Yim

Ocean Engng, Vol. 19, No 3, pp. 219–238, 1992.
Printed in Great Britain.

0029–8018/92 $5.00 + .00
Pergamon Press Ltd

PRACTICAL METHODS OF EXTREME VALUE ESTIMATION BASED ON MEASURED TIME SERIES FOR OCEAN SYSTEMS

SOLOMON C. S. YIM,* R. M. BURTON† and M. R. GOULET†

*Ocean Engineering Program, Department of Civil Engineering, †Department of Mathematics, Oregon State University, Corvallis, OR 97331, U.S.A.

Abstract—Three practical methods for computing the expected maxima of Gaussian time series for ocean system analysis are developed. These methods utilize Pierce's sample scaling concept to overcome the maxima counting and correlation difficulties, but minimize the associated complexity and uncertainties. The first (Direct) method removes the dependence on the envelope for maxima estimation of the time series by directly operating on the time series itself. The second (Poisson clumping) employs the notion of sample scaling factor, but requires neither computing the envelope nor segmenting. The third (Log-fit) is a simple logarithm curve fitting, using the slowly varying, logarithmic growth property of the expected maximum. The accuracy and computational efficiency of these methods are examined. The Direct method and the Poisson clumping method are found to have comparable accuracy. Employment of the envelope does not improve the accuracy of the estimate in practice. Hence, the Direct method and the Poisson clumping method should be preferred. The Poisson clumping method is more efficient than both the Direct method and Pierce's method because of its straightforwardness in implementation. The Log-fit method is the simplest to implement, and computationally the most efficient. Its accuracy is acceptable for many engineering preliminary designs.

1. INTRODUCTION

IN THE design of ocean structural systems, detailed analytical models are often difficult and costly to develop due to the complexity of the systems. General purpose computer programs are usually employed to perform precursory parametric studies to develop preliminary design models. Experimental tests are routinely performed in model basins to provide statistical data for estimating the life-time extreme and fatigue loads of the preliminary structural systems. The time series of the structural responses are analyzed using standard data analysis packages to extract design parameters of the structural systems for selected sea-states. Due to the physical and economical limitations of model tests, the experimental time-series of each sea-state are generally of significantly shorter duration than those anticipated over the field/service life of the structural systems. Analytical and/or numerical methods have to be employed to extrapolate statistical parameters of the responses to the desired durations. The extrapolated extreme load estimates are then used in the final design of the structural systems.

Despite the abundance of research work in this area, a simple practical method for estimating the extreme loads is not yet available. Major difficulties for developing such a method include the lack of a consistent procedure for: (1) identifying "major peaks" (local maxima) and "major zero crossings" (especially for the case of wide-band time series); and (2) accounting for the correlation between nearby maxima.

A method proposed by Pierce (1985) made significant progress towards resolving

these two problems. By analyzing the associated envelope of the time series, the need for identifying and analyzing individual maxima is removed. By determining an equivalent number of independent samples of the envelope, the method accounts for the interdependence among nearby maxima and converts the complicated extreme value distribution of the original time series to that of a simple process with identically and independently distributed sample points. However, Pierce's method is not conceptually simple, and the code development effort and the amount of numerical computations can be quite involved. The envelope function of the time series has to be computed using a digital filter. The expected maxima used in determining the sample scaling factor may be dependent on the detail of the computation procedure (segmenting the original time series). There is no simple procedure to determine an "optimal" segments number.

This paper presents three alternate methods and examines their accuracy compared to Pierce's method. The goal is to develop simple, practical methods that utilize the concepts proposed by Pierce for overcoming the maxima counting and correlation difficulties, and to minimize the complexity and uncertainties in the computations.

2. PROBLEM STATEMENT AND BACKGROUND

The problem of estimating the expected extreme value of a stochastic response of specific duration in the field given a representative record of relatively short length from a model test can be mathematically stated as follows. Given a known (measured) time series $\{x_i\}$ of p points, how can the maximum value M_n of other time series with n points with the same statistical properties be estimated?

This problem is interesting because the results involving extreme values often are asymptotic in nature, that is, the probability density of M_n for large n approaches a specific distribution (Gumbel, 1958; Galambos, 1978; Leadbetter *et al.*, 1983; Castillo, 1988). However, this value of n may often be too large for practical applications.

It is well known that if $\{X_i\}_1^\infty$ is an independent and identically distributed (iid) sequence of random variables with distribution function

$$F(x) = P(X \leq x) \tag{1}$$

then, a new random variable, M_n equal to the maximum of the X_is, that is

$$M_n = \max_{1 \leq i \leq n} X_i , \tag{2}$$

has a distribution function

$$\begin{aligned} G_n(x) &= P(M_n \leq x) \\ &= P(X_1 \leq x, \ldots, X_n \leq x) \\ &= P(X_1 \leq x) \ldots P(X_n \leq x) \\ &= F^n(x) . \end{aligned} \tag{3}$$

By the Fisher and Tippett theorem, the limiting distribution for $G_n(x)$ converges to one of the three (Frechet, Weibull, or Gumbel) extreme value distributions (Leadbetter *et al.*, 1983). To take advantage of the asymptotic properties of the distribution of the

extreme value, the time series considered are usually assumed to be stationary, zero mean, and Gaussian.

To describe the probability distribution of the extreme value of a time series, it is necessary to determine the rate of occurrence and the correlation structure of the local maxima. For a narrow-band process, there is only one maximum per zero up-crossing. Thus in early studies of extreme value distribution, a narrow-band auto-spectrum and statistical independence of maxima of the time series are assumed (Longuet-Higgins, 1952; Cartwright and Longuet-Higgins, 1956). The average number of maxima per unit time can then be computed from the mean zero-crossing period. In addition, the maxima are assumed to be statistically independent, in which case the probability density function of the maxima approaches the Rayleigh distribution. Hence, the extreme value distribution of a time series of given duration can be determined by order statistics in terms of zero-crossing period and mean, which are obtained by straightforward analysis of the given time series.

For wide-band processes, there is no straightforward method for determining the rate of occurrence of the local maxima and their correlation structure due to the multiple number of maxima between zero crossings. Cartwright and Longuet-Higgins (1956) and Ochi (1973) showed that if the maxima are assumed independent, in the limit, the distribution of the maxima approaches the Rice distribution (Rice, 1944, 1945; Ochi, 1990), which is controlled by two parameters—the variance and a bandwidth parameter. If the time series is narrow-band, the bandwidth parameter approaches zero, and the distribution of the maxima approaches the Rayleigh distribution. If the time series is wide-band, the distribution of the maxima approaches a truncated normal distribution. To evaluate the bandwidth parameter, it is necessary to either compute the distribution of the velocity and acceleration histories, or the first three non-zero moments of the spectral density function. However, these higher parameter methods are found to be sensitive to measurement noise due to the need to approximate velocity and acceleration traces (or higher moments in the frequency domain). Because of its sensitivity, the Rice distribution is not often employed in practice.

To circumvent the maxima counting difficulty and to avoid extrapolating high moments, the notion of analyzing the associated envelope function of the time series was examined (Tayfun, 1981; Naess, 1982; Pierce, 1985). Given that the time series is Gaussian, the distribution of the envelope is Rayleigh regardless of bandwidth. The sample distribution is fully characterized by computing the variance of the envelope. To account for the correlation among samples, a sample scaling factor needs to be determined. For this purpose, Pierce proposed to divide the envelope function of the time series into a number of sufficiently long segments such that the maximum from each segment can be considered independent. The maximum value of each envelope segment is then determined and the maxima are averaged to yield an expected maximum. The theoretical number of independent maxima that produce the mean expected maximum is computed from the Rayleigh distribution. The ratio of this theoretical number of samples to the actual number of samples in the envelope segment yields a sample-scaling factor. With both the distribution and sample-scaling factor determined, the expected maximum of an arbitrary time series of given duration can then be computed using order statistics.

While Pierce's envelope method succeeded in circumventing the difficulties of maxima

counting and correlation structure among samples, improvements are needed in two areas. First, the method can computationally be rather involved, requiring a digital filter to compute the Hilbert transform to obtain the associated envelope function. Second, there is no apparent rule for selecting an optimal number of segments in computing the time scaling factor, or a minimum length for each segment to determine the expected maximum. For a time series of fixed duration, increasing the number of segments improves the accuracy of the expected maximum estimate, but at the expense of reducing segment length and hence the independence of the maxima of nearby segments. On the other hand, increasing the segment length improves the independence of the maxima of nearby segments, but at the expense of reducing the number of segments and hence the accuracy of the estimate of the expected maximum. These competing interests have to be balanced. It is possible that for a range of segment lengths, the expected maximum may be insensitive to the exact value used. However, this range may be dependent on the parameters of the time series considered. At present this range, if it exists, has to be determined numerically by trial and error.

In this study, three simple alternate methods for computing the expected maximum of a time series of a given duration are presented. The first method removes the dependence on the envelope for maxima estimation of the time series by directly operating on the time series itself. The second method employs the concept of using sample-scaling factor, but does not require computing the envelope or segmenting. The third method is a simple logarithm curve fitting, which is straightforward and quick to implement. The effectiveness of these methods in estimating the expected value of the maxima of time series are examined. These proposed alternative methods, and a review of Pierce's envelope method, are presented in the following sections.

3. PIERCE'S ENVELOPE METHOD

Pierce's method for estimating the extreme value of an arbitrary bandwidth Gaussian process from a single measured (p-point) time series can be divided into three main components. First an envelope function of a measured time series is computed. Second, using the (p-point) envelope function, a quantity called the sample-scaling factor (SSF) is found. Third, using the sample-scaling factor, the maximum value of a time series of arbitrary length (n-points) from the same ensemble is estimated.

The envelope

If $x(t)$ is a Gaussian time series, let the analytic envelope, $a(t)$, be defined as

$$r(t) = \sqrt{x(t)^2 + y(t)^2} \tag{4}$$

where

$$y(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(s)}{t-s}\,\mathrm{d}s \tag{5}$$

is the *Hilbert transform* of $x(t)$. Since $x(t)$ is Gaussian, $y(t)$ is also Gaussian, being limits of linear combinations of $x(t)$ for various t. It can be shown that $x(t)$ and $y(t)$ are uncorrelated and independent (Bendat and Piersol, 1986). The distribution of $r(t)$ is that of a Rayleigh random variable. This is so because X and Y are iid $N[0,1]$ random

variables (Ochi, 1990). The analytic integral in Equation (5) is usually computed using a digital filter (Pierce, 1985).

The sample-scaling factor

The sample-scaling factor (SSF) can be computed with the following steps:

(I) Divide the envelope function into a number of n_a-point segments and compute an empirical expected maximum value $E[M_a]$ for the selected segment length.

(II) Determine an effective number n_e of independent sample points in a segment based on $E[M_a]$ and the Rayleigh distribution of the envelope function.

(III) Compute the sample-scaling factor n_e/n_a.

Step I is accomplished by first subdividing the analytic envelope (of size p) into m segments (sub-intervals), of length $n_a = p/m$. The maximum value of the time series within each segment is determined and then averaged over all segments to get the $E[M_a]$. In other words, if the envelope function is given by $\{e(n)\}_{n=1}^{p}$, then define $M(i)$, $i = 1, \ldots, m$ by

$$M(i) = \max\left\{e(n) \mid n = 1 + \left[\frac{p}{m}\right](i-1), \ldots, \left[\frac{p}{m}\right] i\right\} \tag{6}$$

where [ ] denotes the greatest integer function.

The expected maximum for a time series with length n_a is given by

$$E[M_a] = \frac{1}{m}\sum_{i=1}^{m} M(i)\,. \tag{7}$$

Note that if m is small then there will be few terms in the average given by Equation (7) and the estimate will be poor. Conversely, if m is very large then the maximal within the subintervals will not be independent.

An effective number of independent sample points is then determined using information about the distribution of the envelope process and the theoretical expected maximum value. Let M be the maximum of the process in a segment. For large numbers of segments (that is, for m large), $E[M_a]$ approaches the theoretical expected maximum, which is given by

$$E[M_e] = \int_0^\infty P(M_e > t)\,\mathrm{d}t\,. \tag{8}$$

If the points in the segment are independent, then this distribution is easily computed. Let the probability distribution of the extreme value (maximum) of the n-point envelope segment be

$$F_{M_e}(t) = P(M_e \leqslant t)\,. \tag{9}$$

Then

$$F_{M_e}(t) = F_X(t)^n \tag{10}$$

where

$$F_X(t) = P(X \leqslant t) \tag{11}$$

is the probability distribution of the envelope function. So, the expected extreme value is given by

$$E[M_e(n)] = \int_0^\infty [1 - F_X(t)^n]\,dt. \tag{12}$$

An effective number of independent sample points, n_e, for the segment length considered is determined by choosing n_e such that the difference between the empirically computed expected value and the theoretical value based on the assumed probability distribution, $D(n)$, given by Equation (13) below, is minimized.

$$D(n) = |E[M_e(n)] - E[M_a]|. \tag{13}$$

The ratio of the theoretical number of independent points n_e to the total number of sample points in the segment, $n_a = p/m$, yields the SSF, which is the quotient of the effective number of samples and the actual number of samples,

$$\text{SSF} = \frac{n_e}{n_a} = \frac{n_e m}{p}. \tag{14}$$

As discussed above, if the time series is Gaussian then the envelope is Rayleigh. In this case

$$F_x(t) = 1 - e^{-t^2/2\sigma^2}, t \geq 0 \tag{15}$$

where

$$\sigma = E[X]\sqrt{\frac{2}{\pi}}. \tag{16}$$

Now substituting Equation (15) into Equation (12) we get

$$E[M_e(n)] = \int_0^\infty [1 - (1 - e^{-t^2/2\sigma^2})^n]\,dt. \tag{17}$$

Figure 1 shows an example of the graph of Equation (17) for $\sigma = 1$. Figure 2 shows the relationship between the expected extreme value $E[M_a]$ and the effective number of independent points n in the form of a graph of $y = E[M_e(n)]$.

Extreme value estimation

To estimate the extreme value of a time series of length n, the effective number of independent points contained in a time series is first obtained by multiplying the length n by the SSF. Equation (17) can then be used to compute the theoretical expected maximum value $E[M_e(n*\text{SSF})]$ based on the effective sample length $n*$SSF.

4. ALTERNATIVE ESTIMATION METHODS

Pierce's method has been shown to yield reasonably accurate estimates for both simulated time series and experimental data. However, the method is conceptually and computationally involved due to the need for an envelope function and segmenting. For wide-band processes, the method tends to over-estimate the expected maximum

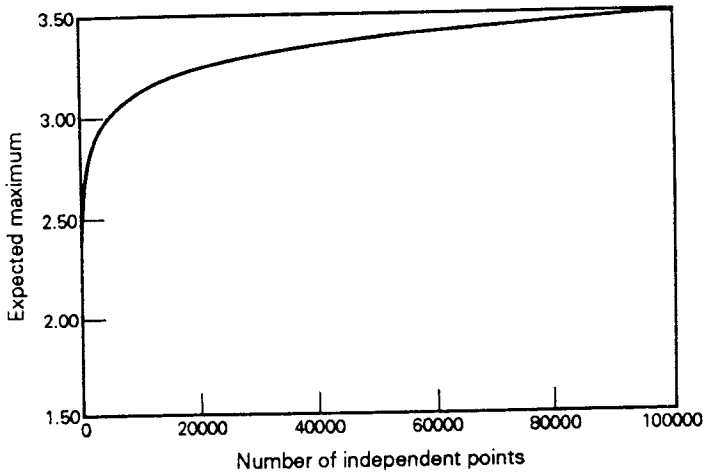

FIG. 1. Expected maximum value vs number of independent points for the envelope process.

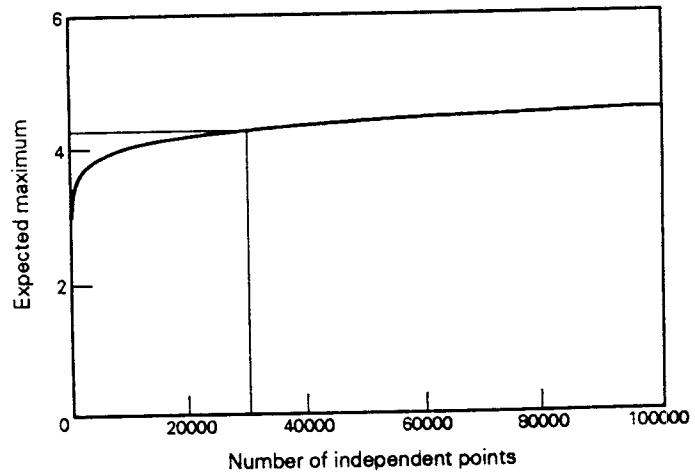

FIG. 2. Determining the number of equivalent independent samples.

due to the fact that the maximum of the associated envelope function of a time series can be significantly greater than the maximum of the time series itself (Pierce, 1985).

Upon re-examining the theoretical basis of the method, it is noted that the technique of analyzing the associated envelope is a convenient, but not necessary, tool. Only the sample-scaling factor is needed to compute the estimate of the maximum. In fact, since the time series is Gaussian, it is better to work directly with the time series itself. Two alternative maximum estimation methods based on rescaling the effective sampling rate directly or indirectly through the sample-scaling factor are proposed in the following sections. Because these methods do not operate on the envelope function, they are

anticipated to be insensitive to the bandwidth of the time series, thus eliminating the tendency to overestimating the maximum for wide-band time series.

Direct method

The Direct method is similar to Pierce's method except that computation of the sample-scaling factor is directly based on the Gaussian time series itself instead of the associated envelope function. The three steps for computing the sample-scaling factor are modified to operate on the time series as follows:

(I) Divide the *time series* into a number of n_a-point segments and compute an empirical expected maximum value $E[M_a]$ for the selected segment length.

(II) Determine an effective number n_e of independent sample points in a segment based on $E[M_a]$ and the *Gaussian distribution* of the *time series*.

(III) Compute the sample-scaling factor n_e/n_a.

As in Pierce's method, the expected maximum value, $E[M_a]$, is the average of the maxima over all segments of the time series is given by Equation (7). However, $M(i)$, $i = 1, \ldots, m$ in this case is defined by

$$M(i) = \max\left\{x(n) \mid n = 1 + \left[\frac{p}{m}\right](i-1), \ldots, \left[\frac{p}{m}\right] i\right\} \tag{18}$$

where $\{x(n)\}_{n=1}^{p}$ is the given time series.

The effective number of independent sample points is determined using information about the distribution of the *time series* and the theoretical expected maximum value. Assuming the sampling points are independent, then $E[M_a]$ is equal to the theoretical expected maximum, which is given by

$$E[M_e(n)] = \int_0^\infty [1 - F_x(t)^n]\, \mathrm{d}t\,. \tag{19}$$

Again choosing that value of n_e which minimizes the difference,

$$D(n) = |E[M_e(n)] - E[M_a]| \tag{20}$$

gives an effective number of independent points. Since the time series is Gaussian, the probability distribution function is given by

$$F_{|x|}(x) = \int_{-x}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp(-y^2/2\sigma^2)\, \mathrm{d}y\,. \tag{21}$$

Substituting Equation (21) into Equation (19) yields

$$E[M_e(n)] = \int_0^\infty \left(1 - \left(\int_{-x}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp(-y^2/2\sigma^2)\, \mathrm{d}y\right)^n\right) \mathrm{d}x\,. \tag{22}$$

The absolute value of X has been used in Equation (21) since one seeks to predict the largest deviation from the mean in any direction. In Pierce's method the envelope accounts for this by definition. The graph of Equation (22) is shown in Fig. 3.

Once the effective number of samples n_e in a segment is determined, the sample-scaling factor can then be computed by

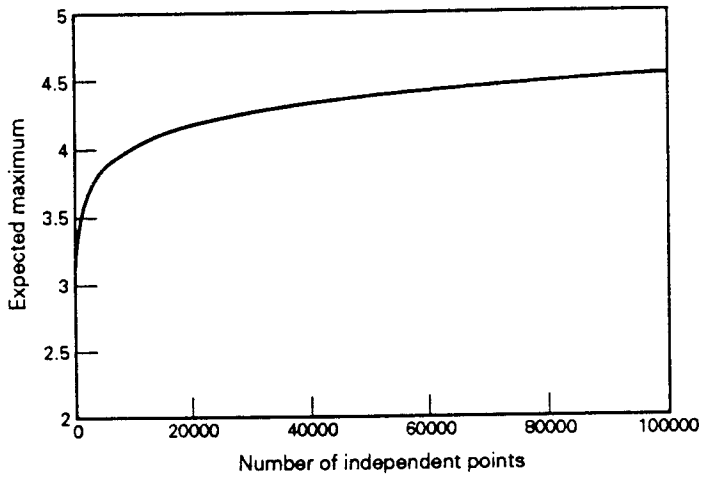

FIG. 3. Expected maximum value as a function of the number of independent points for the non-envelope process.

$$\mathrm{SSF} = \frac{n_e}{n_a} = \frac{n_e m}{p}. \tag{23}$$

The expected extreme value of a time series of arbitrary length n can be computed using Equation (22) with argument $n*\mathrm{SSF}$.

Poisson clumping method

Another alternative to Pierce's envelope method is the Poisson clumping method (Aldous, 1989), which uses the concept of rescaling the sample rate, but does not rely on segmenting the original time series.

An "extreme event" is often meant as an event occurring with small probability as opposed to an event of "large" or "small" size. However, in many cases, such as the problem discussed here involving the maximum of Gaussian time series, these interpretations coincide. When extreme events are considered rare the following heuristic approach may be adopted.

Consider the set of sample times, S_b, for which the time series exceeds the value b, that is,

$$S_b = \{t | X_t \geq b\}. \tag{24}$$

Since X_t is Gaussian, then for large b, S_b is distributed sparsely, that is, a rare or extreme event. This point set is a collection of intervals whose initial points form a Poisson process. This is well known and a consequence of the "Law of Rare Events".

Using a Poisson process to model this phenomenon, an extreme value estimation method may be developed. Let S_b be distributed as a Poisson process with rate λ_b given by the following fundamental identity

$$P(t \in S_b) = P(X_t \geq b) = \lambda_b E[C_b] \tag{25}$$

where C_b (the clump size) is the length of the intervals where the process exceeds the

value b. The parameter λ_b may be interpreted as a scaling factor of the rate of occurrence of independent samples. Now define M_t as the maximum of the time series in the interval $[0,t]$

$$M_t = \max\{X_s | 0 \leq s \leq t\} \tag{26}$$

and T_b as the first (minimum) time that the time series exceeds the value b,

$$T_b = \inf\{s | X_s \geq b\} . \tag{27}$$

Based on these definitions, the following statements

$$[M_t < b] , [T_b > t] , [S_b \cap [0,t] = 0] \tag{28}$$

all indicate the same event. This is so because $[M_t < b]$ means the value b has not been observed by time t, so for $s \in [0,t]$, X_s is smaller than b, or $[S_b \cap [0,t] = 0]$. Furthermore, this is equivalent to the statement that the first time b is observed is beyond time t. By definition T_b is this first time, so $[T_b > t]$.

Now by the Poisson assumption, the probability that the process does not exceed the value b in the time interval $[0,t]$ is

$$P(S_b \cap [0,t] = 0) = e^{-t\lambda_b} . \tag{29}$$

So

$$P(M_t < b) = e^{-t\lambda_b} \tag{30}$$

and by Equation (25) thus

$$P(M_t < b) = \exp\left\{\frac{-tP(X_t \geq b)}{E[C_b]}\right\} . \tag{31}$$

So the remaining task is to estimate $E[C_b]$ as a function of the amplitude b.

$$\begin{aligned} E[M_t] &= \int_0^\infty P(M_t \geq b)\, db \\ &= \int_0^\infty [1 - P(M_t < b)]\, db \\ &= \int_0^\infty \left[1 - \exp\left\{\frac{-tP[X_t \geq b]}{E[C_b]}\right\}\right] db . \end{aligned} \tag{32}$$

As shown in Fig. 4, this can be accomplished numerically by fixing a value of b and finding all the intervals where the time series exceeds this value. Then average over all the lengths of these intervals and plot this average vs b (see Fig. 5). Fitting a curve of the form

$$y = Ab^{-\gamma} \tag{33}$$

to these data and substituting this in for $E[C_b]$ in the above integral for $E[M_t]$ yields an approximation of the expected maximum for given t. This curve with a power law tail is a good fit for this and also in the case of other Gaussian processes.

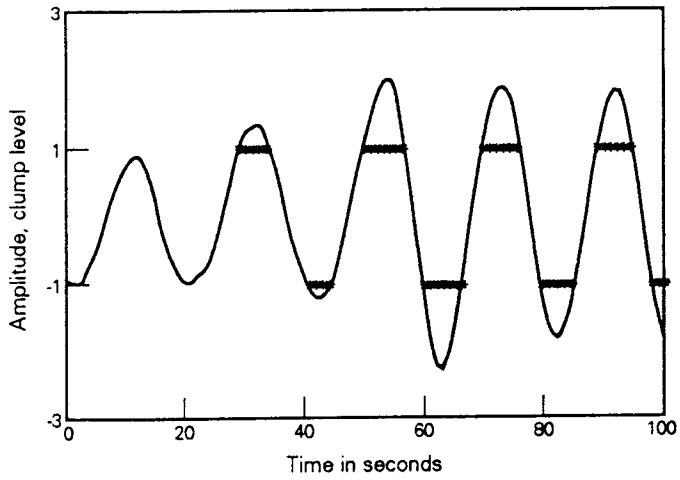

FIG. 4. Interval of exceedence for a given level, b ($= +1$ and -1), of the Poisson clumping method.

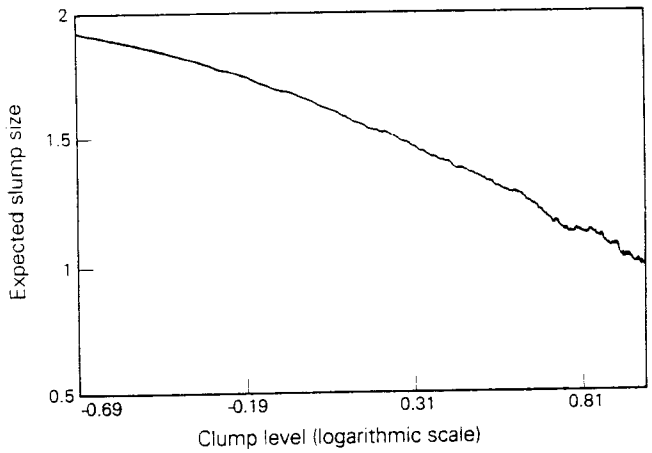

FIG. 5. Expected clump size vs clump level.

Log-fit method

A possible empirical relationship between the time series duration and the expected extreme value is that of a log–linear form. As in Pierce's method the extreme value can be derived by considering sub-intervals, and computing the maximum of the time series over these sub-intervals, then averaging over all sub-intervals.

$$E[M_{L_i}] = \frac{1}{n}\sum_{i=1}^{n} M_{L_i}. \tag{34}$$

Repeating this method for an increasing sequence of sub-interval lengths gives a logarithmically increasing sequence of average maxima. Plotting these average maxima

vs sub-interval length gives points which can be approximated by a logarithm (see Fig. 6)

$$E[M_L] = \beta \ln L \,. \tag{35}$$

The extreme value can then be extrapolated to the relevant value of t using this logarithmic relationship.

To improve the efficiency of computation, a nesting strategy is employed, wherein each iteration of the procedure the number of comparisons are reduced by a factor of one-half. That is, if at stage n, $M_1^{(n)}$ and $M_2^{(n)}$ are the maxima in two adjoining intervals, then at stage $n + 1$, $M^{(n+1)}$ is the maximum over the union of these intervals.

Summary

The main similarities and differences of these four methods can be summarized as follows. Pierce's method involves enveloping the time series to smooth out fluctuations and emphasize maximum values. This method is unique in that it does not deal with the time series directly whereas all the other methods do. All the methods except the Log-fit use information about the distribution of the time series to estimate its expected extreme value. Pierce's method and the Direct method are related in the sense that they both seek to determine a number l where the distribution of

$$M_l = \max_{1 \le i \le l} X_i \tag{36}$$

is the same as

$$M_p = \max_{1 \le i \le p} X_i \tag{37}$$

where in Equation (36) the X_is are taken to be independent, thereby simplifying many of the calculations. The Poisson clumping method uses the concept of rescaling the sample rate to account for dependency of nearby maxima without segmenting the time

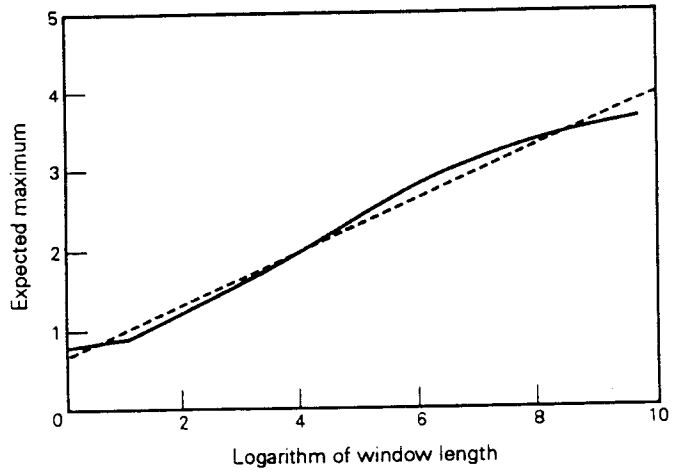

FIG. 6. Line fit to maximum vs log of time series segment length.

series. The Log-fit method uses no distributional information about the X_is and simply uses the slowly varying, logarithmic growth properties of M_n to estimate EM_n. From a practical point of view this technique is crude but the simplest and quickest to implement.

5. EVALUATION OF ALTERNATIVE METHODS

The performance of the alternative methods are evaluated using simulated time series, and the results are compared to Pierce's envelope method in this section.

Evaluation procedure

The prediction capability of the three alternative methods are evaluated as follows. Ensembles of time series of specific duration with various bandwidth parameters and spectral density functions are generated. Each time series is then divided into two segments, with the second segment having a length equal to twice that of the first. The first segment is used to determine the parameters of the particular process using the four methods. The predicted extreme values of a duration equal to the second segment are compared to the simulated maximum of the second segment of the same time series. The mean and standard deviation of the resulting errors from the four methods are computed and compared.

Time series simulation

Two types of Gaussian time series are generated. The first type simulates the response of a linear oscillator to white-noise excitation. The time series of the response are obtained by passing Gaussian white noise through a second-order linear filter. The bandwidth parameter of the time series of the linear oscillator is controlled by varying the normalized damping coefficient of the oscillator (see the Appendix for detail).

The second Gaussian time series simulates random waves in the ocean. They are obtained by superposition of sinusoids with amplitudes specified by: (1) the Pierson–Moskowitz spectrum, and (2) the JONSWAP spectrum (Sarpkaya and Isaacson, 1981). One-thousand cosines with unequal frequency spacings but equal spectral area are employed. The randomness is represented by the random phases in each sinusoid, which are uniformly distributed over the interval $(0, 2\pi)$ (Borgman, 1967). The ensembles of time series with various dominant wave periods are examined. The shape factor for the JONSWAP spectra is assumed to be 3.

For each specific set of parameters an ensemble of 30 independent time series (150,000 sample points for the linear oscillator, and 60,000 points for the ocean waves) are generated. For convenience of comparison, the time series are normalized to be zero mean and unit variance. The sample maximum over the last two-thirds of the time series is found for each of these time series. All the methods are then applied to the first third of the time series and their estimate of the expected value of this measured sample maxima, along with the relative error are found. All these quantities are then averaged across the ensemble and means and standard deviations computed.

Comparisons of results and computational efficiencies

Linear oscillator. Figure 7 shows the mean relative error of each method vs the bandwidth parameter for time series of the response of a linear oscillator, where

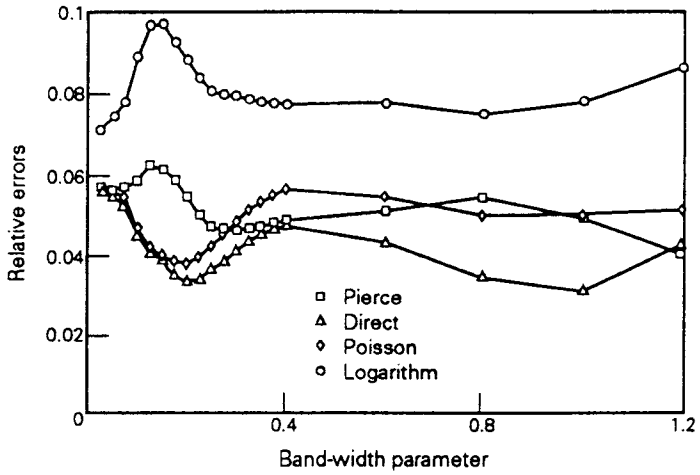

FIG. 7. Comparison of the relative errors; linear oscillator.

$$\text{relative error} = \frac{|\,\text{estimated maximum} - \text{measured maximum}\,|}{(\text{measured maximum})}. \qquad (35)$$

It is observed that the relative errors for Pierce's envelope method, the Direct method, and the Poisson clumping method are all about 6% or less, indicating that all three methods are sufficiently accurate for engineering applications. While Pierce's and the Poisson clumping methods are close to each other throughout the range of bandwidth parameters considered, the Direct method is consistently more accurate than the other two. The Log-fit method, while not as accurate as the other three, produces predictions that are within 10% of the true value, and are about 8% for most cases. Thus, the Log-fit method may be useful for preliminary estimation because of its simplicity.

Figures 8 and 9 show the mean and the standard deviation of the predicted maxima

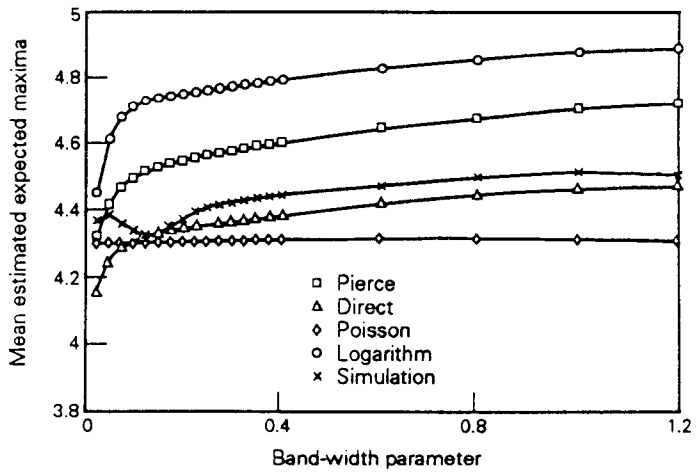

FIG. 8. Comparison of mean estimated expected maxima to simulation; linear oscillator.

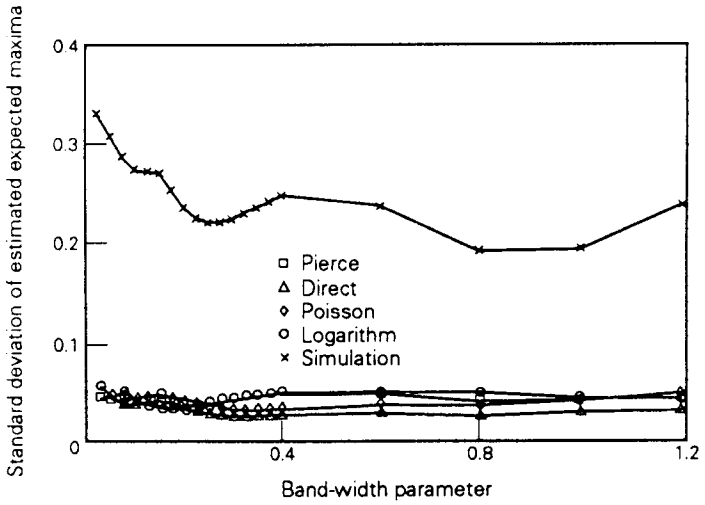

FIG. 9. Comparison of standard deviation of estimated expected maxima to simulation; linear oscillator.

for all the methods as well as the simulated time series. Note that for the system parameters considered, the Direct method consistently produces the best estimate among all the methods, while Pierce's method tends to overestimate the true maxima and the Poisson clumping method tends to underestimate the true maxima. The Log-fit method again produces reasonable, but less accurate, estimates than the other three methods. All four methods appear to be insensitive to the randomness of the time series. The standard deviation of the estimated maximum of the four methods are consistently much smaller than that of the simulated time series.

Ocean waves. Figure 10 shows the mean relative error of each method vs dominant wave period. For the JONSWAP spectrum (with a shape factor equal to 3), the time series are narrow-band. All four methods have similar degrees of accuracy. For the Pierson–Moskowitz spectrum, the time series are more broad-band. The Direct and the Poisson clumping methods for dominate periods of 8 sec and higher are more accurate than Pierce's and the Log-fit methods. In all cases, the Direct method consistently produces the most accurate estimates, and the Direct and the Poisson clumping methods have similar trends.

Figures 11 and 12 show the mean, and standard deviation of the predicted maxima for all the methods and the simulated ocean wave time series. The results indicate that the Direct and the Poisson-clumping methods tend to produce estimates that are lower than the simulation results, while Pierce's and the Log-fit methods tend to overestimate the simulated maxima. As in the linear oscillator case, the estimates computed from Pierce's, Direct and Poisson clumping methods are not very sensitive to the individual time series (small standard deviations), although the simulated maxima have rather high standard deviations. This is due to the fact that the three prediction methods utilize information from all the relative peaks, rather than just the highest peak (absolute

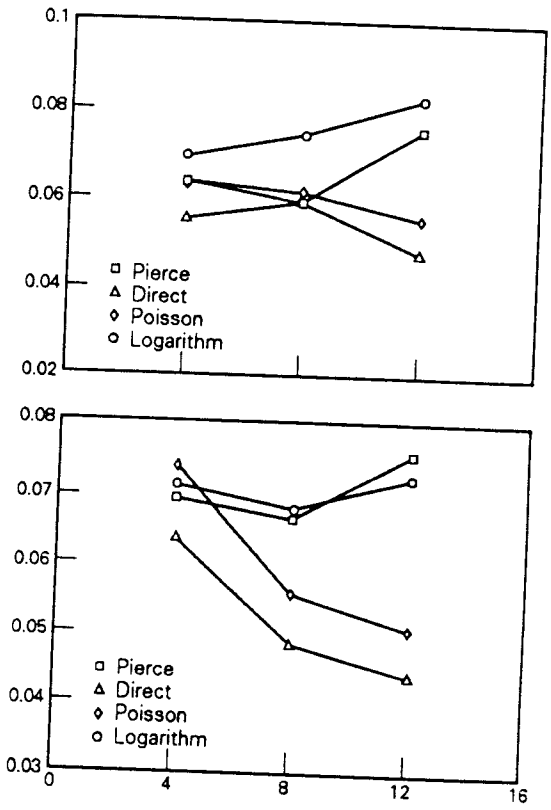

FIG. 10. Comparison of the relative errors; ocean waves: (a) JONSWAP spectrum, and (b) Pierson–Moskowitz spectrum.

maximum), in each time series. The Log-fit method produces estimates with relatively high standard deviations, indicating a strong sensitivity to the absolute maximum.

Overall efficiency

The overall efficiency of the four methods are rated by comparing the relative efforts required to develop the numerical codes and, to a lesser extent, the amount of CPU time required to compute the estimated maxima. As mentioned earlier, the Log-fit method is conceptually the simplest and is the simplest to implement, while the Poisson clumping method requires slightly more coding efforts. The Direct method requires even more coding effort due to the need to determine an acceptable number of segments to ensure accuracy of the method. Pierce's method, which further needs an algorithm for computing the envelope function, requires the most code development efforts. To compare the computational efficiency of the codes, the CPU times for the four methods are determined for a typical run of the examples considered in the above sections (i.e. estimating the extreme value for a 100,000 point time series based on a 50,000 point record). It is found that the Log-fit method takes the least CPU time (approximately 25 sec on a SUN SPARC-Station 1). The Poisson clumping method requires approxi-

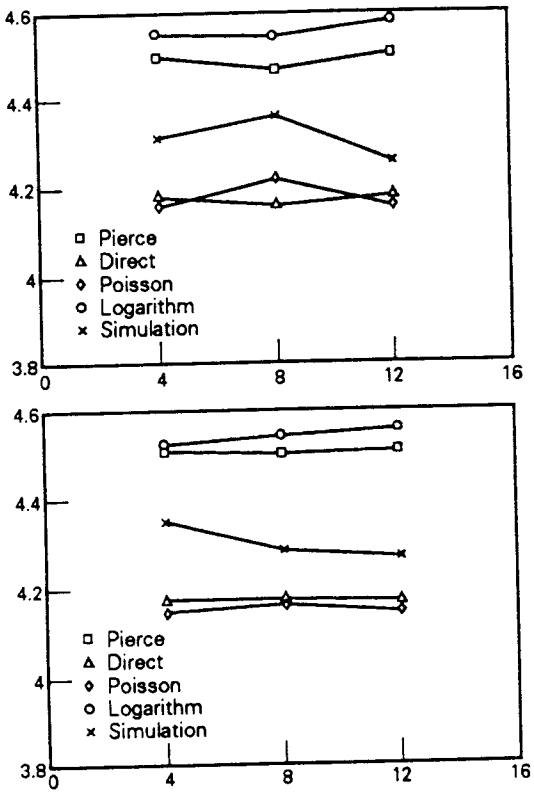

FIG. 11. Comparison of mean estimated expected maxima to simulation; ocean waves: (a) JONSWAP spectrum, and (b) Pierson–Moskowitz spectrum.

mately 500 CPU sec per run. Including five segmenting iterations per run, the Direct method takes half as long as the Pierce's method (350 and 700 sec, respectively). Thus, it is observed that the Log-fit method is computationally an order of magnitude more efficient than the other three methods. However, since these procedures are needed only once for each sea-state or model test, the total amount of computational effort needed for a complete analysis and design of a system is insignificant compared to the code development efforts.

6. CONCLUSIONS

Three practical methods have been presented and their accuracy and efficiency examined in this study. None of the three methods require the identification of "major peaks". The Direct and the Poisson clumping methods rescale the sample rate to take into account the correlation between nearby maxima. The Log-fit method does not utilize distribution information about the time series but simply uses the slowly varying, logarithmic growth property of the expected maximum. Based on the observations from the above section, it can be concluded that:

1. For the simulated time series considered, Pierce's method, the Direct and the Poisson

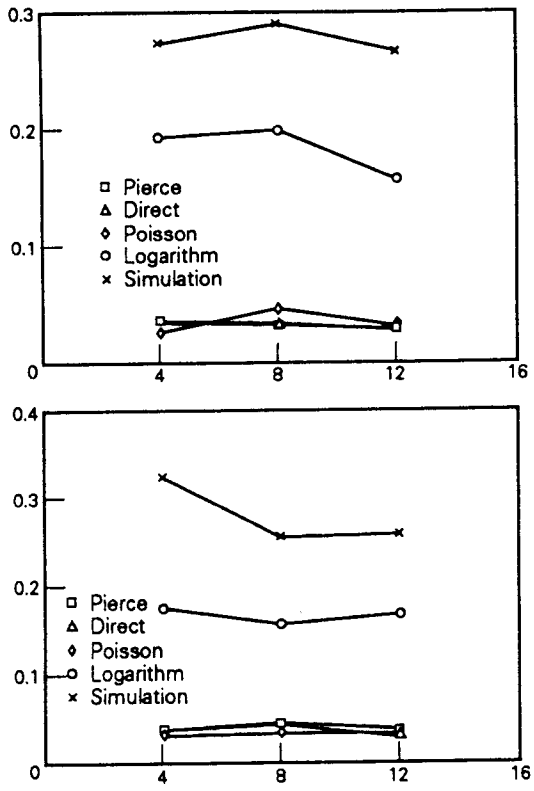

FIG. 12. Comparison of standard deviation of estimated expected maxima to simulation; ocean waves: (a) JONSWAP spectrum, and (b) Pierson–Moskowitz spectrum.

clumping methods have approximately the same accuracy in estimating the extreme value. The Log-fit method is less accurate.

2. Employment of the envelope does not improve the accuracy of the estimate in practice. In fact, it tends to induce a bias to the estimator as the bandwidth of the time series increases. Therefore, the Direct and the Poisson clumping methods, which do not rely on the envelope function, should be preferred in general.
3. The Poisson clumping method is more efficient in the sense that it is straightforward to apply, without having to determine the optimal segment length of the sub-intervals.
4. The Log-fit method is the most straightforward and computationally efficient. Its accuracy is acceptable for most engineering preliminary design applications.

Acknowledgements—Partial support for this research by the ONR Young Investigator Award under grant N00014-88-K-0729 (SCSY), the Air Force of Scientific Research under grant AFOSR 90-0251 (RMB), and the ONR University Research Initiative Program under grant N00014-86-K-0687 (MRG) are gratefully acknowledged. The authors are thankful to Professor David Aldous for implementation suggestions concerning the Poisson clumping heuristic.

REFERENCES

ALDOUS, D. 1989. *Probability Approximations via the Poisson Clumping Heuristic*. Springer, New York.

BENDAT, J.S. and PIERSOL, A.G. 1986. *Random Data, Analysis and Measurement Procedures* (2nd Edn). John Wiley, New York.

BORGMAN, L.E. 1967. Ocean wave simulation for engineering design. University of California Technical Report, HEL-9-13.

BOX, G.E.P. and JENKINS, G.M. 1970. *Time Series Analysis Forecasting and Control*. Holden-Day, San Francisco.

CARTWRIGHT, D.E. and LONGUET-HIGGINS, M.S. 1956. The statistical distribution of the maxima of a random process. *Proc. R. Soc. London* **237A**, 212.

CASTILLO, E. 1988. *Extreme Value Theory in Engineering*. Academic Press, San Diego.

GALAMBOS, J. 1978. *The Asymptotic Theory of Extreme Order Statistics*. John Wiley, New York.

GUMBEL, E.J. 1958. *Statistics of Extremes*. Columbia University Press, New York.

LEADBETTER, M.R., LINDGREN, G. and ROOTZEN, H. 1983. *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

LONGUET-HIGGINS, M.S. 1952. On the statistical distribution of the heights of sea waves. *J. Mar. Res.* **11**, 245.

NAESS, A. 1982. Extreme value estimates based on the envelope concept. *Appl. Ocean Res.* **4**, 181.

NEWLAND, D.E. 1984. *An Introduction to Random Vibrations and Spectral Analysis*. Longman, New York.

OCHI, M.K. 1973. On prediction of extreme values. *J. Ship Res.* **17**, 29.

OCHI, M.K. 1990. *Applied Probability and Stochastic Processes in Engineering and Physical Sciences*. John Wiley, New York.

PIERCE, R.D. 1985. Extreme value estimates for arbitrary bandwidth Gaussian processes using the analytic envelope. *Ocean Engng* **12**, 493.

RICE, S.O. 1944 and 1945. Mathematical analysis of random noise. *Bell Syst. Tech. J.* **23**(1944) and **24**(1945).

SARPKAYA, T. and ISAACSON, M. 1981. *Mechanics of Wave Forces on Offshore Structures*. Van Nostrand Reinhold, New York.

TAYFUN, M.A. 1981. Distribution of crest-to-trough wave heights. *J. WatWay Port Coastal Ocean Div., ASCE* **107**, 149.

APPENDIX: LINEAR OSCILLATOR TIME SERIES GENERATION

The time series of the response of the linear oscillator to white noise excitation are generated as follows:

A normal random variable can be generated using the uniformly distributed variables U_1, $U_2 \sim \mathbf{u}[0,1]$ from the following

$$X_1 = \sqrt{-2\log U_1}\cos(2\pi U_2) \tag{A1a}$$

$$X_2 = \sqrt{-2\log U_1}\sin(2\pi U_2) \tag{A1b}$$

then X_1, X_2 have standard normal distributions, that is, X_1, $X_2 \sim N[0,1]$. A uniform random number generator of the form

$$U(i+1) = 16{,}807 \times U(i) \bmod(2^{31} - 1) \tag{A2}$$

with a period of approximately 3.93 million is employed.

Gaussian time series have been generated via a second-order auto-regressive process

$$x(t) = \phi_1 x(t-1) + \phi_2 x(t-2) + a_t \tag{A3}$$

where $\{a_t\}$ is a sequence of iid normal random variables (see Box and Jenkins, 1970). In order for this process to be stationary it is required that ϕ_1 and ϕ_2 lie in the region

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ \phi_1 + \phi_2 &< 1 \\ |\phi_2| &< 1 . \end{aligned} \tag{A4}$$

For this process to exhibit pseudo-periodic behavior it is required that

$$\phi_1^2 + 4\phi_2 < 0 \,. \tag{A5}$$

Box and Jenkins (1970) showed that the following choices of ϕ_1 and ϕ_2

$$\phi_1 = 2\exp(-2\pi f_0 \xi\delta)\,|\cos 2\pi f_0| \tag{A6a}$$

$$\phi_2 = -\exp(-4\pi f_0 \xi\delta) \tag{A6b}$$

simulates the response of a second-order linear oscillator with natural frequency f_o, and damping ratio ξ; δ is the sampling rate.

All time series generated have a dominant frequency of 0.05 cycles/sec and a sampling rate of 2 sec (resulting in approximately 10 points/period). A damping parameter, which has the effect of determining the dependence structure of the time series is then fixed. The time series are generated with a filter parameter ranging from 2.5% damped (narrow band) to 120% damped (wide band).