International Journal of Marine Energy 14 (2016) 27–40

Contents lists available at ScienceDirect

International Journal of Marine Energy

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

Characterizing biological impacts at marine renewable energy sites

Lauren E. Wiesebron [a,*], John K. Horne [a], A. Noble Hendrix [b]

[a] University of Washington, School of Aquatic and Fishery Sciences, Box 355020, Seattle, WA 98195, USA
[b] QEDA Consulting, LLC 4007 Densmore Ave N, Seattle, WA 98103, USA

ARTICLE INFO

Article history:
Received 9 November 2015
Revised 29 January 2016
Accepted 1 April 2016
Available online 1 April 2016

Keywords:
Extreme value
Impact
Biological monitoring
Marine renewable energy
Hydroacoustics

ABSTRACT

Tidal energy is a renewable resource that can help meet growing energy demands, but uncertainties remain about potential environmental impacts of device installation and operation. Environmental monitoring programs are used to detect impacts and are a mandatory requirement of project operating licenses in the United States. Because tidal technology is new, studies describing environmental change due to tidal devices are rare, limiting information that can be used to characterize environmental impacts for monitoring requirements. Extreme Value Analysis (EVA) was used to characterize infrequent values from monitoring studies that are potentially associated with impact, defined as relevant biological change as a consequence of human activity, at a tidal energy site. EVA was adapted for monitoring aquatic organisms in the water column using an active acoustic dataset from Admiralty Inlet, a proposed tidal energy site. First derivatives were used to identify extreme value thresholds to improve precision of EVA parameters. Return level plots, which indicate the average period that extreme values are expected to appear, and uncertainty estimates of return level predictions, were generated using Markov Chain Monte Carlo (MCMC) simulations. Managers and site developers could use EVA to characterize rare values that may be associated with impacts, and tailor monitoring programs to include operational protocols for conditions under which these events occur.

1. Introduction

Global interest in renewable energy continues to increase due to rising energy demand and environmental concerns. The ocean provides renewable energy resources, in the form of wind, geothermal, and marine hydrokinetic energy [1]. Interest in developing tidal energy projects is growing because tides are a constant and predictable energy source [2]. While marine renewable energy (MRE) is an attractive alternate energy, implementing MRE technology includes uncertainty about how the technology will affect both biological and physical components of the environment [1,3,4]. To quantify and minimize impacts of MRE technologies, environmental monitoring programs are conducted at all sites through the life of a project.

Biological components of monitoring programs focus on the detection of change in variables such as diversity, length composition, or abundance of monitored species [5]. A successful biological monitoring program provides data that will help developers and regulators make informed decisions on adaptive management options [6]. To achieve this goal, it is essential that monitoring detects changes that are biologically relevant. These changes have been described as "effects" or "impacts".

* Corresponding author.
E-mail address: lw78@uw.edu (L.E. Wiesebron).

http://dx.doi.org/10.1016/j.ijome.2016.04.002
2214-1669/

Stewart-Oaten and Bence [7] define an "effect" on abundance as the "difference between the abundance at a site after an alteration and the abundance a site would have if the alteration had not occurred". An environmental impact can be distinguished from an effect by measuring "severity, intensity, or duration of the effect, and also the direction (positive or negative) of the effect" [8]. To detect an impact, baseline data (i.e. data collected before alteration began [7]) must be collected to facilitate comparison after devices are installed and become operational [9]. Determining the maximum level of "acceptable" impact is a high priority when forming a monitoring plan [10]. Observed values above a threshold can determine if a tidal project is allowed to continue operating [11]. Thus it is imperative that setting thresholds and characterizing impacts should be completed before MRE operations and concurrent monitoring begins [12].

At this time, there are no regulations for MRE monitoring procedures, technologies, or metrics for monitoring programs [13]. Monitoring programs are developed prior to the application for an operating license, and the amount of time to develop a monitoring plan that is acceptable to regulators can delay a developer's submission of the license application, adding cost and temporal uncertainty to permitting (e.g. [14]).

During monitoring plan development, characterizations of environmental impacts are often derived from preliminary samples or prior monitoring observations at analogous sites [15]. Because tidal technology is relatively new, studies describing environmental change due to tidal devices are scarce, restricting the information available to inform monitoring programs [2]. To set thresholds for biological monitoring at tidal energy sites, regulators can either estimate thresholds, or use models to characterize change.

Extreme Value Analysis (EVA) is an approach used to model values that are infrequent but are potentially associated with impacts caused by large change [16]. Although commonly used in engineering and hydrology [17,18], EVA has rarely been applied to ecological problems. EVA can be used in environmental monitoring to target rare but potentially significant events. These events are expected to be important to MRE regulators as there may be long-lasting consequences for both the ecosystem and tidal devices. Examples of this type of impact would include a collision between a marine mammal and a device, or altering fish migration patterns. This study evaluates whether EVA can be used to characterize infrequent values that are potentially associated with biological impacts at a tidal energy site.

2. Methods

2.1. Study site description

Admiralty Inlet is the proposed site of the Snohomish Public Utility District 1 (SnoPUD) tidal energy pilot project that received its project license from FERC on March 20th, 2014. The site is tidally dynamic, with an average tidal current speed of 1.4 ms^{-1} and maximum speeds reaching almost 3 ms^{-1}. Environmental concerns at the site include potential impacts on commercially important fish species and southern resident killer whales which are an iconic species in Washington State [19]. The proposed project, now dormant, would deploy two OpenHydro turbines (http://www.openhydro.com/) approximately one kilometer west of Whidbey Island. Two sub-sea power cables would connect the turbines to the onshore electric grid (Public Utility District No. 1 of Snohomish County, 2012).

Acoustic backscatter (i.e. reflected energy) data were recorded using an upward looking, bottom mounted BioSonics DTX echosounder (http://www.biosonicsinc.com/) operating at 120 kHz from May 9th until June 9th, 2011 [20]. The echosounder was placed at 55 m depth about 750 m off Admiralty Head at the SnoPUD tidal turbine site (Fig. 1). The echosounder sampled at 5 Hz for 12 min every 2 h. Because of a 3rd surface echo, data values were constrained to 25 m from the bottom. A −75 dB re 1 m^{-1} threshold was applied to remove noise [20]. Data were binned into 12 min samples and vertically integrated, yielding 361 datapoints [21].

2.2. Biological metrics

Echometrics are a suite of indices that quantify the variability of vertical biomass in the water column over space and through time [22]. Among the suite of seven Echometrics developed by Burgos and Horne [23] and refined by Urmy et al. [22], density and aggregation indices are used to characterize horizontal or vertical changes in biomass distribution, which could be applied to evaluate interactions between pelagic biomass and MRE devices. For the purpose of this study, high aggregation and density are assumed associated with high risk of interaction with MRE devices. The density metric is the mean volume-backscattering strength, or mean Sv (unit: dB re 1 m^{-1} [24] hereafter dB), which is proportional to biomass density. The aggregation index was used to quantify patchiness with values from 0 to 1, with 0 being evenly dispersed and 1 being aggregated.

2.3. Extreme Value Analysis

Extreme Value Analysis [25,16,26] is a statistical technique used to model the probability and periodicity of extreme values, which are rare values in the tail of a probability distribution. Observed extreme values are used to model extremes of greater magnitude [16], making this analysis unusual in that it focuses on the tails and not the mean of a sample distribution.

Fig. 1. Study location within Puget Sound, Washington (upper right), and location of the acoustic package (left) within the SnoPUD proposed turbine location.

2.3.1. Peaks-over-threshold (POT)

In the peaks-over-threshold (POT) method, extreme values are identified as exceedances above a threshold. These exceedances follow a generalized Pareto distribution (GPD) which is given by Pickands [25]:

$$G(z) = \begin{cases} 1 - \left(1 + \frac{\varepsilon(z-u)}{\sigma}\right)^{-1/\varepsilon} & \text{if } \varepsilon \neq 0 \\ 1 - \exp\left\{\frac{-(z-u)}{\sigma}\right\} & \text{if } \varepsilon = 0 \end{cases} \quad (1)$$

where u is the threshold, σ is the scale parameter, ε is the shape parameter, and $\sigma > 0$.

The scale parameter determines the steepness of the GPD. The shape parameter determines whether the GPD is bounded. The sign of the shape parameter determines the behavior of the GPD [16]. If the shape parameter is negative, the GPD is finite. If positive, then the GPD can continue to infinity. To perform a POT analysis, first a threshold (u) is selected, then the scale (σ) and shape (ε) parameters are fit to the data to model extreme values.

2.3.2. Identifying the extreme value threshold

Selecting the threshold for fitting the GPD to a frequency distribution is an important but difficult step in applying the POT method. If the threshold is too low then the model will be biased by including observations from the middle of the frequency distribution. If the threshold is too high then the model will be fit to too few data points and the variance of the GPD parameter estimates will increase [27,28]. The ideal threshold is the lowest value that includes as many excesses as possible, while still meeting the statistical assumptions of the GPD.

The threshold is usually defined visually [29] using mean residual life (MRL) plots and parameter stability plots. An MRL plot shows the mean number of values above a threshold as the threshold value is increased. If a GPD is valid for excesses at a threshold u_0, it should also be valid for thresholds $u > u_0$, with the scale parameter adjusted to the threshold u [16]. So $E(X - u \mid X > u)$ is a linear function of u, and the mean excesses change linearly with u at values of u for which the GPD is appropriate. The optimal GPD threshold is identified as the value where the MRL curve becomes linear. In contrast, the parameter stability plot shows the fit of the GPD scale or shape parameters for successive thresholds. The rationale for this method is that the shape ε and adjusted scale parameter σ^*, with $\sigma^* = \sigma_u - \varepsilon u$, should be constant above u_0, if u_0 is a valid threshold for the GPD [16]. In a parameter stability plot, the threshold is identified as the value where parameter estimates become stable, or near-constant. The adjusted scale and shape parameter plots are often complements of each other, so visual diagnostics on only one is necessary.

Interpretation of MRL and parameter stability plots is challenging. Since the MRL plot is rarely smooth, it is difficult to decide where linearity is achieved. Interpretation of the parameter stability plot is a little easier, but in both cases the choice of threshold is subjective [30,31].

An objective and automated way of selecting a threshold for extreme values is to take the derivative of the threshold diagnostic plots, and identify the value where the derivative first equals zero. Plot functions are smoothed to remove local

variability using a polynomial kernel density smoother [32], implemented using the KernSmooth package in R [33]. Derivatives are then calculated for the smoothed functions, and the inflexion point corresponding to the best threshold estimate from each plot is identified.

To evaluate how dependent the threshold value is on the proportion of data used, thresholds were calculated for random subsets of the Admiralty Inlet mean Sv and aggregation index data. Derivatives were first calculated for an MRL plot of the full dataset ($n = 361$), and then were calculated on random subsets with the sample size for each set decreasing by one datapoint at a time. A threshold value was obtained for 350 ($n = 360$ to $n = 11$, below which a threshold was not identifiable) subsets of Admiralty Inlet data, for both mean Sv and aggregation index metrics.

2.3.3. Fitting the GPD using Bayesian methods

2.3.3.1. Applying Bayesian theory to POT. While the most widespread method for fitting model parameters is maximum likelihood estimation (MLE), we performed a POT analysis using Bayesian inference.

Bayes theorem can be summarized as:

$$P(\theta|\text{data}) = \frac{L(\theta|\text{data}) \times P(\theta)}{P(\text{data})} \tag{2}$$

where $P(\theta|\text{data})$ is a probability distribution for an unknown variable. $L(\theta|\text{data})$ is the likelihood of the variable taking on a value given the data, in this case, the value of the scale or shape parameter, $P(\theta)$ is the prior, and $P(\text{data})$ is the sum of all the possible ways of observing the data. The primary difference between MLE and Bayesian methods is that MLE uses estimate maximization whereas Bayesian analysis uses integration. The Bayesian method will result in a more conservative evaluation of risk since it takes into account parameter uncertainty by integrating over all probable values. A conservative estimate of parameter uncertainty is appropriate for MRE monitoring to ensure that the range of outcomes is not underestimated. As posterior distributions are analytically challenging to compute, they are typically simulated numerically. To obtain the posterior distribution for the GPD scale σ and shape ε parameters, a Markov Chain Monte Carlo (MCMC) simulation [34] was used (see below).

A Bayesian analysis includes a prior, which is information on the probability distribution of parameter values, formed without knowledge or previous experience with the sample data. The use of prior information needs to be justified as it influences the posterior distribution. For this study, the use of an informative prior could not be justified as there is a dearth of information on biotic distributions at tidally dynamic sites. Uninformative or flat priors for the scale and shape parameters were used. These priors were centered at 0 and have high variance to ensure that no bias was introduced to the posterior parameter distributions: $\sigma \sim \text{N}(0,1000)$; $\varepsilon \sim \text{N}(0,100)$.

2.3.3.2. MCMC application. The MCMC method used for this study was a Metropolis–Hastings sampler. The negative log-likelihood (NLL) for the GPD is given by [16]:

$$\text{NLL}(u,\sigma,\varepsilon;x) = -\left\{-n\log\sigma - \left(1+\frac{1}{\varepsilon}\right)\sum_{i=1}^{n}\log\left[1+\varepsilon\frac{(x_i-u)}{\sigma}\right]\right\} \tag{3}$$

where u is the threshold, σ is the scale parameter, ε is the shape parameter, x is the data, n is the number of values in the dataset.

As is customary in MCMC simulations, the first 20% of the chains' accepted draws were discarded as a burn-in period, and then chains were thinned according to the autocorrelation between chain draws [35]. To obtain thinned chains with the same number of draws for both indices (64,000 draws), the starting MCMC chains for the aggregation index were twice as long (4,000,000) as chains for mean Sv (2,000,000). The autocorrelation between aggregation index draws was twice as strong as that for mean Sv values. Two tests, Geweke [36] and Gelman-Rubin [37] were performed on six chains, starting with different pairs of initial values to ensure that the MCMC chain was converging on the same posterior distribution.

Posterior distributions for the GPD scale and shape parameters were produced for both mean Sv and the aggregation index. MCMC jump size (Table 1) was iteratively selected to obtain a well-mixed chain (30–40% draws accepted) [35].

2.3.3.3. GPD parameter sensitivity to threshold value. Because the threshold determines what portion of the data is fit to the GPD, it is important to examine the sensitivity of the GPD scale and shape parameter estimates. Simulated datasets were generated following GPD distributions with known threshold (for all simulations, the true threshold had a value of 20), scale, and shape parameters. The MCMC routine was then used to fit the scale and shape parameters while increasing the threshold value by increments of 0.1 units until it reached a value of 22. The simulations were run on two datasets with the scale and

Table 1
MCMC parameter values for mean Sv and aggregation index simulations.

Metric	Scale jump size	Shape jump size	% draws accepted	% draws out of bounds	Chain length	Thinning interval
Mean Sv	1.54	0.42	35	5	2,000,000	25
Aggregation index	0.18	0.46	36	21	4,000,000	50

shape combinations from the median mean Sv and aggregation index parameter posteriors. For both parameter combinations, 18 separate simulations were run. The mean of these simulations were used to evaluate the sensitivity of the GPD scale and shape parameters to the threshold.

2.3.4. Return level

While it is informative to examine values of location, shape, and scale parameters of the fitted GPD, further insight can be gained from examining return levels. The value q_p is the return level associated with the return period $1/p$, and is the value that is expected to be exceeded on average once every $1/p$ time units [27]. Return levels for data exceedances are generated by inversing the GPD cumulative density function (Eq. (1)). For an arbitrary probability p, the corresponding return level q_p is [27]:

$$q_p = u + \frac{\sigma}{\varepsilon}(p^{-\varepsilon} - 1) \tag{4}$$

where u is the threshold, σ is the scale parameter, ε is the shape parameter, q_p is the return level, and p is the return level probability.

Return levels q_p were plotted as a function of return periods $1/p$, to obtains a return level plot that shows the expected periodicity for data excesses and values extrapolated beyond the range of the sampled data.

2.3.5. Bivariate peaks-over-threshold analysis

The bivariate POT method fits the GPD to two variables as a joint process. This method can be used to examine the correlation of processes underlying extreme values. While multivariate extreme value theory is well-developed, model computation and validation are challenging due to greater independence between high-level extreme event processes [16]. There are several methods to obtain a bivariate model, including a logistic model, a bilogistic model, or an asymmetric logistic model. The logistic model was chosen over the bilogistic model and the asymmetric logistic model through an analysis of deviance, which was possible because the models were nested. The logistic model is given by:

$$G(x,y) = (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha}, 0 < \alpha < 1 \tag{5}$$

where x and y are the fitted univariate GPDs for the x and y variables [16] and α is the correlation between the two variables. Independence between the two variables is obtained when $\alpha = 1$, and inversely, dependence is obtained when α approaches 0. Results of the bivariate analysis are GPD scale and shape parameter estimates for x, y, and an α estimate that indicates the dependence between x and y variables. In this study, following Segers and Vanderwalle [38], a single threshold value for the bivariate distribution was found so that the same number of mean Sv and aggregation index observations were above the threshold. The logistic bivariate model was fitted to mean Sv and the aggregation index using a maximum likelihood estimation (MLE) function provided by the evd R package [39].

3. Results

3.1. Threshold estimation

3.1.1. Mean Sv threshold diagnostics

Visual interpretation of diagnostic plots resulted in a preliminary threshold estimate. The mean residual life plot is approximately linear between $u \approx -75$ dB and $u \approx -71$ dB (Fig. 2a). While it may appear that linearity is not achieved until $u \approx -69$ dB, there are only 6 datapoints above -69 dB which increases uncertainty of an estimate. Patterns in the shape parameter stability plot (Fig. 2b) mimic those of the mean residual life plot (Fig. 2a). The shape parameter appears to be stable until about $u \approx -75$ dB, which is also the value where variance sharply increases. After $u \approx -71$ dB, the sharp increase in variance indicates that there are too few values to estimate parameter stability.

To obtain a more precise estimate of the threshold, the derivative method was applied to mean Sv data (Fig. 2). After smoothing both the MRL (Fig. 2a) and parameter stability plots (Fig. 2b), and taking the derivative (Fig 2c, d), the first point where dY = 0 for the parameter stability plot is $u = -74.48$ dB, and for the mean residual life plot $u = -74.58$ dB. These values are very close and are consistent with the visual diagnostics (Fig. 2a, b). The threshold for the POT analysis was set to the average of the two values from the derivative plots, which is $u = -74.53$ dB. A threshold of $u = -74.53$ dB results in 90 exceedances, which is 25% of the data (Fig. 3).

3.1.2. Aggregation index threshold diagnostics

Diagnostic plots for the aggregation index metric (Fig. 4) differed from the mean Sv threshold diagnostic graphs. The mean residual life plot (Fig. 4a) increased rapidly until about $u \approx 0.14$ m^{-1}, where variance increased and the plot became approximately linear. The linear trend is more visible in the scale parameter stability plot (Fig. 4b), where the shape parameter estimate decreased from $u \approx 0.05$ m^{-1} to $u \approx 0.15$ m^{-1}, then it remained constant until $u \approx 0.2$ m^{-1} where the variance increased steadily.

A threshold estimate was obtained using the derivative method for each plot. The threshold value from the mean residual life plot derivative was $u = 0.135$ m^{-1}. The first value where dY = 0 for the derivative of the parameter stability plot was

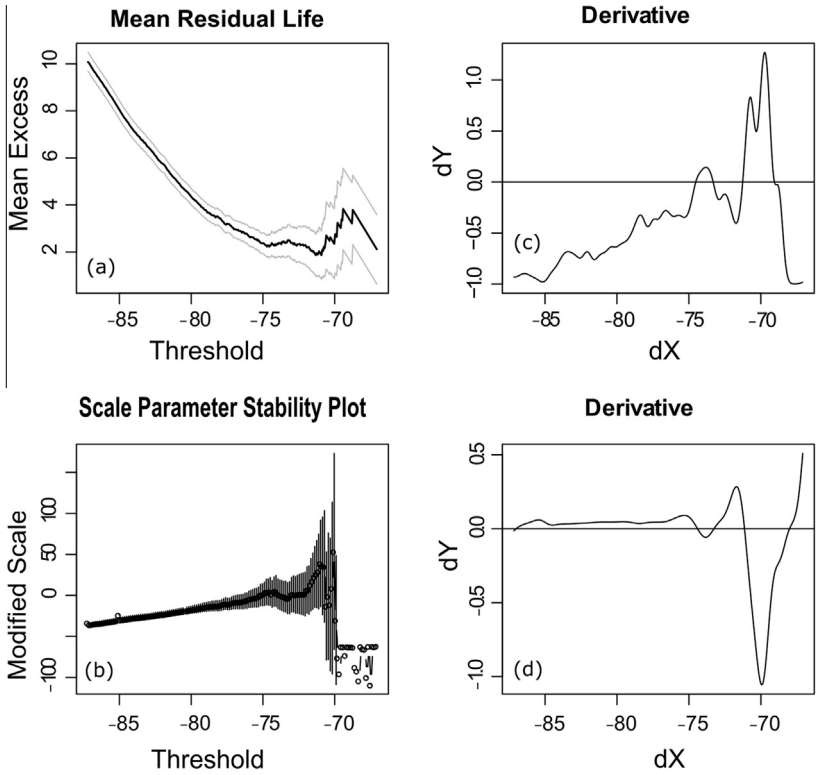

Fig. 2. The mean residual life plot (a) and scale parameter stability plot (b) with the black line representing the smoothed plots, the corresponding derivatives (c and d) of the smoothed plots, with the black line showing dY = 0 for mean Sv.

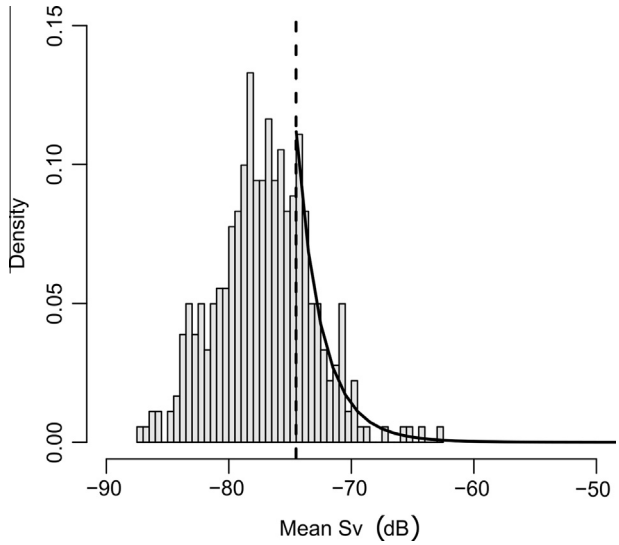

Fig. 3. Histogram of mean Sv values with the threshold marked at −74.53 dB (dotted line) and GPD fit (solid line). The scale and shape parameter values were obtained from posterior medians.

$u = 0.05\ \mathrm{m}^{-1}$, but this value did not stabilize in the shape parameter estimate, which is illustrated by the amplitude of values around this point. The threshold from the parameter stability plot was set to $u = 0.146\ \mathrm{m}^{-1}$, the second inflexion point in the derivative plot, as it also matched the result from the MRL plot (Fig. 4c). The average of the threshold estimates from the MRL and parameter stability plots is $0.140\ \mathrm{m}^{-1}$. As with mean Sv, this value is consistent with visual interpretation of the mean residual life and parameter stability plots. A threshold of $0.140\ \mathrm{m}^{-1}$ results in 26 exceedances for the aggregation index, which is 7% of the data (Fig. 5).

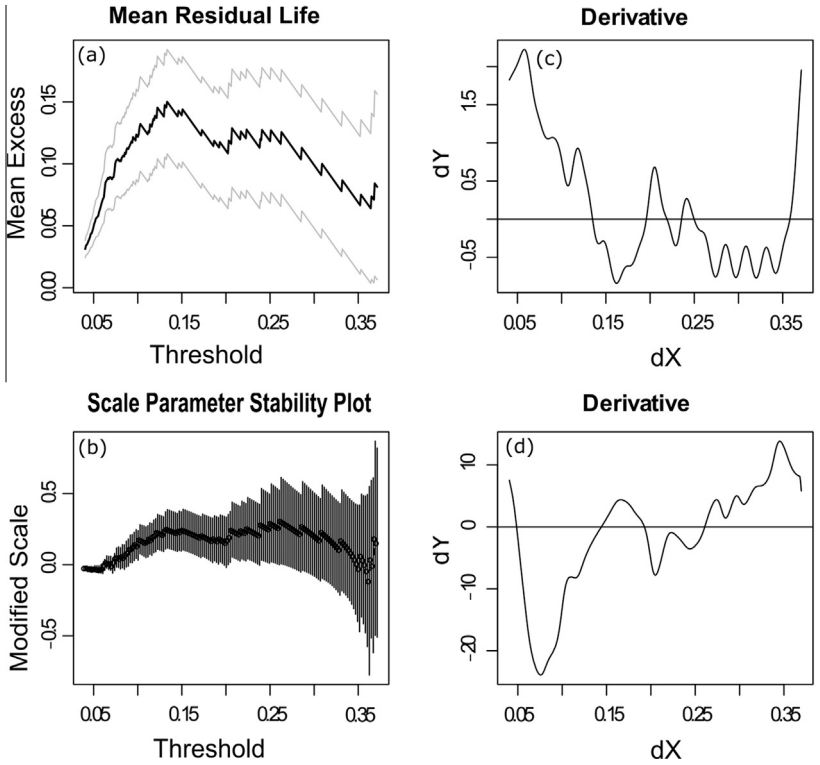

Fig. 4. Graphical and derivative aggregation index threshold diagnostic plots. The mean residual life plot (a) and scale parameter stability plot (b) with the black line representing the smoothed plots, the corresponding derivatives (c and d) of the smoothed plots, with the black line showing dY = 0.

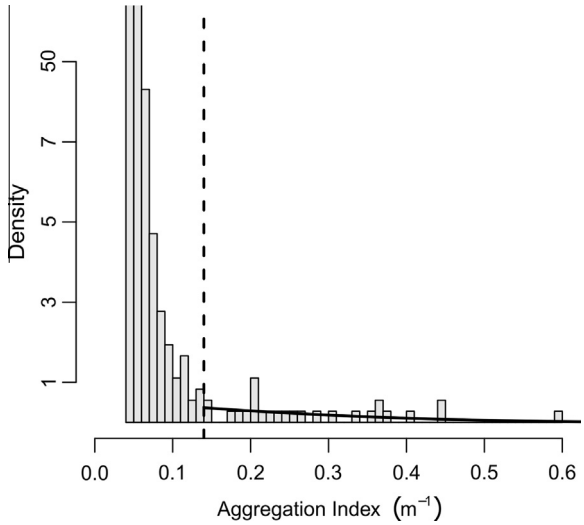

Fig. 5. Histogram of aggregation index values with the threshold marked at 0.140 m^{-1} (dotted line) and GPD fit (solid line). The scale and shape parameter values were obtained from posterior medians.

3.1.3. Threshold robustness

The robustness of the derivative analysis was examined by calculating thresholds for subsets of data of decreasing size. For both the mean Sv and aggregation indices, reduction in the sample size affected the threshold estimate (Fig. 6). The threshold for mean Sv remains stable until about $n = 110$, where the estimated threshold starts to decrease with sample size. The threshold for the aggregation index becomes unstable at a sample size of approximately $n = 235$. The aggregation index threshold is more dependent on sample size as 7% of the data are considered extreme compared to 25% of the mean Sv data.

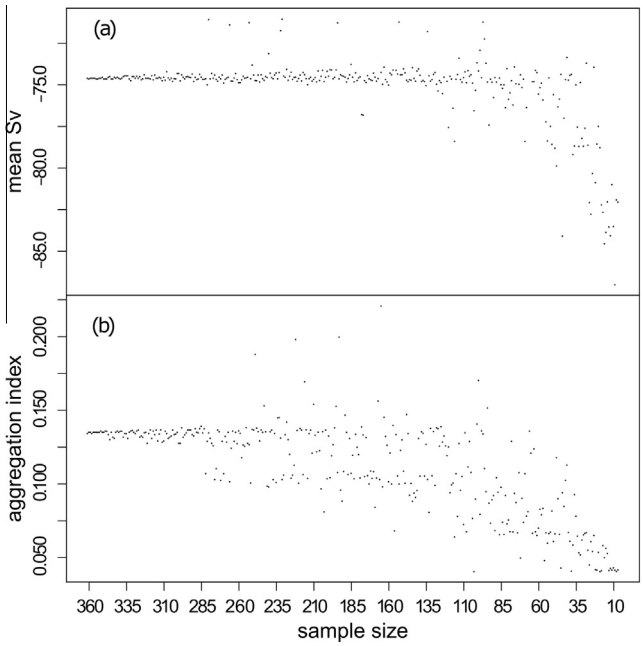

Fig. 6. Threshold estimates from the derivative method plotted against data sample size for (a) mean Sv (b) aggregation index.

In both cases, the estimated threshold is considered robust as the threshold estimate does not immediately change with decreasing sample size.

3.2. Bayesian analysis

3.2.1. Shape and scale posteriors

The first objective of the Bayesian analysis was to estimate posterior distributions for the GPD scale and shape parameters using the MCMC method. The scale and shape posterior distributions for both mean Sv and aggregation index metrics show a slight right-skew (Fig. 7). The median for the mean Sv scale and shape parameters are respectively 2.14 and 0.07, the median for the aggregation index scale and shape parameters are 0.19 and −0.26.

The aggregation index shape parameter posterior distribution has a larger range (−1 to 1) and larger tails compared to the mean Sv shape parameter posterior distribution (range: −0.2 to 1) (Table 2). The greater 95% credible interval range is attributed to the smaller number of points above the aggregation index threshold compared to the mean Sv threshold. The range (−1 to 1) of the shape parameter for both variables contained positive and negative values. There is a 0.27 probability that the mean Sv shape will be negative, and a 0.87 probability that the aggregation shape will be negative. The sign of the shape parameter will affect the shape of return level predictions (i.e. positive shape will result in infinite return levels, whereas a negative shape will lead to finite return levels).

3.2.2. MCMC diagnostics

3.2.2.1. Convergence. Three diagnostics were computed for the MCMC chains to verify that they converged to stationary distributions. The first diagnostic was to start multiple chains from different pairs of scale and shape values to ensure that they converged at the same posterior distribution. Six chains of a million draws, which had a range of starting values (scale: 0 to 15, shape: −2 to 2), converged to the same stationary distribution. The Gelman-Rubin test resulted in point estimates of the potential scale reduction factors that were equal to 1, indicating that convergence had been achieved. The Geweke test resulted in Z-scores that were between −2 and 2 indicating that the first 10% were not significantly different from the last 50% of the scale and shape chains. Collectively, these tests confirmed that the chain was converging to a single stationary distribution.

3.2.2.2. MCMC sensitivity to threshold value. The sensitivity of the GPD scale and shape estimates to the threshold value was examined. MCMC simulations performed on simulated GPD data with known scale and shape parameters returned consistent scale and shape parameters when the threshold input was within 0.2 to 1.2 units to the right of the defined threshold (Fig. 8). These results varied depending on the combinations of scale and shape parameters in the simulated data. In all cases, the MCMC algorithm was unable to fit the true scale and shape parameters once the threshold value deviated greater than 1.2 units from the defined threshold.

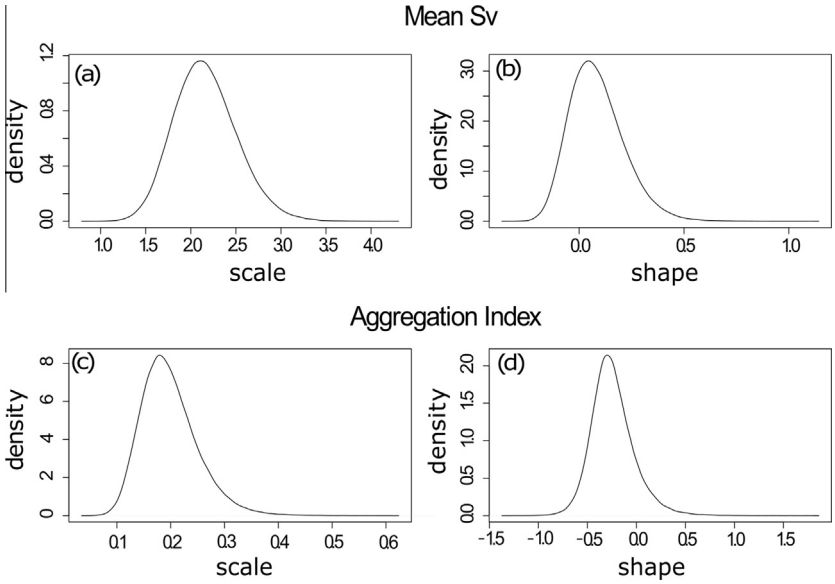

Fig. 7. Posterior distribution for the mean Sv (a) scale and (b) shape parameters, and aggregation index (c) scale and (d) shape parameters.

Table 2
Median parameter values and the 95% credible interval in parentheses (lower, upper bounds).

Metric	Scale	Shape
Mean Sv	2.147 (1.541, 2.913)	0.071 (−0.129, 0.378)
Aggregation index	0.191 (0.114, 0.319)	−0.264 (−0.619, 0.257)

3.2.3. Return levels

Using the GPD parameter posteriors, return level plots with credible intervals were generated to examine how the return period changes as extreme values increased. The median mean Sv return level steadily increased as the return period increased, reaching −40 dB every 10 years (Fig. 9a).

The aggregation index return level exhibited a different behavior than that of mean Sv. The aggregation index median return level increased slowly until it reached about 0.75 m^{-1} at 1 year, and then increased exponentially, reaching an asymptote at 0.75 m^{-1} (Fig. 9b). The aggregation index credible intervals increased at a greater rate than the intervals for mean Sv, with the upper quantile slopes rapidly increasing after a 1 day return period.

3.3. Bivariate peaks-over-threshold

The logistic model for the bivariate POT analysis was fitted to the mean Sv and aggregation index data to examine whether additional information was provided by modeling the two variables as a joint process. The bivariate threshold for mean Sv was −75.3 dB and for aggregation index was 0.058 m^{-1}, which are both lower than the univariate thresholds and result in 115 exceedances or 32% of the data being fit to the bivariate model. There were 39 data points jointly above the mean Sv and aggregation index thresholds. The α-value is 0.95 suggesting weak correlation, and almost independent variables. MLE estimates for the bivariate model are different than the median univariate values estimated by Bayesian inference (Table 3). The difference between the univariate and bivariate parameter estimates (Table 4) is greater for the aggregation index than for mean Sv, which is reflected in the large discrepancy between the univariate and bivariate aggregation index return level curves (Fig 9c, d).

Return levels for mean Sv (Table 3) in the bivariate case are lower than in the univariate case. This is an expected result as high aggregation and densities rarely occur at the same time, and joint events are predicted to occur less frequently than single events. The aggregation index bivariate return levels start lower than the univariate, but quickly exceed them for return periods greater than a day. The shape of the aggregation index bivariate return level is attributed to the threshold value of the bivariate model being lower than that of the univariate model.

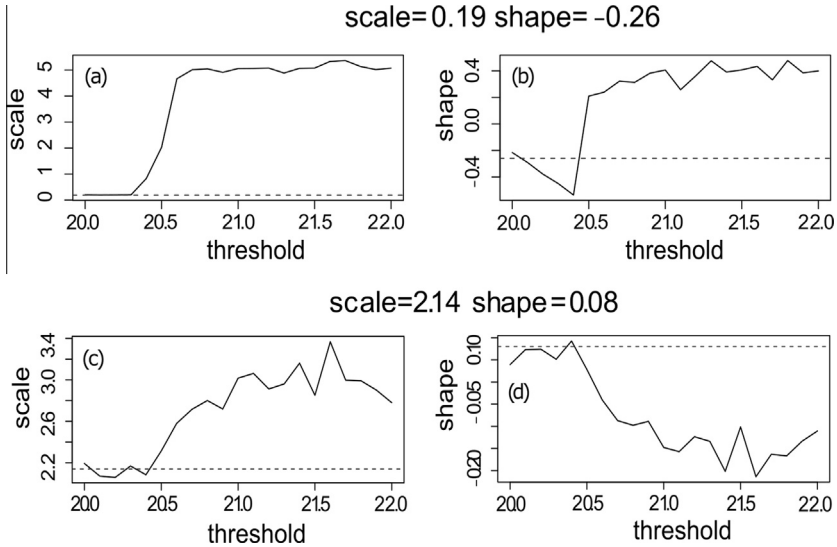

Fig. 8. Median values from MCMC generated posteriors of scale and shape parameters plotted as a function of threshold, for simulated data with known parameters, (a) scale and (b) shape parameter combination for mean Sv and for (c) scale and (d) shape aggregation index. The known scale and shape parameter values are depicted by the dotted line.

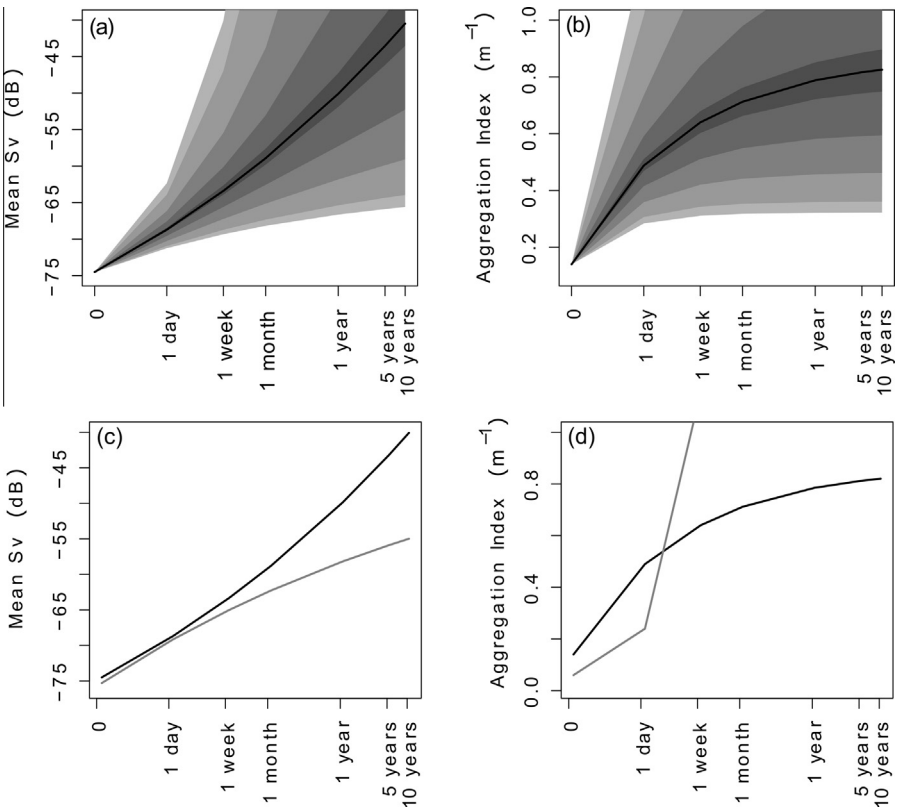

Fig. 9. Median (a) mean Sv and (b) aggregation index return levels. The solid lines is the best fit, and gray colors indicate credible intervals, from 10% (darkest gray) 40%, 80%, 90% lightest gray). Median univariate (black line) and bivariate (gray line) return level for (c) mean Sv and (d) aggregation index.

Table 3
Median values and 95% (lower, upper) credible intervals for mean Sv and aggregation index return levels for univariate GPD fit, MLE return level for bivariate GPD.

	Mean Sv		Aggregation index	
Return period	Univariate	Bivariate	Univariate	Bivariate
1 day	−68.66 (−71.21, −62.43)	−68.86	0.49 (0.29, 1.25)	0.46
1 week	−63.32 (−69.28, −40.64)	−63.93	0.64 (0.31, 2.79)	0.59
1 month	−58.80 (−68.12, −9.68)	−59.92	0.71 (0.32, 4.57)	0.64
1 year	−49.88 (−66.58, 106.6)	−52.31	0.79 (0.32, 9.75)	0.69
5 years	−43.23 (−65.82, 267.7)	−46.88	0.81 (0.32, 15.4)	0.71
10 years	−40.13 (−65.54, 374.2)	−44.41	0.82 (0.32, 18.7)	0.71

Table 4
Parameter estimates for the bivariate peaks-over-threshold analysis.

	Mean Sv		Aggregation Index		
	Scale	Shape	Scale	Shape	Alpha
MLE	2.71	−0.07	0.02	0.89	0.95
Standard error	0.34	0.08	0.05	0.20	0.03

4. Discussion

4.1. Extreme Value Analysis applications for biological monitoring

EVA can be used to establish or refine a monitoring program by providing a way to detect impacts, defined as relevant biological change caused by human activity, and to model extreme events, defined as observations above a GPD threshold. Observing values above a threshold or an increase in the frequency of extreme values compared to baseline measurements could be used to indicate that an impact has occurred. Observations above a threshold are statistically rare and occur when high-risk events are likely to transpire [16]. Defining a threshold for extreme values can aid MRE managers to assess the risk of potentially catastrophic events by establishing a baseline for expected extreme value periodicity.

High correlations between extreme events and biological or environmental covariates indicate conditions when high-risk events may occur, which could be used to increase real-time monitoring effort. For this approach to be successful, it is critical to define EVA metrics within their biological and physical contexts. For example, observations have shown that turbines act as fish aggregation devices during slack tides and that these aggregations disperse as current speed increases [40]. Extreme aggregation events that occur during slack tides when a turbine is not operating probably do not pose a high risk for negative interactions between marine life and tidal turbines. The aggregation index could be modified to account for tidal state and calculated during periods of high flow when the turbine is in operation.

EVA could be applied to other environmental data collected for MRE monitoring in addition to the examples used in this paper. EVA is not limited to high extreme values and could be used to model high or low extremes in physical data relevant to tidal turbine operation, such as current speed and turbulence.

Threshold estimates are not sample size dependent, but are sample dependent. POT analysis allowed estimation of GPD threshold values, but the 'true' threshold may differ from that determined by the POT threshold analysis, as sample data must be representative of the true distribution of outcomes. The absence (not occurring or not sampled) of large events during baseline sampling may result in a threshold value lower than if the full range of events were sampled. For example, it would be important to include observations of large schools of shoaling fish in the baseline dataset as their presence and abundance is highly variable [41]. Accuracy of the GPD threshold estimate is increased by sampling the full range of conditions at a MRE site.

There are no generic guidelines for biological monitoring at tidal energy sites in the US, and there are no guidelines for the temporal sampling resolution or contents of a baseline dataset [13]. Monitoring plans for pilot projects in the US have a wide range of monitoring methods, objectives, and are site-specific. The Admiralty Inlet baseline dataset was collected continuously for one month. In comparison the baseline data for the Cobscook Bay tidal energy project was collected over nine, 24 hour periods in a year [40]. If impacts are to be detected in a BACI analysis, a baseline dataset must be representative for comparison to data collected after any environmental alteration has occurred [42]. The Admiralty Inlet and the Cobscook Bay datasets contain different temporal resolutions and the threshold analysis for the Cobscook Bay dataset may reflect seasonal fluctuations in biomass rather than daily variability. It will be important to determine the appropriate temporal and/or spatial resolution of baseline sampling, both to establish an accurate GPD threshold and to detect biologically relevant change once project installation and operation begins.

When using POT, selecting a GPD threshold is a critical and challenging step in the EVA [30]. Besides examining diagnostic plots, other methods for selecting thresholds have been proposed, but many are computationally intensive or case-specific [18,29]. In this study, a derivative method is used to establish the threshold. This method utilizes traditional diagnostic plots

but with greater precision and objectivity compared to visual examination, and with low computation load. For both mean Sv and aggregation index metrics, values where derivatives equaled zero were similar for both the MRL and parameter stability plots. Consistency of derivative-based threshold estimates from both plots is interpreted that the derivative method accurately estimated the GPD threshold.

A bivariate analysis was incorporated in this study because a joint extreme aggregation and density event (e.g. a large and highly aggregated fish formation in close proximity to the turbine) potentially represents a greater risk for animal interaction with a tidal turbine device. While a joint return level prediction was determined, the results from this analysis also show that extreme values of the two metrics were independent, making simultaneous occurrence of a joint extreme density and aggregation event improbable for this dataset. Furthermore, the discrepancy between the proportion of data fit for the univariate model and the bivariate model was especially large for the aggregation index (7% of data in the univariate case, 31% in the bivariate case) leading to very different parameter estimates for the two models. The difference in threshold estimates between the univariate and bivariate cases as well as the near independence of the mean Sv and aggregation index metrics indicate that the univariate models should be used instead of the bivariate models for this dataset.

Multivariate extremes can be used to model highly correlated processes (e.g. wind, wave amplitude, and current fluxes [43]), and the potential spatial autocorrelation among variables [44]. A multivariate model may be informative if additional data confirms a correlation between biomass density and aggregation, or if correlations exist between extremes of other metrics. A multivariate model could also be used to examine spatial variability of metric values if multiple datasets are collected at a site.

The rapid expansion of credible intervals for return levels of both mean Sv and aggregation index metrics is attributed to sample size. Data were only collected for one month, which increases the uncertainty in return levels for larger return periods [16]. The spread of the 95% credible intervals around return levels shows that return level predictions are uncertain, even at small return periods. This wide range of return level credible intervals is partially due to the conservative uncertainty estimate from the Bayesian computation. Uncertainty could be decreased by collecting baseline data over a longer period, but this may be impractical as data collection surveys are expensive [45]. Baseline surveys could also be supplemented by data collected during project operations; with return levels estimated separately and used to inform project managers about conditions under which increased monitoring may result in extreme events. This approach is useful during ongoing monitoring programs, as more data will increase accuracy of GPD parameter estimates, which in turn, should decrease return level uncertainty.

Return levels must be interpreted with caution. The aggregation index return levels are bounded by the values of 0.14 m^{-1} and 1 m^{-1}, which is the largest possible aggregation index value. Most of the calculated upper 95% credible interval bounds exceed the upper bounds for the metric (Table 3). Exceeding the upper bound occurs because a positive shape parameter makes the GPD infinite. For metrics that are bounded, values that exceed bound limits could be replaced by the true metric bounds (e.g. 1 m^{-1} for aggregation index). For metrics that are unbounded, such as mean Sv, return levels could be constrained to biologically reasonable values. An alternate approach would be to transform values of the dependent variable, and then use the inverse of the transformation to obtain return level values.

Parameter uncertainty could potentially be reduced by performing an MCMC analysis using informative priors. Informative priors have been used in impact detection studies (e.g. [46–48]), and can be formed by soliciting expert opinion on the effects of and relationship to a disturbance. To obtain a prior for a GPD, experts are asked the median and 90% quantile estimates (e.g. high density values) for specific return levels (see [27] for more detail). These estimates can be used to construct priors for the scale and shape parameters. As more baseline and operational biological effect studies are conducted and environmental impacts are better understood, it will be possible to create informative priors. An informative prior is expected to reduce the range of GPD parameter estimates, which will also decrease the range of return level credible intervals. As a cautionary note, interpretation of MCMC analyses using informative priors should be conservative, as they may be biased in favor of expert belief [35]. Informative priors could also be used to predict impact severity under different impact expectations (e.g. tidal turbines will or will not impact fish behavior). Strategic Environmental Assessments have shown that stakeholder groups (e.g. developers, regulators, and fishers) vary greatly in how they perceive MRE development's effects on the environment [49,50]. If in doubt, then non-informative priors will provide the most conservative, even though they will result in the widest, credible intervals in return period prediction.

4.2. Statistical vs. biological significance

The goal of any monitoring is to detect whether a perturbation causes a significant change. Regulators typically set a threshold for the amount of change that is acceptable. Determining this threshold is crucial as regulators may use the amount of environmental change detected to evaluate the success of a project, trigger adaptive management, or terminate the project earlier than scheduled [13]. The debate over how to quantify a biologically significant effect is ongoing [51–53]. Using an extreme value approach, one can set a threshold for extreme events based on statistical significance. A biologically significant change does not necessarily correspond to statistical significance [12]. The choice of biological significance relies on expert, and potentially subjective, judgment [53]. Consensus among stakeholders may be difficult to reach [49] and these decisions should be made before monitoring begins [11], which, given the paucity of the data, is challenging in poorly studied ecosystems.

Statistical thresholds can be used as guidelines for regulators, and supplemented by return level plots to facilitate evaluations of the biological relevance of extreme events. One advantage of EVA in the debate on what constitutes a significant impact is that extreme values are rare, but very detectable. The detectability of extremes is important in variable and energetic environments. Detectability of extreme events coupled with the ability to set thresholds provides a starting point to define an impact, compared to the uncertainty when establishing a biologically-based threshold. Knowledge of a study site's biology helps determine whether statistical thresholds identified by EVA are biologically relevant.

Acknowledgments

We would like to thank Dale Jacques for data management and processing of the Admiralty Inlet data. Funding was provided by the National Oceanographic Partnership Program, the US Bureau of Ocean Energy Management, and the National Science Foundation's Sustainable Energy Pathways Program.

References

[1] R. Pelc, R.M. Fujita, Renewable energy from the ocean, Marine Policy 26 (2002) 471–479, http://dx.doi.org/10.1016/S0308-597X(02)00045-3.
[2] B. Polagye, B. Van Cleve, A. Copping, K. Kirkendall, Environmental Effects of Tidal Energy Development, in: NOAA Technical Memorandum NMFS F/SPO-116, 2011, p. 186.
[3] R. Inger, M.J. Attrill, S. Bearhop, A.C. Broderick, W.J. Grecian, D.J. Hodgson, et al, Marine renewable energy: potential benefits to biodiversity? An urgent call for research, J. Appl. Ecol. 46 (2009) 1145–1153, http://dx.doi.org/10.1111/j.1365-2664.2009.01697.x.
[4] C. Frid, E. Andonegi, J. Depestele, A. Judd, D. Rihan, S.I. Rogers, et al, The environmental interactions of tidal and wave energy generation devices, Environ. Impact Assess. Rev. 32 (2012) 133–139, http://dx.doi.org/10.1016/j.eiar.2011.06.002.
[5] A.I. Bijleveld, J.A. van Gils, J. van der Meer, A. Dekinga, C. Kraan, H.W. van der Veer, et al, Designing a benthic monitoring programme with multiple conflicting objectives, Methods Ecol. Evol. 3 (2012) 526–536, http://dx.doi.org/10.1111/j.2041-210X.2012.00192.x.
[6] A. Stewart-Oaten, Chapter 2 – Goals in Environmental Monitoring, in: R.J. Schmitt, C.W. Osenberg (Eds.), Detecting Ecological Impacts, Academic Press, San Diego, 1996, pp. 17–27. http://www.sciencedirect.com/science/article/pii/B9780126272550500045 (accessed July 23, 2014).
[7] A. Stewart-Oaten, J.R. Bence, Temporal and spatial variation in environmental impact assessment, Ecol. Monogr. 71 (2001) 305–339.
[8] G.W. Boehlert, A.B. Gill, Environmental and ecological effects of ocean renewable energy development: a current synthesis, 2010. http://ir.library.oregonstate.edu/jspui/handle/1957/16152 (accessed June 11, 2014).
[9] A.J. Underwood, On Beyond BACI: sampling designs that might reliably detect environmental disturbances, Ecol. Appl. 4 (1994) 4–15, http://dx.doi.org/10.2307/1942110.
[10] B.D. Mapstone, Scalable decision rules for environmental impact studies: effect size, Type I, and Type II errors, Ecol. Appl. 5 (1995) 401, http://dx.doi.org/10.2307/1942031.
[11] Federal Energy Regulatory Commission, Licensing Hydrokinetic Pilot Projects, 2008.
[12] A. Martínez-Abraín, Statistical significance and biological relevance: a call for a more cautious interpretation of results in ecology, Acta Oecol. 34 (2008) 9–11, http://dx.doi.org/10.1016/j.actao.2008.02.004.
[13] Federal Energy Regulatory Commission, Handbook for Hydroelectric Project Licensing and 5 MW Exemptions from Licensing, Federal Energy Regulatory Commission, 2004.
[14] Ocean Renewable Power Company, Comments on the Draft Pilot Project License Application Cobscook Bay Tidal Energy Project, 2011.
[15] C.W. Osenberg, R.J. Schmitt, Chapter 1 – Detecting Ecological Impacts Caused by Human Activities, in: R.J. Schmitt, C.W. Osenberg (Eds.), Detecting Ecological Impacts, Academic Press, San Diego, 1996, pp. 3–16. http://www.sciencedirect.com/science/article/pii/B9780126272550500033 (accessed July 23, 2014).
[16] S. Coles, An Introduction to Statistical Modeling of Extreme Values, Springer, London, 2001.
[17] A. Agarwal, V. Venugopal, G.P. Harrison, The assessment of extreme wave analysis methods applied to potential marine energy sites using numerical model data, Renew. Sustain. Energy Rev. 27 (2013) 244–257, http://dx.doi.org/10.1016/j.rser.2013.06.049.
[18] F. Mazas, L. Hamm, A multi-distribution approach to POT methods for determining extreme wave heights, Coast. Eng. 58 (2011) 385–394, http://dx.doi.org/10.1016/j.coastaleng.2010.12.003.
[19] Public Utility District No. 1 of Snohomish County, Admiralty Inlet Tidal Project Final Monitoring and Mitigation Plans, 2012.
[20] J.K. Horne, D.A. Jacques, S.L. Parker-Stetter, H.L. Linder, J.M. Nomura, Evaluating acoustic technologies to monitor aquatic organisms at renewable energy sites, U.S. Department of the Interior, Bureau of Ocean Energy Management, 2013.
[21] D. Jacques, Describing and Comparing Variability of Fish and Macrozooplankton Density at Marine Hydrokinetic Energy Sites, University of Washington, 2014.
[22] S.S. Urmy, J.K. Horne, D.H. Barbee, Measuring the vertical distributional variability of pelagic fauna in Monterey Bay, ICES J. Mar. Sci. 69 (2012) 184–196, http://dx.doi.org/10.1093/icesjms/fsr205.
[23] J.M. Burgos, J.K. Horne, Sensitivity analysis and parameter selection for detecting aggregations in acoustic data, ICES J. Mar. Sci. 64 (2007) 160–168, http://dx.doi.org/10.1093/icesjms/fsl007.
[24] D.N. Maclennan, P.G. Fernandes, J. Dalen, A consistent approach to definitions and symbols in fisheries acoustics, ICES J. Mar. Sci. 59 (2002) 365–369, http://dx.doi.org/10.1006/jmsc.2001.1158.
[25] J. Pickands, Statistical inference using extreme order statistics, Ann. Statist. 119–131 (1975).
[26] J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, Statistics of Extremes, John Wiley & Sons, Chichester, England, 2004.
[27] C.N. Behrens, H.F. Lopes, D. Gamerman, Bayesian analysis of extreme events with threshold estimation, Statist. Model. 4 (2004) 227–244, http://dx.doi.org/10.1191/1471082X04st075oa.
[28] P. Jonathan, K. Ewans, Statistical modelling of extreme ocean environments for marine design: a review, Ocean Eng. 62 (2013) 91–109, http://dx.doi.org/10.1016/j.oceaneng.2013.01.004.
[29] C. Scarrott, A. MacDonald, A review of extreme value threshold estimation and uncertainty quantification, REVSTAT-Statist. J. 10 (2012) 33–60.
[30] D.J. Dupuis, Exceedances over high thresholds: a guide to threshold selection, Extremes 1 (1999) 251–261.
[31] P. Thompson, Y. Cai, D. Reeve, J. Stander, Automated threshold selection methods for extreme wave analysis, Coast. Eng. 56 (2009) 1013–1021, http://dx.doi.org/10.1016/j.coastaleng.2009.06.003.
[32] M.P. Wand, M.C. Jones, Kernel Smoothing, CRC Press, 1994.
[33] M.P. Wand, KernSmooth: Functions for kernel smoothing for Wand & Jones (1995), 2012. http://CRAN.R-project.org/package=KernSmooth.
[34] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109, http://dx.doi.org/10.1093/biomet/57.1.97.
[35] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, third ed., CRC Press, 2013.
[36] J. Geweke, Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments, Bayesian Statistics, University Press, 1992, pp. 169–193.

[37] A. Gelman, D.B. Rubin, Inference from Iterative Simulation Using Multiple Sequences, Statist. Sci. 7 (1992) 457–472.
[38] J. Segers, B. Vandewalle, Statistics of Multivariate Extremes, Statistics of Extremes: Theory and Applications, Wiley, UK, 2004.
[39] A.G. Stephenson, Evd: Extreme Value Distributions, R News 2 (2002) 31–32.
[40] H. Viehman, Fish in a Tidal Dynamic Region in Maine: Hydroacoustic Assessments in Relation to Tidal Power Development, University of Maine, 2012.
[41] M.C. Siple, T.B. Francis, Population diversity in Pacific herring of the Puget Sound, USA, Oecologia (2015) 1–15, http://dx.doi.org/10.1007/s00442-015-3439-7.
[42] R.J. Schmitt, C.W. Osenberg, Detecting Ecological Impacts: Concepts and Applications in Coastal Habitats, Academic Press, 1996.
[43] R. Nerzic, C. Frelin, M. Prevosto, V. Quiniou-Ramus, Joint Distributions of Wind/Waves/Current In West Africa And Derivation of Multivariate Extreme I-FORM Contours, International Society of Offshore and Polar Engineers, 2007. https://www.onepetro.org/conference-paper/ISOPE-I-07-298 (accessed June 22, 2015).
[44] P.J. Northrop, P. Jonathan, Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights, Environmetrics 22 (2011) 799–809, http://dx.doi.org/10.1002/env.1106.
[45] Verdant Power, Final Kinetic Hydropower Pilot License Application, 4, 2010.
[46] F.H.J. Crome, M.R. Thomas, L.A. Moore, A novel bayesian approach to assessing impacts of rain forest logging, Ecol. Appl. 6 (1996) 1104–1123, http://dx.doi.org/10.2307/2269595.
[47] P.H. Garthwaite, A. O'Hagan, Quantifying expert opinion in the UK water industry: an experimental study, J. Royal Statist. Soc.: Series D (The Statistician) 49 (2000) 455–477, http://dx.doi.org/10.1111/1467-9884.00246.
[48] T.G. Martin, P.M. Kuhnert, K. Mengersen, H.P. Possingham, The power of expert opinion in ecological models using bayesian methods: impact of grazing on birds, Ecol. Appl. 15 (2005) 266–280, http://dx.doi.org/10.1890/03-5400.
[49] T. Gray, C. Haggett, D. Bell, Offshore wind farms and commercial fisheries in the UK: a study in Stakeholder Consultation, Ethics, Place Environ.: A J. Philos. Geograp. 8 (2005) 127–140.
[50] M. Doelle, Role of Strategic environmental assessments in energy governance: a case study of tidal energy in Nova Scotia's Bay of Fundy, J. Energy Nat. Resour. L. 27 (2009) 112.
[51] J.D. Germano, Ecology, statistics, and the art of misdiagnosis: the need for a paradigm shift, Environ. Rev. 7 (2000) 167–190, http://dx.doi.org/10.1139/a99-014.
[52] N.T. Hobbs, R. Hilborn, Alternatives to statistical hypothesis testing in ecology: a guide to self teaching, Ecol. Appl. 16 (2006) 5–19, http://dx.doi.org/10.1890/04-0645.
[53] D.P. Lovell, Biological importance and statistical significance, J. Agric. Food Chem. 61 (2013) 8340–8348, http://dx.doi.org/10.1021/jf401124y.