MESO-SCALE ESTIMATION OF EXPECTED EXTREME VALUES

Robert M. Burton
Marc R. Goulet
Solomon C.S. Yim

Department of Mathematics Oregon State University Corvallis, OR 97331
Department of Mathematics Oregon State University Corvallis, OR 97331
Department of Civil Engineering Oregon State University Corvallis, OR 97331

Abstract.
We consider algorithms for estimating the expected maximum value of a time series for a period in the future given past observations. This is a "mid-range" problem in which the long term asymptotics of extreme value theory do not apply. There are essentially two approaches, estimating an "extremal index" and the "Poisson clumping heuristic". Variations on these methods are tested with simulated Gauusian data. Similarities in performance are explained rigorously.

Introduction

We consider the following problem. Given a time series from a stationary process $\{X_i\}_{i=1}^{\infty}$, define the expected maximum $E[M_{N,N'}]$, where $N' > N$ and

$$M_{N,N'} = \max_{N+1 \leq i \leq N'} |X_i|.$$

The problem is to find a good estimator of $E[M_{N,N'}]$ based on observations $\{X_i\}_{i=1}^{N}$. We will alway consider Gaussian time series but it will be clear that our methods apply more generally.

Here we describe several estimators of $E[M_{N,N'}]$, present some empirical results and give some theoretical explanations of our results.

Description of the Estimators

Time Rescaling.

The idea is to estimate an *extremal index* of the process for this time scale. We say that ρ is the extremal index for $\{X_i\}_{i=1}^{\infty}$ on the scale of N if M_N has approximately the same distribution as the maximum of $[\rho N]$ independent random variables with the same distribution as X_1, i.e Gaussian.

Key words and phrases. extremal index, poisson clumping, extreme values.

The first and second authors would like to acknowledge support from AFOSR Grant # 91-0215 and NSF Grant # DMS-9103738. The second and third authors would like to acknowledge support from ONR Grant # N00014-92-J1221. .

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TeX

There are essentially three choices to be made in approach.

The first is whether to use the time series itself or an enveloped version of it. Given the data, $\{X_i\}_{i=1}^N$ we may construct the discrete Hilbert transform $\{Y_i\}_{i=1}^N$. The process $\{R_i\}_{i=1}^N$ defined by

$$R_i = \sqrt{X_i^2 + Y_i^2}$$

is called the *analytic envelope*. It covers the "surface" of the time series, smoothing out the oscillations. The maximum of the envelope is close to that of the original process, especially in the narrow band case. It has the further computational advantage of being Rayleigh distributed. This is described in detail in [P,YBG]. We call these choices *direct* and *enveloped.*

The second choice is how carefully to compute the expected maximum of n independent Gaussian random variables (or Rayleigh in the enveloped case) as a function of n. One could either use a good but computationally intensive numerical approximation or an asymptotic formula, $\sqrt{2\log n}$. We will refer to these choices as *strong* and *weak* and call this approximation we use $L(n)$.

The third choice is how carefully to fit the empirical expected maxima as computed from the data to $L(n)$. One possibility is to use one value of n_0, say 50. Find the average of the maximum value in the data for non-overlapping windows of length n_0. This value $\hat{L}(n_0)$ is the empirical expected maximum at n_0. To estimate the extremal index then find ρ so that $\hat{L}(n_0) = L(\rho n_0)$.

The other possibility is to compute multiple window lengths, i.e. to compute $\hat{L}(n) = L(n)$ for various n. If we take n to be powers of 2 then the computation time is not large because we may "nest" the computations of the maxima. We will refer to these choices as *single window* and *multiple window* methods.

We note that the prevailing method among ocean engineers was the enveloped, strong, single window method.

Poisson Clumping. Another possible estimator is suggested by Aldous' use of the Poisson clumping heuristic [A]. This heuristic assumes that the set of t for which $X_t > b$ is given by random sets distributed as a Poisson process. We make the further assumption that these random sets are intervals.

Consider, for b relatively large

$$\{t | X_{t-1} < b, X_t \geq b\}$$

to be distributed as a Poisson process with rate λ_b. The following fundamental relation is assumed

$$P[X_1 \geq b] = \lambda_b E[C_b],$$

where C_b is the random length of an interval (clump) in which the time series spends above a given value b. The event $[M_N < b]$ is equivalent to

$$\{t | X_{t-1} < b, X_t \geq b\} = \emptyset.$$

So by the Poisson assumption,

$$P[M_N < b] = e^{-\lambda_b N},$$

and by the fundamental identity

$$P[M_N < t] = e^{-P[X_t \geq b]N/E[C_b]}.$$

Hence we have

$$E[M_N] = \int_0^\infty (1 - e^{-P[X_1 \geq b]N/E[C_b]})\, db. \tag{1}$$

The work now reduces to estimating $E[C_b]$. To do this we fix a value of b and average the length of the intervals where the time series is above b. Varying b and plotting $E[C_b]$ versus b yields data which is well fit by a curve of the form

$$y = b^{-\gamma}/A.$$

Substituting this curve into (1) yields our estimator,

$$\hat{E}_P[M_{N,N'}] = \int_0^\infty (1 - e^{-AP[X_1 \geq b](N'-N)b^{\gamma}})\, db.$$

As before we may use either the original or enveloped data. Note that the above analysis assumes that the clumps are intervals so one guesses that enveloping narrow band data would be advantageous.

Empirical Results

We have described 10 possible algorithms in all.

In earlier work [BGY, YBG] we investigated several algorithms. The algorithm used by most ocean engineers was due to Pierce [P]. In our terminology this was an enveloped, strong, single window time rescaling method.

We proposed to modify this by removing the envelope, that is to use instead the direct, strong single window time rescaling method, referred to in the figures as *Direct*. In cases where computation ease was paramount we proposed the direct, weak, multiple window rescaling method, here referred to as *Logfit*. These were compared with the direct Poisson clumping algorithm, or *Poisson*. Here we describe the results of these simluations. In a future paper we will also include study of the enveloped Poisson clumping algorithm. Work continues on the other variations.

In this study two types of Gaussian time series are used. The first is a second order autoregressive moving average (ARMA).

$$X_n = aX_{n-1} + bX_{n-2} + Z_n,$$

where the Z_n is are independent identically distributed Gaussian random variables.

The second type is intendend to simulate random waves in the ocean and are obtained by superposition of sinosoids, with amplitudes specified by the Pierson-Moskowitz and JONSWAP spectrums [SI]. One thousand cosines with unequal frequency spacings and uniformly random phases are employed. More detail on these processes will be given in a future paper.

In all cases we have $N' - N = 2N$. For the autoregressive moving averages $N = 50000$, while for the simulated ocean waves $N = 20000$. These time series are run for various parameters and the expected maximum are estimated by the algorithms. The mean relative errors are computed. The parameter for the ARMA model is a damping ratio corresponding to a linear oscillator. This ratio relates directly to the bandwidth of the excitaiton. For the simulated ocean waves we use a dominant wave period [YBG].

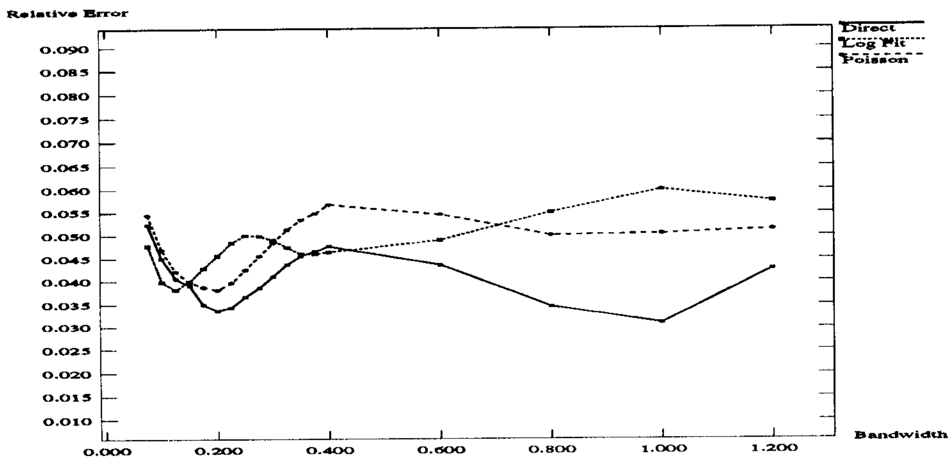

Figure 1

Our results are pictured in Figures 1-3. To summarize the ARMA experiments (Figure 1), the Direct Method consistently gives the estimator with minimal relative error, while the Poisson clumping and Log Fit methods yield estimators with relative errors under 6%. It is interesting to note that the results for Poisson Clumping and the direct, strong, single window time rescaling method follow each other.

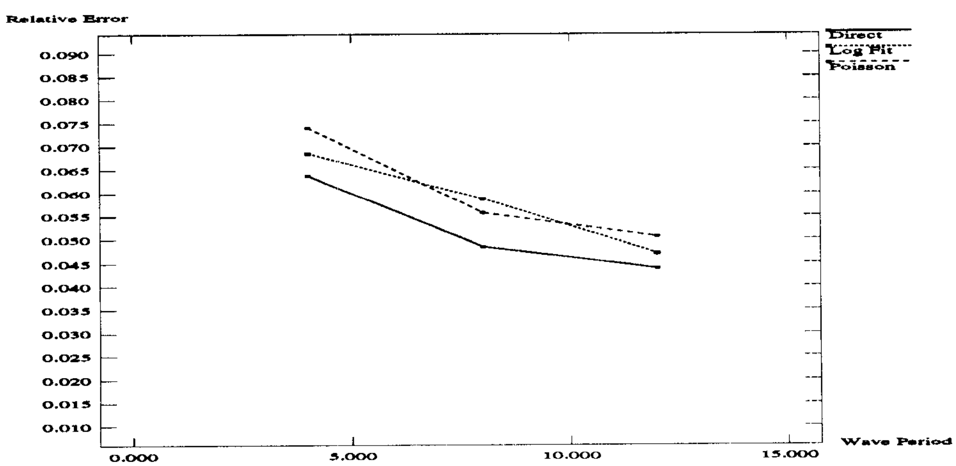

Figure 2

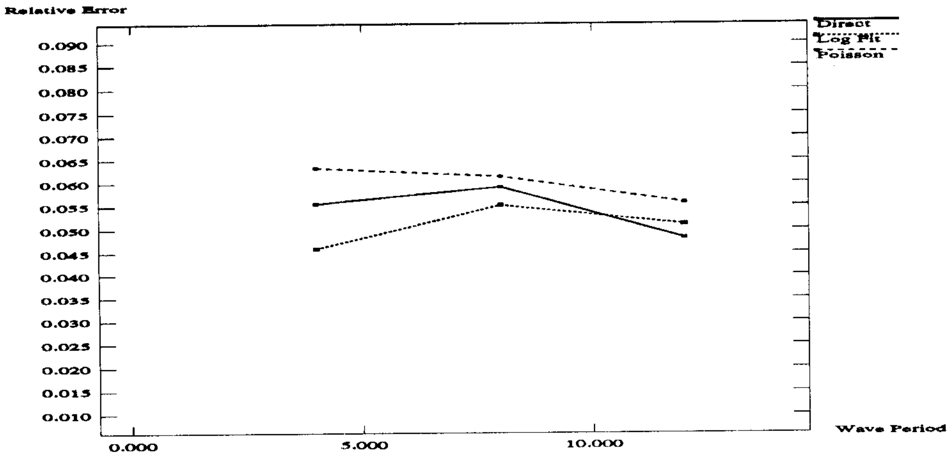

Figure 3

For the simulated ocean waves the results are similar. In the case of the Pierson-Moskowitz spectrum (Figure 2) the Direct Method consistently provides the best estimator, regardless of the dominant wave length. The JONSWAP spectrum (Figure 3) provides a narrow band case and all the techniques yield relative errors between 4% and 7%, while the Log Fit provides the best estimator in two instances.

Conclusion

The theoretical underpinning for these algorithms is given by theorems of O'Brien [O'B] and independently by Rootzen [R]. There it is showed for processes satisfying a strong mixing condition (as ours do) that for long enough time scales there is an extremal index which in turn gives the Poisson clump structure of the exceedance process. Thus it is not surprising to find the direct, strong, single window estimator and the direct Poisson clumping estimator in close agreement.

Perhaps the most striking result in this study is the performance of the direct, weak, multiple window estimator. It is the simplest conceptually and algorithmically, and gives relative errors near 6%.

References

[A] D. Aldous, *Probability Approximations via the Poisson Clumping Heuristic*, Springer-Verlag, New York, 1989.

[BGY] R.B. Burton, M. Goulet and S. Yim, *Mid-range estimators of expected maximum values for Gaussian time series*, submitted.

[LLR] M.R. Leadbetter, G. Lindgren and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York, 1983.

[O'B] G.L. O'Brien, *Extreme Values for Stationary and Markov Sequences*, Annals of Probability **15** (1987), 281-291.

[P] R.D. Pierce, *Extreme Value Estimates for Arbitrary Bandwidth Gaussian Processes Using the Analytic Envelope*, Ocean Engineering. **12** no. 6 (1985), 493.

[R] H. Rootzén, *Maxima and Exceedences of Stationary Markov Chains*, Advances in Applied Probability **20** (1988), 371-390.

[SI] T. Sarpkaya and M. Isaacson, *Mechanics of Wave Forces on Offshore Structures*, Van Nostrand Reinhold, New York, 1981.

[YBG] S.C.S. Yim, R.M. Burton and M.R. Goulet, *Practical Methods of Extreme Value Estimation Based on Measured Time Series for Ocean Systems*, Ocean Engineering **19** no. 3 (1992), 219-238.