1] in fact corresponds to the population model $f(\boldsymbol{x})$, however, the target failure probabilities of offshore structural systems are generally so small ($<10^{-3}$) that only the rare extremes contribute to P_F. Since models fitted to population data are normally biased in the tails, where the most 'interesting' responses are occurring, extreme value, or threshold models are employed. However, this method is not suitable for compliant systems since their response maxima are determined by waves, winds, tide levels etc., and so a different approach is required to allow for the contribution of more than one load process.

The multivariate analysis of the DB1 data, described in Ch. 4, gave a model of the joint distribution for the set $x_L = \{Hs, Tz, Uw, Uc\}$, whilst the directional analysis of the wind and current data in Ch. 5 gave the distributions of the set $x_D = \{D_w, D_c\}$. For this reliability study we add the yield strength of the tendons, $x_R = \{\sigma_y\}$ giving a complete set of 7 random variables $x = \{x_R, x_L, x_D\}$. The long-term failure probability P_F is then given by

$$P_F = \iint_{\forall x} P_f(x) f(x) dx = \iint_{\forall x} p_f(x)^* f(x_R).f(x_L).f(x_{Dw}).f(x_{Dc}) dx \qquad [7.12]$$

Because of the multi-modality of the directional distributions for the winds, and currents, this integral is rewritten as

$$P_F = \int_0^{360} \int_0^{360} f(x_{Dw}).f(x_{Dc}) \left\{ \iint_{\forall x^*} P_f(x) f(x_R) f(x_L) dx \right\} dx_{Dw} dx_{Dc}$$

The integral enclosed by brackets is the time-variant integral discussed above. It gives the failure probability conditioned on the wind and current direction; the total failure probability is then the weighted sum over all combinations of wind and current direction. For simplicity, the wind and current directions can be modelled as fully correlated and uniform on the circle $[0, 2\pi]$. The problem can then be simplified once more by using a discrete summation giving

$$P_F = \frac{1}{N} \sum_{i=1}^{N} \iint_{\forall x^*} P_f(x|x_i).f(x_R) f(x_L) dx$$

This is read as the uniformly weighted sum of the time-variant probabilities for co-linear winds and currents approaching from directions $\theta_i; i = 1, N$. The symmetry of the TLP structure enables this integral to be further simplified and then written as

$$p_F = \frac{4}{N} \sum_{N/4} \iint_{h(u, u_{p+1})<0} f(u^*) f(u_{p+1}) du^* du_{p+1}$$

which requires $N/4$ calculations of the time invariant integral. On the other hand if the relative directionality of the wind and current is to be modelled the single summation must be replaced by a weighted double summation. The weights W_{ij} might be taken as the normalised values in a wind and current direction scatter diagram (Ch. 2 Table 2.13). The long-term failure probability is then given by

$$P_F = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} W_{ij} \iint_{h(u,u_{p+1})<0} f(u^*)f(u_{p+1})du^* du_{p+1}$$

The problem with this summation is that we now require NM reliability optimisations which is computationally expensive.

This chapter examines the reliability of a tension leg platform using a simple set of limit state functions. The DB1 buoy information therefore allows us to model the wind and wave processes as stochastic, however, the currents must be modelled as constant during the three hours. During this event we must calculate the probability that the responses exceed the safety margins, $P_f(x)$. The random load processes must therefore be transformed to response processes for each limit state in order to calculate the failure probability.

7.3 RETURN PERIODS FOR THE ENVIRONMENT AND RESPONSES

If the total failure probability P_F has been calculated the chance of failure in any of the three hour duration events is known. During the life of the structure, T_D (hours), the number of encounters will be $T_D / 3$. Since it is acceptable to assume the events are independent, Bjerager et al (1988), the lifetime failure probability P_{FL} is given by

$$P_{FL} = P_F T_D / 3$$

since we assume only one event can lead to failure during the design life. If the design life is 50 years then $P_{FL} = 146100 P_F$ which means the target failure probability for a single seastate should be of the order of $<10^{-5}$. Alternatively, we can define the life of a structure corresponding to some level of failure probability using $T_D = 3P_{FL}^* / P_F$.

7.4 FORMULATING MODELS FOR THE RELIABILITY ANALYSIS

Transformed normal distributions are required for all of the variables used in this reliability analysis. Providing the mean, variance, and lower bound are known it is possible to synthesise a model using simulation for any variable. Two example are given below for the tendon material yield strength and the JONSWAP shape parameter.

7.4.1 TENDON YIELD STRENGTH MODEL

The yield strength of a TLP's tendon material is generally determined during the design stage after having performed redundancy, strength and fatigue analysis. There seems to be no

conventional wisdom for selecting the material properties, API RP2T (1987) gives no guidance other than the net section stress remains less than 80% of the yield stress F_y or 60% of the ultimate stress F_u, whichever is less. In a recent study by Woo-Sun et al (1991), the mean ultimate strength was taken as 965 N/mm^2 which corresponds to high strength, low alloy steel like AISC A709; 100W. The use of such high strength material is questionable because fatigue damage is proportional to stress range cubed. This suggests a lower grade of steel is more appropriate. Bea et al (1992) for example reports on the 'Methodologies for Comparison of Alternative Production Systems' project undertaken in the USA in which the minimum yield strength was taken as 311 N/mm^2 with a mean value of 345.9 N/mm^2 and standard deviation of 17.4 N/mm^2. This data is consistent with the values used by Lotsberg (1991), and has been used in this work to set the limit states of the tendons (Ch. 6, Section [6.1.1]).

The Weibull distribution is usually found to fit strength data well. Using Bea's (1992) data we have the lower bound (Ch. 3, Section [3.6]) $\xi_w = 311$ N/mm^2, the mean $E[\sigma_y] = 345.9$ N/mm^2 and the variance $v[\sigma_y] = \{302.76$ N/mm$^2\}^2$. These data are sufficient to numerically solve for the shape c_w and scale α_w parameters of a Weibull model, giving

$$\theta^* = \{c_w = 2.175, \alpha_w = 40.537, \xi_w = 311.0\} \qquad [7.13]$$

To use this model in the reliability analysis it must first be converted into a transformed normal variate. This has been done approximately by simulation. A sample of 1000 Weibull random variables with shape, scale and location θ^* were generated by inversion. The simulated sample of yield stresses was then transformed to approximately normal variates using the Box & Cox (1964) method, Ch. 4: Section [4.3.1] . This gave maximum likelihood shape and location parameters $\hat{\lambda} = 1.39$, and $\hat{\xi}$ = -1.39 respectively, and a transformed sample with mean 0.46, and standard deviation 0.03. A comparison of the actual population, and the 'normal' model is shown in Fig. 7.3 which demonstrates the fit is good.

When estimating the transformed normal parameters for the material strength the transformation process failed to converge at the first attempt. Large numbers were identified as the cause which suggests the Box transformation is not scale invariant. The obvious solution was to introduce a scale parameter α in the Box transformation. The value can be chosen arbitrarily to avoiding including it in the likelihood maximisation. However, it should be noted the introduction of a scale parameter α changes the density function. Writing x as the original variate and y as the transformed variate, the model for $f(x)$ becomes

$$f(x)=\frac{\alpha(\alpha x+\xi)^{\lambda-1}}{\sigma_y\sqrt{2\pi}}exp\left\{-\frac{1}{2}\frac{(\alpha x+\xi)^{\lambda}-1-\lambda\mu_y}{\lambda\sigma_y}\right\} \qquad [7.14]$$

where $\{\sigma_y,\mu_y\}$ are the standard deviation and mean value of the normally distributed variate y and $\{\alpha,\xi,\lambda\}$ are the transformation parameters. A sensible value for α is the square root of the maximum likelihood estimate of the sample variance

$$\alpha^2=v[x]=\frac{1}{n}\sum_{i=1}^{n}(x_i-E[x])^2 \qquad [7.15]$$

This re-scales the data x to have unit variance.

7.4.2 SPECTRAL SHAPE PARAMETER

The statistical uncertainty of the JONSWAP spectrum parameters was summarised in Ch. 6, Section [6.1] where the mean value and standard deviation are given as $E(\gamma)$ = 2.65 and $\sigma_\gamma = 1.166$. In addition, we know the lower bound on γ is 1.0 which corresponds to fully developed seas. These three statistics are sufficient to estimate the shape and scale parameters of a Weibull model using the approach given above for the tendon material yield strength. The 'equivalent' Weibull model then has shape, scale and location 1.44, 1.87, 1.0, respectively. Taking these values as the population parameters we can simulate a sample and then find the maximum likelihood Box transformation parameters. The results are:

λ	=	0.14
ξ	=	-0.098
μ_y	=	0.78
σ_γ	=	0.43

The Weibull population and transformed normal models are compared in Fig. 7.4.

7.5 MOTION THRESHOLDS

The lateral motion of the platform is made up of static and dynamic components. Static offset is induced by the wave drift, mean wind, and current forces, and dynamic offset is caused by first order wave, gusting wind, and second order wave effects. A feature of compliant

offshore platforms is the need to restrict the lateral motions to within some threshold. In the case of a TLP the threshold is determined by the design of the tensioning system for the tendons, and risers; and, in the case of a semi-submersible the thresholds are generally determined by the moorings. A further limit on lateral displacements is imposed by the minimum clearance between the crest of an extreme wave, and the underside of the deck. This limit state becomes increasingly important as the water depth increases because the set-down effect increases linearly with water depth.

Without details of the tensioning system design it was not possible to identify the true thresholds on the lateral motions. Furthermore, no details of the still water deck clearance are given in Tan & de Boom (1981). Instead, for this work, an artificial threshold was defined as the sum of the 3 hour most probable maximum displacements due to the first and second order wave forces generated by the 50 year return period seastate (15.5 m) *with zero current and wind speed.* This gave a threshold of 35 m for surge and sway motions which is used in the parametric studies below.

7.6 JOINT DISTRIBUTION OF LOAD AND RESISTANCE VARIABLES

The joint distribution of the loading and resistance random variables $\boldsymbol{x} = \{\boldsymbol{x}_L, \boldsymbol{x}_R\}$ in Equation [7.1] can be formulated using the results in Ch. 4, and the model for the tendon yield strength σ_y. Since the loading and resistance are uncorrelated we can write the covariance matrix for $\boldsymbol{x}$ as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_L & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_R \end{bmatrix}$$

where the transformed variates y_i^* are given by the Box transformation:

$$y_i^* = \left\{ \left(x_i^* + \xi \right)^{\lambda} - 1 \right\} / \lambda \qquad [7.16]$$

The mean vector of the transformed variates is simply

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_L, \boldsymbol{\mu}_R]$$

giving a normal joint density for $\boldsymbol{x}$ in y-space where

$$f(\boldsymbol{y}^*) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{y}^* - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}^* - \boldsymbol{\mu}) \right\} \qquad [7.17]$$

this model is transformed to uncorrelated and standardised variates u^* using the eigenvector transformation in Ch. 4, Section [4.5] where

$$u^* = \left(A^T \Sigma_y A\right)^{-1/2} A^T \left(y^* - \mu_y\right) \qquad [7.18]$$

this gives the standard Normal model, $N(u^*; 0, I)$.

In fact, if we examine the objective function Equation [7.10] and then consider how the optimisation of β is performed, we see the density is never explicitly used. Starting the optimisation at some point in U-space, u_o^*, the unconstrained optimisation algorithm chooses the direction of 'steepest descent' using a *forward difference* estimate of the gradient vector. Therefore, at each iteration we make $2N + 1$ evaluations of the objective function to solve $\Phi^{-1}\left[P_f(u)\right]^2$. The calculation of the failure probability P_f for the current step point u_i, requires a complete frequency domain response analysis for the limit state. This process is repeated until the algorithm has satisfied some convergence criterion on the objective function.

For this work the NAg quadratic programming method was found to give a satisfactory convergence rate usually in less than 6 iterations - that is after making $(\sim 6) * (2N + 1)$ calls to the routine that calculates the responses statistics.

7.7 SHORT-TERM STATISTICS: THE LOAD COMBINATION PROBLEM

The theory of stochastic load combination for the within seastate winds, waves and currents is a complex subject which is made more difficult when the response is non-linear. Madsen et al (1986) summarises the theory of linear stochastic load combination and discusses how problems can be formulated using simple rules for codified design. Non-linear, time-variant, analysis is examined by Wen & Chen (1987); and the International Ship Structures Congress included a special session on stochastic modelling, see Armand et al (1991). Practical applications of the full theory tend to be limited to simple systems since the method is computationally expensive.

For our purposes the load combination problem is made as simple as possible. Each mean force is summed and then used to calculate the static offset response. The first order wave, wave drift, and wind force spectra are then used individually to calculate the response spectra for each limit state function. These response spectra are then summed giving the total response

spectra $S_{T_i}(f)$ for each limit state *i*. These spectra are broad-banded with peaks located at the modal wave frequency, and the structure natural frequency. By assuming the responses are linear broad-band, and Gaussian, the Rice distribution can be used to calculate the distribution of the response maxima. Denoting the normalised response amplitude of each limit state $\xi = a / m_o^{1/2}$, then the distribution of its response maxima is

$$f(\xi|u^*) = \sqrt{1-\alpha^2}\,\varphi\left(\frac{\xi}{\sqrt{1-\alpha^2}}\right) + \alpha\xi e^{-\xi^2/2}\Phi\left(\frac{\alpha\xi}{\sqrt{1-\alpha^2}}\right) \qquad [7.19]$$

Here the regularity factor $\alpha = \sqrt{1-\varepsilon^2}$, where ε is the bandwidth of the response; $\varphi(.)$ is the standard normal density; and $\Phi(.)$ is the cumulative standard normal. The failure probability for the component corresponds to the probability that the limit state function will be negative during the event $u = T(x)$. In Ch. 6 the tendon limit states $G_i(x)$ are defined as the margin between the axial tensile and compressive capacities of the tethers. The compressive capacity is in fact independent of the material properties whereas the tensile capacity is dependent on the yield strength of the tendons. The failure probability is then given by

$$P_f\left(T^{-1}(u)\right) = 1 - F(\xi) = 1 - \Phi\left(\frac{\xi}{\sqrt{1-\alpha^2}}\right) + \alpha\xi e^{-\xi^2/2}\Phi\left(\frac{\alpha\xi}{\sqrt{1-\alpha^2}}\right) \qquad [7.20]$$

Here, the variance of the response m_0, and the regularity factor α are calculated from the response spectrum moments using the relationships:

$$m_n = (2\pi)^{n+1}\int_0^\infty f^n . S(f)\,df \qquad n = 0,1,2,...$$

$$\xi = a / m_o^{1/2}$$

$$\alpha^2 = m_2^2 / (m_o m_4) \qquad [7.21]$$

The thresholds ξ_i are determined for each limit state, for example, the compression limit state gives

$$\xi_c = T_0 + \delta\bar{T}(x) > \delta T(t;x) \qquad [7.22]$$

where T_0 is the tendon pretension, $\delta\bar{T}(x)$ is the mean increase in tendon tension for the seastate $x = T^{-1}(u)$, and $\delta T(t;x)$ is the time varying tendon tension induced by the wind, drift, and wave processes.

For the tension limit state threshold we have

$$\xi_T = A\sigma_y - T_o - \delta\bar{T}(x) > \delta T(t;x) \qquad [7.23]$$

Here: A is the tendon cross sectional area; σ_y is the random yield strength of the tendon material; and the statistics for the amplitudes of the dynamic tension $\delta T(t;x)$ are modelled using the Rice distribution, see Equation [7.19]. The probabilities of not exceeding the tendon tension and compression thresholds ξ_T, ξ_C are then $F(\xi_T)$ and $F(\xi_c)$, respectively, for each encounter with a maximum in the process $\delta T(t;x)$.

During the seastate there will be on average N encounters. We require the probability that all encounters are less than the thresholds ξ_i

$$P_R\left\{\bigcap_{j=1}^{N}\xi_j < \xi_i\right\} \qquad i = 1,2,... \qquad [7.24]$$

By assuming the maxima are all independent, we get

$$Pr\left\{\bigcap_{j=1}^{N}\xi_j < \xi_i\right\} = \prod_{j=1}^{N} Pr(\xi_j < \xi_i) = F^N(\xi_i) \qquad i = 1,2,... \qquad [7.25]$$

The expected number of encounters was estimated using the number of zero-up-crossings for the dynamic load process during the three hour seastate

$$T_Z = 2\pi\sqrt{m_o / m_2}$$

$$N = 3 * 3600 / T_Z \qquad [7.26]$$

The probability of exceeding the threshold ξ_i - which was our failure probability in Equation [7.1] - can then be estimated by

$$P_f\left(x = T^{-1}(u)\right) = 1 - F^N\left(\xi_i\right) \qquad [7.27]$$

This value corresponds to $P_f(x)$ in Equation [7.1] and is evaluated several times during the optimisation for the β distance.

7.8 SENSITIVITY STUDIES

The sensitivity of the within seastate failure probability $P_f(x)$ to changes in the environmental random variables was examined for the tendon stresses and surge motion. In the examples given below the zero-up-crossing period was increased from 6.00 to 11.0 (s) and the significant wave height was taken as the most probable value (conditioned on the T_Z) using

$$H_s = 0.1632 T_Z^2$$

This simplified model was determined by visually fitting a quadratic to the contours of the $\{H_s, T_Z\}$ kernel density plot in Ch. 4. It is not recommended as a general rule.

The effect of changing the spreading model, wind speed, and spectral shape is discussed below.

7.8.1 WINDS AND CURRENTS

The wind and current loads were compared in Ch. 6 where it was shown gusting wind forces are significantly larger than the steady current forces. The influence of wind force on the threshold exceedance probabilities is shown in Fig. 7.5. Note how the higher wind speeds reduce the failure probability for the compression limit state by increasing the mean tendon stress. The effect on the tension limit state is adverse with high wind speeds contributing significantly to the exceedance probabilities.

7.8.2 SPECTRAL SHAPE PARAMETER

The sensitivities of the threshold exceedance probabilities to spectral shape are illustrated by Fig. 7.6. The $\gamma = 1$ case corresponds to a fully developed sea, and the $\gamma = 5$ case corresponds to a fetch limited sea with a narrow-band spectrum. The largest surge responses occur in fully developed seas in which there is a larger amount of energy in the drift force spectrum at low frequencies near the natural frequencies of the structure in surge, sway and yaw.

7.9 DESIGN EVENT SEASTATE RESPONSES

In ch.4 the marginal extreme wind speed, current speed, and significant wave height were estimated using both population and monthly maxima models. As a result the 50 year return period design event environmental variables were defined as:

significant wave height	Hs	15.02	m
wind speed	Uw	33.0	m/s
current speed	Uc	1.0	m/s

These values are close to the Department of Energy (1990a) recommended values. In a design event analysis the winds, waves and currents would be assumed to act colinearly and simultaneously. The range of wave zero-up-crossing period associated with the design event *Hs* is usually determined by the maximum, and minimum wave steepness observed around the British Isles, ie

$$3.2\,(Hs)^{1/2} < Tz < 3.6\,(Hs)^{1/2}$$

For the 50 year return period wave height this gives a range of 12.5 to 14.0 seconds so these values were used as the upper and lower bounds for the Design Event analysis of the TLP model developed in Ch.6.

7.9.1 MOTIONS

The motion and tendon stress responses of the TLP were calculated for colinear winds, waves, and currents using a JONSWAP spectrum with shape factor of 3.3. A zero degree heading was used. The results are summarised in Table 7.1. The time varying response statistics for the combined wind, drift, and first order wave forces were calculated using a 'sum-of-squares' approach (which corresponds to the sum of independent normal time series). In uni-directional waves this gives a combined three hour most probable maximum (mpm) offset of 50.1m in the 12.5 second period sea, and 45.0m in the 14.0 second sea.

7.9.2 TENDON STRESSES

The 3 hour mpm combined tendon stresses have also been calculated for the 12.5 and 14.0 second periods, again in JONSWAP seas with a gamma of 3.3. The stresses for the 12.5 second case have a 3 hr mpm of 101.3 N/mm2 and the stresses for the 14.0 second case were 106.0 N/mm2.

7.10 CALCULATION RETURN PERIOD RESPONSES FROM THE RELIABILITY ANALYSIS

The American Petroleum Institute recommended practice for TLP design stipulates the *responses* be calculated to have a return period of 100 years. This compares with the Department of Energy (1990a) who specify the *environmental parameters* be designed to have a return period of 50 or 100 years.

If a variable x has a return period of 50 years and we have observed it every 3 hours there will be a total of 146100 observations in the 50 years and, on average, one of these will have exceeded some value x_{50}. Assuming the events are independent and identically distributed - and in fact Bjerager et al (1988) has shown using Markov dependence between seastates that the effect of correlation between seastates is not important - the probability of exceeding the level x_{50} in any one event will be $Pe_{50} = 6.8446 * 10^{-6}$.

The probability of a *response* exceeding a threshold x_T can then be calculated using the time-variant reliability methods discussed earlier. As an example, the 50 year return period compression stress with exceedance probability Pe_{50} can be calculated by setting some threshold level and then using the time-variant reliability method to calculate the probability of exceeding the level. The process is then repeated several times using higher thresholds to build a picture of the variation of exceedance probability. Then having obtained

a sufficient number of points the threshold with exceedance probability $Pe = Pe_{50}$ can be found by simple interpolation.

7.10.1 SURGE MOTIONS

Ten surge motion thresholds were examined with values ranging from 10m to 100m. A time-variant reliability analysis was then performed for each threshold in which the winds, waves, and currents were assumed to be colinear. The joint probability model was based on the multivariate normal, Box-transformed, distribution with the parameters taken from Ch. 4 Table 4.9. The JONSWAP spectrum was used with a gamma=3.3 and a single wave heading angle of zero degrees was analysed.

The beta distance and exceedance probabilities for each level of threshold are shown in Fig. 7.7. The beta distance varies almost linearly with increasing motion threshold and the exceedance probability varies exponentially. The level of surge response with a return period of 50 years, ie $Pe_{50}=6.8446\text{x}10^{-6}$, is 35.0m in the unidirectional sea. These values are compared with the design event results in Table 7.2 where the figures in brackets indicate the percentage reduction over the design event surge motions.

7.10.2 TENDON STRESSES

The variation of exceedance probability for a range of tendon stress thresholds is shown in Fig. 7.8 . Interpolating from the graph we get a 50 year return period stress of 102N/mm2 in the uni-directional sea. These values are compared with the upper and lower zero-up-crossing period design event results in Table 7.3. Note in this case the time-varying component of the tension in the tendons is so small compared to the pre-tension it does not show a signoficant change from the design event stress obtained using the 12.5 and 14.0 second zer-up-crossing periods. In a future study using this model the tendons need to be re-designed with smaller cross-sectional area.

7.11 CONCLUSIONS

A new level III method has been developed for the calculation of reliability, based on the transformed normal method outlined in Ch. 4. In this new method the *data are transformed to a normal model* whereas in the conventional Rosenblatt transformation method *the model fitted to the data is transformed.* In most cases the Rosenblatt transformation requires a set of numerical integrations which are time consuming, in addition, it usually means the transformation must be defined in a subroutine which is compiled and then linked to the a main program. The new method on the other hand makes this step redundant and furthermore is simple to implement in a generalised computer code.

The time-variant reliability calculation can be converted to a time-invariant calculation by introducing an auxiliary random variable into the standard normal joint density for the loading and resistance. This conversion allows us to use the simple 'beta' optimisation method to approximate the solution of the survivor probabilities for sets of threshold or limit state functions.

The sensitivity studies indicate the importance of spectral shape, directional spreading, and wind speed for this type of structure. Current forces on the other hand are small for TLP's - but it should be noted they are significant for jacket structures. The responses of the TLP were largest in fully developed seas, with no directional spreading. Wind gusting forces and second order drift forces have the same order of magnitude for this structure and are both some 25 ~ 30% of the first order wave force.

A time-variant reliability analysis has been performed on the tension leg platform specified in Ch. 6 using the statistical models of the DB1 environmental data described in Ch's 3,4, and 5. The transformed normal multivariate model for the joint density of the winds, waves and currents was used successfully with the new method and, when used with a good non-linear optimiser, was very efficient. Generally, the number of within seastate analyses performed for each optimisation for the beta distance was given by (4~8)*(2N+1) where N is the number of variables. For this work N = 4 ~ 5 and the solution of the within seatstate statistics took some 30~40 seconds. This gave run-times of up to

$$10*(8*(2*5+1)) * 40 \text{ sec.} = 35200 \text{ seconds}$$

for each threshold run in which ten optimisations were performed to obtain the variation of survivor probability for a range of thresholds.

The new reliability method has been used to calculate survivor probabilities for a range of motion and tendon stress thresholds. This enables the levels of motion, or stress, corresponding to a given exceedance probability, like the 50 year return period value, to be calculated by interpolation. These values have been compared with the results from a notional design event approach in which the motions and stresses corresponding to concurrent 50 year winds, waves and currents are calculated. The comparisons show that if the return periods are specified on the responses and not the environmental parameters then the motions and stresses in the tendons are reduced by 15~25%. The level of reduction is dependent on the threshold type, ie motion or stress, and the number of variables that are treated as random. In this work the relative directionality of the winds, waves and currents was ignored. It is likely their inclusion in the model would lead to further reductions.

significant wave height	15.02	15.02	m
zero up-crossing period	12.5	14.0	s
wind speed	33.0	33.0	m/s
current speed	1.0	1.0	m/s
steady wind forces	1.31	1.31	MN
steady drift forces	0.505	0.35	MN
current forces	5.95	5.95	MN
resultant steady forces	7.76	7.71	MN
resultant steady offset	22.9	22.3	m
resultant steady setdown	0.617	0.60	m
tendon pretension	85.0	85.0	N/mm2
change in tendon tension caused by steady forces	3.6	3.4	N/mm2
wind motion in x-direc.			
standard deviation	6.62	6.62	m
zero up-crossing period	104.3	104.3	s
significant amplitude	13.24	13.24	m
drift motion in x-direc.			
standard deviation	7.12	4.9	m
zero up-crossing period	97.3	97.3	s
significant amplitude	14.2	9.8	m
first order wave motions in x-direc.			
standard deviation	1.9	2.2	m
zero up-crossing period	15.8	17.3	s
significant amplitude	3.8	4.4	m
combined wind, drift, and wave			
significant amplitude	19.8	17.1	m

Table 7.1 TLP surge responses for colinear 50 year return period winds, waves, and currents. Uni-directional, 0 degree heading waves, JONSWAP sea with gamma = 3.3 and zero crossing periods of 12.5 and 14.0 seconds

surge motions (m)

wave	design wave approach	joint probabilities
z.c.period	long crested	long crested
12.5	50.1	35
	[30%]	
14	45	
	[22%]	

Table 7.2 Comparison of design event and reliability method surge motions in short crested and long crested seas

tendon tensions (N/mm2)

wave	design wave approach	joint probabilities approach
z.c.period	long crested	long crested
12.5	101	102
	[-1%]	
14	105	
	[3%]	

Table 7.3 Comparison of design event and reliability method tendon stresses in short crested and long crested seas

Note the small percentage differences are due to the pre-tension being much larger than the dynamic stress component

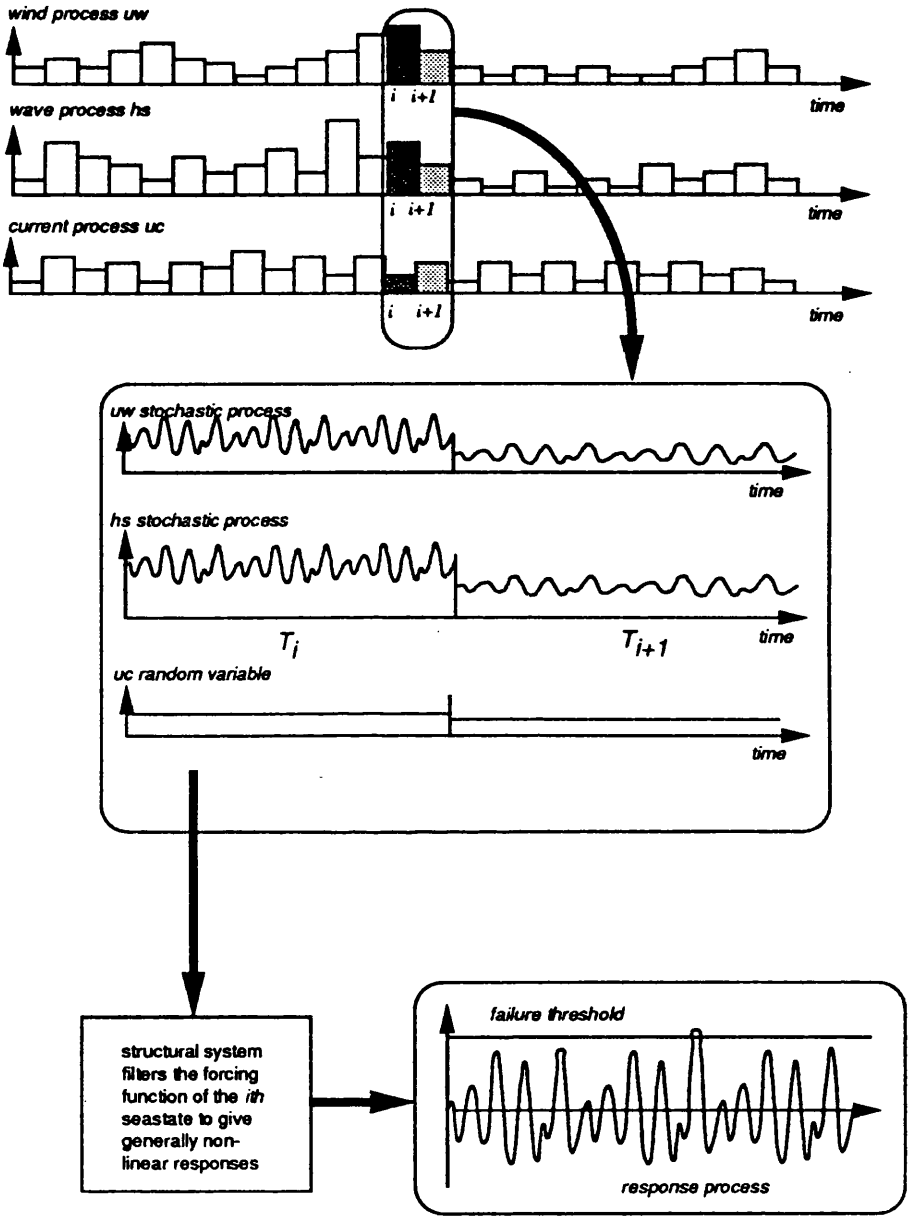


Figure 7.1 Idealisation of the slow and fast random variables and processes for an offshore structure reliability calculation

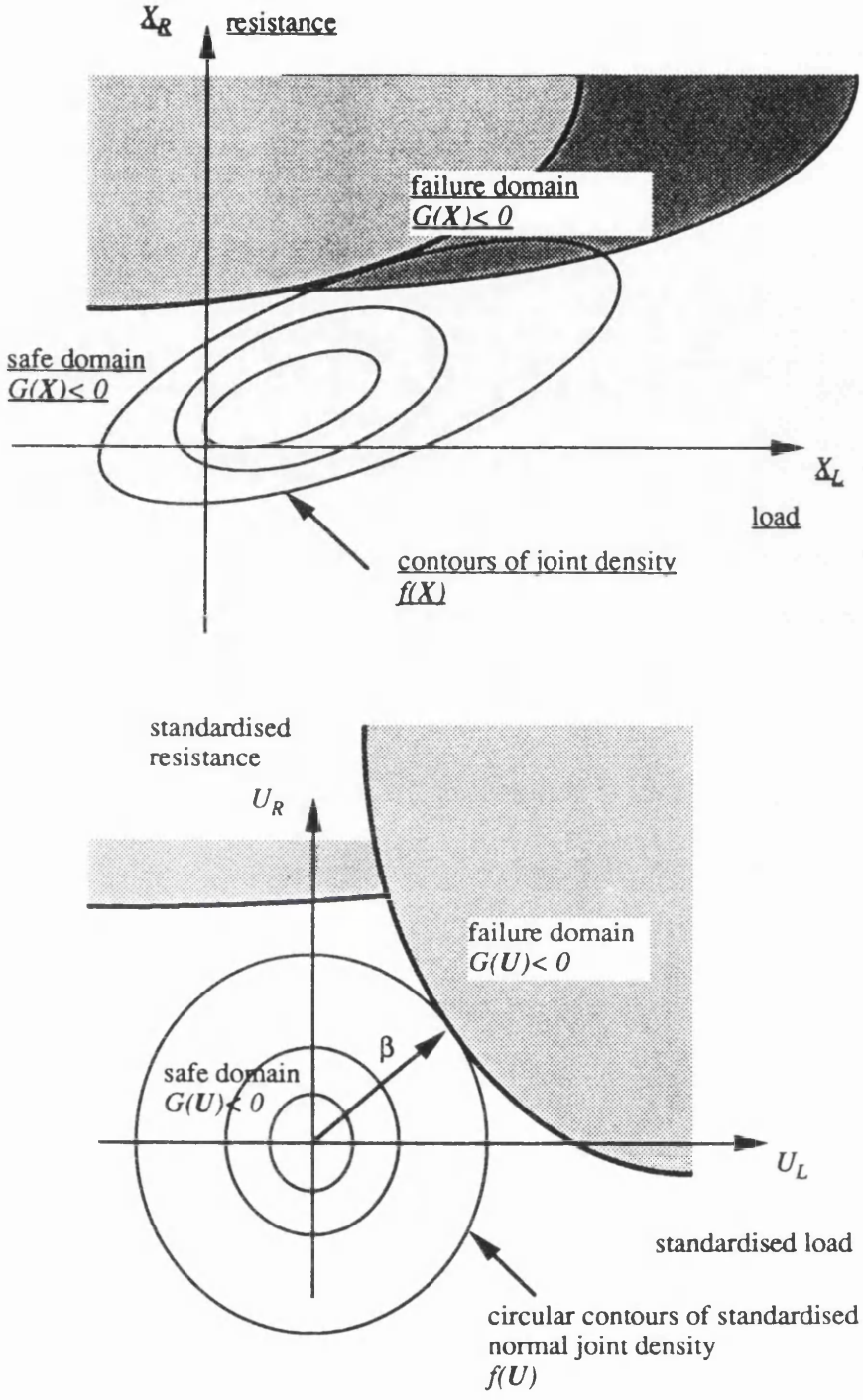


Figure 7.2 Safe domain, failure domain and joint density for a system with one load and one resistance random variable: in basic space (top); and transformed Normal $\boldsymbol{U}$ - space (bottom).

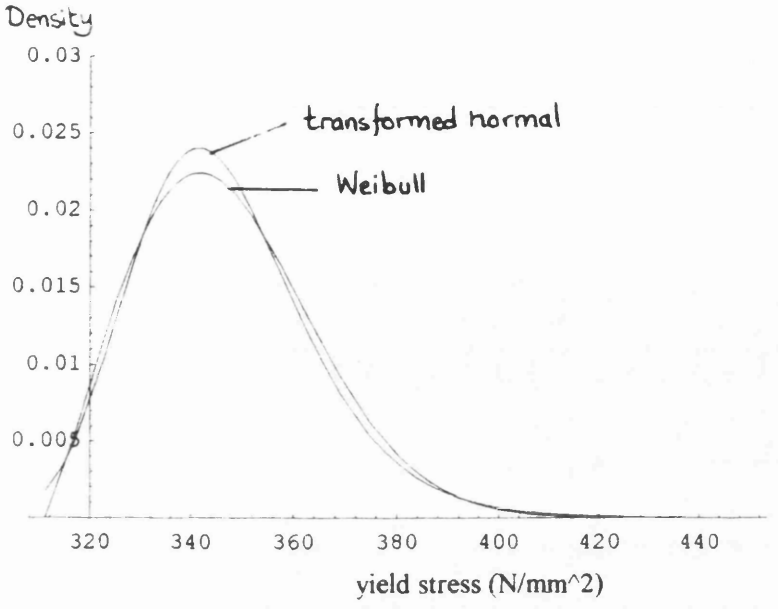


Figure 7.3 Comparison of Weibull population and transformed normal models for the yield strength of the tendons.

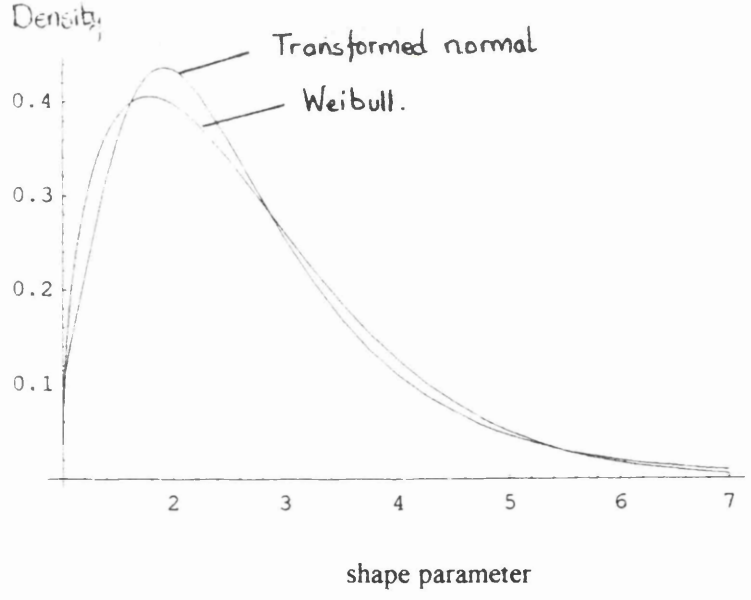


Figure 7.4 Comparison of Weibull population and transformed normal models for the JONSWAP shape parameter.

7.8.1 WINDS AND CURRENTS

The wind and current loads were compared in Ch. 6 where it was shown gusting wind forces are significantly larger than the steady current forces. The influence of wind force on the threshold exceedance probabilities is shown in Fig. 7.5. Note how the higher wind speeds reduce the failure probability for the compression limit state by increasing the mean tendon stress. The effect on the tension limit state is adverse with high wind speeds contributing significantly to the exceedance probabilities.

7.8.2 SPECTRAL SHAPE PARAMETER

The sensitivities of the threshold exceedance probabilities to spectral shape are illustrated by Fig. 7.6. The $\gamma = 1$ case corresponds to a fully developed sea, and the $\gamma = 5$ case corresponds to a fetch limited sea with a narrow-band spectrum. The largest surge responses occur in fully developed seas in which there is a larger amount of energy in the drift force spectrum at low frequencies near the natural frequencies of the structure in surge, sway and yaw.

7.9 DESIGN EVENT SEASTATE RESPONSES

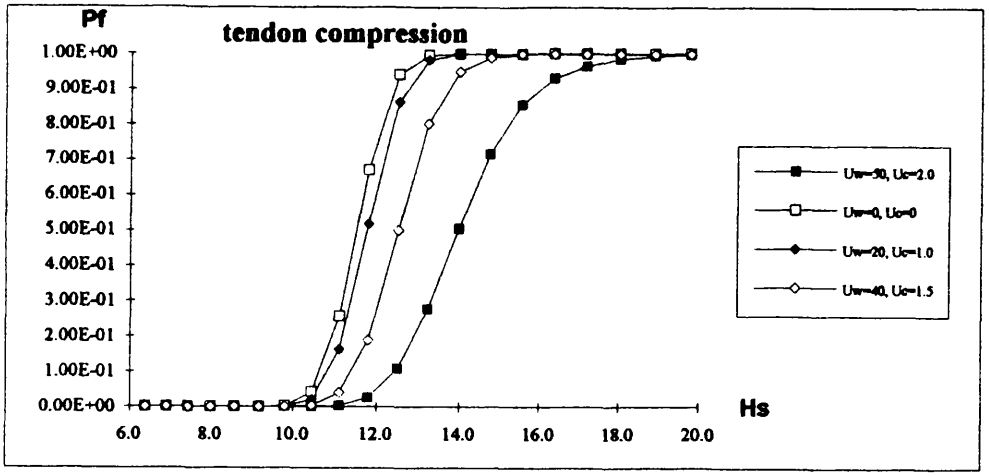
In ch.4 the marginal extreme wind speed, current speed, and significant wave height were estimated using both population and monthly maxima models. As a result the 50 year return period design event environmental variables were defined as:

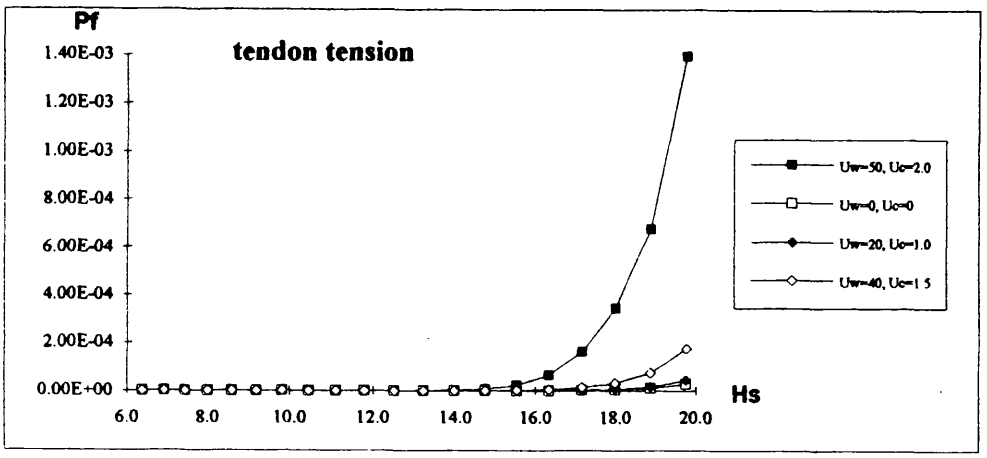
significant wave height	Hs	15.02	m
wind speed	Uw	33.0	m/s
current speed	Uc	1.0	m/s

These values are close to the Department of Energy (1990a) recommended values. In a design event analysis the winds, waves and currents would be assumed to act colinearly and simultaneously. The range of wave zero-up-crossing period associated with the design event *Hs* is usually determined by the maximum, and minimum wave steepness observed around the British Isles, ie

$$3.2\,(Hs)^{1/2} < Tz < 3.6\,(Hs)^{1/2}$$

For the 50 year return period wave height this gives a range of 12.5 to 14.0 seconds so these values were used as the upper and lower bounds for the Design Event analysis of the TLP model developed in Ch.6.





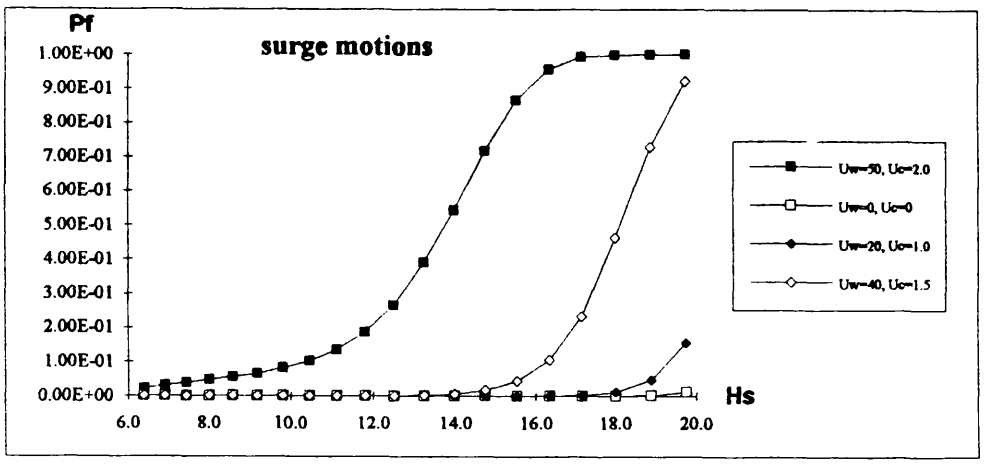


Figure 7.5 Tendon stress and surge motion threshold exceedance probabilities for extreme seastate Hs,Tz and a range of wind and current speeds witha zero degree heading

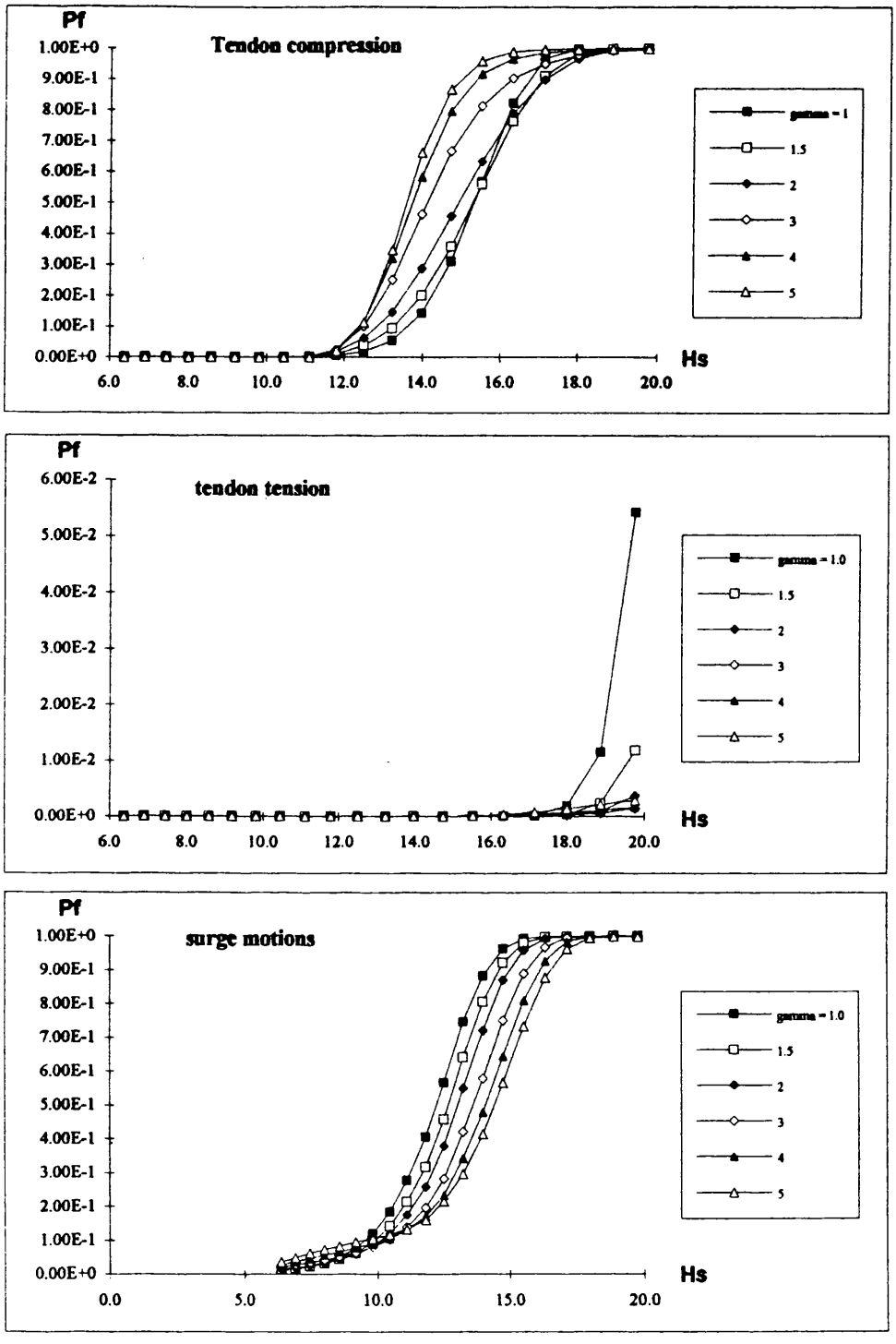
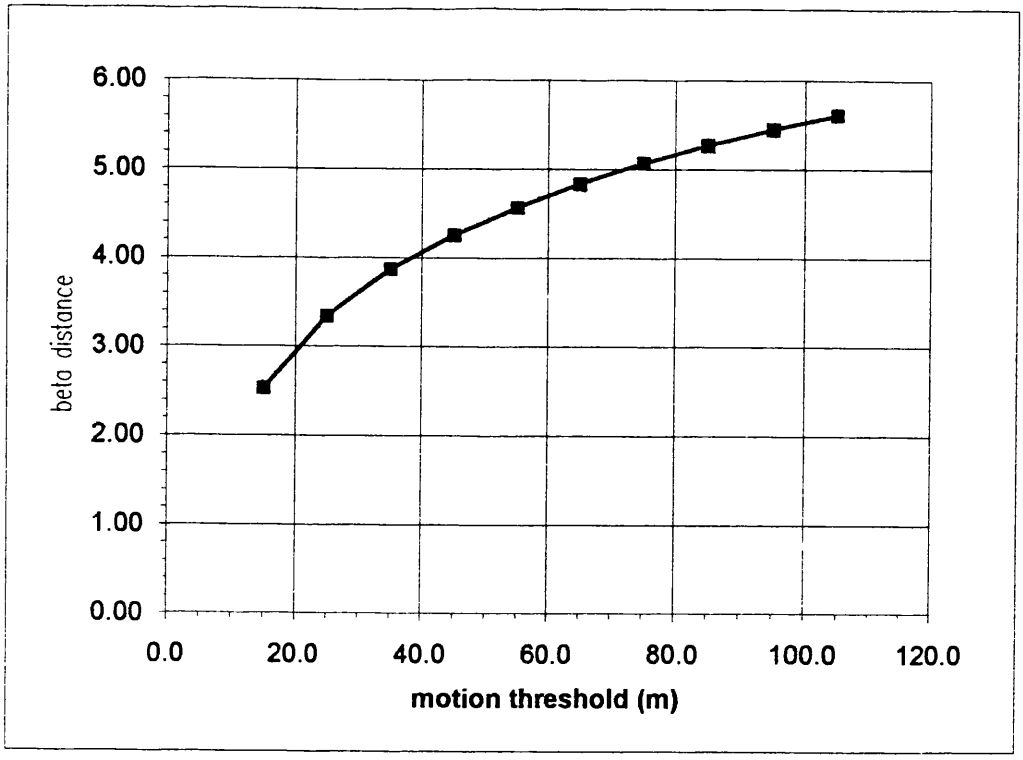


Figure 7.6 Influence of spectral shape parameter on tendon stress and surge motion threshold exceedance probabilities.



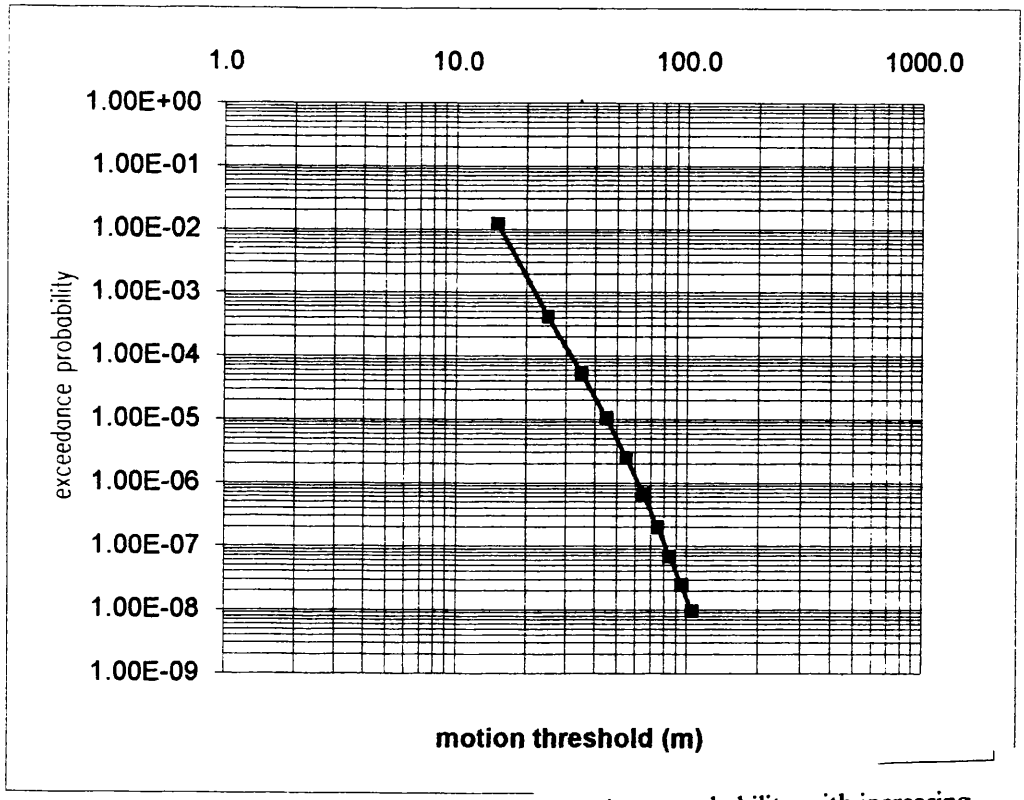
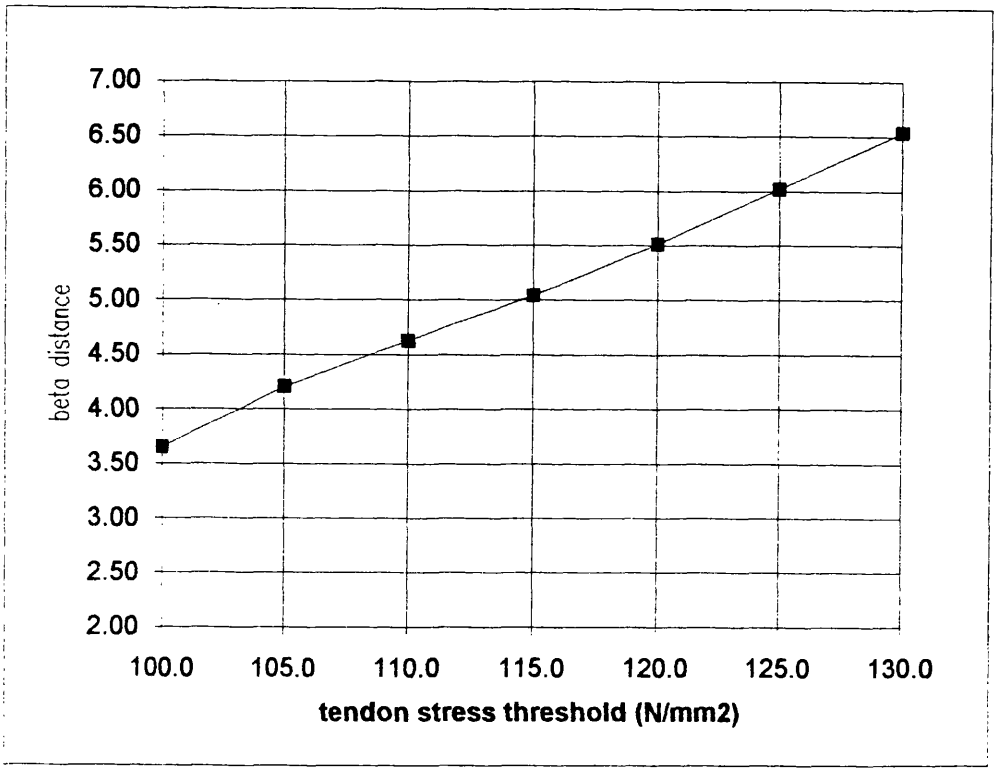


Figure 7.7 Platform surge motions exceedance probability with increasing threshold. 0 degree heading JONSWAP sea with gamma = 3.3



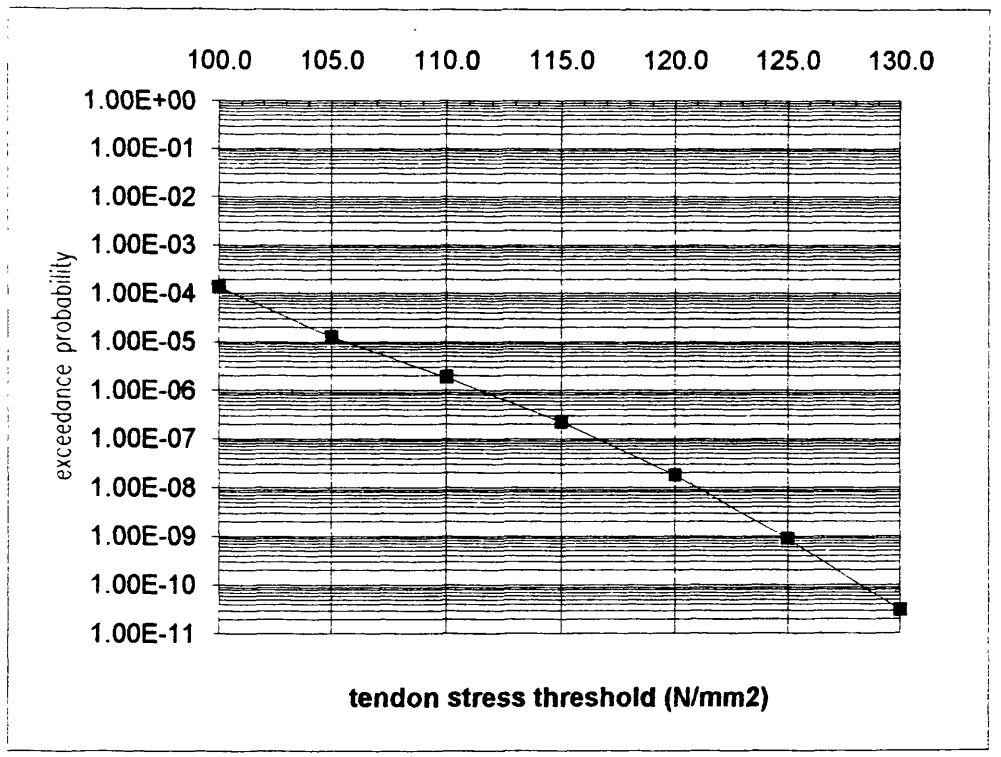


Figure 7.8 Tendon tension exceedance probability with increasing threshold.
0 degree heading JONSWAP sea with gamma = 3.3

Chapter 8

CONCLUSIONS & FUTURE WORK

CONCLUSIONS

Probabilistic methods can be used to assess the reliability of complex structures providing accurate and sufficient models of the random variables are available. This thesis defines some tools for synthesising models from measured data and from the work undertaken the following conclusions can be drawn.

The DB1 wind, wave and current data can be treated as unstructured in the statistical sense for the multivariate modelling. The only evidence of structure appears in the significant wave height and zero-up-crossing period density. This is the result of the breaking wave criterion. Furthermore marginal kernel density estimators show no evidence of bimodality or mixing of different statistical populations in the modal region or indeed for moderately large values. This suggests a population modelling approach is reasonable, providing the correct tail behaviour can be guaranteed by the use of a suitable population model.

The desirable characteristics of a good estimator were stated as: consistency, sufficiency, low bias, low sampling variance, reliability, and simplicity. Maximum likelihood is an optimal estimator in the large sample case (>100) and can be used with population data and extreme value data. ML estimates of a model's parameters are efficient with low bias and the parameter variance is close to the optimum attainable from any estimator. The method can easily be generalised to the multivariate case and the parameter uncertainty can be deduced directly from the Information matrix if the ML solutions can be found.

Both Weibull and GEV population and monthly maxima models have been compared as estimators of return period values. Both the population and monthly maxima models gave the worst results for the significant wave height with the Weibull model underestimating and the GEV model overestimating the value. By comparing only the return period estimates it is not clear if the population method or the extreme value method results in the lowest modelling uncertainty. However, if the model parameter uncertainty is included in the comparison it is clear the population approach results in an estimator with much lower statistical uncertainty. This suggests that the population modelling method is best when only a few years of data are available - providing the correct model can be found.

The Box and Cox transformation is powerful enough to transform even highly non-normal data. Furthermore the shape and location parameters respond strongly to the skewness and kurtosis in the data.

This ensures the models are capable of modelling the extremes since these moments are dominated by the tails of the data. The population model has been assembled from a mixture of ML parameters obtained from the marginal, bivariate,and multivariate datasets. The criterion used for selecting the best set of transformation parameters has been the accuracy with which the resulting population model can predict the marginal 50 year return period values. This study shows the population models obtained from the application of likelihood theory are accurate even for the 50 year return period values. Furthermore, the general structure of the fitted model matches the behaviour seen in the scatter plots given in Ch. 2.

The transformation of the data to a normal model has considerable advantages when used in level III reliability studies. Most important is that it makes the Rosenblatt transformation redundant. Second is that the method is simple to implement and, when used with a good non-linear optimiser, is very efficient. Overall the transformation approach has several advantages over the conventional methods summarised in the introduction to the thesis.

Measures of location, spread, skewness and kurtosis have been taken from the work of Mardia and applied successfully to the DB1 directional wave data. These data were then used to develop a regression model for the frequency dependent - second moment estimate - cosine model spreading powers and von Mises model concentration parameters. The models are largely in agreement with the Hasselmann study results - despite their use of the average of the first and second moment estimates for the cosine spreading power. The major differences are either side of the spectral peak where this study predicts a slower increase in the directional width.

For design it is suggested that the spreading power be taken as frequency independent and that the results at the spectral peak be used for the whole range of f/fm. The results at the spectral peak show considerable scatter which has been modelled by fitting a Weibull distribution to the observed values. The result has a lower bound of 1.2 a modal value of 4.6 and a mean of 17.7. It is suggested the effect of this variability be examined at a later date by including the uncertainty in a reliability analysis.

The use of circular statistical methods is equally valid for describing the long-term directional distributions of the winds, waves and currents. A method based on the use of Fourier-Steiltjes series is presented which enables multimodal directional distributions to be described using only a few terms from the characteristic function.

A new level III method has been developed for the calculation of reliability, based on the transformed normal method outlined in ch. 4. In this new method the data are transformed to a

normal model whereas in the conventional Rosenblatt transformation method the model fitted to the data is transformed. In most cases the Rosenblatt transformation requires a set of numerical integrations which are time consuming, in addition, it usually means the transformation must be defined in a subroutine which is compiled and then linked to the a main program. The new method on the other hand makes this step redundant and furthermore is simple to implement in a generalised computer code.

The time-variant reliability calculation can be converted to a time-invariant calculation by introducing an auxiliary random variable into the standard normal joint density for the loading and resistance. This conversion allows us to use the simple 'beta' optimisation method to approximate the solution of the survivor probabilities for sets of threshold or limit state functions.

The sensitivity studies indicate the importance of spectral shape, directional spreading, and wind speed for this type of structure. Current forces on the other hand are small for TLP's, but it should be noted they are significant for jacket structures. The responses of the TLP were largest in fully developed seas, with no directional spreading. Wind gusting forces and second order drift forces have the same order of magnitude for this structure and are both some 25 ~ 30% of the first order wave force.

A time-variant reliability analysis has been performed on the tension leg platform specified in Ch. 6 using the statistical models of the DB1 environmental data described in Ch's 3,4, and 5. The transformed normal multivariate model for the joint density of the winds, waves and currents was used successfully with the new method and, when used with a good non-linear optimiser, was very efficient. The new reliability method has been used to calculate survivor probabilities for a range of motion and tendon stress thresholds. This enables the levels of motion, or stress, corresponding to a given exceedance probability, like the 50 year return period value, to be calculated by interpolation. These values have been compared with the results from a notional design event approach in which the motions and stresses corresponding to concurrent 50 year winds, waves and currents are calculated. The comparisons show that if the return periods are specified on the responses and not the environmental parameters then the motions and stresses in the tendons are reduced by 15~25%. The level of reduction is dependent on the threshold type, ie motion or stress, and the number of variables that are treated as random. In this work the relative directionality of the winds, waves and currents was ignored. It is likely their inclusion in the model would lead to further reductions.

FUTURE WORK

The 50 year return period estimate of significant wave height obtained from fitting a GEV model to the monthly maxima is inaccurate and has a negative lower bound. The cause is the attraction to a lower bounded FT-II (Frechet) model which has a long upper tail. Two additional constraints in the likelihood optimisation may result in an improvement. The first is to restrict the lower bound to be greater than or equal to zero which effectively constrains to the model to be either an FT-I (Gumbel) or FT-III (Weibull) model; the second is to constrain the lower bound to be zero or positive. This should be examined in a future study.

The kernel densities used for this study were of the simplest type. Improved estimates can be obtained by using adaptive kernels in which the degree of smoothing is adjusted to the local density.

The multivariate model for the winds, waves, and currents has been treated as unstructured. The major problem with this approach is that the joint distribution for significant wave height and zero-up-crossing period is influenced by the breaking wave limit for steep waves. Ideally the model should be improved to allow for this behaviour, perhaps by using alternatives to the Box and Cox transformation or alternatively by using a structured likelihood.

The directional wave models are based on fitting parametric forms to the circular moments at each frequency. This could be improved by fitting a parametric model to all frequencies in one optimisation. The robustness aspects of the model are also worth detailed study using simulation.

The response analysis of the TLP is essentially a simple linearised solution which probably underestimates the non-linear response of for example the tendon stresses. It would be useful if the work could be extended to include a more sophisticated stochastic linearisation in which the seastate parameters are taken into account or alternatively if a quadratic programming method could be used.

References

Abramowitz, M. and Stegun, I.A. 1965. Handbook of mathematical functions with formulas, graphs and mathematical tables, National Bureau of Standards, Applied Mathematics Series.

American Petroleum Institute. 1987. Recommended practice for planning, designing and constructing tension leg platforms (RP2T), First Ed.

Andrews, D.F., Gnanadesikan, R., Warner, J.L. 1971. Transformations of multivariate data. *Biometrics*, **27**.

Ang, G.L., Ang, A.H-S., Tang, W.H. 1992. Optimal importance sampling density estimator. *Journal of Engineering Mechanics*. **118**, **6**.

Armand, J.L. 1991. Stochastic Modelling. In: *Proceedings of the 11th International Ship and Offshore Structures Congress*, ed. by HSU, P.H. Volume 2.

Barltrop, N., Adams, A.J. 1991. *Dynamics of fixed marine structures*. Third ed. Butterworth-Heinemann.

Battjes, J.A. 1970. Long-term wave height distribution at seven stations around the British Isles. Institute of Oceanographic Sciences: Report 44.

Bea, R.G., Cornell, C.A., Vinnem, J.E., Geyer, J.F., Stahl, B., Shoup, G.J. 1992. Comparative risk assessment of alternative TLP systems: Structure and foundation aspects. *Proceedings of the Offshore Mechanics and Arctic Engineering Conference*. Volume 2.

Bitner-Gregersen, E.M., Haver, S. 1991. Joint environmental model for reliability calculations. *Proceedings of the First International Offshore and Polar Engineering Conference.*

Bitner-Gregersen, E.M. 1980. Non-linear effects of the statistical model of shallow-water wind waves. *Applied Ocean Research*, **2**.

Bjerager, P. 1989. Methods for structural reliability computations, Lecture notes for the course: Structural Reliability Methods and Applications, University of California.

Bjerager, P., Loseth, R., Winterstein, S.R., Cornell, C.A. 1988. Reliability method for marine structures under multiple environmental load process. *Proceedings of the International Conference on Behaviour of Offshore Structures.*

Blackman, P.B. Tukey, J.W. 1958. *The measurement of power spectra from the point of view of communications engineering.* Dover Publications.

Borgman, L.E. 1969. Directional spectra models for design use. *Proceedings of the Offshore Technology Conference.*

Borgman, L.E. 1975. Extremal Statistics in Ocean Engineering. *In: Civil Engineering in the Oceans III, American Society of Civil Engineers.*

Box, G.E.P., Cox, D.R. 1964. An analysis of transformations (with discussion) *Journal of the Royal Statistical Society*, **B,26**.

Burns, G.E. 1983. Calculating viscous drift of a tension leg platform. *Proceedings of the 2nd International Offshore Mechanics and Arctic Engineering Symposium*, ASME.

Carter, D.J.T. 1987. A simple model for estimating return value of wave height. *Advances in Underwater Technology, Ocean Science and Offshore Engineering*, **12**.

Carter, D.J.T., Challenor, P.G. 1981. Estimating return values of environmental parameters. *Quarterly Journal of the Royal Meteorological Society*. **107**.

Carter, D.J.T., Challenor,P.G. 1983. Methods of fitting the Fisher-Tippet Type I Extreme Value Distribution. *Ocean Engineering*, **10**.

Cartwright, D.E. 1963. The use of directional spectra in studying the output of a wave recorder on a moving ship. In: *Ocean Wave Spectra*, Prentice-Hall.

Cartwright, D.E., Longuet-Higgins, M.S. 1956. The Statistical Distribution of the Maxima of a Random Function. *Proceedings of the Royal Society* **A, 237**.

Chakrabarti, S.K. 1984. Steady drift force on vertical cylinder viscous vs. potential. *Applied Ocean Research*, **6, 2**.

Chakrabarti, S.K. 1987. *Hydrodynamics of offshore structures*. Springer-Verlag.

Challenor, P.G. 1982. A New Distribution for Annual Extremes of Environmental Variables. *Quarterly Journal of the Royal Meteorological Society*, **108**.

Chan, H-S. 1990. A three dimensional technique for predicting 1st and 2nd order hydrodynamic forces on a marine vehicle advancing in waves. University of Glasgow, August.

Chase, J. et al. 1957. The directional wave spectrum of wind-generated seas as determined from data obtained by the stereo wave observation project. Coll. Eng. Report, New York University, July.

Chatfield,C. 1991. *The analysis of time series: an introduction*. Chapman and Hall.

Cohen, A.C. 1965. Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and Censored Samples. *Technometrics*, **7, 4**.

Cooley, J.W., Tukey, J.W. 1965. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, **19.**

Crowder, M.J., Kimber, A.C., Smith, R.L., Sweeting, T.J. 1991. *Statistical Analysis of Reliability Data.* Chapman & Hall.

Davison, A.C., Smith, R.L. 1990. Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society* **B.**

de Boom, W.C., Pinkster, J.A., Tam, S.G. 1984. Motion and tether force prediction for a TLP. *Journal of Waterway, Port, Coastal and Ocean Engineering,* **100, 4.**

Denis, J.P.F., Heaf, N.J. 1979. A comparison between linear and non-linear response of a proposed tension leg platform. Proceedings of the Offshore Technology Conference, OTC 3555.

Department of Energy, 1990b. Metocean Parameters - Wave Parameters. *Offshore Technology Report,* OTH 89 300.

Department of Energy. 1990a. Offshore Installations: Guidance on design, construction and certification. Fourth ed. HMSO.

Det norske Veritas. 1981. Rules for the classification of mobile offshore units, DnV, Oslo.

Draper, L. 1963. Derivation of a Design Wave from Instrumental Records of Sea States. *Proceedings of the Institution of Civil Engineers,* **26.**

Dubey, S.D. 1967. Normal and Weibull Distributions. *Naval Research Logistics Quarterly,* **14.**

Eatok-Taylor, R., Jeffereys, E.R. 1986. Variability of hydrodynamic load predictions for a tension leg platform. *Ocean Engineering,* **13,5.**

Eidsvik, K.J. 1985. Large sample estimates of wind fluctuations over the ocean. *Boundary-Layer Meteorology,* **32.**

Epanechnikov, V.A. 1969. Nonparametric estimation of a multidimensional probability density. *Theor. Probabl. Appl.,* **14.**

ESDU:Engineering Sciences Data Unit. 1974. Characteristics of atmospheric turbulence near the ground. ESDU International, London.

Ewing, J.A., Laing, A.K. 1987. Directional spectra of seas near full development. *Journal of Physical Oceanography,* **17.**

Exploration & Production Forum. 1985. The application of joint probability of metocean phenomena in the oil industry's structural design work.Workshop held November 1985.

Faltinsen, O.M., Demirbilek, Z. 1989. Hydrodynamic analysis of TLPs. In: *Tension Leg Platform, A state of the art review.* American Society of Civil Engineers.

Fisher, R.A., Tippet, L.H.C. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings Cambridge Phil. Soc.*, 24, 2.

Freathy, P.E., Hooper, A.G., MacDonald, H.W. 1982. Wind and wave directional data obtained from the DB1 in the South Western approaches to the United Kingdom. *Int. Conf. on Wave and Wind Directionality Paris.* Ed. Technip.

Galambos, J. 1978. *The Asymptotic Theory of Extreme Order Statistics.* New York: Wiley.

Galambos, J. 1984. Asymptotics; Stable laws for Extremes; Tail Properties. In: *Statistical Extremes and Applications.* J.T. de Oliveira(editor), Reidel Dordrect.

Gnanadesikan, R. 1977. *Methods for statistical data analysis of multivariate observations.* John Wiley, New York.

Goda , Y. 1985. *Random seas and design of maritime structures.* Univ. of Tokyo Press.

Goda,Y. 1974. Estimation from spectral information. *Intl.symposium on Ocean Wave Measurement and Analysis.* ASCE, 1.

Gumbel, E.J. 1958. *Statistics of Extremes.* New York: Columbia University Press.

Gumbel, E.J. 1960. Bivariate exponential distributions. Journal of the American Statistical Association, 55.

Hasselmann, D.E., Dunkel, M. and Ewing, J.A. 1980. Directional wave spectra observed during JONSWAP 1973. *Journal Phys. Oceanog,* **10.**

Hasselmann, D.E., Dunkel, M., Ewing J.A. 1976a. Directional spectra observed during JONSWAP 1973. *Journal of Physical Oceanography,* **6.**

Hasselmann, K., Ross, D.B., Muller, P., Sell, W. 1976b. A parametric wave prediction model. *Journal of Physical Oceanography.* **6.**

Havelock, T.H. 1940. The pressure of water waves upon a fixed obstacle, *Proceedings of the Royal Society of London, Series A,* , **963.**

Haver, S, Winterstein, S.R. 1990. The effects of a joint description of environmental data on design loads and reliability. *Ninth International Conference on Offshore Mechanics and Arctic Engineering.* ASME.

Haver, S. 1991. Effects of wave directionality and choice of wave spectrum on the response of a deep water jacket. *Marine Structures,* **4, 6.**

Haver, S., Natvig, B. 1991. One some uncertainties in the modelling of ocean waves and their effects on TLP response. *International Journal of Offshore and Polar Engineering*, **1, 2.**

Helvacioglu, I. 1990. Dynamics of coupled articulated tower and floating production systems, Ph.D. thesis 8823, University of Glasgow.

Hoaglin, D., Mosteller, F., Tukey, J.W. 1983. Understanding robust and exploratory data analysis. Wiley Series in Probability and Mathematical Analysis.

Hosking, J.R.M., Wallis, J.R., Wood, E.F. 1985. Estimation of the Generalised Extreme Value Distribution by the Method of Probability Weighted Moments. *Technometrics*, **27, 3.**

Huang, N.E., Chen, D.T., Tung, C.C. 1972. Interactions between steady non-uniform currents and gravity waves with applications for current measurements, *Journal of Physical Oceanography*, **2.**

Huber, P.J. 1981. *Robust Statistics.* Wiley.

Incecik, A., Wu, S-K, Soylemez, M. 1987. Effect of different mathematical models in calculating motion and structural response of offshore platforms. Integrity of Offshore Structures - 3.

Isaacson, M. de St. Q. 1981. Long-term Distributions of Ocean Waves: A Review. *Journal of the Waterway, Port, Coastal and Ocean Division, American Society of Civil Engineers.*

Jaynes, E.T. 1982. On the rationale of maximum entropy methods. Proceedings of IEEE, **70, 939.**

Jeffereys, R.E. 1986. Comparison of three methods for calculation of directional spectra. *Proceedings of the Fifth International Offshore Mechanics and Arctic Engineering Symposium*, Tokyo.

Jenkinson, A.F. 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society.*

Johnson, N.I., Kotz, S. 1970. *Continuous Univariate Distributions - 1.* Wiley.

Johnson, N.I., Kotz, S. 1970. *Continuous Univariate Distributions - 2.* Wiley.

Johnson, N.I., Kotz, S. 1972. *Distributions in Statistics, Continuous Multivariate Distributions.* Wiley.

Kato, S., Kinoshita, T. 1990. Non-linear response of moored floating structures in random waves and its stochastic analysis. Part 1. Theory and model experiment. Ship Research Institute, Ministry of Transport, Tokyo, Japan.

Kraznowski, W.J. 1988. *Principles of Multivariate Analysis: a users perspective.* Oxford Statistical Science Series.

Kuik, A.J. van Vledder, G.Ph., Holthuijsen, L.H. 1988. A method for the routine analysis of pitch-and-roll buoy wave data. *Journal of Physical Oceanography,* **18.**

Labeyrie, J. 1990. Stationary and Transient States of Random Seas. *Marine Structures* , **3.**

Ledermann, W. 1984. *Handbook of Applicable Mathematics, vol VI: Statistics Parts A and B.* Wiley.

Lenth, R.V. 1981. Robust measures of location for directional data. *Technometrics* , **23, 1.**

Leverette, S.J., Bradley, M.S., Bliault, A. 1982. An integrated approach to setting environmental design criteria for floating production facilities. *Proceedings of the Conference on the Behaviour of Offshore structures.*

Lewis, O., Orav, E.J. 1989. Simulation Methodology for Statisticians, Operations Analysts, and Engineers. **1,** Wadsworth.

Long, R.B. Hasselmann, K. 1979. A variational technique for extracting directional spectra from multi-component wave data. *Journal of Physical Oceanography,* **9.**

Longuet-Higgins, M.S. 1963. The effect of non-linearities on statistical distributions in the theory of sea waves. *Journal Fluid Mech.,* **19.**

Longuet-Higgins,M.S.,Cartwright,D.E. and Smith,N.D. 1963. Observations of the directional spectrum of sea waves using the motion of a floating buoy. *Proceedings Conf., Ocean Wave Spectra,* Prentice-Hall.

Lotsberg, I. 1991. Probabilistic design of the tethers of a tension leg platform. *Journal of Offshore Engineering Mechanics and Arctic Engineering,* **113.**

MacCamy, R.C., Fuchs, R.A. 1954. Wave forces on piles: a diffraction theory. Beach Erosion Board, Technical Memorandum, **69.**

Madsen, S.O., Krenk, S., Lind, S.C. 1986. *Methods of structural safety.* Prentice-Hall.

Mardia, K.V. 1972. *Statistics of Directional Data,* Academic Press.

Mathiesen, J., Bitner-Gregersen, E. 1990. Joint distributions for significant wave height and wave zero crossing period. *Applied Ocean Research,* **12, 2.**

Melchers, R.E. 1987. *Structural reliability analysis and prediction.* John Wiley & Sons.

Melchers, R.E. 1991. New methods for structural reliability assessment.*Proceedings of the First International Offshore and Polar Engineering Conference*, vol.1.

Menon, M.V. 1963. Estimation of the Shape and Scale Parameters of the Weibull Distribution. *Technometrics*, **5, 2.**

Miller, K.S. 1964. Multidimensional Gaussian Distributions. SIAM Series in Applied Mathematics, Wiley.

Miller, N.S. 1987. Problems associated with the determination of mooring loads on floating structures. *International Conference on Mobile Offshore Structures*, London.

Mitsuyasu, H., Tasai, F., Suhara, T., Mizuno, S., Ohkuso, M., Honda, T., Rikishi, K. 1975. Observations of the directional spectrum of oceans waves using a clover leaf buoy. *Journal Phys. Oceanography*, **5.**

Naess, A. 1989. Prediction of extremes of combined first-order and slow-drift motions of offshore structures. *Applied Ocean Research.* **11, 2.**

Newton, H.J. 1988.*Timeslab: A time series analysis laboratory.* Wadsworth.

Ochi, M.K., Shin, S.Y. 1988. Wind turbulent spectra for design of offshore structures. *Offshore Technology Conference*, Houston.

Ochoa, J. 1990. Pitfalls in the estimation of wind wave directional spectra by variational principles. *Applied Ocean Research*, **12, 4.**

Patel, M., Witz, J. 1991. Compliant offshore structures. Butterworth - Heinemann.

Patil, G.P., Radhakrishna, Rao, C. 1977. The weighted distributions. In: *Applications of Statistics*, ed. P.R. Krishnaiah, North-Holland Publishing Company.

Pickands, J. 1981. Multivariate extreme value distributions. In: *Proceedings of the 43rd session I.S.I.*, Buenos Aires.

Pierson, W.J., Moskowitz, L.I. 1964. A proposed spectral form for fully-developed seas base on the similarity theory of S.A. Kitaigorodskii. *Journal Geophys. Res.* **69.**

Pinkster, J.A. 1974. Low frequency phenomena associated with vessels moored at sea. Society of Petroleum Engineers of AIME, SPE 48.

Prescott, P., Walden, A.T. 1980. Maximum Likelihood Estimation of the Generalised Extreme Value Distribution. *Biometrika*, **67, 3.**

Press,W.H., Flannery,B.P., Teukolsky,S.A., Vetterling,W.T. 1989. *Numerical Recipes*. Cambridge.

Prince-Wright, R.G., Percival, D. 1989. Directional wave energy models and their influence on offshore structure design. University of Glasgow: Report NAOE-89-23.

Prince-Wright, R.G. 1990. Estimation of Weibull parameters using maximum likelihood estimators. University of Glasgow: Report NAOE-90-23.

Prince-Wright, R.G. 1991a. Analysis of the DB1 wind, wave and current data set . University of Glasgow: Report NAOE-91-33.

Prince-Wright, R.G. 1991b. Estimating extremes using the method of maximum likelihood with censored data. *International Symposium of Marine Structures*, ISMS 1991, Shanghai.

Pugh, D.T. 1987. *Tides surges and mean sea level: A handbook for engineers and scientists.* John Wiley and Sons.

Pugh, D.T., Vassie, J.M. 1980. Applications of the joint probability method for extreme sea level computations. *Proceedings of the Institute of Civil Engineers*, **2, 69.**

Rao, C.R. 1965. On discrete distributions arising out of methods of ascertainment. In: *Classical and Contagious Discrete Distributions, ed G.P. Patil,* Statistical Publishing Soc., Calcutta.

Rao, C.R. 1973. *Linear Statistical Inference and it Applications.* Second Edition, Wiley.

Rice, S.O. 1944/5. Mathematical Analysis of Random Noise. *Bell Technical Journal*, **23/24.**

Rosenblatt, M. 1952. Remarks on a multivariate transformation, *The Annals of Mathematical Statistics*, **23.**

Rosenblatt, M. 1971. Curve Estimates, *The Annals of Mathematical Statistics*, **42, 6.**

Rubinstein, Y. 1981. Simulation and the Monte Carlo method. John Wiley, New York.

Sarpkaya, T., Bakmis, C., Storm, M.A. 1984. Hydrodynamic forces from combined wave and current flow on smooth and rough cylinders at high Reynolds number, *Proceedings of the Sixteenth Offshore Technology Conference*, OTC, 4830.

Scott, D.W., Wand, P.M. 1991. Feasibility of multivariate density estimates, *Biometrika*, **78, 1.**

Shi, W.B. 1991. Stochastic load combinations with particular reference to marine structures. *Marine Structures*, **4.**

Shinozuka, M. 1989. Freudental lecture: developments in structural reliability. *Proceedings of the 5th International Conference on Structural Safety and Reliability*, ASCE.

Sibuya, M. 1960. Bivariate extreme statistics. *American Institute of Statistical Mathematics*, **11.**

Silverman, B.W. 1986. *Density estimation for statistics and data analysis*: Monographs on Statistics and Applied Probability, Chapman and Hall.

Silvey, S.D. 1975. *Statistical Inference, Monographs on Statistics and Applied Probability.* Chapman and Hall.

Singh, V.P., Cruise, J.F., Ming, M.A. 1990. A Comparative Evaluation of the Estimators of the Weibull Distribution by Monte Carlo Simulation. *Journal of Statistical Computation and Simulation*, **36.**

Smith, D., Csenki, A., Ellinas, C.P. 1987. Ultimate limit state analysis of unstiffened and stiffened structural components. *Symposium on Integrity of Offshore Structures.*

Smith, R.L. 1984. Threshold Methods for Sample Extremes, In: *Statistical Extremes and Applications.* J.T. de Oliveira(editor), Reidel Dordrect.

Tagaki, M., Arai, S., Takezawa, S., Tanaka, K., Takarada, N. 1985. A comparison of methods for calculating the motion of a semi-submersible. Ocean Engineering, 12, 1.

Tan, S.G., de Boom, W.C. 1981. The wave induced motions of a tension leg platform in deep water. *Proceedings of the Annual Offshore Technology Conference*, OTC 4074.

Tawn, J.A. 1988a. An Extreme Value Theory Model for Dependent Observations. *Journal of Hydrology*, **101.**

Tawn, J.A. 1988b. Bivariate extreme value theory: models and estimation. *Biometrika*, **75.**

Tawn, J.A. 1990. Modelling multivariate extreme value distributions. *Biometrika*, **77, 2.**

Thoft-Christensen, P., Baker, M.J. 1982. *Structural Reliability Theory and its Applications.* Berlin: Springer-Verlag.

Tiago de Oliveira, J. 1962. Structure theory of bivariate extremes: extensions. *Estudos de Matematica, Estatistica e Econometria.*

Tiago de Oliveira, J. 1980. Bivariate extremes: models and foundations. In: *Multivariate Analysis*, ed. P.R. Krishnaiah.

Tiago de Oliveira, J. 1989a. Extreme Value Distributions of Loads and Responses. In: *ISPRA Courses on Structural Reliability*, Lisboa.

Tiago de Oliveira, J. 1989b. Intrinsic estimation of the dependence structure for bivariate extremes. *Statistics & Probability letters*, 8.

Tucker, M.J. 1963. Analysis of Records of Sea Waves. *Proceedings of the Institution of Civil Engineers*, **26.**

Tucker, M.J. 1989. Interpreting directional data from large pitch-roll-heave buoys. *Ocean Engineering*, **16, 2.**

Tucker, M.J. 1991. *Waves in ocean engineering: measurement, analysis, and interpretation.* Ellis Horwood.

Turner, R.C., Baker, M.J. 1988. A probabilistic wave model for reliability analysis. *Proceedings of the 2nd Working Conference on Reliability and Optimisation of Structural Systems.* London.

Veneziano, D., Grigoriu, M., Cornell, C.A. 1977. Vector process models for systems reliability. *Journal of Engineering Mechanics Division*, ASCE, **103, EM3.**

Weibull, W. 1951. A statistical Distribution of Wide Applicability. *Journal of Applied Mechanics*, **18.**

Wen, I.K., Chen, H-C. 1987. On fast integration for time variant structural reliability. *Probabilistic Engineering Mechanics*, **2, 3.**

Wen, Y.K., Chen, H.C. 1989. Systems reliability under time varying loads - I. *Journal of Engineering Mechanics*, ASCE, **115, 4.**

Woo-Sun, P., Chung-Bang, Y, Byun-Kun, Y. 1991. Reliability analysis of tension leg platforms by domain crossing approach. *Proceedings of the First International Offshore and Polar Engineering Conference.* **1.**

Worton, B.J. 1989. Optimal smoothing parameters for multivariate fixed and adaptive kernel methods, *Journal of Statistical Computation and Simulation*, **32.**

Zanic, V., Grubisic, I., Trincas, G. 1992. Multiattribute decision making system based on random generation of nondominated solutions: an application to fishing vessel design. PRADS. 2.1443.