Ocean Engineering 124 (2016) 241–251

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Extreme value estimation using the likelihood-weighted method

Ryota Wada [a,*], Takuji Waseda [a], Philip Jonathan [b]

[a] Department of Ocean Technology, Policy and Environment, University of Tokyo, Japan
[b] Shell Projects and Technology, Manchester, M22 0RR United Kingdom

ARTICLE INFO

Article history:
Received 6 January 2016
Received in revised form
7 June 2016
Accepted 26 July 2016
Available online 6 August 2016

Keywords:
Likelihood-weighted method
Extreme
Uncertainty
Group likelihood
Bayes

ABSTRACT

This paper proposes a practical approach to extreme value estimation for small samples of observations with truncated values, or high measurement uncertainty, facilitating reasonable estimation of epistemic uncertainty. The approach, called the likelihood-weighted method (LWM), involves Bayesian inference incorporating group likelihood for the generalised Pareto or generalised extreme value distributions and near-uniform prior distributions for parameters. Group likelihood (as opposed to standard likelihood) provides a straightforward mechanism to incorporate measurement error in inference, and adopting flat priors simplifies computation. The method's statistical and computational efficiency are validated by numerical experiment for small samples of size at most 10. Ocean wave applications reveal shortcomings of competitor methods, and advantages of estimating epistemic uncertainty within a Bayesian framework in particular.

1. Introduction

Extreme value estimation characterizes the tail of a probability density distribution, and often requires extrapolation beyond what has been observed (Coles et al., 2001). Extrapolation is motivated by extreme value theory for the asymptotic distribution of large values from any max-stable distribution. A basic assumption in fitting an extreme value model to a sample is that observations are independently and identically distributed. This assumption usually holds for the rarest and severest of ocean wave events (e.g. the storm peaks over threshold of significant wave heights in a tropical cyclone at a location). The trade-off between sample size and adequate tail fit, and the fact that measurement errors on most extreme observations tend to be large, render the analysis problematic.

The increasing availability of high quality measurements and hindcasts means that the metocean engineer is often blessed with huge samples for estimation of return values for design purposes. Extreme value modelling is then a large-scale computational task, within which the effects of non-stationarity and spatial dependence can be estimated (Jonathan and Ewans, 2013). However, there are many other applications where large samples of high quality data are still not available. The metocean engineer is then required to provide design values from small samples of typically poor quality. For such analysis, uncertainties in extreme value parameters and return value estimates are large and often difficult to estimate well. The effective number of influential observations in estimating extreme events with very low probability, such as at the ten thousand year return period level, may be small even in samples corresponding to a hundred years of observations. The goal of this paper is to explore a method for extreme value estimation useful for small samples (of size at most 10) of poor quality data, which provides realistic estimation of epistemic model uncertainty. The approach, called the likelihood-weighted method (LWM), involves Bayesian inference for the group generalised Pareto (or generalised extreme value) likelihood and uniform prior distributions for parameters. Group likelihood provides a straightforward mechanism to incorporate measurement error; adopting flat priors simplifies computation.

Statistical models exhibit two types of uncertainty (Bitner-Gregersen and Skjong, 2009). Aleatory uncertainty represents the inherent randomness of nature and physics; it is intrinsic and cannot be reduced. Epistemic uncertainty represents our limited knowledge, and can be reduced (e.g.) by increasing sample size or reducing sample measurement error. Realistic estimation of epistemic uncertainty is critical to reliable extreme value modelling. We will demonstrate that estimation methods such as maximum likelihood provide poor estimates of epistemic uncertainty from small samples of poor quality.

The organisation of the article is as follows. In Section 2, we review methods in extreme value analysis with emphasis on uncertainty quantification from poor data. A description of LWM, our new estimation method, is given in Section 3. In Section 4, LWM's

* Corresponding author.
E-mail address: r_wada@k.u-tokyo.ac.jp (R. Wada).

http://dx.doi.org/10.1016/j.oceaneng.2016.07.063
0029-8018/

statistical and numerical efficiency is validated through numerical experiments. An application to observed extreme wave height data is considered for further discussion in Section 5, followed by conclusion in Section 6.

2. Extreme value estimation for small samples measured with error

2.1. Extreme value theory

The central limit theorem provides an asymptotic distributional form (the Gaussian distribution) for the mean $A_n \left(= (1/n)\sum_{j=1}^{n} X_j \right)$ of n independent observations of identically-distributed random variables $X_1, X_2, \ldots, X_n$, regardless of the underlying distribution. Analogously, extreme value theory provides an asymptotic distributional form for independent observations from any of a large class of so-called max-stable distributions (Kotz and Nadarajah, 2000). The limiting forms for extreme values of block maxima M_n ($= \max(X_1, X_2, \ldots, X_n)$) were given by Jenkinson (1955), and were later rationalised into one generalised extreme value (GEV) distributional form. Pickands (1975) and Balkema and De Haan (1974) derived the generalised Pareto (GP) distribution for peaks over threshold (POT) by considering the logarithms of the GEV.

GEV and GP are three-parameter distributions, with parameters shape ξ, scale σ and location μ (for GEV) or extreme value threshold ψ (for GP). Cumulative distribution functions (cdfs, F_{GEV} and F_{GP} respectively) for these distributions are given in Eqs. (1) and (2). Other distributional forms are used for extreme value estimation, including the Weibull and log-normal distributions e.g. Ochi (2005), Muir and El-Shaarawi (1986). Here we focus on GEV and GP, given their natural asymptotic motivation and wide application:

$$\Pr(M_n \le x) \overset{\text{large } n}{\approx} F_{GEV}(x) = \exp\left(-\left(1 + \frac{\xi}{\sigma}(x-\mu) \right)^{-1/\xi} \right) \quad \text{for } \xi \neq 0 \qquad = \exp\left(-\exp\left(-\frac{1}{\sigma}(x-\mu) \right) \right) \quad \text{otherwise, and} \tag{1}$$

$$\Pr(X \le x | X > \psi) \overset{\text{large } \psi}{\approx} F_{GP}(x) = 1 - \left(1 + \frac{\xi}{\sigma}(x-\psi) \right)^{-1/\xi} \quad \text{for } \xi \neq 0 \qquad = 1 - \exp\left(-\frac{1}{\sigma}(x-\psi) \right) \quad \text{otherwise.} \tag{2}$$

These distributional forms are correct asymptotically for block maxima and peaks over threshold, but only approximately for finite samples. Increasing sample size for fitting is desirable to reduce estimated parameter bias and uncertainty, but often is achieved at the expense of quality of fit of an extreme value distribution to the largest values in the sample (e.g. by reducing block size for GEV, or reducing extreme value threshold for GP). We do not address this trade-off directly in this work; rather, we assume that the sample is drawn from the extreme value distribution to be estimated, and concentrate on estimating parameters and uncertainty.

2.2. Parameter estimation

There are many possible approaches for parameter estimation in extreme value analysis. Popular schemes include maximum likelihood (ML), the method of moments, probability weighted moments (PWM), L-moments and Bayesian inference (Muir and El-Shaarawi, 1986). Graphical methods have also been proposed, but these are not recommended for quantitative work. Other empirically-derived estimation methods such as Goda's method (Wada and Waseda) lack generality. Methods based on moments or likelihoods are most common in the literature (Palutikof et al., 1999).

For small samples, moment-based methods such as PWM and L-moments, are considered better than ML (in terms of bias and mean square error) for point estimation of parameters (Hosking et al., 1985). Here our interest is not in deriving point estimates, since large epistemic uncertainty is obviously unavoidable, and quantification of the epistemic uncertainty of much greater importance. For both ML and PWM, confidence intervals can be estimated by the so-called delta method, or the profile likelihood method; both are motivated by consideration of asymptotic behaviour, and strictly valid for large samples. Smith and Naylor (1987) considers extreme value estimation for a three-parameter Weibull distribution using maximum likelihood for sample size of over 40, and discusses the resulting unusual likelihood shape. In some applications, even a sample size as small as 20 is difficult to gather. This is the motivation for the current work: we focus on extreme value estimation from sample sizes of at most 10.

Resampling methods such as bootstrapping are also used for uncertainty quantification. The simplest resampling scheme draws random re-samples with replacement from the original sample (Efron, 1979, is easy to implement and widely used. Uncertainty quantification from resampling is rather ad hoc in nature, certainly compared with Bayesian inference. We will illustrate the shortcomings of a simple bootstrap method for small samples in Section 4.

2.3. Bayesian inference

Bayesian methods exploit both the sample likelihood and prior distributions for parameters in inference. The favourable performance of Bayesian inference in extreme value estimation from small samples has been discussed (Coles and Powell, 1996). One advantage of the Bayesian approach is the flexibility offered to estimate unusually shaped likelihood surfaces (Smith and Naylor, 1987).

The basic equations of Bayesian inference are described below. The sample likelihood $L(\theta; D)$ of parameter(s) θ for sample $D = \{x_i\}_{i=1}^{n}$ is interpreted as the probability of the sample given parameters

$$f(D|\theta) = L(\theta; D) = \prod_{i=1}^{n} f(x_i|\theta). \tag{3}$$

The probability of the sample is then

$$f(D) = \int_{\theta} f(D|\theta) dF(\theta), \tag{4}$$

where we can interpret $dF(\theta)$ as $f(\theta)d\theta$ for continuous prior density $f(\theta)$ for θ. We estimate the posterior distribution of θ using Bayes theorem

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{f(D)}. \tag{5}$$

The posterior $f(\theta|D)$ can be used, amongst other things, to estimate credible intervals for parameters. The posterior predictive distribution $g(x|D)$ of any function $g(x|\theta)$ is then the expected value of that function under the posterior distribution $f(\theta|D)$ for θ. The posterior predictive distribution therefore captures both epistemic and aleatory uncertainty

$$g(x|D) = E_{\theta|D}\left(g(x|\theta) \right) = \int g(x|\theta) f(\theta|D) d\theta. \tag{6}$$

Please cite this article as: Wada, R., et al., Extreme value estimation using the likelihood-weighted method. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.063

In spite of its many advantages, there are several drawbacks to Bayesian inference. One objection lies in the difficulty in specifying prior distribution $f(\theta)$. Previous studies have focused on applying informative priors, which impose prior knowledge or belief, compensating for lack of information in a small sample. For example, prior elicitation of extreme rainfall was achieved using observations from neighbouring spatial locations (Coles and Powell, 1996). The inferential value of prior information can be enormous, especially when sample quality is poor. However, expert knowledge of extreme events is often not available or regarded as overly subjective. Uninformative priors are intended to be as objective as possible, by limiting the incorporation of "unintended information" presented by the prior as much as possible. The basic uninformative prior is the uniform or flat prior. A uniform prior $f(\theta)$ allocates the same prior probability to each prior choice of θ. However, even this simple prior has problems: e.g. when defined for parameters with infinite range; moreover, prior uniformity is not transformation invariant. Jeffreys' priors (Jeffreys, 1946) and reference priors (Bernardo, 1979) have been studied to deal with transformation invariance. However, none can theoretically be justified to be objective (Kass and Wasserman, 1996). In this sense, the choice of flat prior can be argued to be more or less subjective (Mori et al., 2010). Computation can also be a problem for Bayesian inference (Coles and Powell, 1996). Only in special cases can the equations above be solved in closed form. The development of Markov Chain Monte Carlo (MCMC) simulation methods has made the computations required for Bayesian inference in general tractable and popular (Smith and Roberts, 1993). However, MCMC requires implementation expertise (Muir and El-Shaarawi, 1986; Jonathan and Ewans, 2013) especially for larger problems.

2.4. Measurement uncertainty

Adequate sample size is critical for good inference. The quality of individuals in the sample is equally important, especially for observations of extreme values. In an ocean engineering context, Soares (1986) assesses the quality of visual observations for wave data, an important source of information for historical wave records. In-situ observations of severe ocean events are likely to be made with large uncertainty, compared to observations of typical events (Jonathan and Ewans, 2013). Moreover, Forristall et al. (1996) notes that maximum values of significant wave height may be overestimated in storms, suggesting additional bias effects for finite samples. Some authors consider the effect of measurement uncertainty for extreme value estimation in ocean engineering settings. Bitner-Gregersen et al. (1990) quantifies measurement uncertainty and explores its impact: data uncertainty is described as the joint effect of observational or instrumental error and sampling variability. Pre-specification of scale and shape parameters for the extreme value distribution is one proposed approach to limit the impact of data uncertainty on inferences.

Measurement uncertainty in general is a combination of systematic bias and random error. Systematic bias can sometimes be reduced by (e.g.) calibration. The impact of random measurement error on extreme value inference is not easily understood. For example, consider a sample of relatively small size containing a single large extreme observation: naive extreme value fitting might suggest a heavy-tailed distribution. However, if the single large value is due to random measurement error of the observation process, the underlying distribution may in fact be short-tailed. Understanding and quantifying the effects of measurement uncertainty in extreme value analysis is clearly critical.

2.5. The case for an improved approach

As outlined above, the combined effect of small sample size and large measurement uncertainty is a real challenge in many ocean engineering applications of extreme value analysis. Naive adoption of results from asymptotic statistical theory, appropriate only for large samples, cannot be justified for uncertainty quantification with small samples. Moreover, measurement error cannot be ignored, and should be accommodated appropriately in any extreme value model.

It is common engineering practice to provide a point estimate of a return value estimated under an extreme value model, typically by assuming that the best-fitting combination of model parameters is a correct and certain inference from the extreme value fit. For example, the conventional closed-form definition of a return value typically corresponds to a particular quantile of the distribution (due to aleatory uncertainty alone) of the maximum value which would be observed during the return period under consideration, near the mode of that distribution. Such a point estimate already ignores the fact that a larger value than the return value might be expected to occur routinely during the return period due to natural variability. Introducing additional epistemic uncertainty due to uncertain model parameter estimates exasperates the issue. Uncertainty in point estimates is mitigated in structural design by incorporation of safety factors. These are calibrated using historical analysis and expert knowledge, and sometimes tuned for specific ocean basins. Yet the magnitude of epistemic uncertainty of a point estimate from any study is dependent on the available data for that study, and may not therefore always be appropriately accounted for in safety factors. We conclude that the point estimate of return value may well not be a wise estimate, e.g. in the light of a preference for conservatism in structural design. Jonathan and Ewans (2007) proposes that a high quantile of the distribution of the maximum value during the return period might be more an appropriate choice, particularly considering the influence of uncertain GP shape parameter estimate. This proposal is made to account for aleatory and epistemic uncertainty in extreme value estimation. Using a high quantile value seems rational for reliable design, yet for small samples, the value of the (e.g.) 90% percentile is likely to be unrealistically large due to epistemic uncertainty. High quantiles have also been recommended for other reasons: Det Norske Veritas (2010) concludes that a quantile in the order of 85–95% is a reasonable choice for return values of environmental conditions and structural loads for use in design, to account for cases when the short-term variability of a process (such as structural loading) within a sea-state is not otherwise being considered.

Our aim therefore is to develop a practical method for extreme value estimation to a small sample of relatively poor quality, which provides reasonable estimation of extreme value models and return values, and allows realistic quantification of epistemic uncertainty. Given that the posterior predictive distribution outlined above provides an intuitive framework to quantify both aleatory and epistemic uncertainties, it seems natural to use the framework of Bayesian inference to achieve this.

3. The likelihood-weighted method

The likelihood-weighted method (LWM) is a straightforward Bayesian approach to extreme value estimation for small samples of poor quality. Its purpose is to provide reasonable estimates of extreme value model parameters and their uncertainties to estimate return values for subsequent structural design calculations, from small samples of poor quality. The LWM model has two distinct features: a group likelihood (see Section 3.1, as opposed to a standard likelihood), and near-uniform prior distributions (see Section 3.2). Section 3.3 provides brief comments on computation.

Please cite this article as: Wada, R., et al., Extreme value estimation using the likelihood-weighted method. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.063

3.1. Group likelihood

Group likelihood is a simple approach to incorporate measurement uncertainty in (Bayesian) inference. Group likelihood was originally proposed to overcome non-regularity issues in ML. We focus on the group likelihood of Giesbrecht and Kempthorne (1976). Suppose we sample $D = \{x_i\}_{i=1}^n$ independently from identically-distributed random variables $X_1, X_2, \ldots, X_n$, related to underlying identically-distributed random variables $Y_1, Y_2, \ldots, Y_n$ of interest to us, such that in terms of the conditional density $f(X_i|Y_i)$

$$f(X_i = x_i|Y_i = y_i) = \frac{1}{2\delta} \quad \text{for } x_i - \delta \le y_i < x_i + \delta = 0 \quad \text{otherwise.} \tag{7}$$

That is, we observe a discretised version X_i of each Y_i; the underlying values of Y_i is uniformly distributed on the interval $[x_i - \delta, x_i + \delta)$. Observation of X_i allows us to make posterior predictive inferences about Y_i using Eq. (6). If the density $f(Y_i|\theta)$ of Y_i is given in terms of parameters θ, the conditional density at $X = x_i$ with respect to θ is

$$\begin{aligned} f(x_i|\theta) &= \int f(x_i|y_i) f(y_i|\theta) dy_i = \frac{1}{2\delta} \int_{x_i-\delta}^{x_i+\delta} f(y_i|\theta) \\ & dy_i \qquad = \frac{1}{2\delta}\big(F(x_i + \delta|\theta) - F(x_i - \delta|\theta)\big), \end{aligned} \tag{8}$$

where F is the cdf of Y_i, and for the full sample

$$L_G(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \frac{1}{2\delta} \prod_{i=1}^{n} \big(F(x_i + \delta|\theta) - F(x_i - \delta|\theta)\big). \tag{9}$$

Group likelihood takes account of data uncertainty, whereas typically extreme value estimation is made assuming no measurement error. In most cases, our knowledge of measurement error will be approximate; it is important to keep this in mind (Cheng and Iles, 1987). Of course, many forms of error distributions $f(X_i|Y_i)$ might be considered, e.g. the case where the range $[x_i^{min}, x_i^{max}]$ of possible values for Y_i corresponding to observation x_i is known. Now

$$L_G(\theta) = \prod_{i=1}^{n} \frac{1}{x_i^{max} - x_i^{min}} \big(F(x_i^{max}; \theta) - F(x_i^{min}; \theta)\big). \tag{10}$$

Given that sample size is small and measurement error an issue, it is essential that every available piece of information is well used. Even knowledge that an extreme event occurred, but that no observation was possible, can be incorporated by use of suitable $[x_i^{min}, x_i^{max}]$. Equivalently, we might be able to specify δ differently for each observation, such that

$$L_G(\theta) = \prod_{i=1}^{n} \frac{1}{2\delta_i} \big(F(x_i + \delta_i|\theta) - F(x_i - \delta_i|\theta)\big) \tag{11}$$

in the obvious notation. Of course, yet more general error structures might be considered, but these would require more sophisticated methods (e.g. MCMC) for estimation. In particular, we note that δ here represents uncertainty due to instrument precision alone. In practice, we might expect additional sources of uncertainty to contribute to the overall measurement error, as mentioned in Section 2.4.

3.2. Near-uniform prior

The second feature of LWM is the use of near-uniform priors; improper uniform priors are avoided (Coles and Tawn, 1996; Scotto and Soares, 2007) and near-uniform Gaussian prior distributions $N(\alpha, \beta)$ with mean α and large variance β adopted. The corresponding density is near-uniform near the mean, yet the distribution provides a proper prior in that it integrates to unity. For studies below, unless otherwise stated, we proceed as follows, assuming a priori that $\xi \sim N(0, 10^2)$, $\log\sigma \sim N(0, 10^4)$ and $\psi = 0$ as suggested by Pickands (1994) for peaks over threshold (GP), and $\xi \sim N(0, 10^2)$, $\log\sigma \sim N(0, 10^4)$, and $\mu \sim N(0, 10^4)$ recommended by Coles and Tawn (1996) for block maxima (GEV).

We note that both the uniform and Gaussian distributions are not conjugate to GP (nor to GEV). Therefore, a numerical procedure is required for parameter estimation.

3.3. Inference scheme

For sample $D = \{x_i\}_{i=1}^n$ and specified δ, LWM inference proceeds as follows for peaks over threshold.

Estimation of group likelihood: Define an index set $\{\theta_j^G\}_{j=1}^m$ of m combinations of parameters on a rectangular grid covering a plausible 2-dimensional domain for parameters ξ and $\log\sigma$, with suitable grid resolution for each parameter, with threshold ψ assumed to be the minimum value in the sample. The plausible domain for ξ and $\log\sigma$ is estimated from a prior trial analysis using the full prior parameter domain and course grid resolution; based on the trial, a sensible restricted grid domain and increased grid resolution are specified.

Compute the group likelihood at each θ_j^G on the index set. Sum the group likelihood over the index set, then divide the group likelihood at θ_j^G by the sum. The result is the estimate for posterior density $f(\theta_j^G|D)$ on the index set

$$f(\theta_j^G|D)\Delta = \frac{L_G(\theta_j^G) f(\theta_j^G)}{\sum_{j'=1}^m L_G(\theta_{j'}^G) f(\theta_{j'}^G)} \overset{\text{prior}}{\underset{\text{uniform}}{\approx}} \frac{L_G(\theta_j^G)}{\sum_{j'=1}^m L_G(\theta_{j'}^G)}, \tag{12}$$

where the near-uniform prior $f(\theta_j^G)$ is assumed constant over the index set, and Δ is (constant) grid cell volume. The posterior is seen to be a weighted likelihood, motivating the choice of name "likelihood-weighted method". The obvious analogous scheme (with 3-dimensional rectangular grid for ξ, $\log\sigma$ and μ) is employed for inferences with block maxima data.

Estimation of credible regions: Credible regions for parameters are estimated by sorting the values in $\{f(\theta_j^G|D)\}_{j=1}^m$ in decreasing order to yield $f(\theta_{r(j)}^G|X)$ for sorting array $\{r(j)\}_{j=1}^m$. Given a probability level p, the subset of the index set contributing to the credible region is then simply $\{\theta_{r(j)}\}_{j=1}^k$, where

$$k = \arg\min_{\kappa} \sum_{j=1}^{\kappa} f(\theta_{r(j)}^G|D)\Delta \ge p. \tag{13}$$

Boundaries of credible regions can be further refined if necessary by a simple interpolative scheme. Marginal credible intervals are estimated analogously.

Estimation of posterior predictive distributions: Distributions $g(x|D)$ for arbitrary functions $g(x|\theta)$, including return value distributions, are also trivially estimated as

$$g(x|D) = \sum_{j=1}^{m} g(x|\theta_j^G) f(\theta_j^G|D)\Delta. \tag{14}$$

4. Evaluation of LWM

In this section we evaluate the performance of LWM in three ways. First, for small samples of GP- and GEV-distributed data with known parameter values, we estimate coverage probabilities of credible regions and compare them with expected values. Then, we compare estimation of credible regions for parameter from LWM, Bayesian inference using the Metropolis–Hastings algorithm and a maximum likelihood scheme with bootstrap uncertainty

Please cite this article as: Wada, R., et al., Extreme value estimation using the likelihood-weighted method. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.063

Table 1
Description of the Monte Carlo experiment.

	Shape	Scale	Threshold/ Location	Sample size	Number of cases
GP	$\xi = -0.5, 0, 0.5$	$\sigma = 4$	$\psi = 0$	$N = 10, 20, 50$	$N_R = 1000$
GEV	$\xi = -0.5, 0, 0.5$	$\sigma = 4$	$\mu = 1$	$N = 10, 20, 50$	$N_R = 1000$

estimation, in terms of quality of inference and computational efficiency of inference. Finally, we assess whether the LWM method is able to identify a known measurement δ in simulated truncated samples.

For large samples and small measurement δ, we can assume that LWM and ML (using the standard likelihood) have similar statistical efficiency. Since ML is considered asymptotically efficient (Coles and Dixon, 1999), LWM is approximately so also. However, here we focus on small samples, for which we might speculate that LWM and ML would yield different statistical and computational efficiencies.

4.1. Coverage probabilities for credible regions

We simulate $N_R = 1000$ random realisations of samples of GP and GEV data with parameters listed in Table 1, with measurement $\delta = 0.005$ imposed. 3 different sample sizes N for each of 3 GP cases and 3 GEV cases are considered. For each realisation r of each sample size for each case, we evaluate credible region $C(\theta_r; p)$ corresponding to probability p for $p = 0.01, 0.02, \ldots, 0.99$. Using the N_R realisations, we estimate a coverage probability $\hat{p}$,

$$\hat{p} = \frac{1}{N_R}\sum_{r=1}^{N_R} I(\theta_r \in C(\theta_r; p)) \quad \text{for } p = 0.01, 0.02, \ldots, 0.99. \tag{15}$$

where I is the obvious indicator function. The discrepancy $\hat{p} - p$ in coverage probability is plotted against p for all GP cases in Fig. 1, and for all GEV cases in Fig. 2. The dashed horizontal lines in each panel of Figs. 1 and 2 correspond to $\alpha = 0.025$ and $\alpha = 0.975$ quantiles for the distribution of the Kolmogorov–Smirnov (KS) statistic $D_N = \sup_{p\in[0,1]}|\hat{p} - p|$. For sample size N, critical values Q for the KS statistic D_N are calculated using $Q = N^{-1/2}k_\alpha$, where k_α is a quantile of the Kolmogorov distribution with non-exceedance probability $1 - \alpha$

$$\Pr(K \le k_\alpha) = 1 - 2\sum_{j=1}^{\infty}(-1)^{(j-1)}\exp(-2j^2k_\alpha^2) = 1 - \alpha. \tag{16}$$

From the figures, we see that excellent agreement between actual and estimated coverage probabilities for critical regions is obtained in all cases. Of course, performance in general depends critically on the value of δ relative to the spread of samples values before truncation. As δ increases, inference becomes increasingly difficult and critical regions for parameters inflate.

4.2. Comparison of estimated critical regions from LWM with competitor methods

Here we compare estimated for parameter credible regions from LWM with 2 competing approaches, in terms of quality of estimate and computational efficiency of the estimation, based on a sample size $N=20$ from a GP distribution. The competitor methods considered are (a) Bayesian inference using a simple Metropolis-Hastings (MH) scheme (Metropolis et al., 1953), and (b) a simple maximum likelihood estimation with bootstrap resampling for uncertainty estimation. We also briefly compare credible regions for a sample size $N=200$, and compare estimates for marginal tail quantiles based on a sample size $N=20$.

Some care was taken in specifying schemes (a) and (b) so that reasonably fair comparison with LWM was possible. In LWM, we evaluate the posterior density on a rectangular grid of m prespecified parameter combinations $\{\theta_j^G\}_{j=1}^m$ as described in Section 3.3. The Bayesian MH (a) is an iterative scheme in which the product of the group likelihood and the near-uniform prior is evaluated for candidate parameter combinations corresponding to a Gaussian random walk with respect to the current state, and accepted with a certain probability (to achieve a specified proposal acceptance rate). With the variance of the random walk step adjusted to achieve reasonable acceptance rate of around 0.35 per candidate, the total number of accepted parameter combinations m_{MH} therefore represents a reasonable measure of the computational burden of the Bayesian inference, although the number of candidate posterior densities evaluated is larger than this. In the ML-bootstrap method (b), ML estimation is undertaken for a large number of bootstrap resamples of the original sample, with each ML estimation involving a function minimisation step. We use the number m_{BS} of bootstrap resamples as a measure of computational burden, but realise that the actual burden is larger due to ML

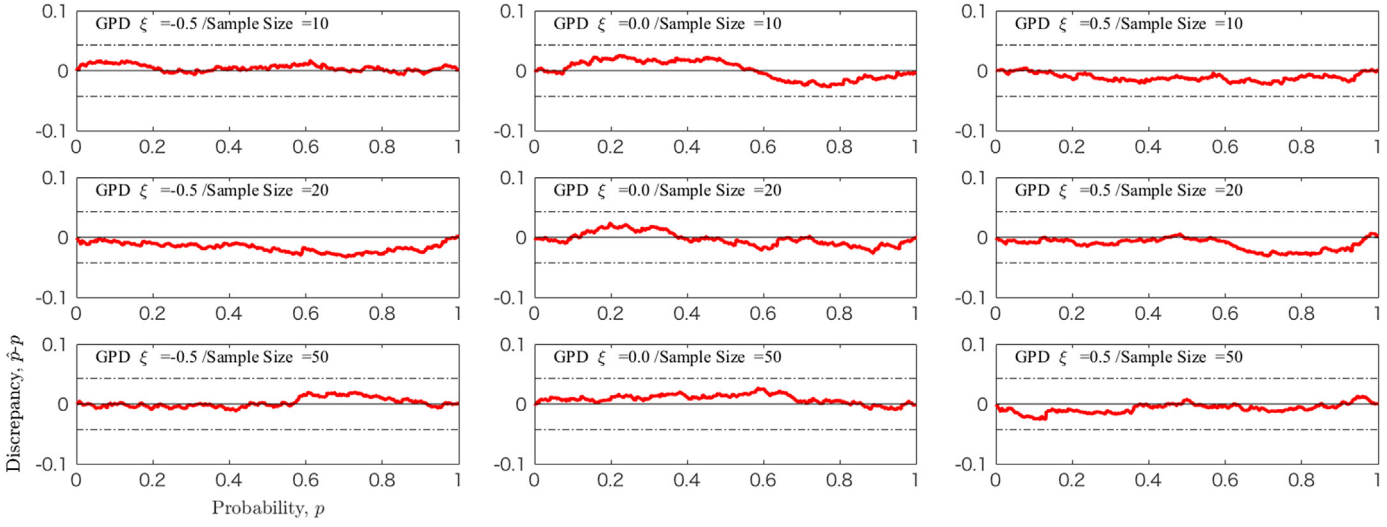


Fig. 1. Discrepancy $\hat{p} - p$ in coverage probabilities for credible regions, as a function of probability p, for GP samples. Rows represent different sample sizes $N = 10, 20, 50$ and columns represent different values of $\xi = -0.5, 0, +0.5$.

Please cite this article as: Wada, R., et al., Extreme value estimation using the likelihood-weighted method. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.063

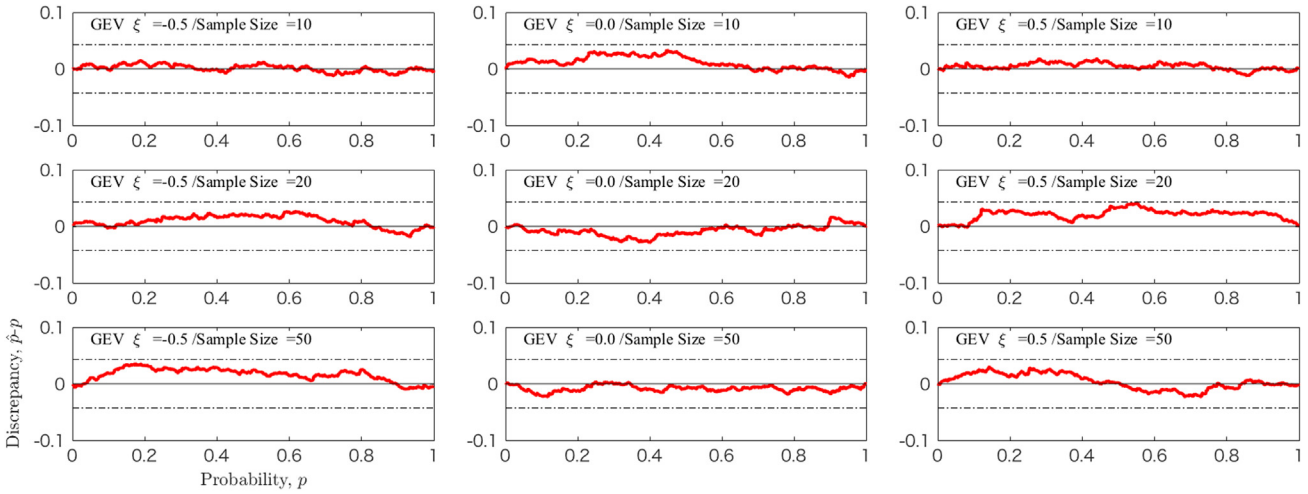


Fig. 2. Discrepancy $\hat{p} - p$ in coverage probabilities for credible regions, as a function of probability p, for GEV samples. Rows represent different sample sizes N = 10, 20, 50 and columns different values of ξ = − 0.5, 0, + 0.5.

estimation. Whenever ML failed, estimation as suggested by Pickands (1975) was performed.

The detailed comparison was set up as follows. We assume a GP-distributed sample of size N=20 with known parameters ξ = 0 and 0.5, σ = 1, ψ = 4, measurement δ = 0.005 and evaluate credible regions for parameters using LWM, Bayesian MH and ML-boostrap. For LWM, appropriate parameter index sets corresponding to $m = 10^4$ and $m = 10^6$ were specified. For Bayesian MH, estimates from chain lengths $m_{MH} = 10^4$ and $m_{MH} = 10^6$ were obtained. Care was taken that the chain converged to its stationary distribution, and that parameter combinations corresponding to MCMC burn in were not used for inference. To estimate credible regions, kernel density estimation (Botev et al., 2010) was used, involving still further computation relative to LWM. For ML-bootstrap, $m_{BS} = 10^4$ and $m_{BS} = 10^6$ resamples were generated and ML estimates obtained for each. Again, kernel density estimation was used to estimate credible regions.

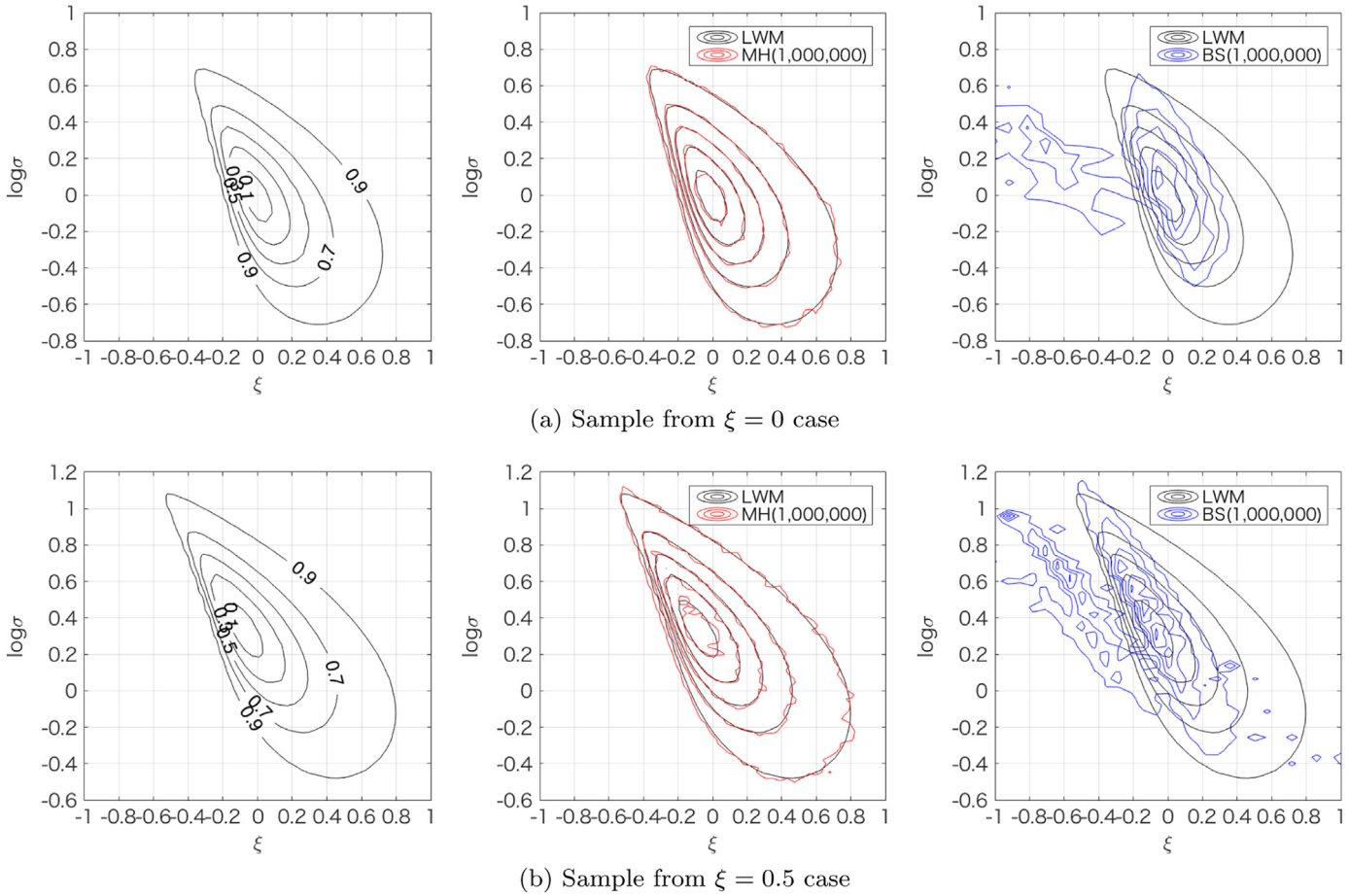


Fig. 3. Credible regions for GP parameter estimates from LWM (right), Bayesian MH (centre) and ML-bootstrap (left) for probabilities 0.1, 0.3, 0.5, 0.7 and 0.9. Results for 10^4 computations (LWM) and 10^6 computations (Bayesian MH and ML-boostrap). Sample size N=20 from GP distribution with ξ = 0 and 0.5, σ = 1, ψ = 4.

Please cite this article as: Wada, R., et al., Extreme value estimation using the likelihood-weighted method. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.063

Fig. 3 shows estimated credible regions corresponding to probabilities 0.1, 0.3, 0.5, 0.7 and 0.9 for the 10^4 case (LWM) and the 10^6 case for Bayesian MH and ML-bootstrap for both $\xi = 0$ and 0.5. For LWM, estimates based on $m = 10^4$ and $m = 10^6$ are indistinguishable. We conclude that $m = 10^4$ is sufficient for evaluation of credible regions for this sample. We observe that credible regions are not symmetric in ξ or $\log\sigma$, and that the domain of parameters is constrained by the identity $\xi(x^+ - \psi) + \sigma = 0$ when $\xi < 0$, where x^+ is the finite upper end point of the GP distribution. Credible regions are clearly not symmetric in parameters, and parameter estimates are clearly not multivariate normally distributed. A maximum a posteriori (MAP) estimate for GP shape near zero is found for both $\xi = 0.0$ and $\xi = 0.5$. This corresponds to a large estimation error for $\xi = 0.5$, and shows the importance of considering epistemic uncertainty. Specifically, the credible region for LWM clear does not exclude a GP shape of 0.5 in the case $\xi = 0.5$. 10^4 estimates for Bayesian MH and ML-boostrap (not shown) are very poor, since the number of parameter combinations available to describe credible regions corresponding to higher probabilities in particular is small. In this case, the somewhat arbitrary choice (e.g.) of kernel width for kernel density estimation will have a large undesirable influence on estimated credible regions. Estimates for credible intervals using Bayesian MH with $m_{MH} = 10^6$ are in much better agreement with LWM, but again uncertainties in the location the boundary of the credible region, especially for higher probabilities, is larger that for LWM. This is despite the fact that Bayesian MH uses at least 2 orders of magnitude more function evaluations. The figure illustrates also that the ML-bootstrap method is inadequate for estimation of credible regions, regardless of the number of bootstrap resamples used. Specifically, when a bootstrap resample fails to include the largest observed value in the sample, the estimation suggests a shorter-tailed distribution. As a result, the posterior density partitions as shown in the figure. We also note as expected that the maximum a posteriori (MAP) estimates for shape parameter ξ are biased towards more negative values for all inference methods.

As validation for a larger sample, estimation was repeated for a sample of size $N=200$ from the same GP distribution with $m = m_{MH} = m_{BS} = 10^6$. As can be seen from Fig. 4, inferences from LWM and Bayesian MH are in relatively good agreement. We see that posterior densities of parameters approach their multivariate normal asymptote. We also note the relative improvement in ML-bootstrap performance, but this inference still exhibits bimodality for $\xi = 0$. The negative bias of MAP estimates for ξ is also reduced for all inference methods.

Using the posterior densities illustrated in Fig. 3, estimates for the marginal GP quantiles with non-exceedance probabilities 0.9, 0.95 and 0.99 were also found, and are shown in Table 2. Since considerable variability between estimates based on 10^4 computational steps were observed for Bayesian MH and ML-bootstrap, these inferences were repeated 100 times (for the same underlying sample). Values quoted are means, and values following in parentheses are standard deviations from the 100 replicates. For other inferences, it was confirmed that uncertainty in estimates of quantiles was zero to two decimal places. Results are in line with expectations following the discussion of Fig. 3 above. For $m_{MH} = 10^6$, there is good agreement between LWM and Bayesian MH. However, inferences from ML-bootstrap are misleading.

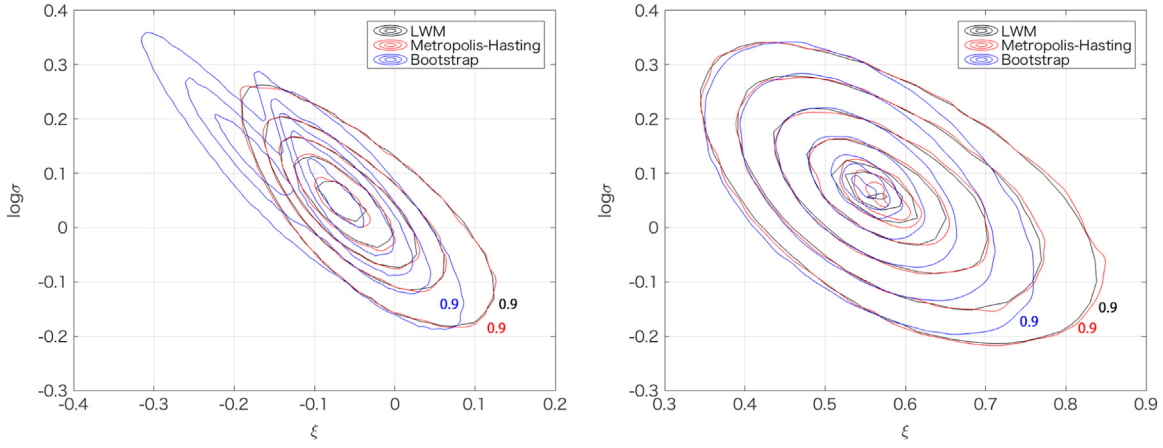
4.3. Validation of group likelihood

We now confirm that LWM correctly identifies a known value of measurement truncation δ used in the group likelihood. To achieve this, we conducted a simple validation exercise for the GP sample of size $N=20$ for $\xi = 0$, $\sigma = 4$, $\psi = 1$, with values truncated to whole numbers such that $\delta = 0.5$. For each of 1000 random realisations of the sample, we estimate discrepancies in coverage probabilities for credible regions with probabilities 0.01, 0.02, …, 0.99 for the parameters with a single assumed value for δ drawn from the set illustrated in Fig. 5. Then, as in Section 4.1, we estimate the KS statistic D_N for the discrepancy, and a 95% confidence level $Q = N^{-0.5}k_\alpha$ for the KS statistic, and record the ratio D_N/Q for each of the 1000 random sample realisations, for each values of δ. When the value of δ is specified appropriately, we expect that the ratio D_N/Q should be <1. The mean of D_N/Q as a function of δ is illustrated in Fig. 5. The figure shows that values of δ near the true value of 0.5 give the lowest values of D_N/Q as expected.

5. Application

5.1. Application to wave data

We now apply the LWM method to estimation of return values for significant wave height (H_S) from a small sample of 21 observations of peaks of H_S over a threshold of 4 m collected over a period of 10.74 years in a Japanese harbour (Goda, 1988). The sample will be referred to henceforth as Goda's sample for brevity. The data are given in decreasing order in Table 3. We compare



(a) Sample from $\xi = 0$ case (b) Sample from $\xi = 0.5$ case

Fig. 4. Credible regions from LWM (black), Bayesian MH (red) and ML-bootstrap (blue) for probabilities 0.1, 0.3, 0.5, 0.7 and 0.9. Results from 10^6 computations for sample size $N=200$ from the GP distribution with $\xi = 0$ and 0.5, $\sigma = 1$, $\psi = 4$. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

Please cite this article as: Wada, R., et al., Extreme value estimation using the likelihood-weighted method. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.063

Table 2
Estimation of marginal quantiles by LWM, Bayesian MH and ML-bootstrap for a sample size $N=200$ from the GP distribution with $\xi = 0$ and 0.5, $\sigma = 1$, $\psi = 4$.

	$\xi = 0$						$\xi = 0.5$					
	LWM		Bayesian MH		ML-bootstrap		LWM		Bayesian MH		ML-bootstrap	
	10^4	10^6	10^4	10^6	10^4	10^6	10^4	10^6	10^4	10^6	10^4	10^6
$F=0.90$	6.71	6.71	6.73(0.05)	6.71	6.10(0.02)	6.11	7.61	7.61	7.61(0.05)	7.61	6.91(0.00)	6.91
$F=0.95$	7.81	7.81	7.83(0.06)	7.81	6.87(0.05)	6.91	8.91	8.91	8.91(0.08)	8.91	7.78(0.05)	7.81
$F=0.99$	12.31	12.31	12.35(0.40)	12.31	8.78(0.05)	8.81	15.11	15.12	15.2(0.60)	15.22	9.92(0.06)	9.91

extreme value estimation from LWM (with group likelihood and $\delta = 0.005$ and the near-uniform priors specified in Section 3.2) and three approaches based on ML with different methods for quantification of uncertainty, assuming that data are drawn from a GP distribution with unknown shape and scale, but known threshold of 4 m. For ML, the three approaches used for uncertainty quantification are profile likelihood, the delta method and bootstrapping. We also compare inferences with those of Goda's method (Goda, 1988) which assumes a Weibull model for the sample. Note that the value of δ for LWM was set to 0.005 m since sample values are specified in metres to two decimal places, to capture uncertainty due to instrument precision. As discussed earlier, other sources of measurement uncertainty are also likely, and might be incorporated by increasing the value of δ.

Estimated 50-year return values are given in Table 4. Return value estimates from LWM and Goda are in reasonable agreement, but estimates from ML are lower. The 95% uncertainty bands from ML-delta method and ML-profile likelihood are narrower than for the other approaches. The ML-bootstrap uncertainty band is implausibly wide. Estimated extreme value tails from LWM, ML-profile likelihood and Goda's method are depicted in Fig. 7.

Credible regions with probabilities 0.5, 0.9 and 0.95 for GP parameters estimated using LWM, ML-delta method and ML-profile likelihood are illustrated in Fig. 6. MAP estimates from LWM, ML-delta method and ML-profile likelihood are similar, but the shapes and sizes of credible regions are quite different.

Measurement uncertainty δ may be larger than that corresponding to just instrument precision. To explore this possibility further, Table 5 gives the results of an investigation into the choice of δ appropriate for analysis of the Goda sample. Different choices (0.005 m, 0.05 m, and 0.5 m) of δ were considered. The table shows that for $\delta \leq 0.5$ m, estimates for the 50-year return value and its uncertainty are stable. However, the choice $\delta = 1.5$ m results in a reduction the return value estimate, although its uncertainty is relatively unchanged. These results illustrate how LWM allows us to deal explicitly with data uncertainty.

Fig. 8 shows that credible regions for GP parameters are stable as expected for $\delta \leq 0.5$ m.

Table 3
Sample of 21 H_S values (in metres) from Goda (1988).

n-th largest	1	2	3	4	5	6	7	8	9	10	11
H_S	8.36	7.02	6.94	6.85	6.74	6.20	5.92	5.68	5.57	5.42	5.34
n-th largest (*cont.*)	12	13	14	15	16	17	18	19	20	21	
H_S	5.10	5.09	4.95	4.81	4.77	4.63	4.61	4.41	4.34	4.11	

Table 4
Estimated 50-year return values (in metres) with corresponding 95% uncertainty bands.

	LWM	ML-delta method	ML-profile likelihood	ML-bootstrap	Goda
50-year RP with 95% intervals	10.21 (7.53, 20.36)	8.34 (7.75, 8.94)	8.34 (7.65, 10.95)	9.10 (5.70, 298.95)	10.38 (6.99, 13.77)

GP shape ξ and scale parameter σ estimates are negatively correlated, so that the observed sample can be equally well estimated using different combinations of ξ and σ corresponding to longer-tailed distributions (with smaller scale) or shorter-tailed distributions (with large scale). Return value estimates from these distributions will be different in general, and differences will increase with increasing return period. The problem of parameter identifiability increases as sample size decreases.

In summary, we note that uncertainty intervals from ML, estimated using both of the delta and profile likelihood methods, are too narrow. Uncertainty intervals from the bootstrap are too wide. Yet LWM gives statistically sound estimates of intervals. From an engineering perspective however, the uncertainty interval for the 50-year return period wave height estimated from LWM remains

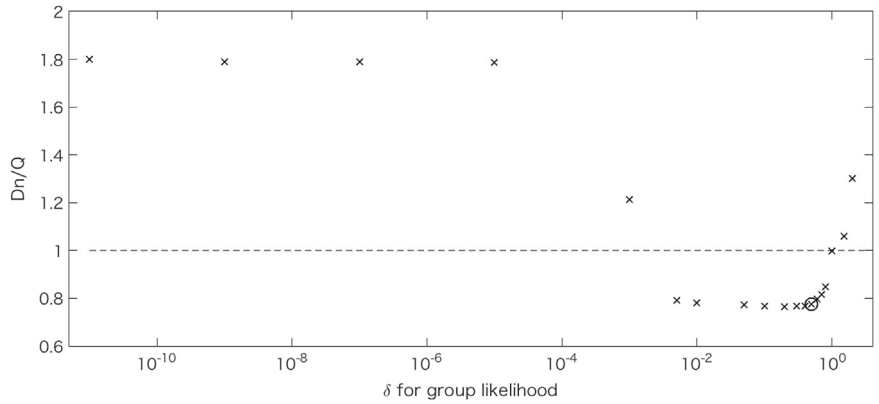


Fig. 5. D_n/Q as a function of δ. True value of δ is 0.5, shown as circle. Values of $D_n/Q < 1$ indicate reasonable fit of LWM.

Please cite this article as: Wada, R., et al., Extreme value estimation using the likelihood-weighted method. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.063

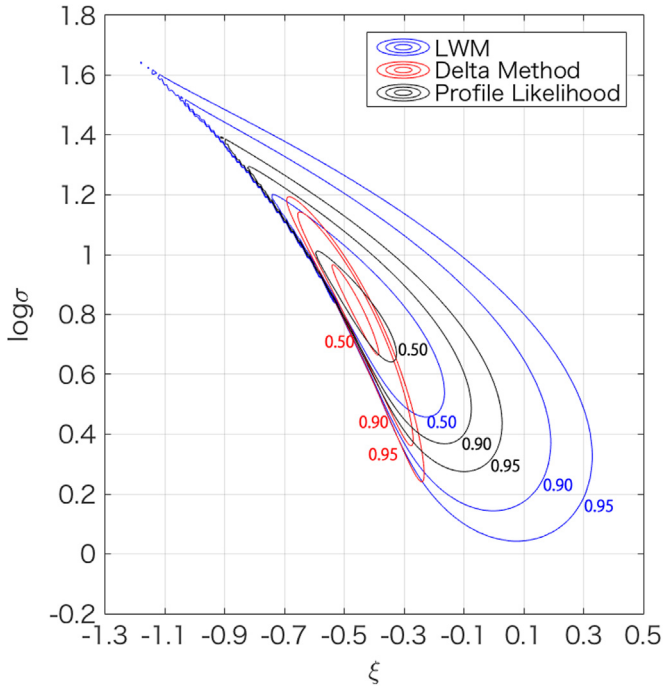


Fig. 6. Credible regions with probabilities 0.5, 0.9 and 0.95 for GP parameters from LWM, ML-delta method and ML-profile likelihood for Goda's sample of 21 H_S values.

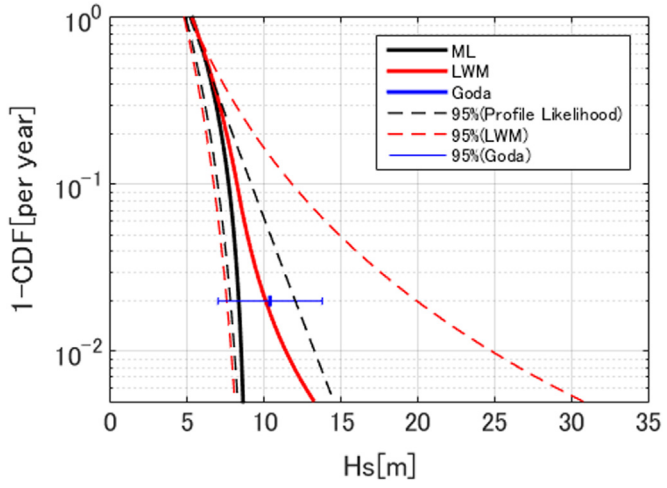


Fig. 7. Estimated cdf of extreme value distribution for LWM, ML-profile likelihood and Goda's method.

Table 5
Estimated 50-year return values (in metres) with corresponding 95% uncertainty bands using different values of δ (in metres).

	$\delta = 0.005$	$\delta = 0.05$	$\delta = 0.5$	$\delta = 1.5$
50-year RP with 95% CI	10.21 (7.53, 20.36)	10.21 (7.53, 20.50)	10.01 (7.38, 21.68)	9.01 (6.74, 20.18)

implausibly wide. This merely reflects the large epistemic uncertainty present in the estimation: LWM is a data-driven method, and wide credible intervals for return values cannot be avoided from small sample sizes. Sample size must be increased, or information from other sources incorporated if the interval is to be reduced. Since LWM is implemented here as a Bayesian procedure, it is straightforward to incorporate prior information, such as in Coles and Tawn (1996). We note however that satellite

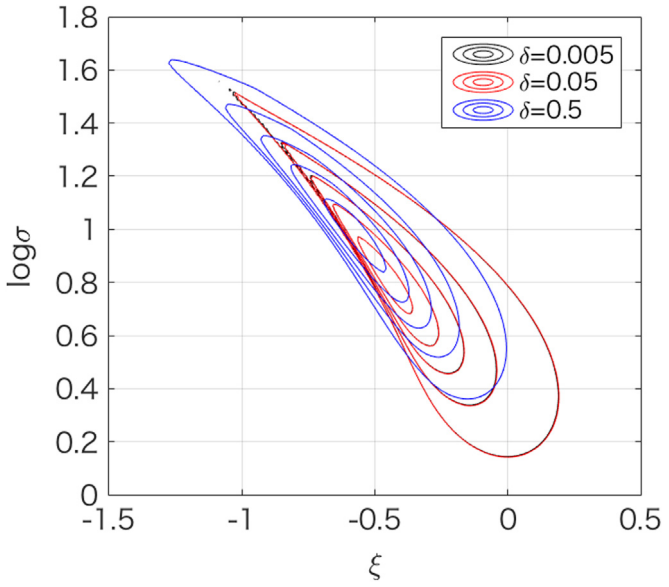


Fig. 8. Credible regions with probabilities 0.1, 0.3, 0.5, 0.7 and 0.9 for GP parameters from inferences with $\delta = 0.005$ m, 0.05 m and (0.5 m. Credible regions for $\delta = 0.005$ m and $\delta = 0.05$ m are almost superimposed.

observations of sea significant wave height in excess of 20 m have been reported (Hanafin et al., 2012), and that imposition of physical constraints on the characteristics of rare and extreme events is not always possible or appropriate. We might surmise that the occurrence of apparently implausibly wide credible intervals raises questions regarding the appropriateness of some existing practices in the field of ocean engineering, and obviates the need for careful incorporation of different sources of uncertainty in design.

5.2. Incorporating threshold uncertainty

The analysis above assumes that threshold ψ for GP estimation is fixed at 4 m. In most applications, threshold specification is a difficult issue, due to the trade off between the need for a high threshold to justify fitting an asymptotic model, and the need for a low threshold to increase sample size. Here we explore different LWM inferences from each of a set of pre-specified thresholds, and propose a weighted LWM scheme. The latter can be viewed as placing a uniform prior over each of a set of threshold choices.

Samples of peaks over a sufficiently high threshold can be assumed to follow the GP distribution approximately, and the threshold choice itself should not affect the estimated extreme value. Threshold choice can be based on stability of the estimated parameters (Coles et al., 2001). Usually, the lowest threshold value ψ_0 that gives near-constant estimation for all larger thresholds is chosen for subsequent inference. Assuming that the GP model is valid for threshold ψ_0, we can also express the GP distribution with respect to any other threshold $\psi > \psi_0$. In the modified distribution, GP shape ξ remains unchanged, but scale σ changes linearly according to

$$\sigma = \sigma_0 + \xi(\psi - \psi_0) \tag{17}$$

where σ_0 is the scale corresponding to threshold ψ_0. If we are to compare inferences for different thresholds, or to combine them appropriately, it is important to adjust scale estimates to that they refer to a common threshold choice, such as ψ_0. For the Goda data, we estimated credible regions for ξ and σ_0 using $\psi_0 = 4$ m, for each of $\psi = 4.0, 4.2, \ldots, 5.0$ m. Results are shown in Fig. 9. Credible regions are consistent across thresholds, but the magnitude of epistemic uncertainty increases as the ψ increases and sample size

Please cite this article as: Wada, R., et al., Extreme value estimation using the likelihood-weighted method. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.063

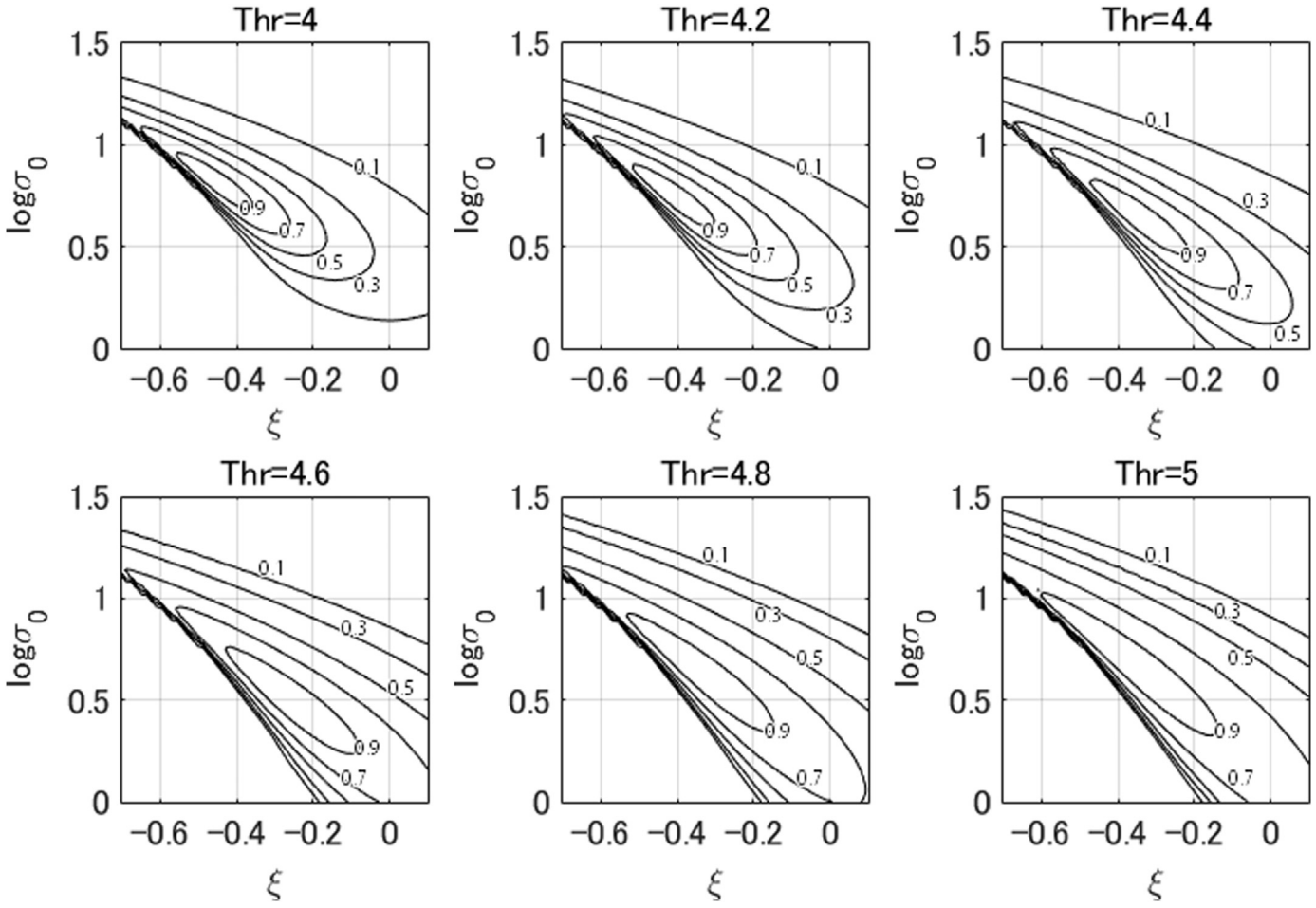


Fig. 9. Credible regions with probabilities 0.1, 0.3, 0.5, 0.7 and 0.9 for ξ and σ_0 with $\psi = 4, 4.2, \ldots, 5.0$ m for Goda's sample.

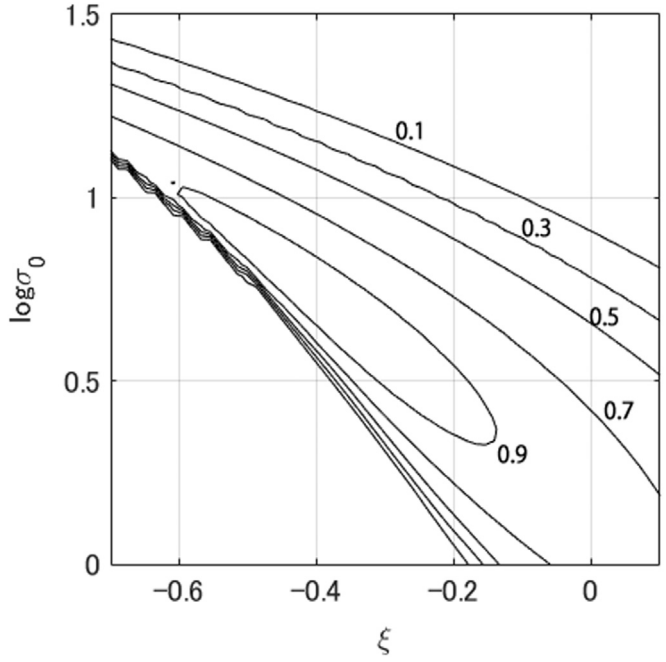


Fig. 10. Credible regions with probabilities 0.1, 0.3, 0.5, 0.7 and 0.9 for ξ and σ_0 from the threshold-aggregated posterior density for Goda's sample.

decreases.

Fig. 10 illustrates credible regions for parameters $\theta = (\xi, \sigma_0)$ estimated from the aggregated posterior density $f(\theta|D)$ over thresholds, expressed in terms of the posterior densities $f(\theta|D, \psi)$ for each of a set of n_ψ thresholds ψ as

$$f(\theta|D) = \int_\psi f(\theta|D, \psi) f(\psi) d\psi = \frac{1}{n_\psi} \sum_\psi f(\theta|D, \psi) \tag{18}$$

where prior $f(\psi)$ has point masses of weight $1/n_\psi$ at each of the n_ψ thresholds ψ considered.

Fig. 11 shows estimates of the 50-year return value with 95% credible interval for individual choices of ψ, and for the threshold-aggregated model. Again we observe that uncertainty in return value increases in general as threshold level increases. However, 50-year return value estimated from posterior distribution is stable for all threshold.

6. Conclusion

A straightforward likelihood-weighted method (LWM) to estimate extreme value models and return values from small samples of low quality data is proposed and demonstrated for samples of simulated and observed data. The method allows computationally efficient and accurate estimation of credible regions for model parameter estimates and posterior predictive distributions for return values. LWM exploits Bayesian inference for a group extreme value likelihood and near-uniform prior distributions for parameters, directly evaluating the posterior density on an index set of pre-specified parameter combinations. We demonstrate the

Please cite this article as: Wada, R., et al., Extreme value estimation using the likelihood-weighted method. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.063

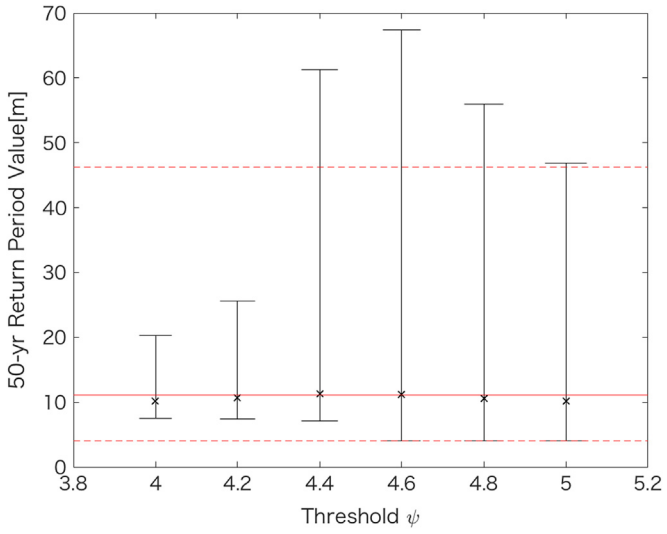


Fig. 11. 50-year return value (in metres) with 95% credible interval for individual choices of $\psi = 4, 4.2, \ldots, 5.0$ m for Goda's sample (in black). Also shown (in red) are the corresponding threshold-aggregated estimates. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

performance of LWM in simulation studies, and find that LWM provides superior inferences for small samples compared with Bayesian inference using the Metropolis–Hastings algorithm, and maximum likelihood estimation with bootstrap uncertainty quantification. We propose a threshold-aggregated LWM procedure for applications where threshold selection is problematic.

Attempting extreme value analysis from samples of less that 50 observations would be considered foolhardy by most. However, in reality, metocean engineers are often required to estimate return values in such circumstances. Given this, it is essential to do this as well as possible, and in particular to incorporate the effects of huge epistemic uncertainty sensibly in estimates of return values. LWM provides a simple, rational, consistent and computationally efficient means to achieve both these objectives. LWM suffers the same difficulties as any other extreme value model, and attempts to address a very difficult problem. However, in comparison with competitors, LWM exploits sound statistical methods to the full, including Bayesian inference with proper near-uniform priors and group likelihood. LWM provides an objective measure of uncertainty in extreme value estimation based strictly on data alone. We hope that LWM provides a useful addition to the metocean engineer's toolbox.

Acknowledgement

The authors thank colleagues at Lancaster University, the University of Tokyo and Shell for useful discussions. This work was supported by JSPS KAKENHI Grant Number JP15K18290.

References

Balkema, A.A., De Haan, L., 1974. Residual life time at great age. Ann. Probab., 792–804.
Bernardo, J.M., 1979. Reference posterior distributions for Bayesian inference. J. R. Stat. Soc. Ser. B, 113–147.
Bitner-Gregersen, E.M., Skjong, R., 2009. Concept for a risk based navigation decision assistant. Mar. Struct. 22 (2), 275–286.
Bitner-Gregersen, E.M., et al., 1990. Uncertainties in data for the offshore environment. Struct. Saf. 7 (1), 11–34.
Botev, Z.I., Grotowski, J.F., Kroese, D.P., et al., 2010. Kernel density estimation via diffusion. Ann. Stat. 38 (5), 2916–2957.
Cheng, R., Iles, T., 1987. Corrected maximum likelihood in non-regular problems. J. R. Stat. Soc. Ser. B, 95–101.
Coles, S.G., Dixon, M.J., 1999. Likelihood-based inference for extreme value models. Extremes 2 (1), 5–23.
Coles, S.G., Powell, E.A., 1996. Bayesian methods in extreme value modelling: a review and new developments. Int. Stat. Rev., 119–136.
Coles, S.G., Tawn, J.A., 1996. A Bayesian analysis of extreme rainfall data. Appl. Stat., 463–478.
Coles, S., Bawa, J., Trenner, L., Dorazio, P., 2001. An Introduction to Statistical Modeling of Extreme Values vol. 208. Springer, London.
Det Norske Veritas, DNV-RP-C205, 2010. Environmental conditions and environmental loads.
Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat., 1–26.
Forristall, G.Z., Heideman, J.C., Leggett, I.M., Roskam, B., Vanderschuren, L., 1996. Effect of sampling variability on hindcast and measured wave heights. J. Waterw. Port Coast. Ocean Eng. 122 (5), 216–225.
Giesbrecht, F., Kempthorne, O., 1976. Maximum likelihood estimation in the three-parameter lognormal distribution. J. R. Stat. Soc. Ser. B, 257–264.
Goda, Y., 1988. On the methodology of selecting design wave height, Coast. Eng. Proc. 1 (21).
Hanafin, J.A., Quilfen, Y., Ardhuin, F., Sienkiewicz, J., Queffeulou, P., Obrebski, M., Chapron, B., Reul, N., Collard, F., Corman, D., et al., 2012. Phenomenal sea states and swell from a North Atlantic storm in February 2011: a comprehensive analysis. Bull. Am. Meteorol. Soc. 93 (12), 1825–1832.
Hosking, J., Wallis, J.R., Wood, E.F., 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27 (3), 251–261.
Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. Proc. R. Soc. Lond. A: Math. Phys. Eng. Sci. 186, 453–461.
Jenkinson, A.F., 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Q. J. R. Meteorol. Soc. 81 (348), 158–171.
Jonathan, P., Ewans, K., 2007. Uncertainties in extreme wave height estimates for hurricane-dominated regions. J. Offshore Mech. Arct. Eng. 129 (4), 300–305.
Jonathan, P., Ewans, K., 2013. Statistical modelling of extreme ocean environments for marine design: a review. Ocean Eng. 62, 91–109.
Kass, R.E., Wasserman, L., 1996. The selection of prior distributions by formal rules. J. Am. Stat. Assoc. 91 (435), 1343–1370.
Kotz, S., Nadarajah, S., 2000. Extreme Value Distributions vol. 31. World Scientific, Imperial College Press, London.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21 (6), 1087–1092.
Mori, H., Hoshi, N., Yoshida, T., 2010. Evaluation of prior information in Bayesian inference. J. Jpn. Stat. Soc. 40 (1), 1–22.
Muir, L.R., El-Shaarawi, A., 1986. On the calculation of extreme wave heights: a review. Ocean Eng. 13 (1), 93–118.
Ochi, M.K., 2005. Ocean Waves: The Stochastic Approach vol. 6. Cambridge University Press, London.
Palutikof, J., Brabson, B., Lister, D., Adcock, S., 1999. A review of methods to calculate extreme wind speeds. Meteorol. Appl. 6 (02), 119–132.
Pickands III, J., 1975. Statistical inference using extreme order statistics. Ann. Stat., 119–131.
Pickands III, J., 1994. Bayes quantile estimation and threshold selection for the generalized Pareto family. Extreme Value Theory and Applications 1. Springer, Gaithersburg Maryland, pp. 123–138.
Scotto, M., Soares, C.G., 2007. Bayesian inference for long-term prediction of significant wave height. Coast. Eng. 54 (5), 393–400.
Smith, R.L., Naylor, J., 1987. A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. Appl. Stat., 358–369.
Smith, A.F., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J. R. Stat. Soc. Ser. B, 3–23.
Soares, C.G., 1986. Assessment of the uncertainty in visual observations of wave height. Ocean Eng. 13 (1), 37–56.
Wada, R., Waseda, T., 2013. Confidence interval of 3 parameter Weibull distribution in extreme value estimation. J. Jpn. Soc. Nav. Archit. Ocean Eng. 18, 135–142.

Please cite this article as: Wada, R., et al., Extreme value estimation using the likelihood-weighted method. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.063