



## Modelling and Statistical Analysis of Ocean-wave Data Using Transformed Gaussian Processes

I. Rychlik<sup>a</sup>, P. Johannesson<sup>a</sup> & M. R. Leadbetter<sup>b</sup>

<sup>a</sup>Department of Mathematical Statistics, Lund University, P.O. Box 117, Lund, Sweden

<sup>b</sup>Statistics Department, University of North Carolina, Chapel Hill, NC 27599, USA

(Received 11 September 1995; revised version received 8 July 1996;  
accepted 8 September 1996)

### ABSTRACT

*In this paper we present a statistical analysis of wave characteristics in oceanographic data, using a transformed Gaussian random process for modelling, and to compare theoretical distributions of wave period and amplitude with observations. A natural transform (estimable from the data) is used throughout, the model compared with that based on purely (untransformed) Gaussian assumptions. The data are measurements of a sea state in deep and shallow water, at different geographical locations.*

*One of the purposes of the paper is to use an appropriately sophisticated method to test the correctness of a Gaussian hypothesis in modelling wave data for evaluation of extreme values and for fatigue analysis. Copyright © 1996 Elsevier Science Ltd*

*Key words:* crest-trough distribution, crossings, extremes, Gaussian processes, local maxima and minima, non-Gaussian processes, ocean-waves.

### INTRODUCTION

Let  $X(t)$  be the height of the sea level at a fixed point as a function of time  $t$ . In oceanographic applications,  $X(t)$  is often viewed as a sequence of waves where each wave can be described by means of its highest and lowest values (crest, trough), or by means of its ‘amplitude’ (=crest–trough) and ‘wave period’, describing the duration of a single wave. There is no general agreement about the formal definition of a wave and two useful alternatives are

considered here. The first of these is the commonly used ‘mean downcrossing wave’, where a wave is considered as a part of a function between the consecutive downcrossings of the mean sea level, as follows:

Definition 1. Let  $X(t)$ ,  $0 \leq t \leq T$ , be a smooth function, in the sense of having a finite number of local extremes, and let  $u_*$  be a fixed reference level, usually the mean of  $X$ . Denote by  $t_i$ ,  $0 \leq t_1 < t_2 < \dots < t_n \leq T$ , the times of downcrossings of  $u_*$ . The crest and trough  $M_i^*$ ,  $m_i^*$ , say, of the  $i$ th wave are the global maximum and the global minimum of  $X(t)$ ,  $t_i < t < t_{i+1}$ , respectively. Of special interest is the so called ‘crest front amplitude’  $H_i^*$ , defined as the difference between the crest and the trough, i.e.,

$$H_i^* = M_i^* - m_i^* = \max_{t_i < t < t_{i+1}} X(t) - \min_{t_i < t < t_{i+1}} X(t),$$

and the ‘half wave period’  $T_i^*$  is defined as the distance between a downcrossing and the following upcrossing of  $u_*$  (see Fig. 1).

The important applications of wave analysis are in prediction of large waves and hence small oscillations superimposed on major waves are neglected, explaining the frequent use of ‘mean crossing waves’ considered above. However, for fatigue accumulation in marine structures, it is well known that even small oscillations can contribute to the damage and hence have to be analyzed. To do this, the distribution of the so called ‘rainflow cycles’, which are local maxima and minima of  $X$ , paired using the hysteresis properties of the material are often studied (see Rychlik [1] for a detailed discussion). We here use a simpler definition, the so called min-max cycle which is useful for such purposes and also for approximation.

Definition 2. As before, let  $X(t)$ ,  $0 \leq t \leq T$ , be a smooth function. The sequence of heights of local minima and the following maxima in  $X$ , denoted by  $m_i$ ,  $M_i$ ,  $i = 1, 2, \dots$ , respectively, is called a sequence of turning points. Then the waveform between  $m_i$  and  $M_i$  will be termed the ‘ $i$ th min-max cycle’ and referred to for brevity as the cycle  $(m_i, M_i)$ . Further, the period and the amplitude of the  $i$ th cycle is the pair  $(T_i, H_i = M_i - m_i)$ , where  $T_i$  is the time

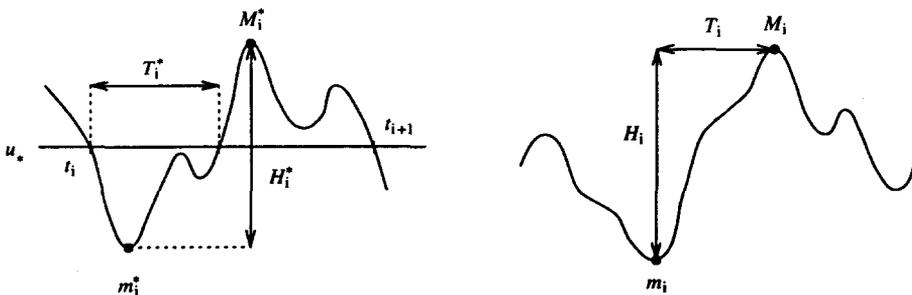


Fig. 1. Definition of mean downcrossing waves and min-max cycles.

between local minimum  $m_i$  and the following local maximum  $M_i$  (see Fig. 1).

Note that, for clarity, the properties in Definitions 1 and 2 will, as far as possible, be distinguished by the words 'wave' and 'cycle', respectively.

Some important statistical characteristics of waves are the (crest, trough)- and (wave period, amplitude)-distributions. The distributions can obviously be described empirically by fitting arbitrary parametric distributions to the observations or using nonparametric estimation techniques. However, it is also important to find a statistical model for sea state data based on more mathematical principles and it is natural to model  $X(t)$  as a stochastic processes. An important problem is to compute the distributions of wave characteristics from the properties of  $X(t)$ .

The aim of this paper is to present methods, programmed in the form of a toolbox WAVE in MATLAB<sup>®</sup>, for the analysis of wave characteristics of marine data. (The toolbox is presented in Ref. 2). We examine sea-waves from both deep and shallow water at different geographical locations. The data sets were supplied by M. Olagnon, IFREMER, France and P. Palo, US Naval Facilities Engineering Service Center.

## STOCHASTIC MODELLING OF SEA DATA

The standard assumptions for the sea state under stationary conditions is that the model  $X(t)$  is a stationary and ergodic stochastic process with mean  $E[X(t)]$  assumed to be zero and spectral density  $S(f)$ . Standard classes of spectral densities  $S(f)$  which are suitable to describe sea state data are well established from experimental studies. The important problem is the computation of the wave characteristic distributions when  $S(f)$  is given. This is not possible in general, but if  $X(t)$  is assumed to be a Gaussian process, then  $S(f)$  fully defines the statistical properties of  $X(t)$  and the wave characteristic distributions. Even for Gaussian processes, there are no known explicit exact solutions to the problem, except in some degenerate cases, but very accurate numerical approximations do exist, (e.g., Refs 3–7). These are based on Slepian model processes and a regression method. The Slepian model is an explicit random function representation of a process after a level crossing and consists of a regression term and one residual process. The regression approximation of a crossing variable is defined as the corresponding variable in the regression term and its distribution can be evaluated numerically as a finite-dimensional integral. For further details on the Slepian model and regression approximation [7]–[9]. The wave characteristic distributions derived using the regression approximation method will be referred to as the 'theoretical distributions', since they differ from the distributions defined in the mathematical model by a numerical error which can be made arbitrarily small. There are other approa-

ches to obtain the distributions based on simplifications of the model for  $X$  to the degree that the distributions can be computed analytically. The two most frequently used distributions of that type are derived by Longuet-Higgins [10] and by Cavanié *et al.* [11]. It is well known that the distributions proposed by Longuet-Higgins, Cavanié *et al.* agree well with the theoretical distributions for a narrow band Gaussian model [5]. However, since the distributions depend only on a few spectral moments the agreement is lost for wider spectra. Further, since the Gaussian model may not be an accurate description of the observed sea elevation  $X$ , it is not obvious that theoretical distributions derived under the Gaussian assumption agree better with data than the simpler analytical formulas. In Srokosz & Challenor [12] a comparison between distributions proposed by Longuet-Higgins, Cavanié *et al.* and Lindgren [3] with measured wave periods and amplitudes have been reported with the conclusion that Lindgren's method was the most accurate one. This uses a few terms in the so called Longuet-Higgins series of factorial moments to construct upper and lower bounds for the period and amplitude distribution in a Gaussian process and hence can be regarded as an alternative way to compute the theoretical distribution. Note also that Lindgren's distribution used in Ref. 12 coincides with the simplest case of regression approximation for min-max period and amplitude distribution. The Longuet-Higgins and Cavanié *et al.* distributions will not be used in this paper.

Real data  $X(t)$  seldom perfectly support the Gaussian assumption for the process  $X(t)$ . However, since the Gaussian case is well understood, the sea state is often modelled using Gaussian processes. Obviously, if the wave characteristics derived from  $S(f)$ , using the Gaussian assumption, do not agree with observations, then a broader class of models is needed. In previous work [9] we have extended the regression approximation method to processes which are functions of a vector Gaussian process,

$$X(t) = G(\tilde{X}_1(t), \dots, \tilde{X}_n(t)),$$

where  $\tilde{X}_i(t)$  are zero-mean stationary Gaussian processes with variance one and  $G$  is a deterministic function, such as the Cramér-Leadbetter envelope of a function or Morisson force. However, the complexity of the methods grows rapidly with  $n$  and for  $n > 2$  the amount of numerical computations in the approximations and data necessary to estimate the transformation,  $G$  become prohibitive.

In this paper, we shall use the simplest (but widely effective) model where  $X(t)$  is a function of a single Gaussian process  $\tilde{X}(t)$

$$X(t) = G(\tilde{X}(t)), \tag{1}$$

where  $G$  is a continuously differentiable function with positive derivative. Note that, once the distributions of crests, troughs, amplitudes or wave

periods in  $\tilde{X}(t)$  are computed, then the corresponding wave distributions in  $X(t)$  are obtained by simple variable transformations involving only the inverse of  $G$  which we shall denote by  $g$ .

## ESTIMATION OF THE TRANSFORMATION

Obviously, in order to use the model eqn (1) it is necessary to estimate the transformation  $G$  and the spectral density function  $S(f)$  from the observed sample  $X(t)$ . Since we are mainly interested in the wave characteristic distributions for the process  $X$  and not in simulations of  $X(t)$ , we need to estimate the inverse function  $g(u) = G^{-1}(u)$ . There are some parametric formulae for  $g$  proposed in the literature, e.g., by Ochi & Ahn [13]. The validation of such formulae on a wide collection of data sets from different geographical locations is of great interest and we plan such an approach in the future. Here, a natural nonparametric method will be used to estimate  $g$ .

As mentioned before, two major uses of wave characteristic distributions are to predict extreme waves and fatigue lifetimes of marine structures. In both cases, the most fundamental information about waves is the number of times a sea level crosses different levels. The expected crossing 'rate' as a function of the level  $u$ , say, will be denoted by  $\mu(u)$  and called the crossing intensity. More precisely, (writing  $\mathcal{E}$  for mean, i.e., 'expected value') let

$$\mu(u) = \mathcal{E}[\text{number of upcrossings of the level } u \text{ by } X(t), t \in [0, 1]] \quad (2)$$

$$\approx \frac{1}{T} \text{ number of upcrossings of the level } u \text{ by } X(t), t \in [0, T]$$

$$= \frac{1}{T} \sum_i 1_{(m_i, M_i)}(u) = \hat{\mu}(u),$$

where  $\hat{\mu}(u)$  is the empirical upcrossing rate observed in time  $T$ . Furthermore,  $1_{(m_i, M_i)}(u) = 1$ , if  $m_i < u < M_i$ , and zero otherwise, and  $(m_i, M_i)$  is again the sequence of min-max cycles (see Rychlik [1] for more detailed discussion). Note that the observed intensity of upcrossings  $\hat{\mu}(u)$  is a stepwise constant discontinuous function while  $\mu(u)$  is usually a smooth (differentiable) function. Consequently, the definition of an upcrossing is modified so that the empirical crossing intensity becomes a left continuous function

$$\hat{\mu}(u) = \frac{1}{T} \sum_i 1_{(m_i, M_i]}(u). \quad (3)$$

We wish to find  $g = G^{-1}$  so that the crossing intensity of our model  $X(t) = G(\tilde{X}(t))$  agrees well with the empirical crossing intensity  $\hat{\mu}(u)$ . This strategy is motivated by the following observations:

- The empirical crossing intensity, defined by eqn (3), is linked to the distribution of local extrema by the following: in ocean data, after customary smoothing in order to eliminate small oscillations, it is often observed that there are no local minima above a high level  $u$  so that each upcrossing of  $u$  corresponds to a local maximum above  $u$  and vice-versa. Hence,  $\mu(u)$  is approximately proportional to the density of high local maxima (above  $u$ ). (Similar reasoning is valid for low local minima.) If high maxima do not cluster an accurate approximation can be given, based on  $\mu(u)$ , for the distribution of a global maximum. This so called Poisson approximation, estimates the probability that there are crests above  $u$  in an interval  $[0, T]$  by  $1 - \exp(-T\mu(u)) \approx T\mu(u)$ , if  $T$  is fixed and  $T\mu(u) \rightarrow 0$  (see Leadbetter *et al.*) [14]
- In fatigue analysis one often uses a wave amplitude, such as rainflow- or min-max-amplitude  $H$  (see Definition 2) to describe the variability of a function relevant for a fatigue process. If  $H_i$  denote a sequence of amplitudes in a load, then the fatigue life time predictor is often based on the sum  $1/T \sum E[H_i^\beta]$ ,  $\beta \geq 1$ . For the rainflow and min-max-amplitudes it can be shown that  $1/T \sum E[H_i] = \int \mu(u) du$ . Similarly, we have that  $1/T \sum H_i = \int \hat{\mu}(u) du$ . For this result and a discussion of other applications of crossing intensity to fatigue analysis, see Refs 1, 15.

Consequently, if the proposed theoretical model accurately predicts the crossing intensities of measured sea states, both extreme properties of wave crests as well as the average amplitudes derived from the model will agree with observations.

We turn now to the estimation of  $g$ . It is well known that, for a zero mean Gaussian process  $\tilde{X}(t)$ , the crossing intensity  $\tilde{\mu}(u)$ , say, is given by Rice's formula,

$$\tilde{\mu}(u) = \frac{1}{2\pi} \frac{\sigma_{\tilde{X}'}}{\sigma_{\tilde{X}}} \exp\left(-\frac{u^2}{2\sigma_{\tilde{X}}^2}\right), \quad (4)$$

where  $\sigma_{\tilde{X}}^2, \sigma_{\tilde{X}'}^2$ , are the variances of  $\tilde{X}(t)$  and the derivative  $\tilde{X}'(t)$ , respectively (see Leadbetter *et al.* [14]). In the following computations, the observed data  $X(t)$  will be normalized so that the maximum of the crossing intensity is  $(2\pi)^{-1}$ . Hence, we may assume that  $\sigma_{\tilde{X}}^2 = \sigma_{\tilde{X}'}^2 = 1$ . (However, all figures presented in this paper will have correct units.) For a transformed process  $X(t) = G(\tilde{X}(t))$ , we thus have that

$$\mu(u) = \tilde{\mu}(g(u)) = \frac{1}{2\pi} \exp\left(-\frac{g(u)^2}{2}\right), \quad (5)$$

where  $g = G^{-1}$ . Note that  $\mu(u)$  has only one local maximum, at  $u = G(0)$ . Consequently, if the crossing intensity of sea waves is unimodal, as is usually observed in practice, then eqn (5) can be inverted, giving the transformation  $g$ .

More precisely, assume that the crossing intensity  $\mu(u)$  is continuously differentiable and unimodal with maximum at  $u = u_0$ . Then the transformation  $g(u)$  given by

$$g(u) = \begin{cases} \sqrt{-2\ln(2\pi\mu(u))} & \text{if } u \geq u_0, \\ -\sqrt{-2\ln(2\pi\mu(u))} & \text{if } u < u_0, \end{cases}$$

satisfies eqn (5). However, the crossing intensity  $\mu(u)$  is usually unknown and has to be estimated from  $X$ .

The algorithm used is as follows: first estimate (from  $X$ ) a mean ( $m$ ) and standard deviation ( $\sigma$ ) of  $X$ . Scale the empirical crossing intensity to have maximum  $(2\pi)^{-1}$ , and define

$$\psi(u) = -2\ln(2\pi\hat{\mu}(\sigma u + m)), \quad -5 \leq u \leq 5.$$

The range of  $\psi$  is limited, since crossings of levels more than five standard deviations away from the mean are not expected. (This can be simply modified.) Using the  $\psi(u)$ -function, define the renormalized empirical transformation

$$\hat{g}(u) = \begin{cases} \sqrt{\psi(u)} & \text{if } \hat{u}_0 \leq u \leq 5, \\ -\sqrt{\psi(u)} & \text{if } -5 \leq u < \hat{u}_0, \end{cases} \quad (6)$$

where  $\hat{u}_0$  is the value of  $u$  which minimizes  $\psi(u)$ .

Typically  $\hat{\mu}(u)$  converges to  $\mu(u)$  as the number of observed waves increases, so that

$$g(u) \approx \hat{g}\left(\frac{u - m}{\sigma}\right).$$

However, the observed crossing intensity  $\hat{\mu}$  is not a continuous function and usually has many local maxima and minima. Hence,  $\hat{g}$  is not continuous or one-to-one and hence must be smoothed in order to define the process  $X$ .

This may be done in two steps. First,  $\psi$  is slightly smoothed (using a small window) so that the smoothed version  $\psi^s(u)$  is differentiable but still close to  $\psi$ . Next, a crude estimate  $g^*(u)$  of  $g(\sigma u + m)$  is computed from

$$g^*(u) = \begin{cases} \sqrt{\psi^s(u) - \psi^s(\hat{u}_0)} & \text{if } \hat{u}_0 \leq u \leq 5, \\ -\sqrt{\psi^s(u) - \psi^s(\hat{u}_0)} & \text{if } -5 \leq u < \hat{u}_0, \end{cases}$$

where  $\hat{u}_0$  is the value of  $u$  which minimizes  $\psi^s(u)$ . Now, by smoothing (with a broad window) the function  $g^*$  and extrapolating linearly outside the interval  $[-5,5]$ , we obtain a smooth approximation  $g^s(u)$  to  $g(\sigma u + m)$ . Finally, the estimate of the transformation  $g$ , (also denoted by  $g$ ) is defined by

$$g(u) = g^s\left(\frac{u - m}{\sigma}\right), u \in R. \quad (7)$$

Two remarks conclude this section. First, it is obvious that a Gaussian model for the sea data  $X(t) = \sigma\tilde{X}(t) + m$  is obtained if  $g^s(u) = u$ . Consequently, the difference between the empirical crossing intensity  $\hat{\mu}$  and the theoretical intensity for a Gaussian model can be described using a function  $g^s(u) - u$ . The  $L^2$  distance between the observed  $g^s$  and that given by Gaussian assumption

$$e(g^s) = \left( \int_{-5}^5 (g^s(u) - u)^2 du \right)^{1/2}, \quad (8)$$

will be used as a relevant measure of deviation of the data from a Gaussian model.

Finally we stress that the important problem of extrapolating the transformation  $g^s$  to the levels which were not crossed by  $X$ , requires further study. This is a difficult problem closely related to statistical extreme value theory and the estimation of high quantiles. For the present, as mentioned above, we extrapolate  $g^s$  linearly.

## ESTIMATION OF SPECTRAL DENSITY

Assume that we have found a transformation  $g$  such that the crossing intensity  $\mu(u)$  agrees well with the observed intensity of upcrossings  $\hat{\mu}(u)$ . The sea level may then be modeled by a stochastic process  $X(t) = G(\tilde{X}(t))$ , where  $\tilde{X}(t)$  is a zero-mean Gaussian process with the spectral density  $S(f)$ . Obviously, the spectral density has to be estimated from  $\tilde{X}(t) = g(X(t))$ . This has to be done cautiously, since the data are usually sampled in both time and space. The space sampling may cause the function  $\tilde{X}$  to be constant in some intervals, which adds high frequencies to the estimate of  $S(f)$ . For the spectral estimates used in this paper, a very crude method is used to resolve this problem; first an 'irregularity factor' is estimated from the data (by dividing the maximum of  $\hat{\mu}(u)$  by the intensity of local maxima), and then high frequencies of  $S(f)$  are eliminated so that  $\tilde{X}$  has the same irregularity factor as the data.

## THEORETICAL WAVE CHARACTERISTIC DISTRIBUTIONS

In previous sections, we have proposed methods to estimate the transformation  $g$  and spectral density  $S(f)$ , so that the model  $G(\tilde{X})$  for  $X(t)$  accurately predicts the observed average wave period and amplitude. The first natural question is whether the marginal probability density of  $G(\tilde{X})$  also agrees with the data. This is not necessarily the case, it may be achieved by a suitable nonlinear time transformation (proposed by Holm & de Maré [16]). This transformation does not change the crossing intensity. We shall not discuss this transformation, since it is complicated and affects only the distribution of wave period, whereas in the examples presented in this paper, the wave period distributions computed for the process  $G(\tilde{X}(t))$  already agree well with the data. In addition, note that wave periods in  $X(t)$  are exactly equal to wave periods in the Gaussian model  $\tilde{X}(t)$ .

We turn now to the problem of computing the wave characteristic distributions for the  $X(t)$ -process. As mentioned before, if the theoretical wave characteristic distribution for a Gaussian process  $\tilde{X}(t)$  is known then the corresponding distribution in  $X(t)$  is given by a simple variable transformation. For a Gaussian process  $\tilde{X}(t)$ , the following distributions can be accurately approximated using the 'regression method': the densities of wave period  $f_{T^*}$  and min-max wave period  $f_T$ ; the joint density of min-max wave period and amplitude  $f_{TH}$ ; the joint density of a minimum and the following maximum  $f_{mM}$ ; the probability that the local maximum  $M$  or the crest of a wave  $M^*$  is higher than a fixed level  $u$ . Since the programs to compute the listed densities are only slightly modified algorithms already presented in papers [17, 7], we shall not discuss them here. We turn directly to the computation of the joint density of a trough and a crest  $f_{m^*M^*}$ . (Obviously the crest-front amplitude  $H^* = M^* - m^*$  and hence  $f_{H^*}$  can be obtained from  $f_{m^*M^*}$  by means of a numerical integration.) The algorithm to compute  $f_{m^*M^*}$  is new and we shall present the main idea in the following section, leaving details for Appendix A.

### The density of trough and crest $f_{m^*M^*}$

As before, the random model assumed for the sea state is  $G(\tilde{X}(t))$ , where  $\tilde{X}(t)$  is a zero-mean Gaussian process (taken to be ergodic for ease of interpretation) with given spectral density function  $S(f)$ . We assume that the inverse function  $g$  is known.

As in Definition 1, let  $u_*$  be the reference level. Further, let  $u, v$  be fixed levels,  $u < u_* < v$ ,  $m^*, M^*$  the heights of a trough and crest of a wave. To compute the probability

$$P(u, v) = P(m^* < u < v < M^*)$$

(and then the density  $f_{m^*M^*}(u, v) = -\partial^2/\partial u\partial v P(u, v)$ ) it is necessary to be precise about how  $P(m^* < u < v < M^*)$  is defined.

Often, in practice, only one sample) of a process  $X(t)$  is observed. All waves in  $X(t)$  are then found and, using a suitable statistical technique, the distributions of wave characteristics are estimated. This ('Palm distribution'-cf. Ref. 14) approach motivated by the ergodicity of  $X$ , defines  $P(u, v)$  simply as the proportion of waves for which  $m^* < u < v < M^*$ . Specifically, let  $X$  be a sample of  $X$  and  $0 < t_0 < t_1 < \dots$  be the upcrossing times of the reference level  $u_*$  and let  $m_i^*, M_i^*$  be the trough and the crest of the  $i$ th wave, respectively. Then the appropriate (Palm) definition of  $P(u, v)$  is

$$P(u, v) = \lim_{\tau \rightarrow \infty} \frac{\#\{t_i \in [0, \tau] : m_i^* < u < v < M_i^*\}}{\#\{t_i \in [0, \tau]\}} \quad (9)$$

$$= \frac{\text{intensity of waves with trough } < u \text{ and crest } > v}{\text{intensity of waves}}$$

(see Refs 9 and 14 for more detailed discussion). Here  $\#\{\}$  means the number of elements in the set  $\{\}$ . As noted above, differentiation of  $P(u, v)$  with respect to  $u$  and  $v$  gives the desired joint (Palm) density for the trough and crest heights of a wave.

Our approximation to  $P(u, v)$  will use the fact that the intensity in eqn (9) can be computed from the (simpler) sequence of min-max cycles  $(m_i, M_i)$  (see Definition 2).

If  $m_i < u < M_i$ , ( $m_i^* < u < M_i^*$ ), for a fixed level  $u$ , it will be convenient to say that the  $i$ th cycle (wave) crosses the level  $u$ ; similarly, if  $m_i < u < v < M_i$ , ( $m_i^* < u < v < M_i^*$ ) we shall say that the  $i$ th cycle (wave) crosses levels  $(u, v)$ . Using this convention, the  $P(u, v)$  function defined by eqn (9) is a ratio of the intensity of waves crossing levels  $(u, v)$  to the intensity of waves. Further, the intensity of waves is clearly equal to the intensity of cycles crossing the reference level ( $u_*$ ).

It can be seen that the intensity of waves crossing levels  $(u, v)$ ,  $u < v$ , can be written as the following infinite sum

$$v^*(u, v) = \lim_{\tau \rightarrow \infty} \frac{\#\{t_i \in [0, \tau] : m_i^* < u < v < M_i^*\}}{\tau} = \sum_{m, n \geq 0} v_{mn}(u, v) \quad (10)$$

where  $v_{00}(u, v)$  is the intensity of cycles crossing levels  $(u, v)$ . The remaining terms of the sum are positive and will be approximated in Appendix A. Their exact definitions can be omitted without loss of understanding of the further presentation, but are given for completeness;  $v_{01}(u, v)$  is the intensity of cycles crossing levels  $(u, u^*)$  but not  $v$ , and such that the following cycle crosses level  $v$  but not  $u^*$ , and  $v_{10}(u, v)$  is the intensity of cycles crossing levels  $(u^*, v)$ , but not  $u$ , and such that the previous cycle crosses level  $u$  but not  $u^*$ .

Furthermore, for  $n, m \geq 1$ ,  $v_{mn}(u, v)$  is the intensity of cycles crossing the reference level  $u_*$ , but not levels  $(u, v)$ , and such that the  $m$ th cycle to the left crosses level  $u$  but not the reference level  $u_*$ , with all the intermediate cycles not crossing the levels  $u$  or  $u_*$ , and the  $n$ th cycle to the right crosses level  $v$  but not  $u_*$ , with all the intermediate cycles not crossing the levels  $u_*$  or  $v$ .

Obviously, by dividing both sides in eqn (10) by the intensity of waves we obtain the probability  $P(m^* < u < v < M^*)$ . Since the intensity of local minima as well as the irregularity factor (Section 4) is assumed to be given, the joint probability density function of  $2(m+n+1)$  consecutive local extremes is needed in order to compute the intensity  $v_{m,n}(u, v)$ . However, as mentioned before, only the joint probability density  $f_{mM}$  can be accurately computed at present, giving the first term  $v_{00}(u, v)$  in the expansion eqn (10). Consequently, we write eqn (10) as

$$v^*(u, v) = v_{00}(u, v) + v_r(u, v), \quad (11)$$

where  $v_{00}$  is readily obtained, but the remainder function  $v_r(u, v)$  is usually difficult to compute. However, if the sequence of turning points form a stationary Markov chain,  $v_v(u, v)$  can be computed from the intensity  $v_{00}(u, v)$ . Usually ocean spectra are broad band and the Markov chain approximation gives very accurate results. A detailed description of the computation of  $v_r(u, v)$  using the Markov chain approximation is given in Appendix A (see also Refs 18, 19 for other applications of Markov chain approximation).

## EXAMPLES

It is generally accepted that a stationary fully-developed sea state at a deep water location can be satisfactorily modelled by a Gaussian process. As an example of such measurements, we use a data record referred to below as 'Deep water, USA'. However, even at the deep water location during a severe storm, the sea state data may show non-Gaussian characteristics. This will be illustrated by the second data set, 'Deep water, North Sea'. If measurements are at a shallow water location, one usually observes asymmetry in the heights of crests and troughs. These cannot be modelled well by a Gaussian process. Shallow water measurements will be illustrated by two data sets, 'Shallow water, USA', where crests are smaller than predicted by a Gaussian model and 'Shallow water, Africa', where both crests and troughs are higher than predicted by the Gaussian model. Figure 2 shows 300 seconds of the sample paths, and Table 1 below presents some simple statistics about the data: the sampling interval  $T_s$ , the total time of measurement  $T$ , the maximum of empirical crossing intensity  $\hat{\mu}_{\max}^* = \max \hat{\mu}^*(u)$ , the irregularity factor

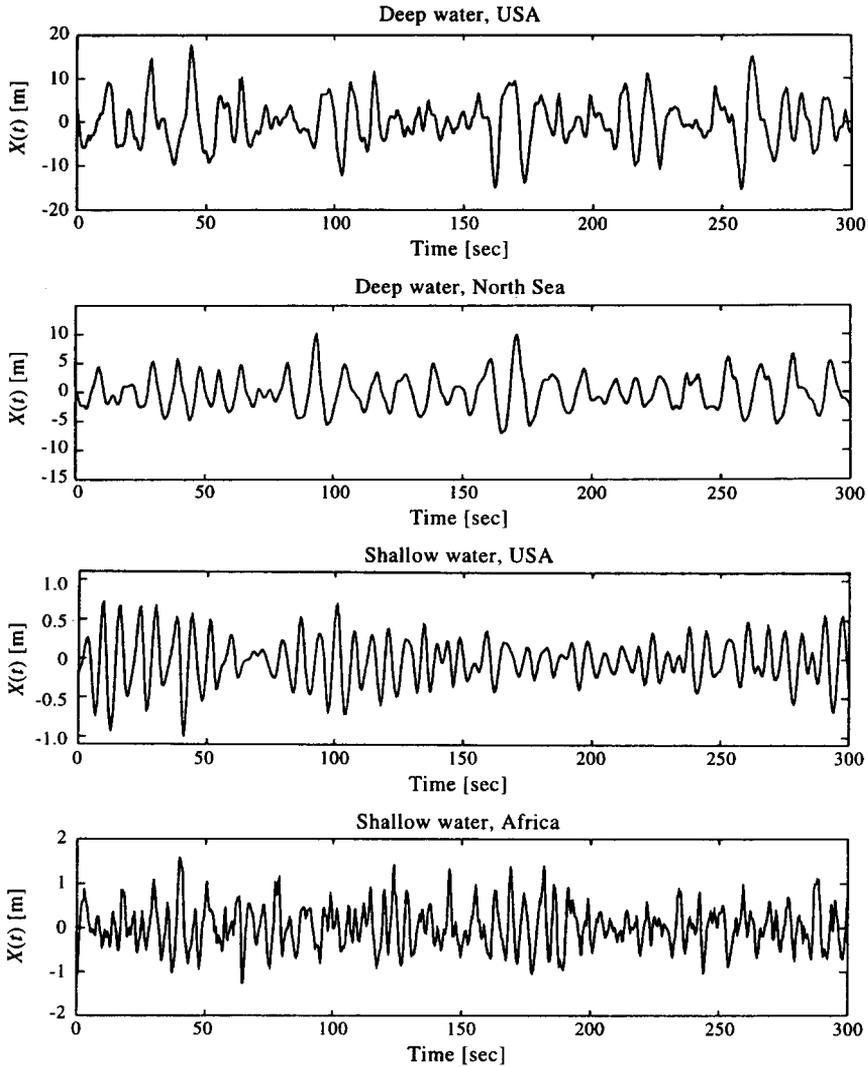


Fig. 2. Part of the sample paths.

**TABLE 1**  
Statistics about the data

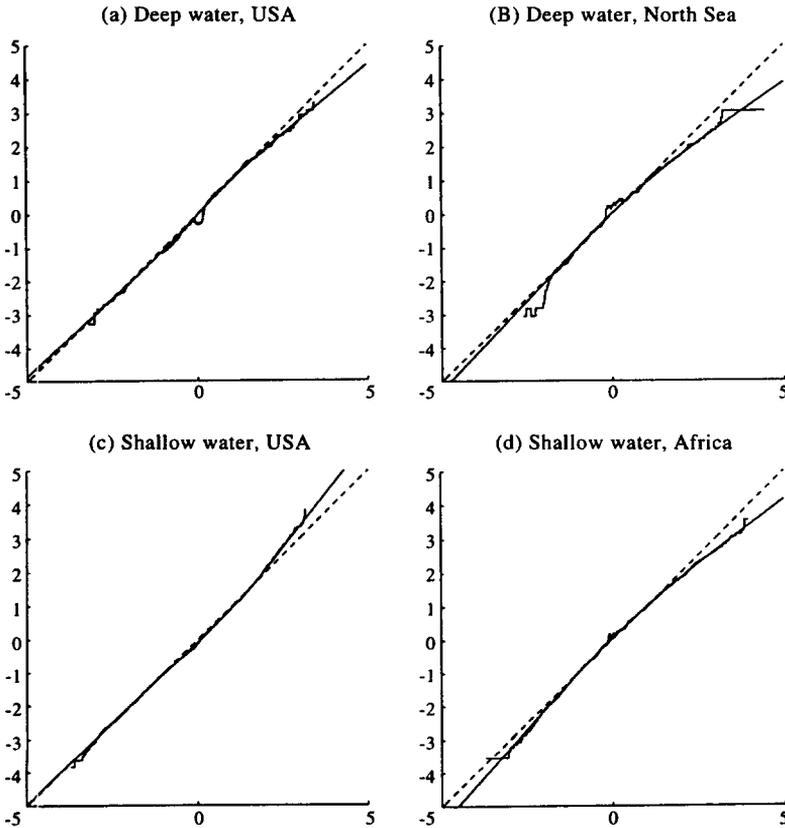
|                       | $T_s$ [sec] | $T$ [sec] | $\hat{\mu}_{\max}$ | $\alpha$ | $\sigma$ |
|-----------------------|-------------|-----------|--------------------|----------|----------|
| Deep water, USA       | 0.3         | 1809      | 0.1194             | 0.5243   | 5.03     |
| Deep water, North Sea | 0.5         | 1122      | 0.0927             | 0.6887   | 3.1193   |
| Shallow water, USA    | 0.5         | 10 240    | 0.1514             | 0.91     | 0.3280   |
| Shallow water, Africa | 0.25        | 2380      | 0.2268             | 0.4977   | 0.4730   |

$$\alpha = \frac{\text{maximum of empirical crossing intensity}}{\text{intensity of local maxima in } X}$$

and the standard deviation  $\sigma$ , (the mean is zero for all data sets).

### Fitting the transformed Gaussian model to data

In this section, we present estimates of the transformation  $g$  for the four data sets. In Fig. 3, the estimators  $g^s(u)$ , (defined just before eqn (7) and given by solid lines) are compared with a linear transformation  $g^s(u) = u$  (representing the Gaussian model and given by dashed lines) and with the empirical transformation  $\hat{g}(u)$ , (defined by eqn (6)), for the four data sets. As in all statistical analyses the accuracy of the estimators depends on the amount of data available. In the case of wave analyses, there is the additional difficulty



**Fig. 3.** A comparison between the empirical transformation  $\hat{g}$  (irregular line), the fitted transformation  $g^s$  (solid line) and the theoretical transformation for a Gaussian model, i.e.,  $g^s(u) = u$ , (dashed line).

that waves form a dependent sequence, which is particularly apparent for narrow band data.

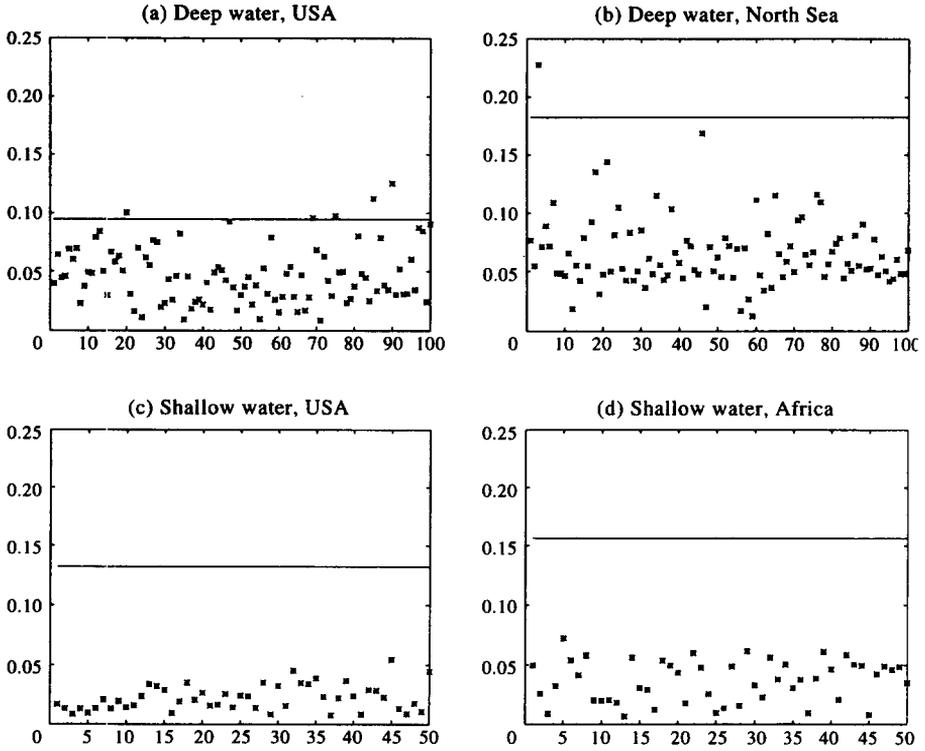
The sequences analyzed contained 216, 104, 1550 and 540 individual waves, respectively. All transformations  $g^s$ , (shown in Fig. 3), deviate from a straight line, representing a Gaussian model, which indicates that the transformed Gaussian model should be considered. The departure of  $g^s$  from the straight line is measured by  $e(g^s)$ , defined by eqn (8), and will be used as a test quantity for the non-Gaussian nature of the data. For the four data sets,  $e(g^s)$  is equal to 0.0946, 0.1823, 0.1323, and 0.1565, respectively, and one may ask how significant these values are. The following Monte-Carlo method was used to check their significance.

A Gaussian model was assumed for the data, (that is  $g^s(u) = u$ ), and four spectral densities  $S(f)$  estimated, as described in Section 5. Next, independent samples of zero-mean Gaussian processes with spectra  $S(f)$ , were simulated using an FFT-algorithm. The simulated samples, contained the same number of waves, on average, as the data sets. The function  $g^s$  was then estimated for each sample and the departure measure  $e(g^s)$  computed. The values of  $e(g^s)$  are plotted as stars in Fig. 4. Obviously, for an ergodic Gaussian process,  $e(g^s)$  is zero for an infinitely long sample and hence, the simulated values of  $e(g^s)$  (stars) show the variability of the estimator  $g^s$ , due to the limited number of observed waves. It can be seen in Fig. 4 that the values of  $e(g^s)$  decrease as the number of waves increases. Now, the horizontal solid lines in Fig. 4 mark the observed  $e(g^s)$  in the respective data sets. The significance of the departure from Gaussian assumptions in the data may now be analyzed, beginning with 'Deep water, USA' data.

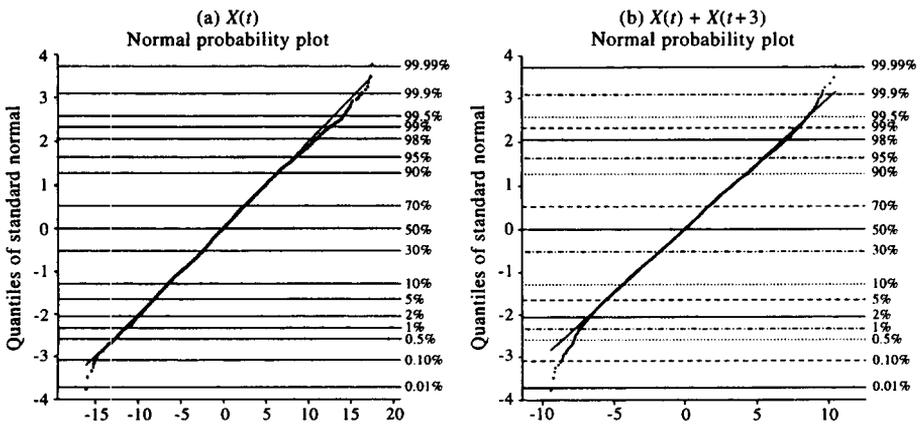
In Fig. 4a, five (of the hundred) stars lie above the observed  $e(g^s)$  for 'Deep water, USA', which could be regarded as marginal for rejection of the Gaussian assumption. However since the verdict is not clear we check the Gaussian model by using normal probability plots for the sequence  $X(t)$  and the linear combination  $X(t) + X(t+3)$  (which was the linear combination giving the highest departure from the straight line in the normal plot). The normal plots are given in Fig. 5 and do not show a significant departure from the Gaussian distribution. Consequently, on the basis of available data it seems reasonable to model 'Deep water, USA' measurements by a Gaussian process.

For the 'Deep water, North Sea' data there are only 104 waves and hence the estimated  $g^s$ , (see Fig. 3b) is not very reliable. However, the measure  $e(g^s)$  is large and indicates that one should not use a Gaussian process to model the data, see Fig. 4b.

Finally, both shallow water data series are longer and Fig. 3 and Fig. 4 show a significant departure from Gaussian assumptions. Furthermore, the estimated transformations seem to be quite reliable.



**Fig. 4.** A study of variability of  $e(g^s)$ , where stars are simulated  $e(g^s)$  for Gaussian processes with the same spectral densities as observed in data and the solid lines are the observed  $e(g^s)$ .



**Fig. 5.** 'Deep water, USA' data plotted on normal probability paper.

The data will be analyzed in detail in the following four subsections. Since ‘Shallow water, USA’ data is the longest record, it will be used to validate the transformed Gaussian model by means of a resampling technique. This is the most important example, while the three other examples are given primarily to demonstrate that the approach is effective for different types of wave data.

### ‘Shallow water, USA’ data

In order to validate the transformed Gaussian model the transformation  $g^s$  and the spectrum  $S(f)$  was estimated using only part of data, beginning at a random point and containing a third of the data record. This data length was selected since for the whole series the departure from a Gaussian assumption  $e(g^s)$  is equal to 0.1323 and we would like to have a short time sufficiently long so that 0.1323 is still significantly larger than  $e(g^s)$  values simulated for a Gaussian model. Figure 4 suggests that an appropriate length is approximately 500 waves, i.e., roughly a third of the ‘Shallow water, USA’ data. The selected part contained 588 waves. The transformation  $g^s$  is given in Fig. 6 and should be compared with that presented in Fig. 3c. It is evident that low values are transformed somewhat differently. The estimated spectral density function  $S(f)$  is also shown in Fig. 6. Note that, in the following, all theoretical computations are for the transformed Gaussian model defined by the spectral density function and transformation obtained from the part of the data and given in Fig. 6. The theoretically computed densities will be compared with the whole data set.

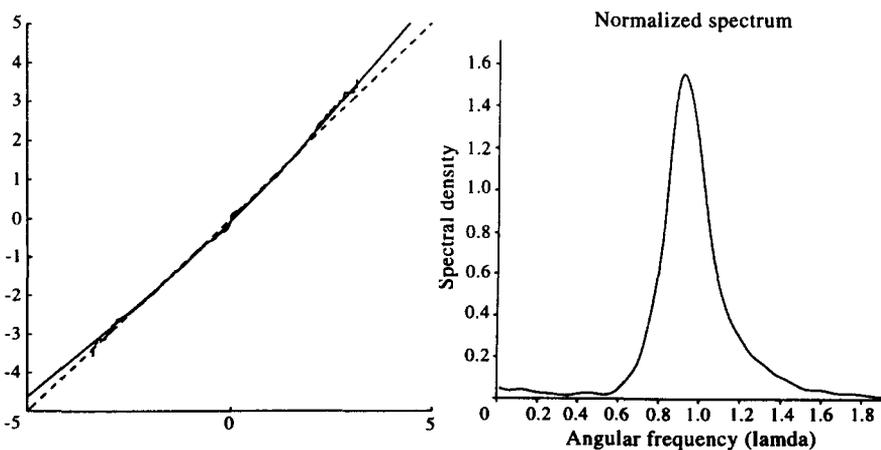
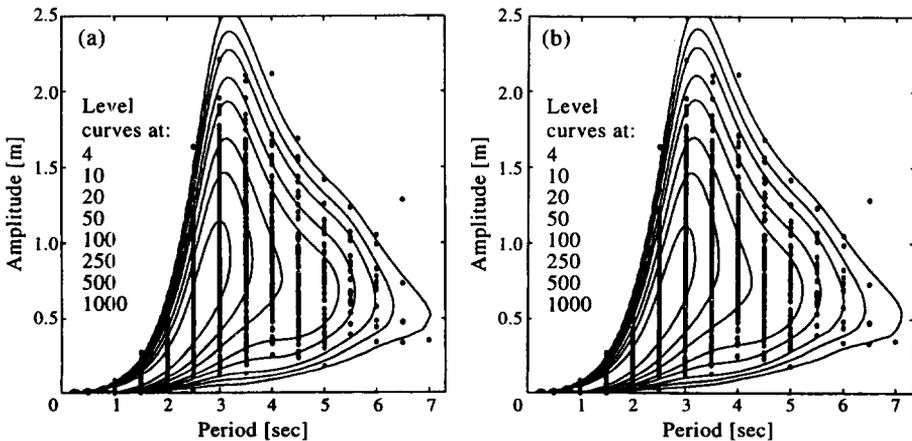


Fig. 6. The transformation  $g^s$  and a spectral density  $S(f)$  estimated from 1/3 of the ‘Shallow water, USA’ data.

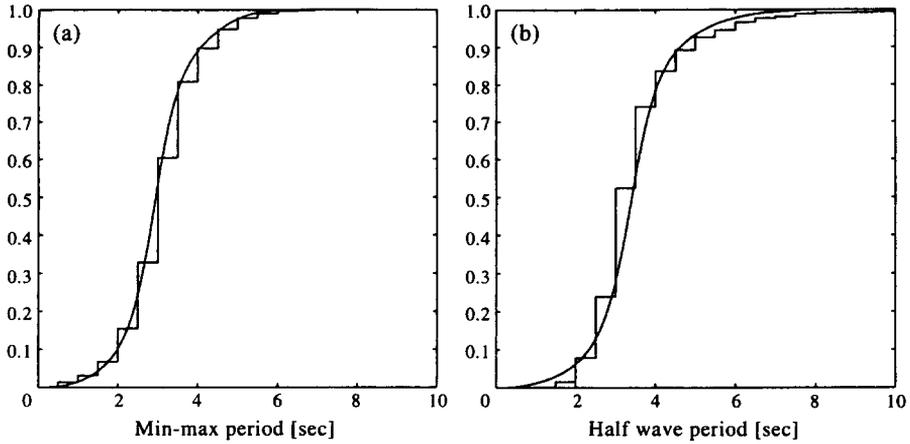
It is important to note the following two properties of the data. First, the spectral density  $S(f)$  is narrow band; this is also measured by the irregularity factor  $\alpha$ , here equal to 0.91. A high or moderate maximum with height  $u$  in a zero-mean, narrow band, Gaussian process is usually followed by a low local minimum, with height approximately  $-u$ . Next, the transformation  $g^s$  shown in Fig. 6 stays on the same side of the straight line which implies that  $g^s(u) - g^s(-u) \approx 2u$ . Consequently, wave periods and amplitudes in the transformed Gaussian model and in the Gaussian model are close to each other. This is illustrated in Fig. 7 where the isolines of the theoretical intensity of min-max cycle period and amplitude  $T, H$  is compared with the observed periods and amplitudes in the data (dots). In Fig. 7a, we have the intensity derived using the transformed Gaussian model, while in Fig. 7b the intensity as obtained using a Gaussian model, i.e.  $g^s(u) = u$ . The differences are very small.

Intensities are used instead of probability densities in these examples, since intensities are easier to compare visually with observations. (The probability densities can still be obtained from the intensities by a simple renormalization.) The intensity function has the following interpretation: the integral of an intensity over a region is the expected number of dots in the region. Thus the accuracy of the approximation could be checked by counting the dots in suitably chosen regions. However, the time sampling of a signal introduces a bias into the counts or empirical distributions, mainly because short or small waves can be misclassified but also because some moderate waves have wave periods which are too long. This occurs if, for example, local extremes have been smoothed out by the relatively large sampling intervals.

The accuracy of the theoretically computed intensities is checked in



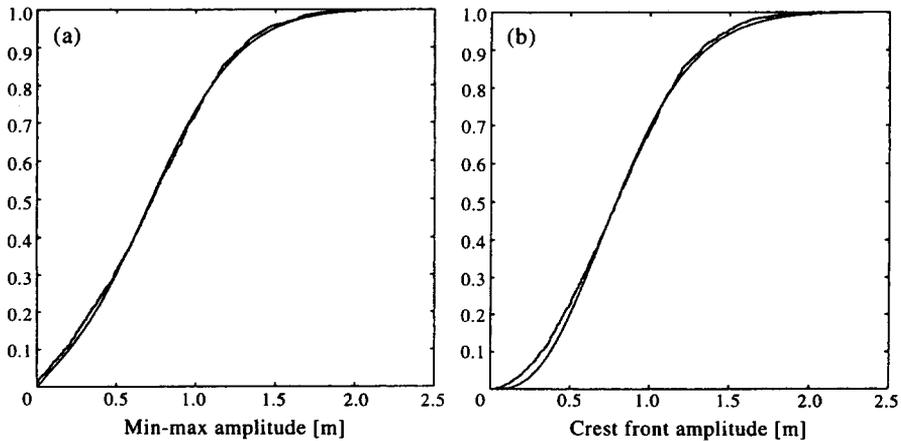
**Fig. 7.** Isolines of the intensity of min-max period and amplitude in 'Shallow water, USA' data computed using (a) the transformed Gaussian process  $X = G(\tilde{X})$  and (b) the Gaussian process  $\tilde{X}$  together with the observed periods and amplitudes in the data (dots).



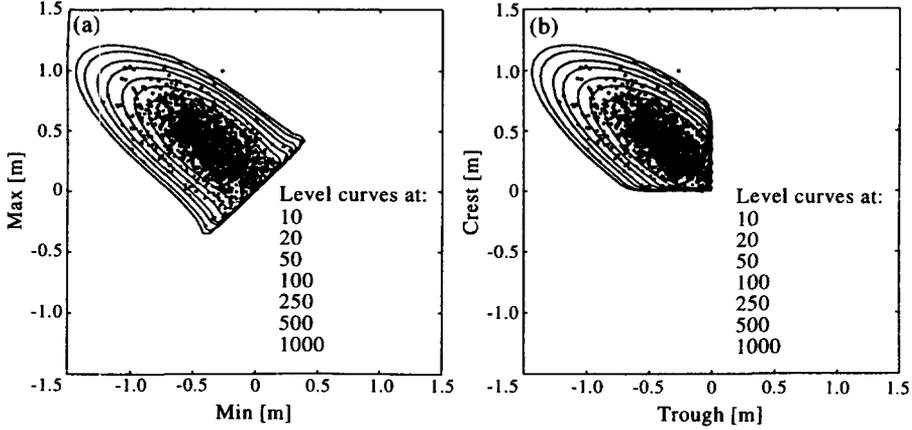
**Fig. 8.** The empirical distributions of min-max period  $T$  and half wave period  $T^*$  observed in 'Shallow water, USA' data compared with those computed for the transformed Gaussian process  $X = G(\tilde{X})$ , (a) and (b), respectively.

Fig. 8a and Fig. 9a where the marginal distributions of min-max period  $T$  and amplitude  $H$  are compared with empirical distributions obtained from the whole 'Shallow water, USA' data set. The accuracy is very good.

The density of the half wave period  $T^*$  is computed for the transformed Gaussian process using the regression method. The theoretical distribution of  $T^*$  is compared with the empirical distribution of wave period in the whole data set, in Fig. 8b, with good agreement.



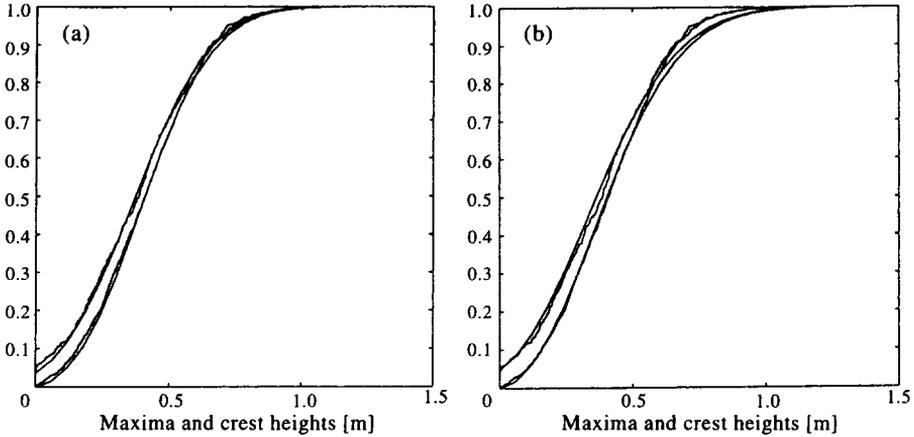
**Fig. 9.** The empirical distributions of min-max amplitude  $H$  and crest front amplitude  $H^*$  observed in 'Shallow water, USA' data compared with those computed for the transformed Gaussian process  $X = G(\tilde{X})$ , (a) and (b), respectively.



**Fig. 10.** Isolines of intensities of min-max cycles ( $m, M$ ) (a) and troughs and crests ( $m^*, M^*$ ) (b) in ‘Shallow water, USA’ data computed using the transformed Gaussian process  $X = G(\tilde{X})$ , together with the observed ( $m_i, M_i$ ) and ( $m_i^*, M_i^*$ ) in the data (dots).

The joint intensity of troughs and crests, given in Fig. 10b, was obtained using the Markov chain approximation to the sequence of turning points in  $X$  (see Appendix A for a presentation of the algorithm). The input to the algorithm is the joint intensity of each minimum with the following maximum in  $X$ . This is computed for the transformed Gaussian process and presented in Fig. 10a. Isolines of the intensities are compared with pairs ( $m_i, M_i$ ) and ( $m_i^*, M_i^*$ ) observed in the whole data set and graphically visualised as dots. One can see in both plots that the intensities are asymmetrical with crests relatively smaller than troughs which seems somewhat surprising. As before the integral of the intensity over a fixed region gives an expected number of the dots in that region. From the intensity of ( $m_i^*, M_i^*$ ) we integrate the density of crest front amplitudes  $H_i^* = M_i^* - m_i^*$ . The distribution of  $H^*$  is compared with the empirical distribution, showing the accuracy of the method (see Fig. 9b). The deviation between the distributions for small amplitudes can be explained by the fact that these amplitudes are the most biased by the time sampling of the signal.

As mentioned earlier, equally accurate approximations of wave period and amplitude densities may be obtained by using the Gaussian model. However, the important properties of waves, such as the height of a local maximum and the height of crests, are not modeled well by a Gaussian process. In Fig. 11a the empirical distributions of local maxima and crest heights are compared with the computed distributions for the transformed Gaussian process, with very good agreement. Fig. 11b presents similar distributions to Fig. 11a, with the difference that the theoretical distributions are computed



**Fig. 11.** The empirical distributions of local maxima  $M$  and crest  $M^*$  heights,  $M \leq M^*$ , observed in ‘Shallow water, USA’ data compared with those computed (a) for the transformed Gaussian process  $X = G(\tilde{X})$  and (b) for a Gaussian process, i.e.,  $g^s(u) = u$ .

using a Gaussian model. We can see that these theoretical distributions overestimate the heights of crests and local maxima.

### ‘Shallow water, Africa’ data

The ‘Shallow water, Africa’ data contains 540 waves, which seems to be sufficient for the reliable estimation of the transformation  $g$ . As mentioned before, the departure of the ‘Shallow water, Africa’ data from Gaussian assumptions is clearly significant and hence we use the transformed Gaussian model with the  $g^s$  transform presented in Fig. 3d to model the data. The estimated spectral density function  $S(f)$  is given in Fig. 12a and is seen to be broad band. The transformation  $g$  makes local maxima and minima higher and, as in the previous subsection,  $g^s(u) - g^s(-u) \approx 2u$ . Since the spectrum  $S(f)$  is much broader than in the previous example a greater difference would be expected between amplitudes in a transformed Gaussian model and in a Gaussian model. However, as can be seen in Fig. 13, where the isolines of the theoretical intensity of min-max period and amplitude  $T, H$  are compared with the observed periods and amplitudes in the data (dots) for the transformed Gaussian model (Fig. 13(a)) and the Gaussian model (Fig. 13(b)), the difference between the densities is not large. This is also illustrated in Fig. 12b where the marginal distributions of min-max amplitude, obtained by means of numerical integration of the joint densities in Fig. 13, are given in the graphs starting at zero. The theoretical distributions shown using dashed lines are computed for the Gaussian model. The line starting

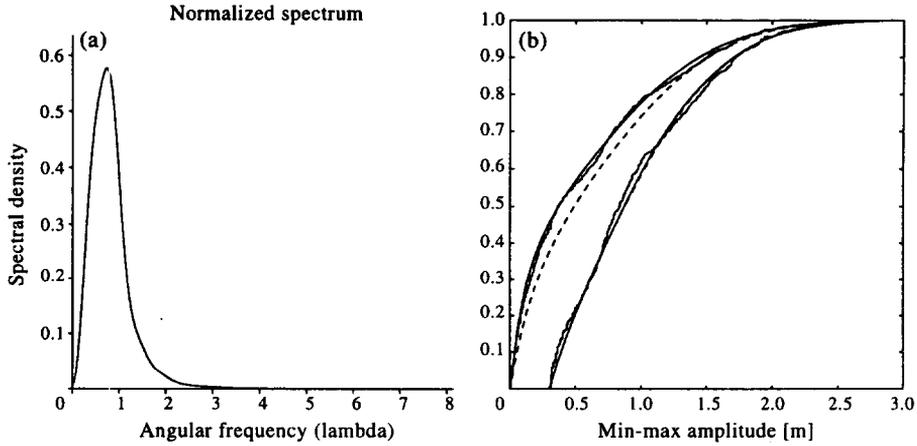


Fig. 12. (a) The spectrum  $S(f)$  estimated from ‘Shallow water, Africa’ data. (b) Comparison between the empirical distribution of min-max amplitude  $H$  and the theoretical distributions based on transformed Gaussian model (solid line) and a Gaussian model (dashed lines).

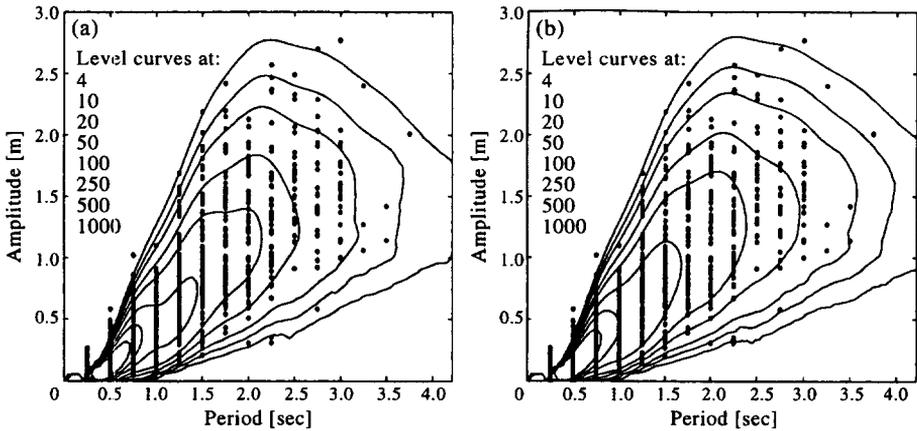
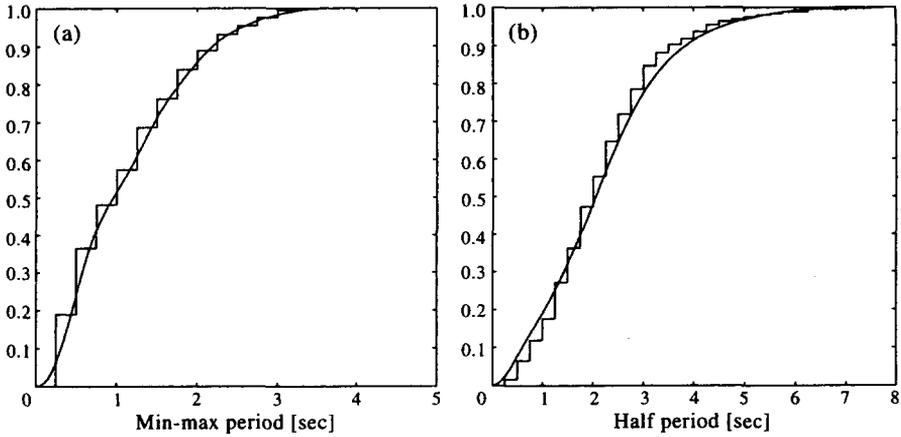
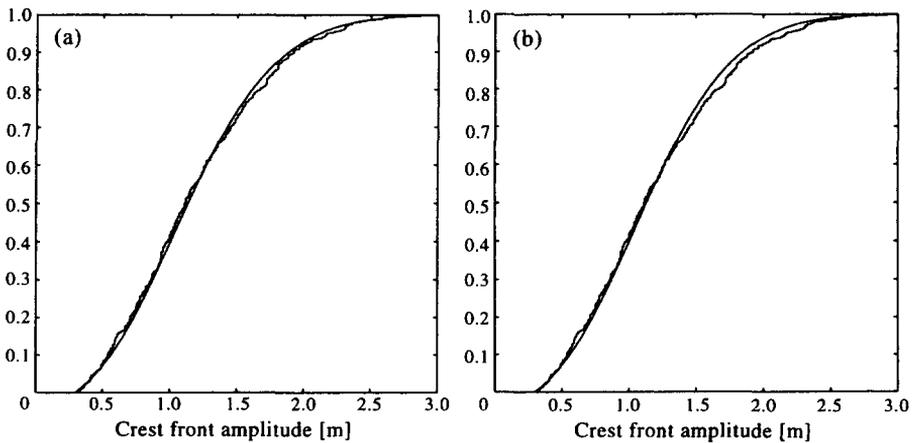


Fig. 13. Isolines of the intensity of min-max wave period and amplitude in ‘Shallow water, Africa’ data computed (a) using the transformed Gaussian process  $X = G(\tilde{X})$  and (b) for the Gaussian process  $\tilde{X}$ , together with the observed wave periods and amplitudes in the data (dots).

at zero seems to deviate significantly from the empirical data. This departure is due to the joint density of  $T, H$  having a peak for short waves with small amplitudes (see Fig. 13b) and hence, the numerical integration can have low accuracy for small amplitudes. In order to check this hypothesis, the conditional distribution of amplitudes bigger than 0.3m was computed (see the second dashed line in (Fig. 12b) and compared with the empirical distribution of amplitudes greater than 0.3m, the



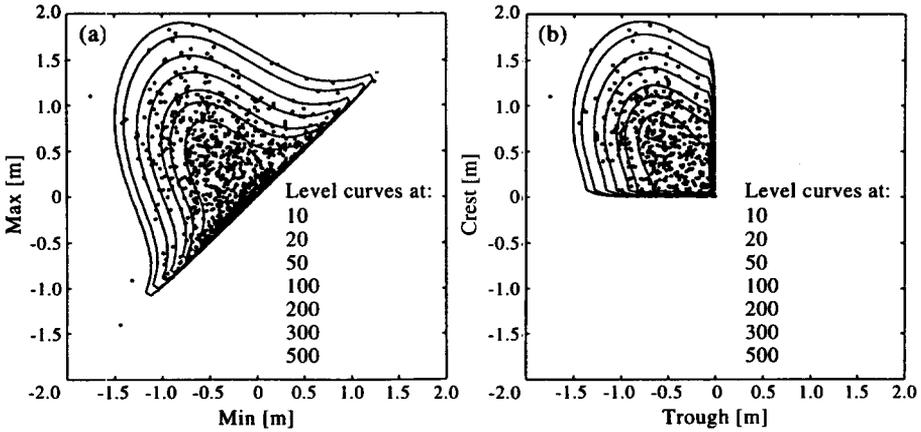
**Fig. 14.** The empirical distributions of min-max period  $T$  and half wave period  $T^*$  observed in 'Shallow water, Africa' data compared with those computed for the transformed Gaussian process  $X = G(\tilde{X})$  ((a) and (b), respectively).



**Fig. 15.** The empirical distributions of crest front amplitude  $H^*$  observed in 'Shallow water, Africa' data compared with those computed (a) for the transformed Gaussian process  $X = G(\tilde{X})$  and (b) for the Gaussian process  $\tilde{X}$ .

second irregular line. Both lines are now close to each other. Fig. 15 shows that the distributions of crest front amplitudes predicted by a transformed Gaussian process and a Gaussian process are also very similar. Consequently, we conclude that as long as the densities of wave periods and amplitudes are of interest, a Gaussian model is appropriate.

The accuracy of the theoretically computed intensities is also checked in Fig. 14, where the marginal distributions of min-max period  $T$  and half wave period  $T^*$  are compared with empirical distributions. The accuracy is very



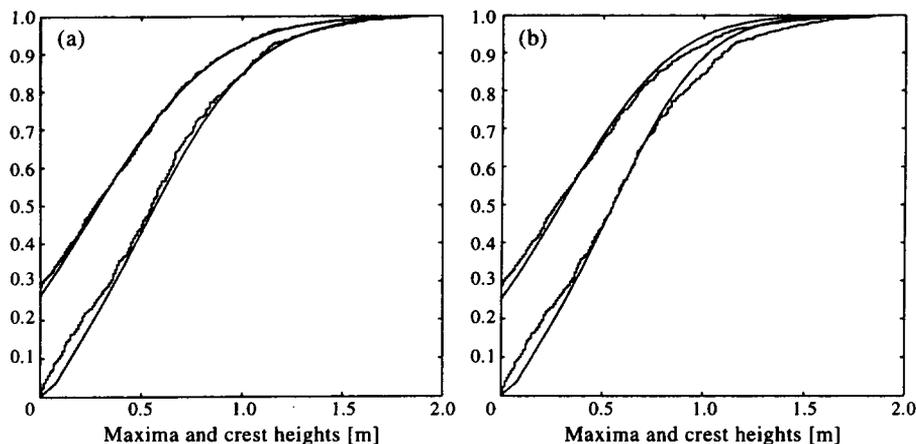
**Fig. 16.** Isolines of intensities of (a) min-max cycles ( $m, M$ ) and (b) troughs and crest ( $m^*, M^*$ ) in 'Shallow water, Africa' data computed using the transformed Gaussian process  $X = G(\tilde{X})$ , together with the observed ( $m_i, M_i$ ) and ( $m_i^*, M_i^*$ ) in the data (dots).

good. (Recall that both Gaussian and transformed Gaussian process give the same period densities.)

The joint intensity of troughs and crests is given in Fig. 16b. The joint intensity of minimum and the following maximum in  $X$ , which is the input for the algorithm computing the intensity of troughs and crests heights, is presented in Fig. 16a. Isolines of the densities are compared with pairs ( $m_i, M_i$ ) and  $m_i^*, M_i^*$ , visualised as dots. It can be seen that the intensities are asymmetrical with crests relatively higher than troughs. This is the opposite behaviour to that observed in the previous example.

From the intensity of  $m_i^*, M_i^*$ , we integrate the density of crest front amplitude  $H^* = M^* - m^*$ . The distribution of  $H^*$  is compared with the empirical distribution, showing the accuracy of the method, see Fig. 15a. Since the small amplitudes are corrupted by the sampling frequency, we have used only the amplitudes larger than 0.3m to compute the empirical distribution and compared it with the conditional distribution of  $H^*$  given that  $H^* > 0.3$ . Fig. 15b compares the same empirical distribution with the theoretical, but computed for the Gaussian model. The difference is larger than in the previous example but probably not significantly so. (Fig. 16)

As before the height of a local maximum and the height of crests are not modelled well by a Gaussian process. In Fig. 17a, the empirical distributions of local maxima and crests heights are compared with the computed distribution for the transformed Gaussian process, giving very good agreement. Fig. 17b presents the similar distributions as in Fig. 17a with the difference that theoretical distributions are computed using a Gaussian model. It can



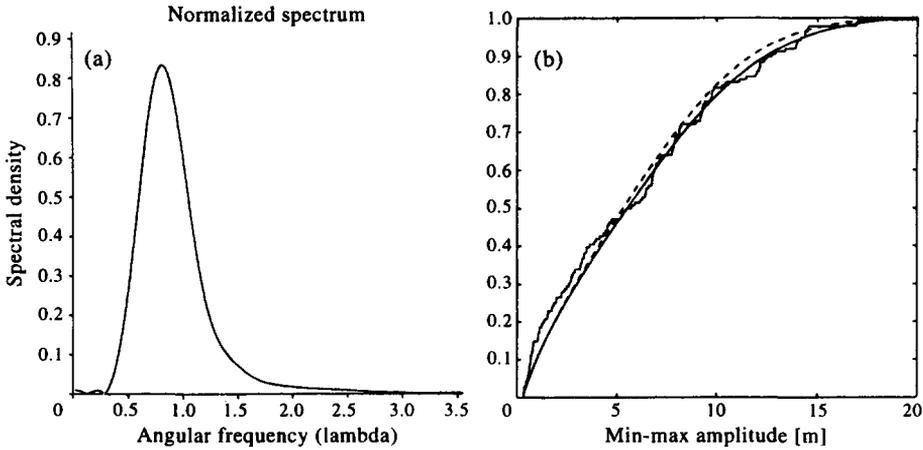
**Fig. 17.** The empirical distributions of local maxima  $M$  and crest  $M^*$  heights,  $M \leq M^*$ , observed in ‘Shallow water, Africa’ data compared with those computed for the transformed Gaussian process  $X = G(\tilde{X})$ , (a) and a Gaussian process, i.e.,  $g^*(u) = u$ , (b), respectively.

be seen that those theoretical distributions underestimate the heights of crests and local maxima.

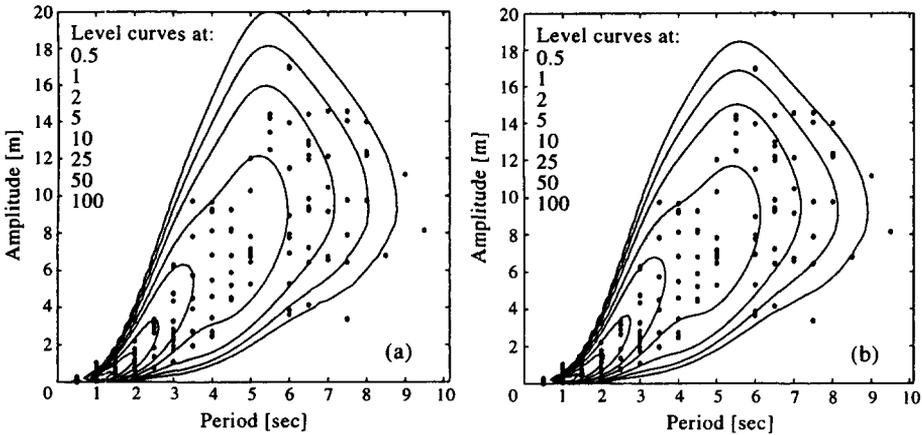
#### ‘Deep water, North Sea’ data

As we have concluded before, the ‘Deep water, North Sea’ data is clearly non-Gaussian. It contains only 104 waves which is not sufficient for reliable estimation of the transformation  $g$ . While, therefore, the use of the transformation  $g^s$ , given in Fig. 3b, is questionable, we will see that the transformed Gaussian model describes the wave characteristic distributions remarkably well.

The estimated spectral density function  $S(f)$  is given in Fig. 18a. The transformation  $g$  makes crests much higher than the Gaussian model would predict, but the negative minima are unchanged and hence one expects a more transparent difference between the amplitudes in a transformed Gaussian model and in a Gaussian model than in previous examples. This can also be seen in Fig. 19, where the isolines of the theoretical intensity of min-max period  $T$  and amplitude  $H$  is compared with observed periods and amplitudes in the data (dots) for the transformed Gaussian model (Fig. 19(a)) and a Gaussian model (Fig. 19(b)). Clearly, the Gaussian model predicts somewhat smaller waves. This is also illustrated in Fig. 18b where the marginal densities of min-max amplitudes, obtained by means of numerical integration from the joint densities in Fig. 19, are given. Further, in Fig. 21b it can be seen that even the crest front amplitudes obtained using Gaussian model are shorter than obtained for the transformed Gaussian



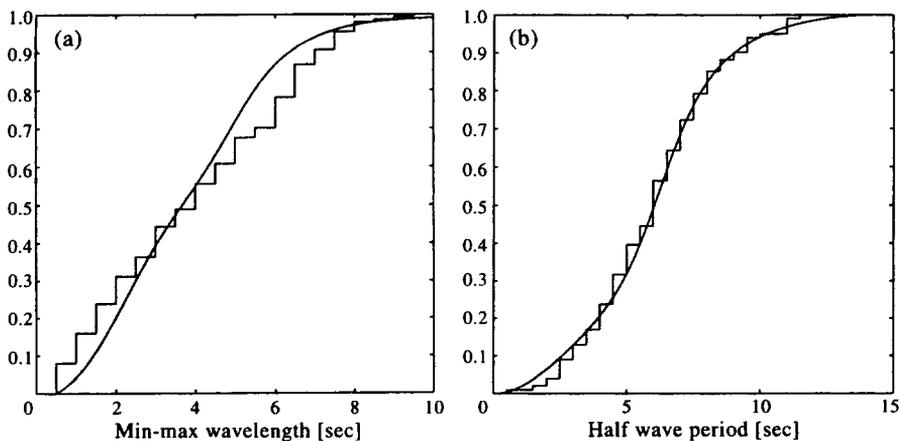
**Fig. 18.** The spectrum  $S(f)$  estimated from 'Deep water, North Sea' data (a). The comparison between the empirical distribution of min-max amplitude  $H$  and the theoretical distributions based on transformed Gaussian model (solid line) and a Gaussian model (dashed line) (b).



**Fig. 19.** Isolines of the intensity of min-max period and amplitude in 'Deep water, North Sea' data computed using the transformed Gaussian process  $X = G(\tilde{X})$  (a) and the Gaussian process  $\tilde{X}$  (b) together with the observed periods and amplitudes in the data (dots).

model (see Fig. 21a). However, taking into account that the transformation  $g$  is based only on 104 waves, and that the differences are not so large, it still seems reasonable to conclude that, as long as the densities of wave periods and amplitudes are of interest, a Gaussian process could be used to model the data.

The accuracy of the theoretically computed intensities is also checked in



**Fig. 20.** The empirical distributions of min-max period  $T$  and half wave period  $T^*$  observed in 'Deep water, North Sea' data compared with those computed for the transformed Gaussian process  $X = G(\tilde{X})$ , (a) and (b), respectively.

Fig. 20, where the marginal distributions of min-max period  $T$  and half wave period  $T^*$  are compared with the empirical distributions. The agreement is acceptable for this type of data.

We turn now to the joint intensity of troughs and crests, which is given in Fig. 22b. The joint intensity of minimum and the following maximum in  $X$ , which is the input for the algorithm computing the intensity of troughs and crests heights, is presented in Fig. 22a. Isolines of the densities are compared with pairs  $(m_i, M_i)$  and  $m_i^*, M_i^*$ , visualised as dots. It is clear that the intensities are asymmetrical with crests relatively higher than troughs.

From the intensity of  $m_i^*, M_i^*$  we integrate the density of crest front amplitudes  $H_i^* = M_i^* - m_i^*$ . The distribution of  $H^*$  is compared with the empirical distribution, showing the accuracy of the method, see Fig. 21a. Fig. 21b compares the same empirical distribution with the theoretical, but computed using the Gaussian model. The difference is greater than in the previous example but not large enough to be important. (Fig. 22)

As before the height of a local maximum and the height of crests are not modelled well by a Gaussian process. In Fig. 23a the empirical distributions of maxima and crests heights are compared with the computed distribution for the transformed Gaussian process, showing very good agreement. Fig. 23b differs from Fig. 23a in that the theoretical distributions are computed using a Gaussian model. It is apparent that these theoretical distributions underestimate the heights of crests and local maxima.

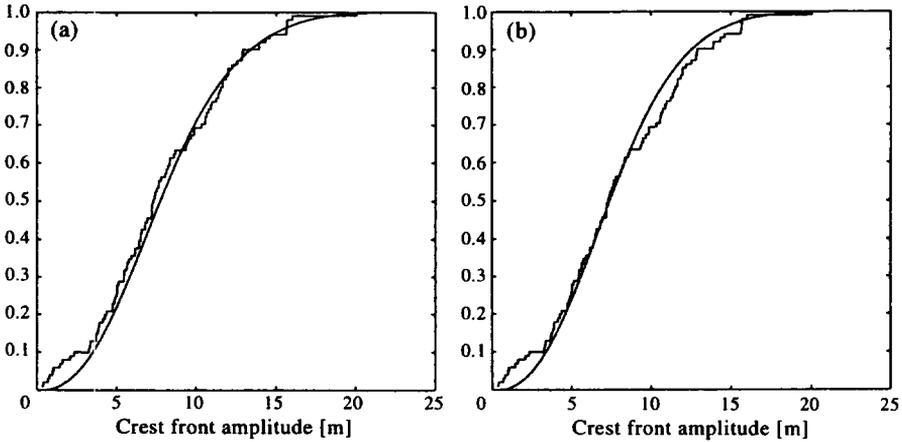


Fig. 21. The empirical distributions of crest front amplitude  $H^*$  observed in ‘Deep water, North Sea’ data compared with those computed for (a) the transformed Gaussian process  $X = G(\tilde{X})$  and (b) the Gaussian process  $\tilde{X}$ .

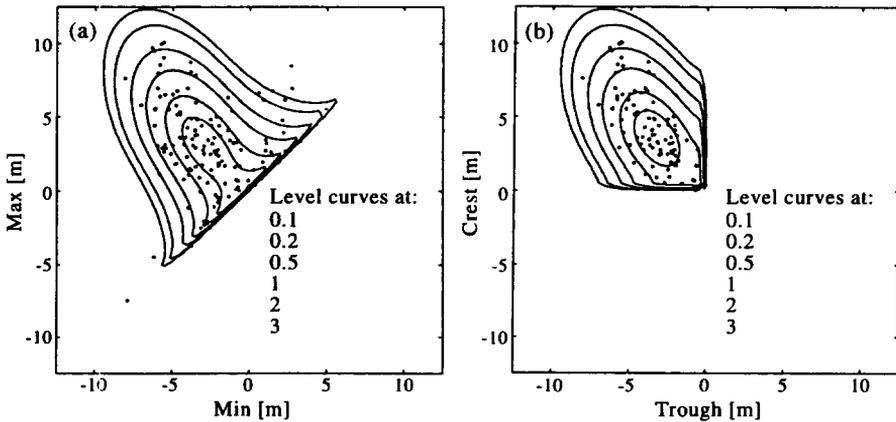
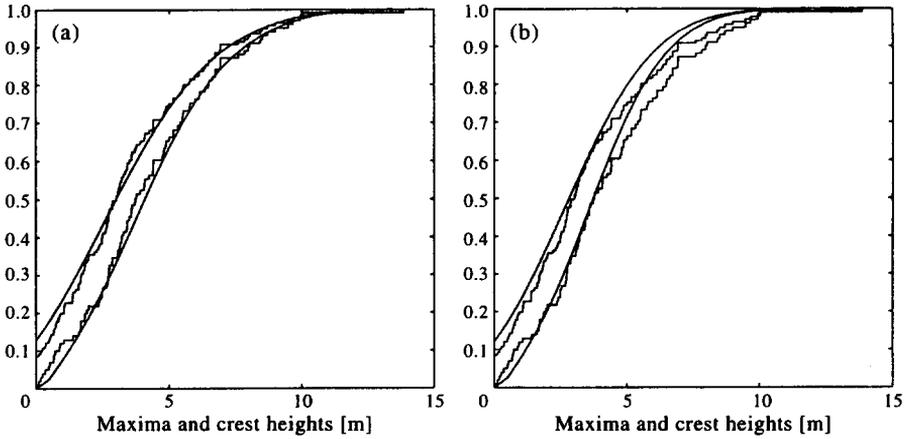


Fig. 22. Isolines of intensities of (a) min-max cycles ( $m, M$ ) and (b) troughs and crests ( $m^*, M^*$ ) in ‘Deep water, North Sea’ data computed using the transformed Gaussian process  $X = G(\tilde{X})$ , together with the observed ( $m_i, M_i$ ) and ( $m_i^*, M_i^*$ ) in the data (dots).

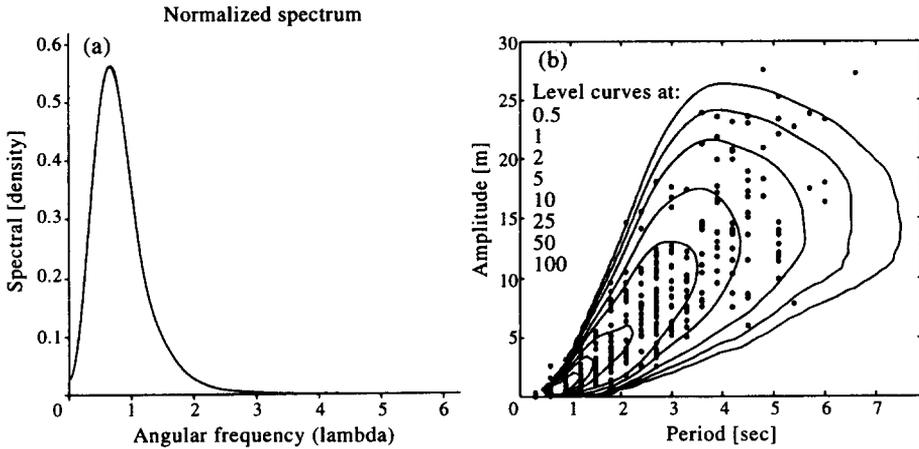
‘Deep water, USA’ data

As has been previously concluded, a Gaussian model cannot be rejected for the ‘Deep water, USA’ data. However, since the estimated  $g^s$  function deviates considerably from the straight line, one cannot expect good agreement between the empirical and theoretical distributions based on a Gaussian model.

Figure 24b shows the isolines of the theoretical intensity of min-max

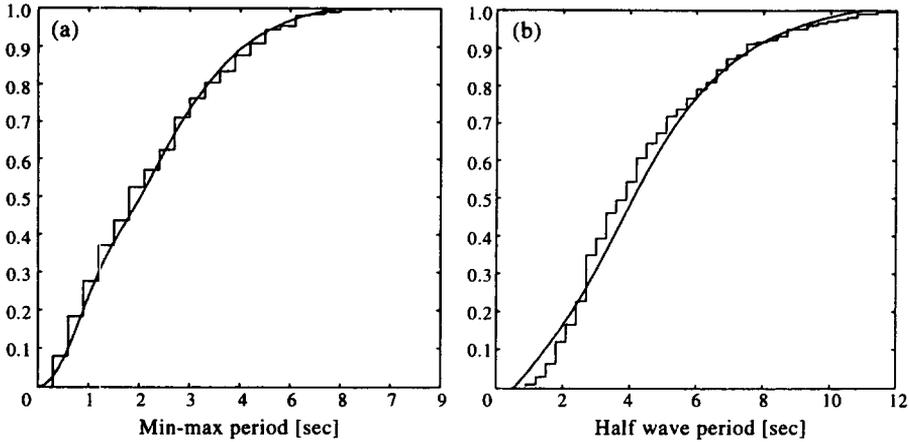


**Fig. 23.** The empirical distributions of local maxima  $M$  and crest  $M^*$  heights,  $M \leq M^*$ , observed in 'Deep water, North Sea' data compared with those computed for the transformed Gaussian process  $X = G(\tilde{X})$  (a) and for a Gaussian process, i.e.,  $g^*(u) = u$ , (b).

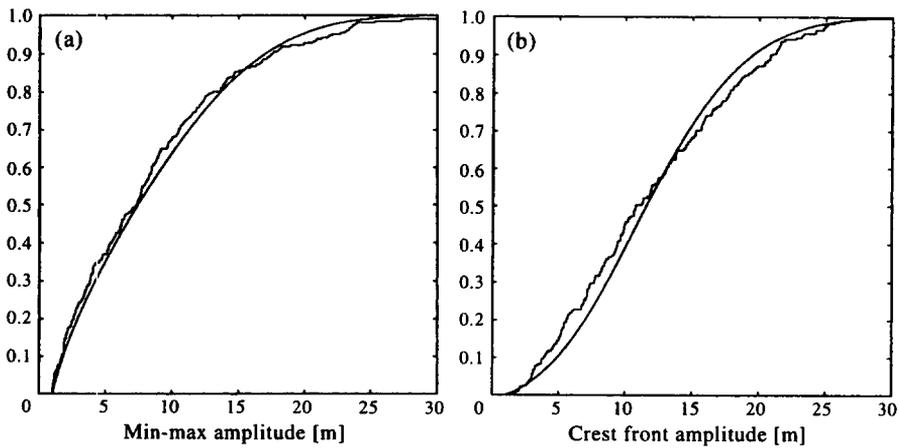


**Fig. 24.** The spectrum  $S(f)$  estimated from 'Deep water, USA' data (a). Isolines of the intensity of min-max wave period and amplitude in 'Deep water, USA' data computed using a Gaussian model together with the observed wave periods and amplitudes in the data (dots) (b).

period  $T$  and amplitude  $H$ , together with the observed wave periods and amplitudes in the data (dots). The accuracy of the Gaussian model is checked by comparing the theoretical distribution of min-max amplitudes with the empirical one (see Fig. 26a). As in previous examples, after excluding small amplitudes the agreement between the lines is acceptable. The accuracy of the theoretically computed intensities is found to be very good in Fig. 25, where the marginal distributions of min-max period  $T$  and half wave period  $T^*$  are compared with empirical distributions. (Fig. 26)



**Fig. 25.** The empirical distributions of min-max period  $T$  and half wave period  $T^*$  observed in 'Deep water, USA' data compared with distributions computed for a Gaussian models (a), (b), respectively.



**Fig. 26.** The empirical distributions of min-max ( $H$ ) and crest front ( $H^*$ ) amplitudes observed in 'Deep water, USA' data compared with those computed for a Gaussian model ((a), (b), respectively).

We turn now to the joint intensity of troughs and crests which is given in Fig. 27b. The joint intensity of minimum and the following maximum in a Gaussian process, which is the input for the algorithm computing the intensity of trough and crest heights is presented in Fig. 27a. Isolines of the densities are compared with pairs  $(m_i, M_i)$  and  $m_i^*, M_i^*$ , visualised as dots.

From the intensity of  $m_i^*, M_i^*$  we integrate the density of crest front amplitudes  $H_i^* = M_i^* - m_i^*$ . The distribution of  $H^*$  is compared with the

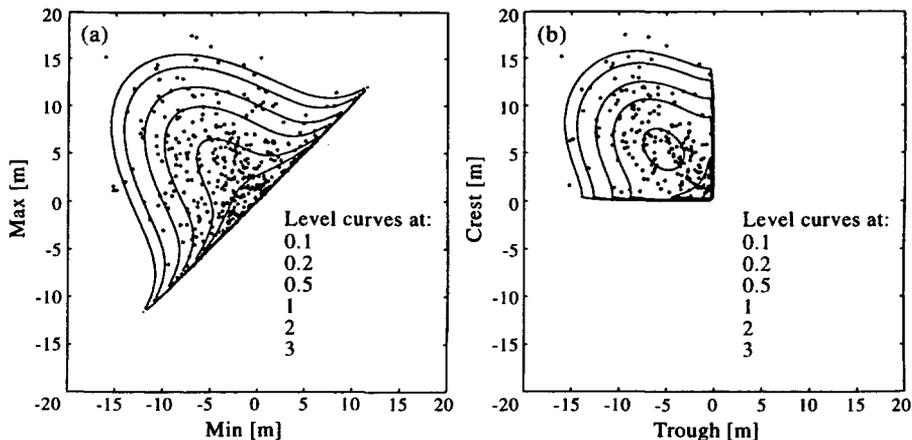


Fig. 27. Isolines of intensities of min-max cycles  $(m, M)$  (a) and troughs and crests  $(m^*, M^*)$  (b) in 'Deep water, USA' data computed using a Gaussian model together with the observed  $(m_i, M_i)$  and  $(m_i^*, M_i^*)$  in the data (dots).

empirical distribution, in Fig. 26b. However, there is less agreement in this case between the theoretically computed and observed distributions.

Finally, we present the theoretical distributions of heights of local maxima and crests for a Gaussian process, in Fig. 28a and for a transformed Gaussian process in Fig. 28b. As expected, the transformed Gaussian process predicted the distributions more accurately than a Gaussian model.

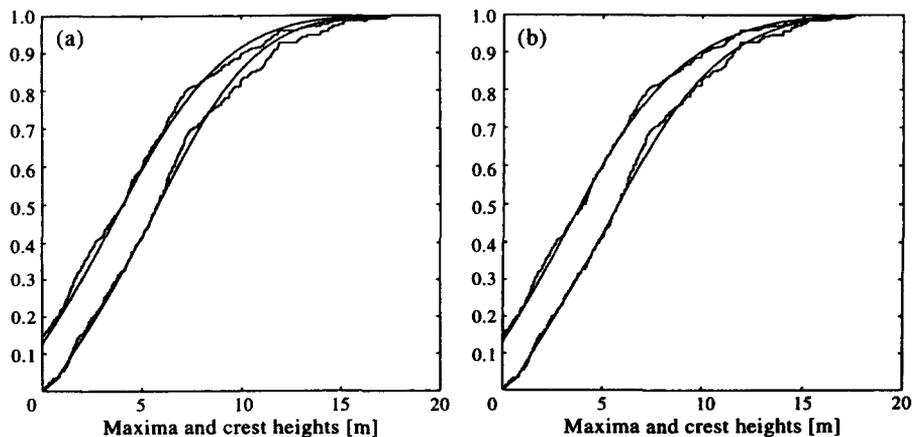


Fig. 28. The empirical distributions of local maxima  $M$  and crest  $M^*$  heights,  $M \leq M^*$ , observed in 'Deep water, USA' data compared with those computed for a Gaussian model, i.e.,  $g^x(u) = u$ , (a) and the transformed Gaussian process  $X = G(\tilde{X})$ , (b), respectively.

## CONCLUSIONS

We have examined four data sets from shallow and deep water locations and found that the transformed Gaussian process predicts the distributions of different wave characteristics very well, especially the heights of local maxima and wave crests, for which simpler Gaussian models give less satisfactory approximations. This implies that the transformed Gaussian model should be used if the extreme wave properties have to be modeled. Further it is evident that the wave period and amplitude distributions are satisfactorily approximated by a simpler Gaussian model, at least for the data sets examined here. This indicates that, if the fatigue life analysis is to be performed, then a Gaussian process could be an acceptable model. Finally, we stress that if one needs to model a sea state as input to a linear or nonlinear system it may happen that the transformed Gaussian model is still too simple and one needs to consider a more general class of processes, e.g.,  $X(t) = G(\tilde{X}_1(t), \tilde{X}_2(t))$  where  $\tilde{X}_i(t)$  are zero-mean Gaussian processes.

## ACKNOWLEDGEMENTS

The authors are grateful to Georg Lindgren and Michel Olagnon for very helpful discussions. Furthermore, Igor Rychlik partially worked on this paper during his stay as research scholar at the Department of Mathematics, University of Queensland for which he is very grateful.

IR's research was supported in part by the Office of Naval Research under Grant N00014-93-1-0841; PJ was supported by the Swedish Research Council for Engineering Sciences grant 908911 TFR 91-747; MRL's research was supported in part by the Office of Naval Research under Grant N00014-93-1-0043.

## REFERENCES

1. Rychlik, I., Note on cycle counts in irregular loads. *Fatigue Fracture Engineering Materials & Structures*, 1993, **16**, 377–390.
2. Rychlik, I. and Lindgren, G., WAVE Analysis Toolbox—a tutorial. Statures Research Report, Dept. of Math. Statist., Lund University, 1996.
3. Lindgren, G., Wave-length and amplitude in Gaussian noise. *Advances in Applied Probability*, 1972, **4**, 81–108.
4. Lindgren, G., Discrete wave-analysis of continuous stochastic processes. *Stochastic Process Applications*, 1973, **1**, 81–108.
5. Lindgren, G. and Rychlik, I., Wave characteristic distributions for Gaussian waves—wave-length, amplitude, and steepness. *Ocean Engineering*, 1982, **9**, 411–432.

6. Rychlik, I., Regression approximations of wavelength and amplitude distributions. *Advances in Applied Probability*, 1987, **19**, 396–430
7. Rychlik, I. and Lindgren, G., CROSSREG—a computer package for extreme value and wave analysis, with reliability applications. *Probability Engineering & Information Science*, 1993, **7**, 125–148.
8. Slepian, D., On the zeros of Gaussian noise. In *Time Series Analysis*, ed. M. Rosenblatt Wiley, New York, 1963, pp.104–115.
9. Lindgren, G. and Rychlik, I., Slepian models and regression approximations in crossings and extreme value theory. *International Statistics Review*, 1991, **59**, 195–225.
10. Longuet-Higgins, M.S., On the joint distribution of wave periods and amplitudes in random wave field. *Proceedings of the Royal Society*, 1983, **389**, 241–258.
11. Cavanié, A., Arhan, M. and Ezraty, R., A statistical relationship between individual heights and periods of storm waves. In *Proc. Conf. On Behaviour of Offshore Structures, Trondheim*, Norwegian Institute of Technology, Trondheim, Norway, 1976, pp. 354–360.
12. Srokosz, M.A. and Challenor, P.G., Joint distributions of wave height and period, a critical comparison. *Ocean. Engineering*, 1987 **14**, 295–311.
13. Ochi, M.K. and Ahn, K., Probability distribution applicable to non-Gaussian random processes. *Probabilistic Engineering Mechanics*, 1994, **9**, 255–264.
14. Leadbetter, M. R., Lindgren, G. and Rootzén, H., *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag, New York, 1983.
15. Rychlik, I., On the ‘narrow-band’ approximation for expected fatigue damage. *Probabilistic Engineering Mechanics*, 1993, **8**, 1–4.
16. Holm, S. and de Maré, J., Generation of random processes for fatigue testing. *Stochastic Processes & Applications*, 1985, **20**, 149–156.
17. Rychlik, I., The two barriers problem for continuously differentiable processes. *Advances Applied Probability*, 1992, **24**, 71–94.
18. Freundahl, M. and Rychlik, I., Rainflow analysis—Markov method. *International Journal of Fatigue*, 1993, **15**, 265–272.
19. Rychlik, I., Lindgren, G. and Lin, Y.K., Markov based correlations of damages in Gaussian and non-Gaussian loads. *Probabilistic Engineering Mechanics*, 1995, **10**, 103–115.

## APPENDIX A: MARKOV APPROXIMATION OF TROUGHS AND CRESTS INTENSITY

Let  $m_1, M_1, m_2, M_2, m_3, \dots$ , be a sequence of turning points of a waveform assumed to have been digitally recorded in a finite number of discrete levels  $u_1 > u_2 > \dots > u_n$ . (This is the usual case for the sea wave data). Let the reference level  $u_*$  be also one of these levels  $u_N$ , say,  $1 < N < n$ . The input to the algorithm is the index  $N$  and the  $n \times n$  intensity matrix  $F^{mM} = \lambda_{ij}$ , where

$$\lambda_{ij} = \lim_{\tau \rightarrow \infty} \frac{\{t_k^m \in [0, \tau] : m_k = u_i, M_k = u_j\}}{\tau}$$

$$= \text{intensity of local maxima} \cdot P(m = u_i, M = u_j)$$

$$\approx \text{intensity of local maxima} \cdot \frac{f_{mM}(u_i, u_j)}{\sum_{i,j} f_{mM}(u_i, u_j)},$$

where  $f_{mM}(u_i, u_j)$  is the original (continuous) Palm density of minima and the following maxima. Further,  $t_k^m$  is a time for the  $k$ th local minimum. The intensity matrix  $F^{mM}$  is first used to compute the transition matrices  $P$  and  $\hat{P}$  from a minimum to a maximum and a maximum to a minimum, respectively, i.e.  $P = (p_{ij})$  and  $\hat{P} = (\hat{p}_{ij})$ , where

$$p_{ij} = P(M_k = u_j | m_k = u_i) = \frac{\lambda_{ij}}{\sum_l \lambda_{il}} \text{ and } \hat{p}_{ij} = P(m_{k+1} = u_j | M_k = u_i) = \frac{\lambda_{ji}}{\sum_l \lambda_{li}}.$$

The sequence of ‘discretized turning points’  $m_1, M_1, m_2, M_2, \dots$ , is called a Markov chain of turning points if it is a time homogenous Markov chain with transition matrices  $P, \hat{P}$ . Before turning to the description of the algorithm, it is important to specify more exactly how the discrete levels for the local extremes are chosen from  $u_1, \dots, u_n$ , in order to properly define crossings of levels  $(u_i, u_j)$  by the discretized cycles  $(m_k, M_k)$ . (However, if the grid is fine any scheme can be used.) Here we use the following scheme: if the local minimum is in the interval  $[u_i, u_{i-1}]$  then the discrete value is  $u_i$ , and if the local maximum is in  $(u_i, u_{i-1}]$  then the discrete value is  $u_{i-1}$ . (Note that  $u_i < u_{i-1}$ .) Consequently, the cycle amplitudes are systematically overestimated. Let  $(u_i, u_j)$ ,  $u_i \leq u_N \leq u_j$  and  $u_i < u_j$ , be fixed levels, note  $(i > j)$ . For simplicity, write  $u = u_i$  and  $v = u_j$ . To calculate the intensity  $v^*(u, v)$  defined by eqn (10), the following submatrices of  $P, \hat{P}$  are needed: if  $u_N < v$ , i.e.,  $j < N$ , then

$$A^+ = (p_{kl}), \quad k = j + 1, \dots, N, \quad l = j, \dots, N - 1, \quad (12)$$

$$B^+ = (\hat{p}_{kl}), \quad k = j, \dots, N - 1, \quad l = j + 1, \dots, N,$$

and if  $u < u_N$ , i.e.,  $N < i$ , then

$$(A^-)^T = (p_{kl}), \quad k = N + 1, \dots, i, \quad l = N, \dots, i - 1, \quad (13)$$

$$(B^-)^T = (\hat{p}_{kl}), \quad k = N, \dots, i - 1, \quad l = N + 1, \dots, i.$$

Define also the following vectors,

$$(e^+)^T = \left[ \sum_{l=1}^{j-1} p_{(j+1)l} \dots \sum_{l=1}^{j-1} p_{Nl} \right], \quad e^- = \left[ \sum_{l=n}^{i+1} \hat{p}_{Nl} \dots \sum_{l=n}^{i+1} \hat{p}_{(i-1)l} \right]. \quad (14)$$

The matrix  $A^+$  contains the conditional probabilities of transitions from a minimum in  $[u_N, u_j)$  to a maximum in  $(u_N, u_j]$ ,  $B^+$  contains the conditional probabilities of transitions from a maximum in  $(u_N, u_j]$  to a minimum in  $[u_N, u_j)$ . Similarly,  $(A^{-T})$  contains the conditional probabilities of transitions from a minimum in  $[u_i, u_N)$  to a maximum in  $(u_i, u_N]$ , while  $(B^-)^T$  is a matrix of transition probabilities from a maximum in  $(u_i, u_N]$  to a minimum in  $[u_i, u_N)$ . The vector  $e^+$  contains the conditional probabilities that given a minimum in  $[u_N, u_j)$  the following maximum is higher than  $u_j$ , while  $e^-$  contains the conditional probabilities that given a maximum in  $(u_i, u_N]$  the following minimum is smaller than  $u_j$ . Further, we require the intensity of cycles which cross the critical level  $u_N$  but not the levels  $u$  and  $v$ . Those are given by the submatrix of  $F^{mM}$ ,  $F$ , say,

$$F = (\lambda_{kl}), k = N + 1, \dots, i, l = j, \dots, N - 1. \quad (15)$$

Now, using the Markov property of the chain of turning points, it is not too difficult to see that for  $m, n \geq 1$  the intensities  $v_{mn}(u, v)$ , defined in eqn (10), can be computed as follows,

$$v_{mn}(u, v) = e^-(A^-B^-)^{m-1}A^-FB^+(A^+B^+)^{n-1}e^+,$$

and by standard matrix algebra, we obtain the following formula for  $\sum_{m,n \geq 1} v_{mn}(u, v)$

$$\sum_{m,n \geq 1} v_{mn}(u, v) = e^-(I - A^-B^-)^{-1}A^-FB^+(I - A^+B^+)^{-1}e^+, \quad (16)$$

where  $I$  is an identity matrix. Turning now to formulae for  $\sum_{m \geq 1} v_{0m}(u, v)$ ,  $\sum_{m \geq 1} v_{m0}(u, v)$ , first define the following two vectors  $d^+$  and  $d^-$  which contain the intensities of cycles crossing levels  $(u_N, u_j)$  but not  $u_i$  and intensities of cycles crossing levels  $(u_i, u_N)$  but not  $u_j$ , respectively,

$$(d^+)^T = \left[ \sum_{l=1}^{j-1} \lambda_{(N+1)l} \dots \sum_{l=1}^{j-1} \lambda_{il} \right], d^- = \left[ \sum_{l=n}^{i+1} \lambda_{lj} \dots \sum_{l=n}^{i+1} \lambda_{l(N-1)} \right]. \quad (17)$$

In a similar way to eqn (16) it follows that

$$\sum_{m \geq 1} v_{0m}(u, v) = d^-B^+(I - A^+B^+)^{-1}e^+ \quad (18)$$

$$\sum_{m \geq 1} v_{m0}(u, v) = e^-(I - A^-B^-)^{-1}A^-d^+. \quad (19)$$

Consequently, by adding eqn (16), eqn (18) and eqn (19) we obtain the the term  $v_r(u, v)$ . Next, the term  $v_{00}(u, v)$  is evaluated from the matrix  $F^{mM}$ , i.e.,

$$v_{00}(u, v) = \sum_{l=1}^{j-1} \sum_{k=n}^{i+1} \lambda_{kl}, \quad (20)$$

and the function  $v^*(u, v) = v_{00}(u, v) + v_r(u, v)$  can now be computed. Note that matrices and vectors defined in eqns 12–17 and eqn (20), used to compute  $v^*(u, v)$ , depend on indices  $i, j$  and have to be recomputed for each pair of levels  $u = u_i, v = u_j$ . Finally, when the matrix with values of  $F^* = (v^*(u_i, u_j))$ , say, is evaluated, then the intensity of troughs and crests is computed by means of numerical differentiations, i.e.,

$$\begin{aligned} \text{intensity of waves : } m^* &= u_i, M^* = u_j \\ &= v^*(u_i, u_j) + v^*(u_{i-1}, u_{j+1}) - v^*(u_{i-1}, u_j) \\ &\quad - v^*(u_i, u_{j+1}). \end{aligned}$$