

EVALUATING BIOLOGICAL CHARACTERISTICS OF MARINE  
RENEWABLE ENERGY SITES FOR ENVIRONMENTAL MONITORING

Lauren E. Wiesebron

A thesis  
submitted in partial fulfillment of the  
requirements for the degree of

Master of Science

University of Washington

2015

Committee:

John K. Horne

James J. Anderson

Noble Hendrix

Program Authorized to Offer Degree:  
Aquatic and Fishery Sciences

© Copyright 2015  
Lauren E. Wiesebron  
ALL RIGHTS RESERVED

University of Washington

## **Abstract**

# **EVALUATING BIOLOGICAL CHARACTERISTICS OF MARINE RENEWABLE ENERGY SITES FOR ENVIRONMENTAL MONITORING**

Lauren Wiesebron

Chair of the Supervisory Committee:  
John K. Horne  
School of Aquatic and Fishery Sciences

Tidal energy is a renewable resource that helps meet growing energy demands, but uncertainties remain about environmental impacts of device installation and operation. Monitoring programs are used to detect impacts caused by anthropogenic disturbances 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 scarce, limiting the 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 estimation precision. 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. To characterize the generality of tidal energy sites, metrics describing temporal and spatial distributions of fish and macrozooplankton at the Admiralty Inlet site and a second tidal energy site from the Fall of Warness, Scotland were compared using statistical methods (t-test, F-test, linear regression), spectral analysis, and EVA. General biological characteristics were similar enough that generic biological monitoring programs could be implemented at these two sites, which would streamline the permitting process as well as facilitate site comparison and detection of environmental impact due to tidal technology deployment.

# Table of Contents

List of Figures	ii
List of Tables	iii
Acknowledgements	iv
1 Introduction	1
1.1 Impact detection during biological monitoring programs	1
1.2 Tidal energy development	2
1.3 Objectives	3
2 Characterizing biological impacts at marine renewable energy sites	5
2.1 Introduction	5
2.2 Methods	6
2.3 Results	12
2.4 Discussion	22
3 Comparison of biological characteristics in distribution of fish and macrozooplankton at two tidal energy sites	26
3.1 Introduction	26
3.2 Methods	27
3.3 Results	32
3.4 Discussion	44
4 Summary and Significance	49
References	52

# List of Figures

1.1 Impact categories	2
2.1 Admiralty Inlet map	6
2.2 Mean Sv threshold diagnostic	12
2.3 Mean Sv GPD threshold and fit	13
2.4 Aggregation threshold diagnostic	14
2.5 Aggregation GPD threshold and fit	15
2.6 Sample size effect on threshold estimate	16
2.7 Posterior densities for scale and shape parameter	17
2.8 Six mean Sv GPD parameter posteriors	18
2.9 Scale and shape estimates as a function of the threshold	19
2.10 Mean Sv and aggregation index return level plots	20
2.11 Univariate and bivariate return levels	22
3.1 Admiralty Inlet and Fall of Warness study sites	28
3.2 Boxplots for ten mean Sv series	33
3.3 Distribution of threshold for series containing increasing data proportions.	34
3.4 The time series for four metrics	36
3.5 The metrics' variability by hour	37
3.6 The metrics as a function of tidal speed	38
3.7 Periodograms for the metrics with significant frequencies	39
3.8 GPD fit and return level plots for mean Sv	42
3.9 GPD fit and return level plots for aggregation index	43

## List of Tables

2.1 MCMC parameter values for mean Sv and aggregation simulations	10
2.2 Median GPD parameter values and the 95% quantiles.	17
2.3 Return levels for univariate GPD fit and bivariate GPD	21
2.4 Parameter estimates for the bivariate peaks-over-threshold analysis	21
3.1 Acoustic sampling parameters	29
3.2 Means and standard deviations of biological characteristics	35
3.3 Significant periods rounded to the nearest hour	40
3.4 The covariates and corresponding p-values for metric linear regressions	41
3.5 Summary of GPD fit for mean Sv and aggregation index	43
3.6 Comparison of Admiralty Inlet and Fall of Warness ecosystem attributes	44

## Acknowledgements

I would like to express my sincere thanks

To my Committee John Horne, Noble Hendrix, Jim Anderson for all their support and insight; especially my advisor, John Horne, for his vision, guidance, and enthusiasm,

To the National Science Foundation Sustainable Energy Pathways for providing the funding for this research,

To Dale Jacques, whose thesis was a stepping stone for this one,

To members of the Fisheries Acoustics Lab, especially Emily Runnells, Dale Jacques, Hannah Linder, Beth Phillips, Dave McGowan, and Mei Sato for their unflagging support and cheer,

To Benjamin Williamson for generously sharing the FLOWBEC data and his expertise,

To Colin Lamont, Dale Jacques, and Hannah Linder for helping me wallow through many quantitative quagmires,

To Steve Scherba and Megan Dethier for showing me how to teach, and to Emily Runnells for convincing me that I could,

To my friends in Seattle and abroad who have kept me happy, sane, and snappy these past two years,

And, most of all, to my family, for their love and constant encouragement. Thank you.

## Chapter 1:

# *INTRODUCTION*

### **1.1 Impact detection during biological monitoring programs**

While scientists and managers have sought to understand the consequences of natural resource use and landscape modification, measuring environmental change as a result of human activity remains a challenge because of uncertainties in how to detect change (Schmitt and Osenberg, 1996). Environmental regulations are implemented to mitigate and understand the consequences of human activity, and the monitoring program is a common approach used to address environmental concerns. Biological monitoring programs focus on the detection of change in biological variables such as diversity, size, or abundance of monitored species (Bijleveld et al., 2012). A successful biological monitoring program provides data that will help regulators make informed decisions on adaptive management options (Schmitt and Osenberg, 1996). Despite the wide use of monitoring programs, optimal design and interpretation of monitoring data are widely debated. As a result, monitoring methods and sampling can vary among similar sites.

Part of the problem in designing and interpreting data from environmental monitoring programs, is that there is a lack of consensus on how to define biological change. Both “effect” and “impact” have been used to designate significant change in environmental monitoring programs. Stewart-Oaten and Bence (2001) 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” (Boehlert and Gill, 2010). Some authors do not make any distinction between effect and impact, use the terms interchangeably (e.g. Osenberg et al., 1994; Underwood, 1996; Hewitt et al., 2001), or use “effect” to designate the consequence of a disturbance and “impact” as the disturbance itself (e.g. Underwood, 1994). Impacts or effects can also be described as “weak” or “strong” (Mapstone, 1995; Hewitt et al., 2001). For the purpose of this study, we only use “impact” to designate relevant biological change as a consequence of human activity. We define impact as:

Impact = frequency of occurrence x magnitude x duration

“Frequency of occurrence” refers to the number of times that an outcome is observed during the period of interest. The “magnitude” of an observation is measured by quantifying characteristics of the biological response to a perturbation, (e.g. mortality or change in behavior), or by quantifying characteristics of the disturbance, (e.g. increases in noise or pollutant concentration) (Polagye et al., 2011). “Duration” is the period of the impact. Increases in any of these three components will increase the severity of an impact. This definition can be used to encompass a wide range of impacts that regulators may be interested in monitoring, and can be used to categorize impacts based on their attributes (Fig. 1.1). Low or High magnitudes of each impact component (i.e. frequency, magnitude, and duration) can be used to characterize overall impacts of anthropological perturbations on an ecosystem.



Fig 1.1: Impact categories as a combination of Low/High quantities of Frequency of Occurrence, Magnitude, and Duration. The color code indicates severity, from green to dark red. The first letter in the two letter pairs corresponds to Frequency, the second to Magnitude.

Deciding on the maximum level of “acceptable” impact is a high priority when forming a monitoring plan (Mapstone, 1995), and thresholds are often used when defining regulations. Definitions of ecological thresholds are diverse, and ecologists do not often agree on threshold values. Examples of variables with detrimental outcomes once thresholds are exceeded include the abundance of a population at collapse (Wilkie and Carpenter, 1999), a large change in percent coverage of benthic macro-fauna (Lundquist et al., 2010), or the temperature at which corals bleach (Jones, 2001). These thresholds are context specific and usually require prior knowledge about relevant and important environmental variables, which makes determining the value of a threshold difficult (Martínez-Abraín, 2008; Duinker et al., 2013), especially for data-poor sectors. In this case, impact thresholds can be modeled or estimated from baseline datasets.

## 1.2 Tidal energy development

Tidal energy is an attractive source of renewable energy because tides are consistent and predictable, resulting in energy that can be extracted at a constant and dependable rate (Polagye et al., 2011). However, extracting energy from tides does not come without

challenges. Because of high currents that pervade areas where tidal energy extraction is optimal, there is a high risk of damaging devices and increased difficulty in the deployment, retrieval, and maintenance of devices (Cada et al., 2007; Copping et al., 2013).

Currently, there are no requirements from the Federal Energy Regulatory Commission (FERC) detailing priorities for environmental monitoring at tidal energy sites (Copping and Geerlofs, 2011). Yet the development and implementation of environmental monitoring is part of every project license. All US prospective developers of tidal energy sites have to obtain a pilot project license from FERC. The FERC application must contain a monitoring plan that includes strategies to detect environmental effects of the project (Commission, 2008). When the license is issued, the monitoring plan that was included in the application becomes part of the license, and the developer is contractually obligated to meet the specified monitoring requirements.

Determining monitoring objectives and methods is protracted and inefficient because the developer needs to iteratively revise the monitoring program until regulators and stakeholders are satisfied. As a result the monitoring objectives for these plans are driven by three main factors: legislation like the Endangered Species Act or Marine Mammal Protection Act, site characteristics, and budget constraints. Main focuses of biological monitoring are: interaction between marine fauna and devices, noise impacts of devices on marine fauna, and physical effects of energy removal (Copping et al., 2014). Perceived importance of biological concerns varies among sites. The consequence of lack of standardization is site-specific objectives, sampling, and analysis.

Designing biological monitoring programs and defining impact thresholds for new technology are particularly challenging, as impact characterizations during monitoring plan development are often derived from preliminary samples or prior observations on similar disturbances (Schmitt and Osenberg, 1996). Tidal technology is a way to mitigate fossil fuel use, but studies describing environmental change due to tidal devices are scarce, which restricts information available to develop monitoring programs (Polagye et al., 2011). Therefore, to characterize impacts for biological monitoring programs at tidal energy sites, regulators must estimate impact thresholds or use models to characterize change.

### **1.3 Objectives**

This study seeks to address some of the difficulties in developing environmental monitoring plans, particularly for data-poor sectors like tidal energy development. By characterizing and comparing biological attributes of tidal energy sites, a method is

developed for impact characterization using baseline data (i.e. data collected before the alteration began (Stewart-Oaten and Bence, 2001) ), and provides evidence for generic monitoring methods at tidal energy sites. Specific objectives include:

1. Adapt EVA to establish thresholds for monitoring metrics using data from Admiralty Inlet.
2. Use Bayesian methods to predict extreme value return levels and their uncertainty.
3. Compare the biological characteristics of the Admiralty Inlet and Fall of Warness sites to determine whether they are similar enough to justify generic monitoring programs.
4. Evaluate the generality of using EVA for biological monitoring at tidal energy sites.

## Chapter 2:

# *CHARACTERIZING BIOLOGICAL IMPACTS AT MARINE RENEWABLE ENERGY SITES*

## 2.1 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 (Pelc and Fujita, 2002). Interest in developing tidal energy projects is growing because tides are a constant and predictable energy source (Polagye et al., 2011). 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 (Pelc and Fujita, 2002; Inger et al., 2009; Frid et al., 2012).

At this time, there are no regulations for MRE monitoring procedures, technologies, or metrics for monitoring programs (FERC, 2004). Monitoring programs are developed prior to the application for an operating license, and the time to develop a monitoring plan that is acceptable to regulators can delay the developer's submission of the license application, adding cost and temporal uncertainty to permitting (e.g. Ocean Renewable Power Company, 2011).

To detect an impact, baseline data (i.e. data collected before alteration began (Stewart-Oaten and Bence, 2001) ) must be collected to facilitate a comparison after devices are installed and become operational (Underwood, 1994). Determining the maximum level of "acceptable" impact is a high priority when forming a monitoring plan (Mapstone, 1995). Impact above a threshold can determine if a tidal project is allowed to continue operating (FERC, 2008). Thus it is imperative that setting thresholds and characterizing impacts should be completed before MRE operations and concurrent monitoring begins (Martínez-Abraín, 2008).

Extreme value analysis (EVA) is an approach used to model values that are infrequent but are potentially associated with impacts caused by large change (Coles, 2001). Used in engineering and hydrology (e.g. Mazas and Hamm, 2011; Agarwal et al., 2013), EVA can be applied to environmental monitoring to target rare but potentially significant impact events. These events are expected to be important to MRE regulators as these may

have 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. The overall goal of this chapter is to evaluate whether extreme value analysis can be used to characterize infrequent values that are potentially associated with biological impacts at a tidal energy site.

## 2.2 Methods

### 2.2.1 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 20<sup>th</sup>, 2014. 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 operating at 120 kHz from May 9th until June 9th, 2011 (Horne et al., 2013). The echosounder was placed at 55m depth about 750m off Admiralty Head at the SnoPUD tidal turbine site (Fig. 2.1). The echosounder sampled at 5Hz for 12 minutes every 2 hours. Because of a 3<sup>rd</sup> surface echo, data values were constrained to 25 m from the bottom. A -75 dB re 1m<sup>-1</sup> threshold was applied to remove noise (Horne et al., 2013). Data were binned into 12 minute samples, and vertically integrated, yielding 361 datapoints (Jacques, 2014).



Fig. 2.1. Study location within Puget Sound, Washington (upper right), and location of the acoustic package (left) within the SnoPUD proposed turbine location. The blue dot indicates the echosounder location.

### 2.2.2 Echometrics

Echometrics are a suite of indices that quantify variability of vertical biomass in the water column over space and through time (Urmy et al., 2012). Among the suite of seven Echometrics developed by (Burgos and Horne, 2007) and refined by Urmy et al. (2012), density and aggregation indices are used to reflect horizontal or vertical changes in biomass distribution, which can be used 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 collision with MRE devices. The density metric is the mean volume-backscattering strength, or mean Sv (unit: dB re 1 m<sup>-1</sup>(Maclennan et al., 2002) hereafter dB), which is proportional to biomass density. The aggregation index was used to quantify vertical patchiness with values from 0 to 1, with 0 being evenly dispersed and 1 being aggregated.

### 2.2.3 Extreme Value Analysis

Extreme value analysis (Pickands, 1975; Coles, 2001; Beirlant et al., 2004) 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 (Coles, 2001), making this analysis unusual in that it focuses on the tails and not the mean of a sample distribution.

#### 2.2.3.1 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, 1975):

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

where  $u$  is the threshold,  $\sigma$  is the scale parameter,  $\varepsilon$  is the shape parameter, and  $\sigma > 0$ . The shape parameter determines whether the GPD is bounded. The sign of the shape parameter determines the behavior of the GPD (Coles, 2001). 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 ( $\sigma$ ) and shape ( $\varepsilon$ ) parameters are fitted to the data to model extreme values.

### 2.2.3.2 Identifying the extreme value threshold (first step of POT)

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 fitted to too few data points and the variance of the GPD parameter estimates will increase (Behrens et al., 2004; Jonathan and Ewans, 2013). The ideal threshold is the lowest value that includes as many excesses as possible while still achieving model fit.

The threshold is usually defined visually (Scarrott and MacDonald, 2012) 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 increases. If a GPD is valid for excesses at a threshold  $u_0$ , it should also be valid for the thresholds  $u > u_0$ , with the scale parameter adjusted to the threshold  $u$  (Coles, 2001). So  $E(X - u | X > u)$  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 curve becomes linear. 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  $\varepsilon$  and adjusted scale parameter  $\sigma^*$ , with  $\sigma^* = \sigma_u - \varepsilon u$ , should be constant above  $u_0$ , if  $u_0$  is a valid threshold for the GPD (Coles, 2001). On a parameter stability plot, the threshold is identified as the value where parameter estimates become stable, or near-constant. The adjusted scale parameter plots and the 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 (Dupuis, 1999; Thompson et al., 2009).

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 were smoothed to remove local variability using a polynomial kernel density smoother (Wand and Jones, 1994), implemented using the KernSmooth package in R. Derivatives were calculated for the smoothed functions and the inflexion point corresponding to the best threshold estimate from each plot was 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 the Admiralty Inlet data, for both mean Sv and aggregation index metrics.

### 2.2.3.3 Fitting the GPD using Bayesian methods

#### 2.2.3.3.1 Applying Bayesian theory to POT

While the most widespread method for fitting model parameters is maximum likelihood estimation (MLE), we performed a peaks-over-threshold analysis using Bayesian inference.

Bayes theorem:

$$P(H_i|data) = \frac{L(H_i|data) \times Prior(H_i)}{P(data)} \quad (2)$$

Bayes theorem allows the generation of a posterior distribution,  $P(H_i|data)$ , which is a probability distribution for an unknown variable.  $L(H_i|data)$  is the likelihood of the variable taking on a value given the data, with  $H_i$  as the hypothetical value of the variable (in this case, the value of the scale or shape parameter), and  $P(data)$  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 estimate of parameter values and distribution since it represents an integration over all probable values instead of selecting a maxima. A conservative estimate of parameter uncertainty is appropriate for MRE monitoring to ensure that the range of outcomes is not underestimated. As posteriors are analytically challenging to compute, these are typically simulated numerically. To obtain the posterior distribution for the two GPD parameters, scale  $\sigma$ , and shape  $\varepsilon$ , a Markov chain Monte Carlo (MCMC) simulation (Hastings, 1970) was used.

A Bayesian analysis includes the utilization of a prior, which is information on the probability of the parameters, formed without knowledge of or previous experience with the sample data. The use of prior information needs to be justified as it influences the distribution of the posterior. 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. Flat priors for the scale and shape parameters were used. These priors were centered at 0 and have high variance which ensures that no bias is introduced to the posterior parameter distributions:  $\sigma \sim N(0,1000)$ ;  $\varepsilon \sim N(0,100)$ .

### 2.2.3.3.2 MCMC application

The MCMC method used for this study was a Gibbs sampler. The negative log-likelihood (NLL) for the GPD is given by Coles (2001):

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

where  $u$  is the threshold,  $\sigma$  is the scale,  $\varepsilon$  is the shape,  $x$  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 (Gelman et al., 2013). Three tests (starting the chain from different initial values, the Geweke test (Geweke, 1992), and Gelman-Rubin test (Gelman and Rubin, 1992)) were performed 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 method attributes (jump size, chain length, thinning interval) (Table 2.1) were iteratively selected to obtain a well-mixed chain (30%-40% draws accepted) (Gelman et al., 2013).

Table 2.1: 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	200 000	25
Aggregation Index	0.18	0.46	36	21	400 000	50

### 2.2.3.3.3 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 estimates. Simulated data were generated following GPD distributions with known threshold, 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 0.01 units. This procedure was repeated on different combinations of the median, 0.025 and 0.975 quantiles for the scale and shape estimates for density (Mean Sv) and aggregation (Aggregation Index) using the Admiralty Inlet data.

#### 2.2.3.4 Return Level

While it is informative to examine values of location, shape, and scale parameters of the fitted GPD, further inference can be gained from examining return levels.  $q_p$  is the return level associated with the return period  $1/p$ , and  $q_p$  is the value that is expected to be exceeded on average once every  $1/p$  time units (Behrens et al., 2004). Return levels for data exceedances are generated by inverting the GPD cumulative density function (Equation 1). For an arbitrary probability  $p$ , the corresponding return level  $q_p$  is (Behrens et al., 2004):

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

where  $u$  is the threshold,  $\sigma$  is the scale,  $\varepsilon$  is the shape,  $q_p$  is the return level, and  $p$  is the return level probability.

By plotting the return level  $q_p$  against the return period  $1/p$ , one obtains a return level plot that shows the expected periodicity for data excesses and values extrapolated beyond the range of the sampled data.

#### 2.2.3.5 Bivariate peaks-over-threshold analysis

The bivariate POT method fits the GPD to two variables as a joint process. This method can be used as a supplement to the univariate analysis as it 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 (Coles, 2001). There are several methods to obtain a bivariate model, including a logistic model:

$$G(x, y) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^\alpha, 0 < \alpha < 1 \quad (4)$$

where  $x$  and  $y$  are the fitted univariate GPDs for the  $x$  and  $y$  variables (Coles, 2001).

Complete independence of the two variables is obtained when  $\alpha = 1$ , and inversely, dependence is obtained when  $\alpha$  approaches 0. Results of the bivariate analysis are GPD estimates for  $x$ ,  $y$ , and an alpha calculation that follows a bivariate model. In this study, 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 (A.G. Stephenson, 2002). The metric thresholds obtained from the univariate analysis were used to define the bivariate distribution, which is an accepted method of determining thresholds for the bivariate analysis (Coles, 2001).

## 2.3 Results

### 2.3.1 Threshold estimation

#### 2.3.1.1 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. 2.2a). 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. 2.2b) mimic those of the mean residual life plot (Fig. 2.2a). 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.

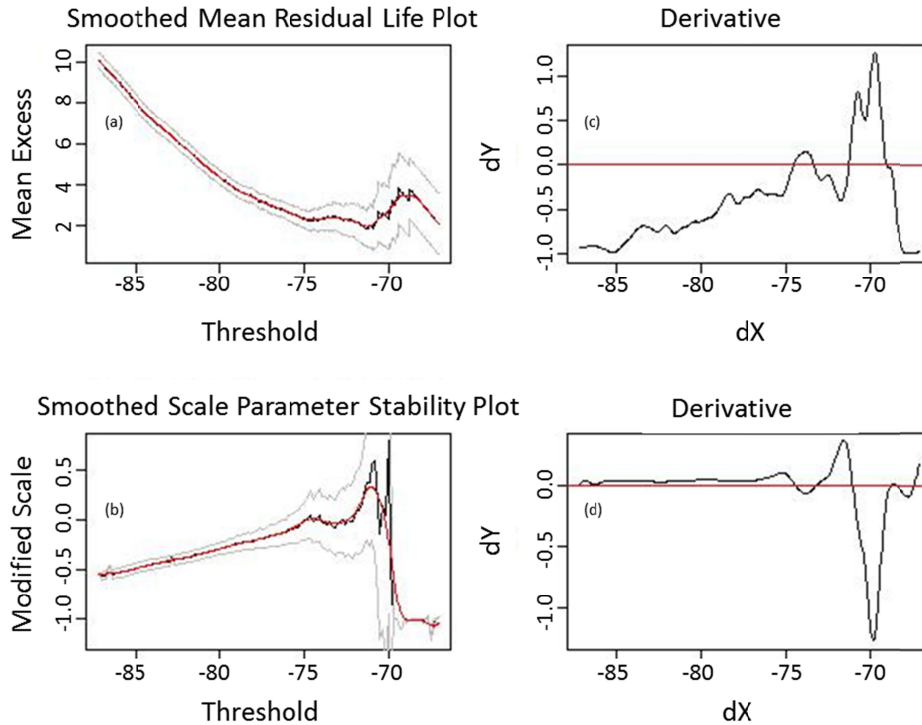


Fig. 2.2: The mean residual life plot (a) and scale parameter stability plot (b) with the red line representing the smoothed plots, the corresponding derivatives (c,d) of the smoothed plots, with the red line showing  $dY=0$  for Mean Sv.

To obtain a more precise estimate of the threshold, the derivative method was applied to mean Sv data (Fig. 2.2). After smoothing both the MRL (Fig. 2.2a) and parameter

stability plots (Fig. 2.2b), and taking the derivative (Fig 2.4c, 2.4d), the first point where  $dY = 0$  for the parameter stability plot is  $u = -74.48$  dB, and for the mean residual life plot it is at  $u = -74.58$  dB. These values are very close and are consistent with the visual diagnostics (Fig. 2.2 a, b). We set the threshold for the POT analysis 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. 2.3).

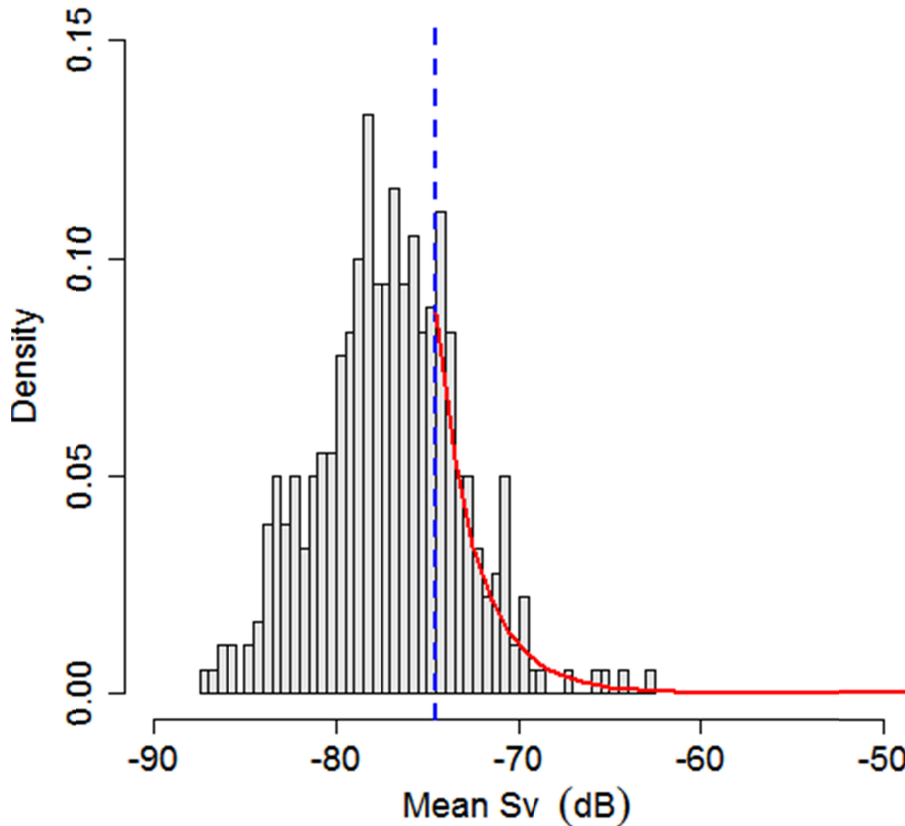


Fig. 2.3: Histogram of mean Sv values with the threshold marked at -74.53 dB (blue line) and GPD fit (red line). The scale and shape parameter values were obtained from posterior medians.

### 2.3.1.2 Aggregation index threshold diagnostics

Diagnostic plots for the aggregation index metric (Fig. 2.4) differed from the mean Sv threshold diagnostic graphs. The mean residual life plot (Fig. 2.4a) increased rapidly until about  $u \approx 0.14 \text{ m}^{-1}$ , where variance increased and the plot became approximately linear. The linear trend is more visible in the scale parameter stability plot (Fig. 2.4b), where the shape parameter estimate decreased from  $u \approx 0.05 \text{ m}^{-1}$  to  $u \approx 0.15 \text{ m}^{-1}$ , then it remained constant until  $u \approx 0.2 \text{ m}^{-1}$  where the variance increased steadily.

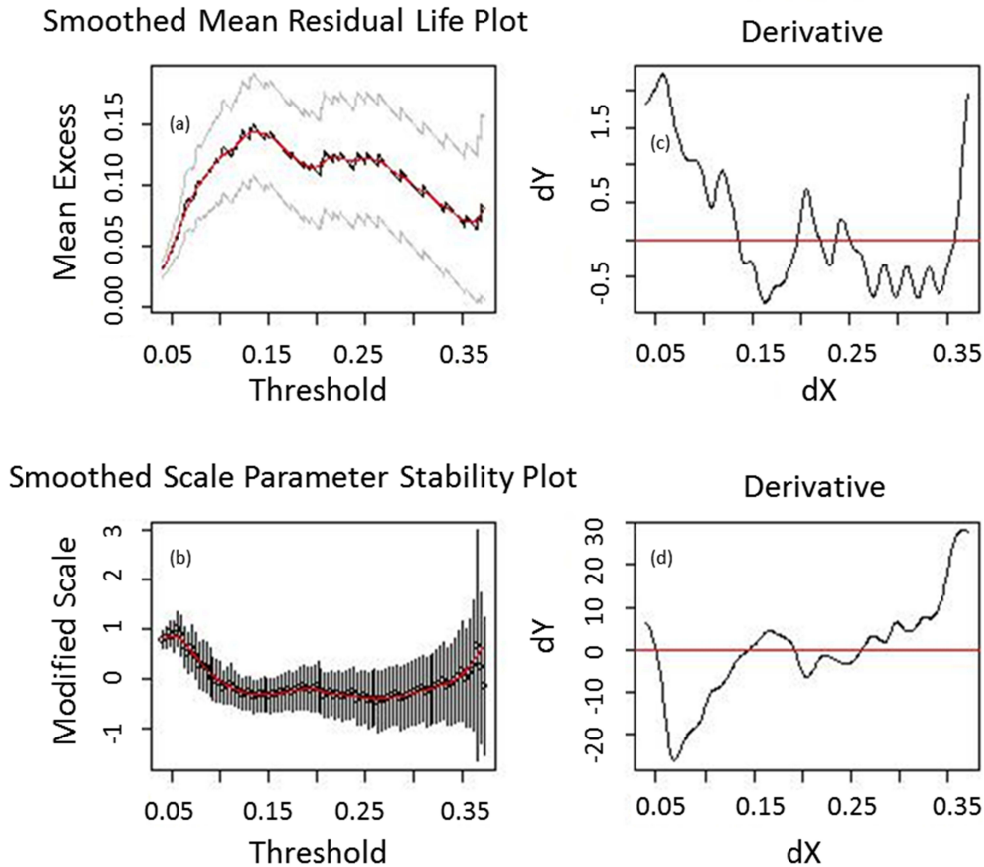


Fig. 2.4: Graphical and derivative aggregation index threshold diagnostic plots. The mean residual life plot (a) and scale parameter stability plot (b) with the red line representing the smoothed plots, the corresponding derivatives (c, d) of the smoothed plots, with the red line showing  $dY=0$

A threshold estimate was obtained using the derivative method for each plot. The threshold value from the mean residual life plot derivative was  $u = 0.135 \text{ m}^{-1}$ . The first value where  $dY=0$  for the derivative of the parameter stability plot was  $u = 0.05 \text{ m}^{-1}$ , but did not correspond to stabilization 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  $u = 0.146 \text{ m}^{-1}$ , the second inflexion point in the derivative plot, as it also matched the result from the MRL plot (Fig. 2.4c). The average of the threshold estimates from the MRL and parameter stability plots is  $0.140 \text{ m}^{-1}$ . As with mean Sv, this value is consistent with visual interpretation of the mean residual life and parameter stability plots. A threshold of  $0.140 \text{ m}^{-1}$  results in 26 exceedances for the aggregation index, which is 7% of the data (Fig. 2.5).

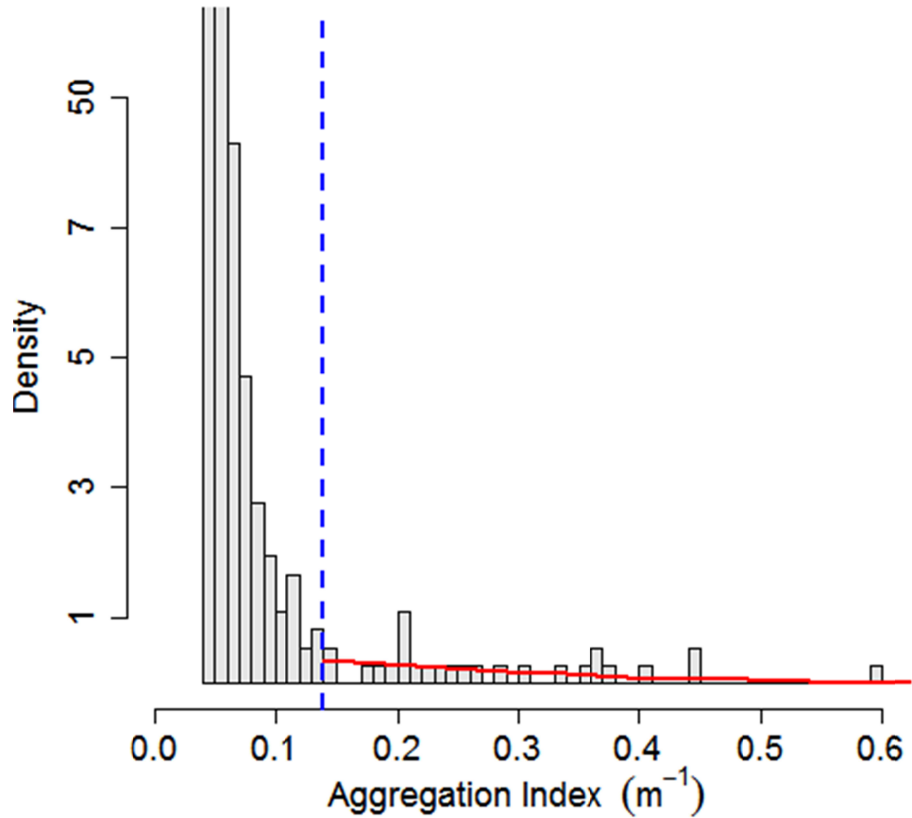


Fig. 2.5: Histogram of aggregation index values with the threshold marked at  $0.140 \text{ m}^{-1}$  (blue line) and GPD fit (red line). The scale and shape parameter values were obtained from posterior medians.

### 2.3.1.3 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. 2.6). 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 datapoints are considered extreme compared to 25% of the mean Sv data. In both cases, the estimated threshold is considered robust as the threshold estimate does not immediately change with decreasing sample size.

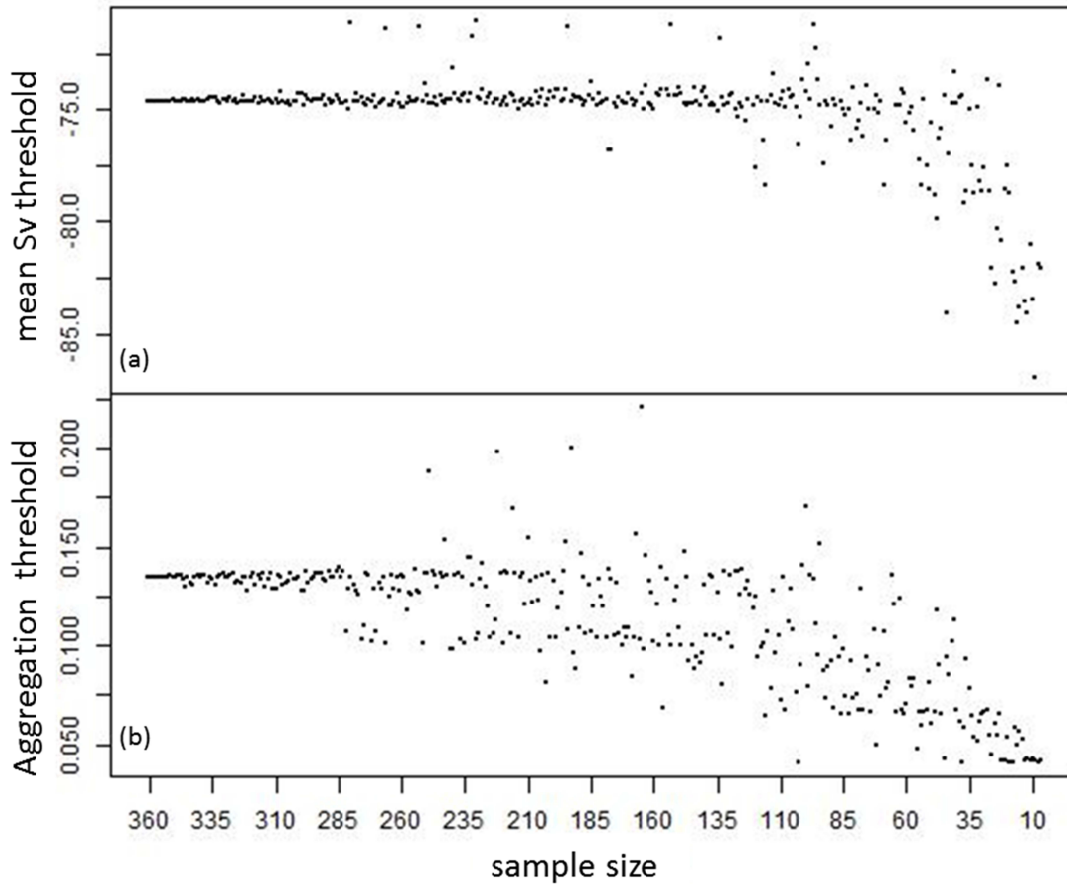


Fig. 2.6: Threshold estimates from the derivative method plotted against data sample size for (a) mean Sv (b) aggregation index.

## 2.3.2 Bayesian analysis

### 2.3.2.1 Shape and scale posteriors

The first objective of the Bayesian analysis was to generate 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. 2.7). The range (-1 to 1) of the shape parameter for both variables contained positive and negative values. Given that the shape parameter controls the shape of the GPD, this will affect return level predictions.

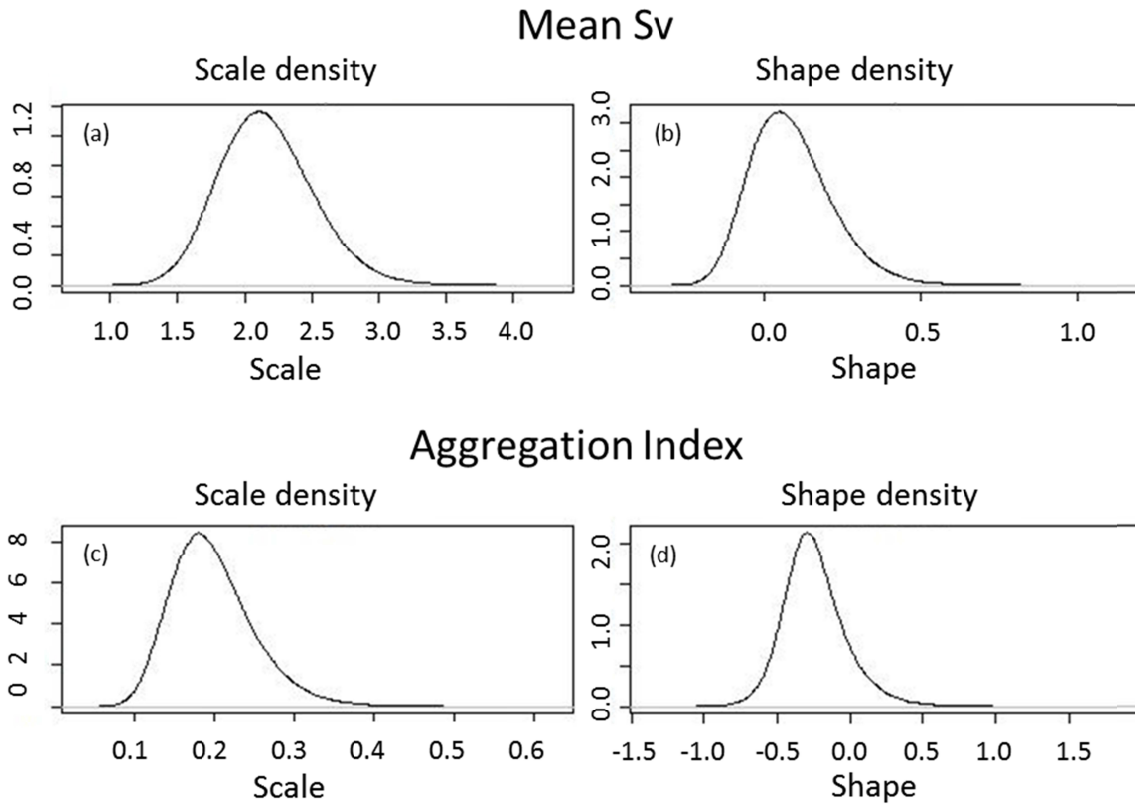


Fig. 2.7: Posterior densities for the mean Sv (a) scale and (b) shape parameters, and aggregation index (c) scale and (d) shape parameters

The aggregation index shape parameter density has a larger range (-1 to 1) and larger tails compared to the mean Sv shape parameter posterior density (range: -0.2 to 1) (Table 2.2). The greater 95% quantile parameter range is attributed to the smaller number of points above the aggregation index threshold compared to the mean Sv threshold.

Table 2.2: Median parameter values and the 95% quantiles in parentheses (lower, upper bounds).

Metric	Scale	Shape
Mean Sv	2.147 (1.541, 2.913)	0.0712 (-0.129, 0.378)
Aggregation Index	0.191 (0.114, 0.319)	-0.264 (-0.619, 0.257)

### 2.3.2.2 MCMC diagnostics

#### 2.3.2.2.1 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 (Fig. 2.8). 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, and 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.

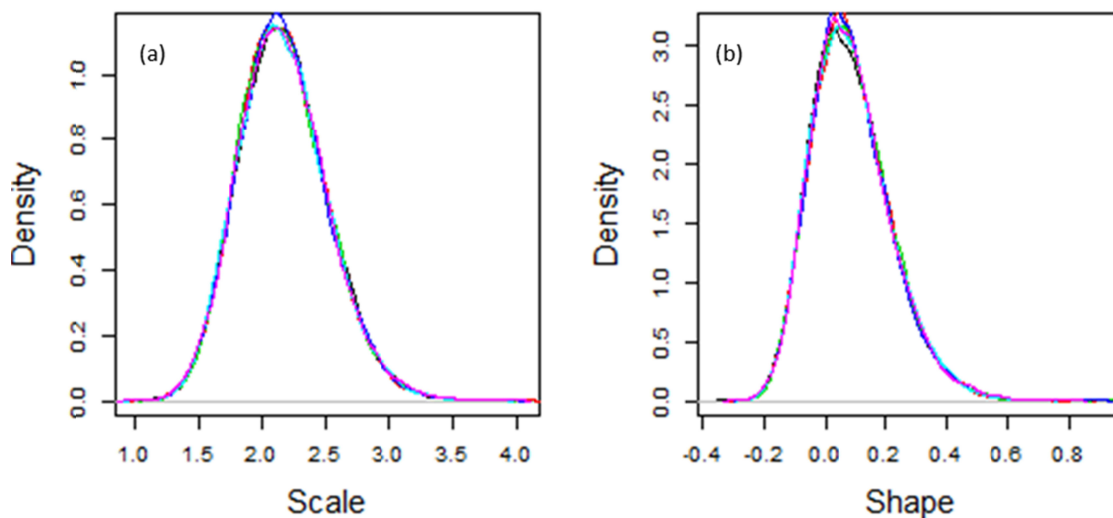


Fig. 2.8: (a) scale and (b) shape posteriors for the mean Sv metric from six chains with six different pairs of starting values.

#### 2.3.2.2.2 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 of the defined threshold (Fig. 2.9). 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 true threshold.

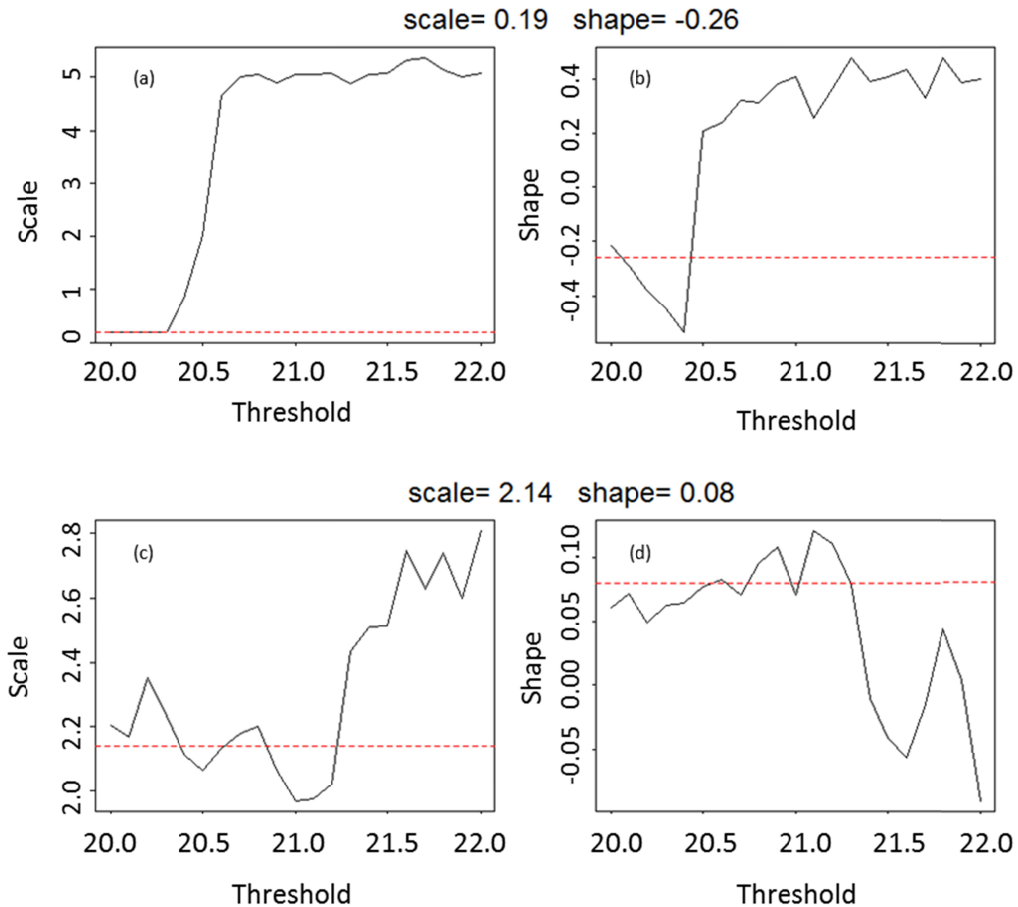


Fig. 2.9: Median values from MCMC generated posteriors of scale and shape parameters plotted as a function of threshold, for simulated data with known parameters, (a) scale and (b) shape parameter combination for mean Sv and for (c) scale and (d) shape aggregation index. The known scale and shape parameter values are depicted by the red line.

### 2.3.2.3 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. 2.10 a).

The aggregation index return level exhibited a different behavior than mean Sv. The median return level increased slowly until it reached about  $0.75 \text{ m}^{-1}$  at 1 year (Fig. 2.10

b), and then increased exponentially, reaching an asymptote at  $0.75 \text{ m}^{-1}$ . 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.

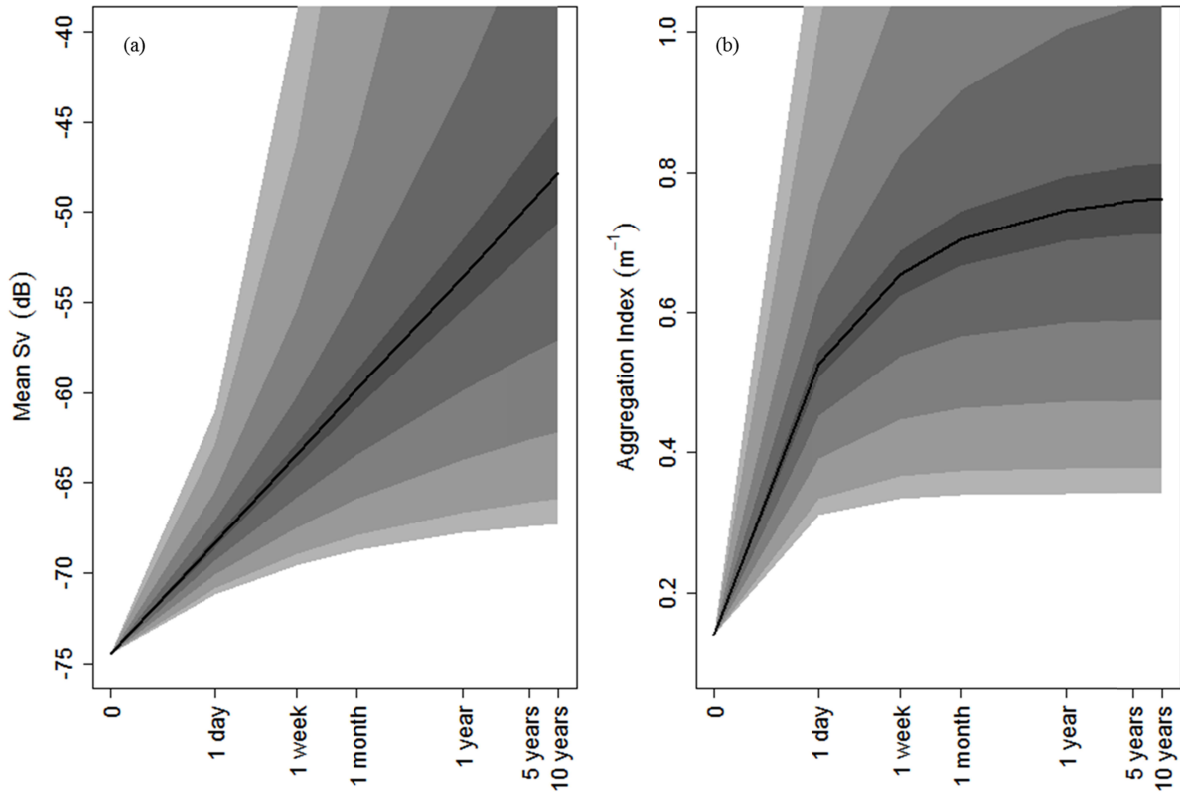


Fig. 2.10: Median (a) mean Sv and (b) aggregation index return levels. The solid lines is the best fit, and grey colors indicate credible intervals, from 10% (darkest grey) 40%, 80%, 90% lightest grey).

The rapid expansion of the credible intervals for the return levels of both metrics is attributed to sample size. The data were collected for one month, which increases the uncertainty in return levels for larger return periods (Coles, 2001).

Return levels must be interpreted with caution. The aggregation index return levels are bounded by the values of  $0.14 \text{ m}^{-1}$  and  $1 \text{ m}^{-1}$ , 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 2.3). This occurs because a positive shape parameter makes the GPD infinite. For metrics that are bounded, the values that exceed the bound limits could be replaced by the true metric bounds (e.g.  $1 \text{ m}^{-1}$  for the aggregation). For metrics that are unbounded, such as Mean Sv, the return levels could be constrained to biologically reasonable values.

Table 2.3: Median values and 95% (lower, upper) credible intervals for mean Sv and aggregation return levels for univariate GPD fit, MLE return level for bivariate GPD.

Return period	Mean Sv		Aggregation Index	
	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

### 2.3.3 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 variables as a joint process. Only 8 data points are jointly above the mean Sv and aggregation index thresholds. The  $\alpha$ -value is 0.95 suggesting weak correlation, and almost independent variables. MLE estimates for the bivariate model are close to the median values estimated by Bayesian inference (Table 2.3), and the differences changed the shape of the return level curves slightly (Fig 2.11 a, b). Standard errors for the MLE estimates are large (Table 2.4), reflecting the difficulty in fitting a bivariate model to this data due to the limited number of joint extreme datapoints ( $n=8$ ) and the near independence of the variables.

Table 2.4: Parameter estimates for the bivariate peaks-over-threshold analysis

	mean Sv		Aggregation Index		
	Scale	Shape	Scale	Shape	Alpha
MLE	2.13	0.05	0.19	-0.32	0.95
Standard Error	0.34	0.12	0.05	0.17	0.04

Return levels for both the mean Sv and aggregation index (Table 2.3) in the bivariate case are lower than in the univariate case. This is an expected result as high aggregation and density rarely occur at the same time and joint events occur less frequently than single events.

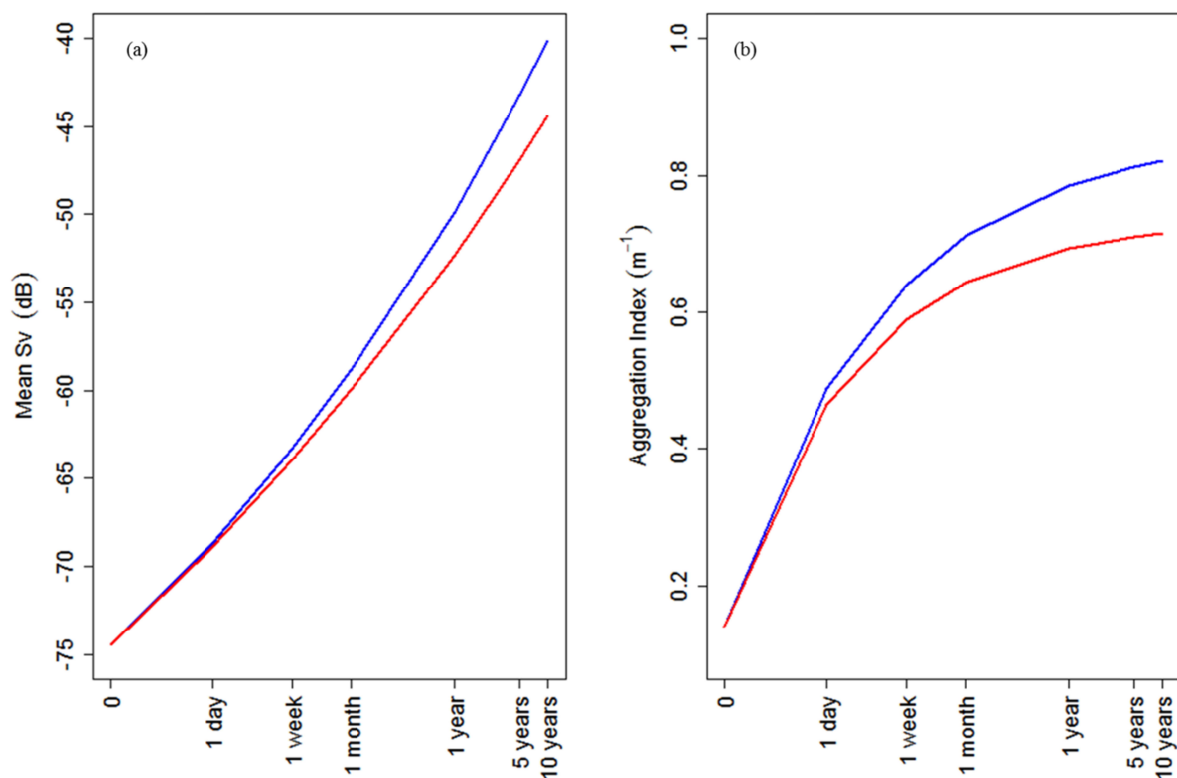


Fig. 2.11: Median univariate (blue) and bivariate (red) return level for (a) mean Sv and (b) aggregation index.

## 2.4 Discussion

### 2.4.1 Extreme value analysis applications for biological monitoring

EVA can be used to analyze a baseline dataset and make inferences about conditions that may be associated with biological impacts. These inferences are useful to refine environmental monitoring programs. Observations above a threshold are statistically rare and occur where high-risk events are likely to transpire (Coles, 2001). Observing values above a threshold or an increase in the frequency of extreme values, compared to baseline measurements, could be used as an indication that an impact has occurred. Defining a threshold for extreme values will help MRE managers assess the risk of impacts, as well as establish a baseline for expected extreme value periodicity.

EVAs are also useful for understanding conditions under which an impact is likely. Extreme events can be tested for correlation with biological covariates, such as a metric tracking vertical distribution in the water column (e.g. center of mass), or environmental

covariates such as tidal speed or weather structures. Correlation with covariates could lead to a greater understanding of conditions under which impacts may occur, which could be used to increase monitoring effort during these conditions.

#### **2.4.2 Utility of Peaks-over-threshold**

Selecting a threshold is a critical and challenging step in the POT analysis (Dupuis, 1999). Besides examining diagnostic plots, other methods for selecting thresholds have been proposed, but many of these are computationally intensive or case-specific (Mazas and Hamm, 2011; Scarrott and MacDonald, 2012). In this study, a derivative method is used to establish the GPD threshold. This method utilizes traditional diagnostic plots but with greater precision and objectivity than 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. Matching threshold estimates from derivatives of both plots is interpreted that the derivative method accurately estimated the GPD threshold.

Even though the threshold is not sample size dependent, it is sample dependent. POT analysis allowed estimation of threshold values, but the true GPD threshold may differ than the one determined by the POT threshold analysis, depending on how the data represents the true distribution of outcomes. Precision of the GPD threshold value is increased by collecting more baseline data and 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 (FERC, 2004). 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 (Viehman, 2012). In a BACI analysis, a representative baseline dataset is crucial for comparison to data collected after the environmental alteration has occurred if impacts are to be detected (Schmitt and Osenberg, 1996). The Admiralty Inlet and the Cobscook Bay datasets contain different temporal resolutions and the threshold analysis for the Cobscook Bay dataset would likely reflect seasonal fluctuations rather than daily biotic variability. It will be important to determine the appropriate temporal resolution of baseline sampling, both to establish an accurate GPD threshold and to detect biologically relevant change once project installation and operation begins.

Results from the bivariate peaks-over-threshold analysis indicated that joint extreme density and aggregation events were too rare to accurately model. A bivariate analysis was incorporated in this study because a joint extreme aggregation and density event potentially represents a greater risk for animal collision with a tidal turbine device. A precise estimate of a joint return level prediction was not determined, and extreme values of the two metrics were independent. The simultaneous occurrence of an extreme density and an extreme aggregation event is improbable for this dataset. Because the mean Sv and aggregation index metrics were independent, the bivariate model does not provide additional information to the univariate models and the mean Sv and aggregation index metrics should be considered separately in a monitoring program.

Multivariate extremes can be used to model highly correlated processes, such as wind, wave amplitude, and current fluxes (Nerzic et al., 2007) and spatial dependence in extremes (Northrop and Jonathan, 2011). A multivariate model may be informative if additional data confirms 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 spread of the 95% credible intervals around return levels show 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 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 (e.g. Verdant Power, 2010). Baseline surveys could also be supplemented by data collected during the project operation period; with return levels estimated separately and used to inform project managers about conditions under which increased monitoring is necessary (i.e. conditions that may result in extreme events). This approach is useful during ongoing monitoring programs, as more data will increase accuracy of GPD parameter estimates, which will, in turn, decrease return level uncertainty.

Parameter uncertainty can also be reduced by performing an MCMC analysis using informative priors. Informative priors have been used in impact detection studies (e.g. Crome et al., 1996; Garthwaite and O'Hagan, 2000; Martin et al., 2005), and can be formed by soliciting expert opinion on the effects of and relationship to a disturbance (e.g. Behrens et al., 2004). To illustrate by example: MRE development has no negative impact on the environment, MRE development has a strong negative effect on fish abundance, the negative effects of MRE devices are linearly related to noise production (Polagye et al., 2011). 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 decrease the range of the 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 (Gelman et al., 2013). 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 (Gray et al., 2005; Doelle, 2009). 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.

### **2.4.3 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 (FERC, 2004). The debate over how to quantify a biologically significant effect is ongoing (Germano, 2000; Hobbs and Hilborn, 2006; Lovell, 2013). 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 (Martínez-Abraín, 2008). The choice of biological significance relies on expert, and potentially subjective, judgment (Lovell, 2013). Consensus among stakeholders may be difficult to reach (Gray et al., 2005). Martínez-Abraín (2008) advocates that these decisions have to be made before monitoring begins, 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 effect is that extreme values are rare, but very detectable with active acoustics. 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.

## **Chapter 3:**

# ***COMPARISON OF BIOLOGICAL CHARACTERISTICS IN DISTRIBUTION OF FISH AND MACROZOOPLANKTON AT TWO TIDAL ENERGY SITES***

## **3.1 Introduction**

Biological monitoring programs are important to identify and minimize impacts from anthropogenic disturbances. These programs focus on the detection of change in biological variables such as diversity, size, or abundance of monitored targets (Bijleveld et al., 2012). Many monitoring programs require the collection of baseline data before any alteration to an ecosystem (Stewart-Oaten and Bence, 2001). At a single site, biological characteristics before and after an alteration can be compared to detect change, as in a classic Before – After – Control – Impact (BACI) design (Underwood, 1994). Among sites, standard sampling protocols permit monitoring datasets to be compared to evaluate if observed changes are site-specific or generic.

Biological monitoring programs are mandatory for marine renewable energy (MRE) tidal energy projects in the US, although no regulations for monitoring procedures, technologies, or metrics currently exist (FERC, 2004). This lack of standardization has resulted in different and site-specific monitoring programs for each tidal energy pilot project in the US. Standardization of a portion or all monitoring components would enable monitoring plans to be proposed in a time-efficient manner, and make monitoring datasets comparable across sites.

Determining what the maximum level of “acceptable” impact, or biologically significant change, is a high priority when forming a monitoring plan (Mapstone, 1995). Impact above a threshold can determine if a tidal project is allowed to continue operating (FERC, 2008). Thus it is imperative that setting thresholds and characterizing impacts should be completed before monitoring begins (Martínez-Abraín, 2008). Extreme value analysis (EVA, described in Chapter 1) is an approach used to model values that are infrequent but are potentially associated with impacts (Coles, 2001). This approach also provides a

threshold for infrequent values and could provide regulators with statistically significant thresholds for biological monitoring (Chapter 1, Section 4).

Little data have been gathered at tidally dynamic sites because they are difficult to sample. One option for studying biota in the water column is active acoustic technology. Acoustic instruments use sound to evaluate distributions, abundances, and behavior of fish (Horne, 2000; Simmonds and MacLennan, 2005). Acoustic instruments offer non-invasive methods to continuously sample large volumes of water, regardless of current speed or light levels. Instruments can be deployed on autonomous or cabled platforms that are suited for short or long-term monitoring at high spatial and temporal resolution (Urmy et al., 2012), and low cost (Handegard et al., 2013).

It is important to evaluate and compare MRE site characteristics so the potential for standardized programs can be developed and then implemented. While MRE tidal sites may be chosen for similar physical characteristics (e.g. high tidal flux), it is unknown whether tidal energy sites have similar biological characteristics. In this study we describe and compare biological characteristics of pelagic fish and zooplankton distribution at two proposed tidal energy sites, using datasets collected by active acoustics, to examine whether fish and zooplankton density distributions are similar or site-specific. We also evaluate whether an EVA is an appropriate general approach to determine impact thresholds at tidal energy sites.

## **3.2 Methods**

### **3.2.1 Site descriptions**

The active acoustic datasets used for this study were collected at two tidal energy sites without any deployed turbines. Admiralty Inlet, on the west side of Whidbey Island in Puget Sound, Washington State, was the proposed site of the Snohomish Public Utility District 1 (SnoPUD) tidal energy pilot project that received its project license from FERC on March 20, 2014. The project, now dormant (<http://www.openhydro.com/>), would have deployed two OpenHydro turbines one kilometer off Whidbey Island (Fig. 3.1 a). Two sub-sea power cables would have connected the turbines to the electric grid onshore (Public Utility District No. 1 of Snohomish County, 2012). The second dataset was collected at the European Marine Energy Council (EMEC) test facility in Fall of Warness, located centrally in the North Isles of Orkney, Scotland (Fig. 3.1 b). The data was collected as a control for the FLOWBEC project (<http://noc.ac.uk/project/flowbec>).

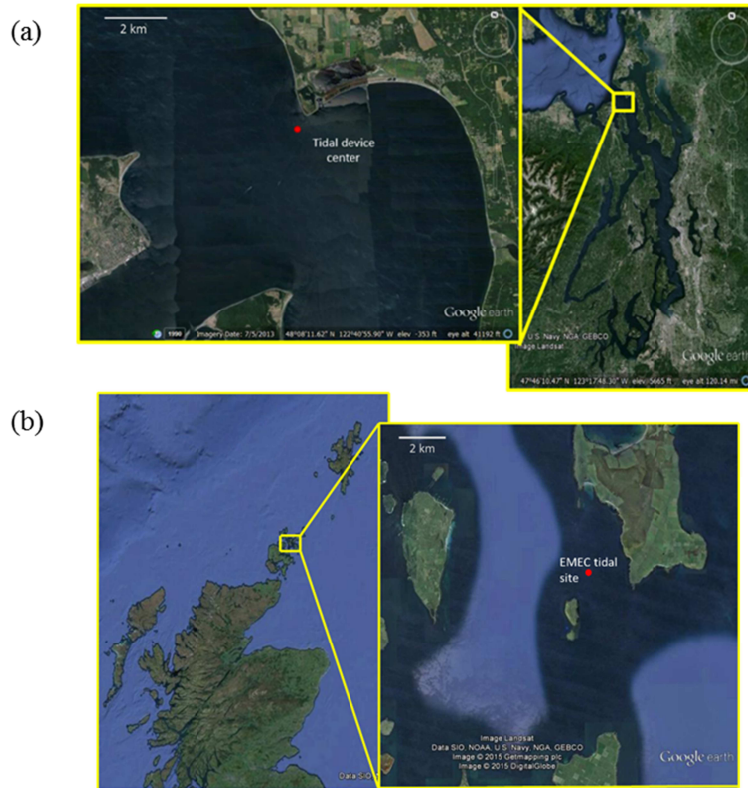


Fig. 3.1 (a) Admiralty Inlet study location within Puget Sound, Washington, and location of the acoustic package within the SnoPUD proposed turbine location, (b) EMEC tidal test site within the Orkney Islands, Scotland, and location of Fall of Warness deployment site.

### 3.2.2 Data acquisition

Acoustic backscatter (i.e. reflected energy) data were recorded at Admiralty Inlet using a bottom mounted BioSonics DTX echosounder operating at 120 kHz from May 9 to June 9, 2011 (Horne et al., 2013). The echosounder was placed on the bottom at 55 meters depth about 750 meters off Admiralty Head at the SnoPUD tidal turbine site. The echosounder sampled at 5Hz for 12 minutes every 2 hours. Tidal velocity data were collected once every 10 minutes by a NORTEC acoustic Doppler current profiler set 10 meters off the bottom (Table 1).

At Fall of Warness, a bottom-mounted acoustic package was placed at 35m depth containing a multibeam sonar and an EK-60 Echosounder (Williamson et al., 2015). It was deployed over an 18 day period from June 18 to July 5 2013. The echosounder collected data at three frequencies: 38 kHz, 120 kHz, and 200 kHz and operated continuously at 1 Hz (Table 3.1). Water column mean tidal speeds were modeled from

tidal velocity data that were collected every minute using an SonTek/YSI ADVOcean acoustic Doppler velocimeter (ADV) (Williamson et al., 2015).

Table 3.1. Acoustic sampling parameters

Deployment	Fall of Warness	Admiralty Inlet
Technology	Echosounder	Echosounder
Manufacturer	Simrad	BioSonics
Model	EK-60	DTX
Frequency	120kHz	120kHz
Beam Angle	7°	7°
Pulse length	1024 $\mu$ s	500 $\mu$ s
Ping Rate	1 Hz	5 Hz

### 3.2.3 Data processing

#### 3.2.3.1 Admiralty Inlet

Data processing of the Admiralty Inlet data was done prior to this study and is described in Jacques (2014) and Horne et al., (2013). Because of a 3<sup>rd</sup> surface echo, data values were constrained to 25 m from the bottom. A -75 dB re 1m<sup>-1</sup> (hereafter dB) threshold was applied to remove noise (Horne et al., 2013). Data were binned into 12 minute samples, and vertically integrated, yielding 361 datapoints (Jacques, 2014).

An 18 day period needed to be sampled from the Admiralty Inlet dataset to closely align it to the 18 day period of the Fall of Warness data. Using historical tide charts from NOAA (National Ocean Service, 2011, 2013), the 18 day portion of the Admiralty Inlet data was selected so that the lunar phases of the two datasets matched. The start and end times of the Admiralty Inlet dataset were selected to match the start and end times of the Fall of Warness dataset.

#### 3.2.3.2 Fall of Warness

Acoustic data from Flowbec’s Fall of Warness site was processed in Echoview (version 6.0). The background noise was removed by using a post-processing TVG-based noise reduction (Watkins and Brierley, 1996). The noise estimates were obtained from three empty water recordings (i.e. low water column integrated Sv value), and the average of the three values, -105.44 dB, was subtracted from the data (Nunnallee, 1990). The data were constrained to 25m from the bottom to ensure complete turbulence removal and to match the depth of the processed Admiralty Inlet data portion. A 12 minute temporal

block was used to accommodate Echoview memory capacity and to match the sample block size used in Admiralty Inlet data.

Turbulence in the water was detected using the SHAPES algorithm (Coetzee, 2000), as implemented in Echoview. This algorithm is normally used to detect fish and macrozooplankton aggregations but can be also used to detect turbulence features (Jacques, 2014). This algorithm detects aggregations by searching for adjacent pixels with density values above a threshold, and applying a minimum size criterion to groups of pixels. The user determines the acoustic threshold, minimum aggregation size, and amalgamation parameters. Virtual positions, necessary for the implementation of this algorithm, were created using flow rates derived from ADV data by matching the start and end times of the echogram to the ADV data, and then indexing each second of the data samples to a corresponding flow speed. The turbulence detection threshold was set to -75 dB, to include all backscatter attributed to pelagic nekton and to exclude particulates. After aggregations were detected, they were classified as turbulence or non-turbulence regions using depth and length of detected aggregation as criteria.

The two datasets had to have identical sample blocks to enable a direct comparison. To match the Admiralty Inlet data (12 minutes of continuous sampling every two hours), the Fall of Warness data needed to be binned to 12 minutes and one of 10 options of 12 minute bins chosen to represent each 2 hour block. The continuous sampling at Fall of Warness also facilitated an analysis of how representative 12 minute samples are of a continuous two hour dataset. Each of these 10 series were comprised of bins selected at regular intervals (i.e. the first series was made up of the first 12 minute bin in every block, the second series was made of the second 12 minute bin in every block ...).

The ten mean Sv series of 12 minute bins from the Fall of Warness data were compared to each other using an ANOVA and by examining the fit of a Generalized Pareto Distribution (GPD) for all series. The GPD is used to model extreme values, which are exceedances above a threshold. The threshold is selected through examination of diagnostic plots, which was done in this study through automated derivative selection (Chapter 1, Section 2.3.2).

The effect of varying the amount of data on the threshold estimate for the GPD was examined. Series containing nine different amounts of data were generated, starting at 10% (one bin randomly selected from each 2 hour block) to 90% of the data (nine bins randomly selected from each two hour block). The random selection was repeated 500 times (total 4500 series) and derivative based thresholds (Chapter 1, Section 2.3.2) were estimated for each series.

### **3.2.4 Data Analysis**

#### **3.2.4.1 Echometrics**

To quantify variability in vertical biomass distribution in the water column over space and through time a suite of metrics were used (Urmy et al., 2012). Among the metrics, density (mean Sv), center of mass, dispersion, and aggregation index were selected to characterize biomass distributions at the two sites, as these metrics describe a wide range of distribution attributes (Jacques, 2014). The density metric is the mean volume-backscattering strength, or mean Sv (unit: dB re 1 m<sup>-1</sup>, Simmonds and MacLennan, 2005), which is proportional to biomass density. The aggregation index (unit: m<sup>-1</sup>) quantifies vertical patchiness and is calculated on a relative scale of 0 to 1, with 0 being evenly dispersed and 1 being aggregated. Mean Sv and aggregation index are used for extreme value analyses. High aggregation and density values are assumed to be associated with a high risk of collision with MRE devices. The center of mass (unit: m) measures the location of the mean weighted acoustic backscatter relative to the bottom. The dispersion (unit: m<sup>2</sup>) metric measures the spread of biomass around the center of mass, and is analogous to the variance.

#### **3.2.4.2 Tidal site comparison**

Spectral analysis (Box and Jenkins, 1976; Chatfield, 1989) is used to compare dominant periodicities in the metrics of the two datasets. Periodograms can be used to examine how the variance of a time series is distributed over its frequency components (Chatfield, 1989). To interpret the periodogram, one plots the frequency against the power. Peaks identify frequencies that contribute more to the variance of the time series than other unpeaked areas. The total area under the curve, for a log-log plot, equals the total variance in the time series. Periodograms were generated for the biological metrics at both Admiralty Inlet and Fall of Warness. A red noise spectrum (a red noise process is an auto-regressive process with the memory of 1) was calculated for each periodogram (Shumway and Stoffer, 2011) and the significant frequencies were defined as those above the red noise spectrum. The coherence is measured from 0 to 1, with 0 signifying that the two time series are significantly different, and 1 being that the phase differences and amplitudes are the same for all frequencies. Coherence between the power spectra of metric pairs was calculated to compare the amplitudes and phases of frequencies in the metric time series between the two sites.

Means and variances of the mean Sv, center of mass, dispersion metrics, and tidal speed were compared using t-tests and F-tests with alpha = 0.05. The aggregation index was not normally distributed and could not be compared with parametric tests, so a Kolmogorov-Smirnov test (Massey Jr, 1951) was used to compare means and a Bartlett's test (Bartlett,

1947) was used to compare variances. Differences in metrics between day (Admiralty Inlet: 06:00 – 20:00, Fall of Warness: 04:00 – 22:00) and night (Admiralty Inlet: 22:00 – 04:00, Fall of Warness: 00:00 – 02:00) were examined in the mean using t-tests and in the variance using F-tests.

Linear models were fit to the four metrics at both sites. A group of covariates were tested in each model: tidal speed ( $\text{ms}^{-1}$ ), hour of day, Julian day, a Fourier series defined by the 4 hour period, a Fourier series defined by a 12 hour period, a Fourier series defined by a 24 hour period. Fast Fourier transforms provides the amplitudes and phases for the Fourier series. The models were fit by forward selection, and the fit was evaluated using the Akaike Information Criterion (AIC) (Akaike, 1987). The model with the lowest AIC was selected as the best model. Residual plots were examined to evaluate model fit and multicollinearity was examined through the variance inflation factor (VIF) (Belsley et al., 1980). Linear regression was also used to evaluate the metrics' relationship with tidal speed as a single covariate.

Extreme value theory (Pickands, 1975; Coles, 2001; Beirlant et al., 2004) 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. In the peaks-over-threshold (POT) method, extreme values are identified as exceedances above a threshold (Coles, 2001), which follow a generalized Pareto distribution (GPD). To fit the GPD to data, first a threshold is selected, and then scale and shape parameters are fit to the data to parametrize the GPD. The GPD was fit to the mean Sv and aggregation index metrics calculated using Admiralty Inlet and Fall of Warness backscatter data. The threshold for the GPD fit was selected by computing the derivative of diagnostic plots (Chapter 1, Section 2.3.2). Posterior distributions were obtained for the scale and shape parameters through Markov chain Monte Carlo (MCMC) simulation (Gelman et al., 2013), as described in Chapter 1 (section 2.3.4). The fit of the GPD was evaluated by computing the sums of squares between the observed density function and the corresponding GPD. Return level plots (the return level is the value expected to be exceeded on average once every its associated return period) with credible intervals were generated for the mean Sv and aggregation indices at both sites.

### **3.3 Results**

#### **3.3.1 Fall of Warness sample block selection**

Results of ANOVA comparing the ten mean Sv series of regularly spaced 12 minute bins per two hour block showed that means of the series were not significantly different ( $p =$

0.7024), suggesting that any of the ten series could be used as a representative dataset for the Fall of Warness site.

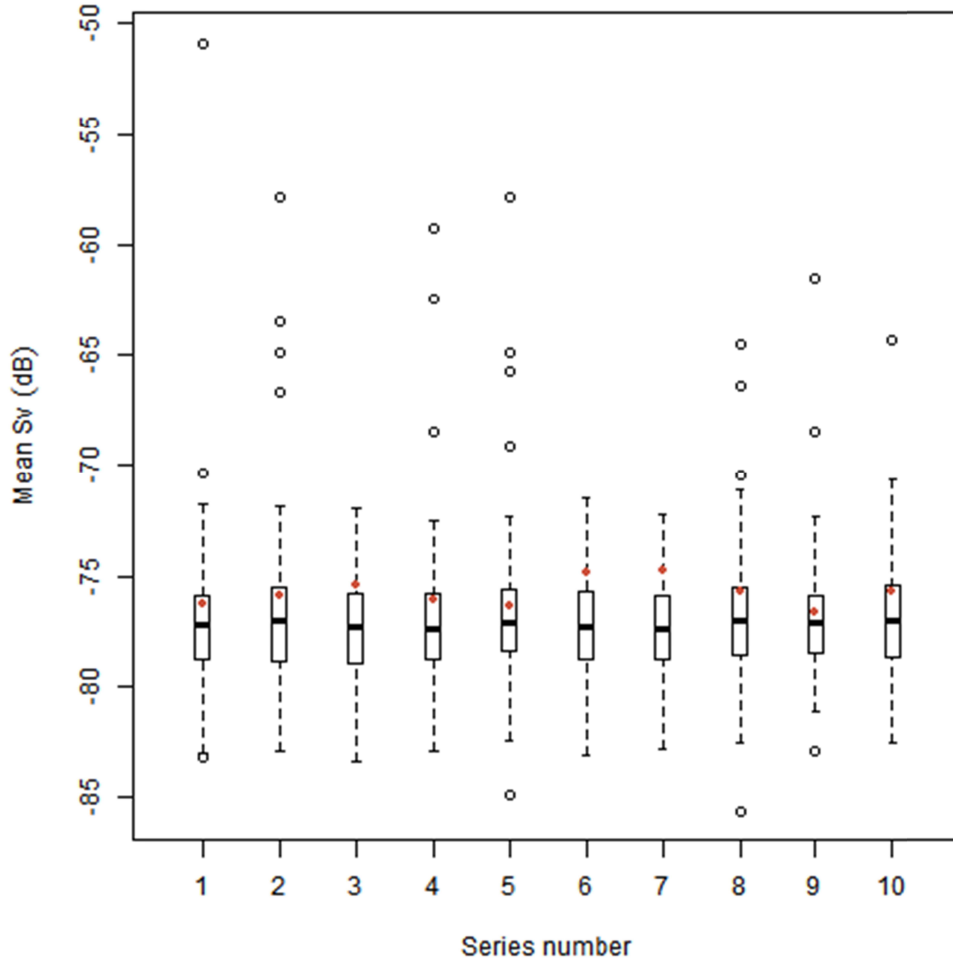


Fig. 3.2: Boxplots for the ten Mean Sv series, with the thresholds as orange dots.

Corresponding GPD threshold values differed between the 10 series. Standard deviations (standard deviation range: 1.95 dB to 2.87 dB) and number of significant outliers varied among series (1 to 4 outliers) (Fig. 3.2) which could affect the GPD fit as the GPD is modeled using extreme values as determined by the threshold. The choice of the threshold value was dependent on how the derivative values were rounded. The MCMC routine to fit the GPD did not converge for the series 3, 6, and 7, which have few outliers and low variance. The mean for the thresholds of the ten series was -75.41 dB. Series 8, with a threshold of -75.63 dB was selected as the Fall of Warness dataset to be used in further analysis as it had the closest threshold to the mean of the thresholds of the 10 series and successfully converged on a GPD posterior under the MCMC routine.

The proportion of data used in the derivative method did not greatly affect the mean value of the GPD threshold, but it affected the variance in the threshold estimate (Fig. 3.3). The threshold when all the data is included is -75.68 dB. The mean threshold value increases slightly as the proportion of data used to estimate the threshold increases (e.g. 1 bin per block = -75.72 dB, 9 bins per block = -75.67 dB), but the overall increase is less than 0.1 dB. The standard deviation of the threshold estimate decreases with an increasing proportion of data used, from 0.63 dB at one bin per block, to 0.045 dB at 9 bins per block. Over 500 draws, when 10% of the data is used (1 bin per block), the threshold estimates range from -77.84 dB to -73.79 dB. This analysis suggests that increasing the amount of data used decreases the variance in the threshold estimate.

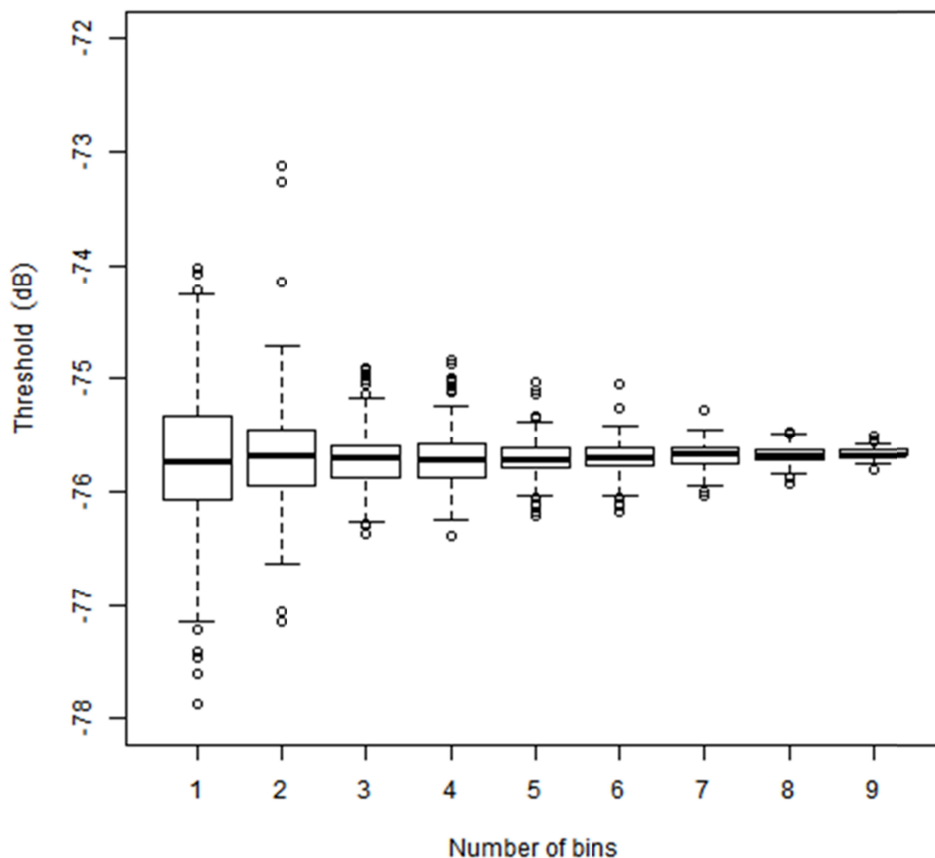


Fig. 3.3: Boxplots showing the distribution of threshold estimates for 500 draws of mean Sv series containing 1 bin (10%) to 9 bins (90%) of the available data.

### 3.3.2 Comparison of tidal site characteristics

Similarities of biological characteristics were evaluated by comparing metrics describing biological distribution of fish and macrozooplankton. Mean Sv, center of mass, and dispersion metrics at both sites displayed a saw-toothed pattern with a low-frequency

sinusoidal component (Fig. 3.4). Values at Admiralty Inlet had larger amplitudes compared to metric values from the Fall of Warness. F-tests show that standard deviations for the three metrics were significantly greater at Admiralty Inlet than at Fall of Warness (Table 3.2). The aggregation index series had a spiky appearance and was mainly composed of low values ( $\sim 0.04$ ) with occasional high aggregations. Metric means, except for mean Sv, were significantly different between sites. The center of mass was, on average, higher in the water column at Fall of Warness (14.07 m) than at Admiralty Inlet (11.35 m). There was greater dispersion at Admiralty Inlet ( $46.13 \text{ m}^2$ ) than Fall of Warness ( $41.97 \text{ m}^2$ ). Aggregation was greater at Fall of Warness ( $0.064 \text{ m}^{-1}$ ) than at Admiralty Inlet ( $0.062 \text{ m}^{-1}$ ).

Table 3.2: Means and standard deviations of biological characteristics at Admiralty Inlet (AI) and Fall of Warness (FoW). The p-values are for t-tests (left) and F-tests (right), for the aggregation index, the p-values are for a Kolgomorov-Smirnov test (left) and Bartlett's test (right).

	Mean			Standard Deviation		
	AI	FoW	p-value	AI	FoW	p-value
Mean Sv	-76.52 dB	-76.94 dB	0.2041	3.869 dB	2.524 dB	2.46E-09
Center of Mass	11.35 m	14.07 m	2.2E-16	2.27 m	1.93 m	1.62E-06
Aggregation	$0.062 \text{ m}^{-1}$	$0.064 \text{ m}^{-1}$	5.12E-06	$0.047 \text{ m}^{-1}$	$0.047 \text{ m}^{-1}$	0.9957
Dispersion	$46.13 \text{ m}^2$	$41.97 \text{ m}^2$	2.21E-05	$10.46 \text{ m}^2$	$8.96 \text{ m}^2$	0.028

Daily patterns in metric values also varied between sites. On average there was greater variability between day and night mean Sv at Admiralty Inlet than at Fall of Warness (Admiralty Inlet , difference =2.63 dB, t-test p value  $p = 9.15\text{E-}07$ ; Fall of Warness difference = 1.35 dB, t-test p values = 0.0005535). At Admiralty Inlet, the center of mass was higher in the water column at night (mean: 13.8 meters from the bottom) and descended during the day to 10 meters from the bottom (Fig. 3.5 b). At Fall of Warness the center of mass increased from 13.8 meters from the bottom at night to 15.2 m during the day. The aggregation index at both sites had two peaks on average, though these occurred at different times. At Fall of Warness, the peaks occurred at: 06:00 ( $0.09 \text{ m}^{-1}$ ) and 18:00 ( $0.12 \text{ m}^{-1}$ ), while at Admiralty Inlet the peaks occurred at 08:00 ( $0.09 \text{ m}^{-1}$ ) and 14:00 ( $0.1 \text{ m}^{-1}$ ). Dispersion was significantly different between night and day at Admiralty Inlet (difference =  $4.25 \text{ m}^2$ ,  $p = 0.002624$ ), but not at Fall of Warness (difference =  $0.15 \text{ m}^2$ ,  $p = 0.972$ ).

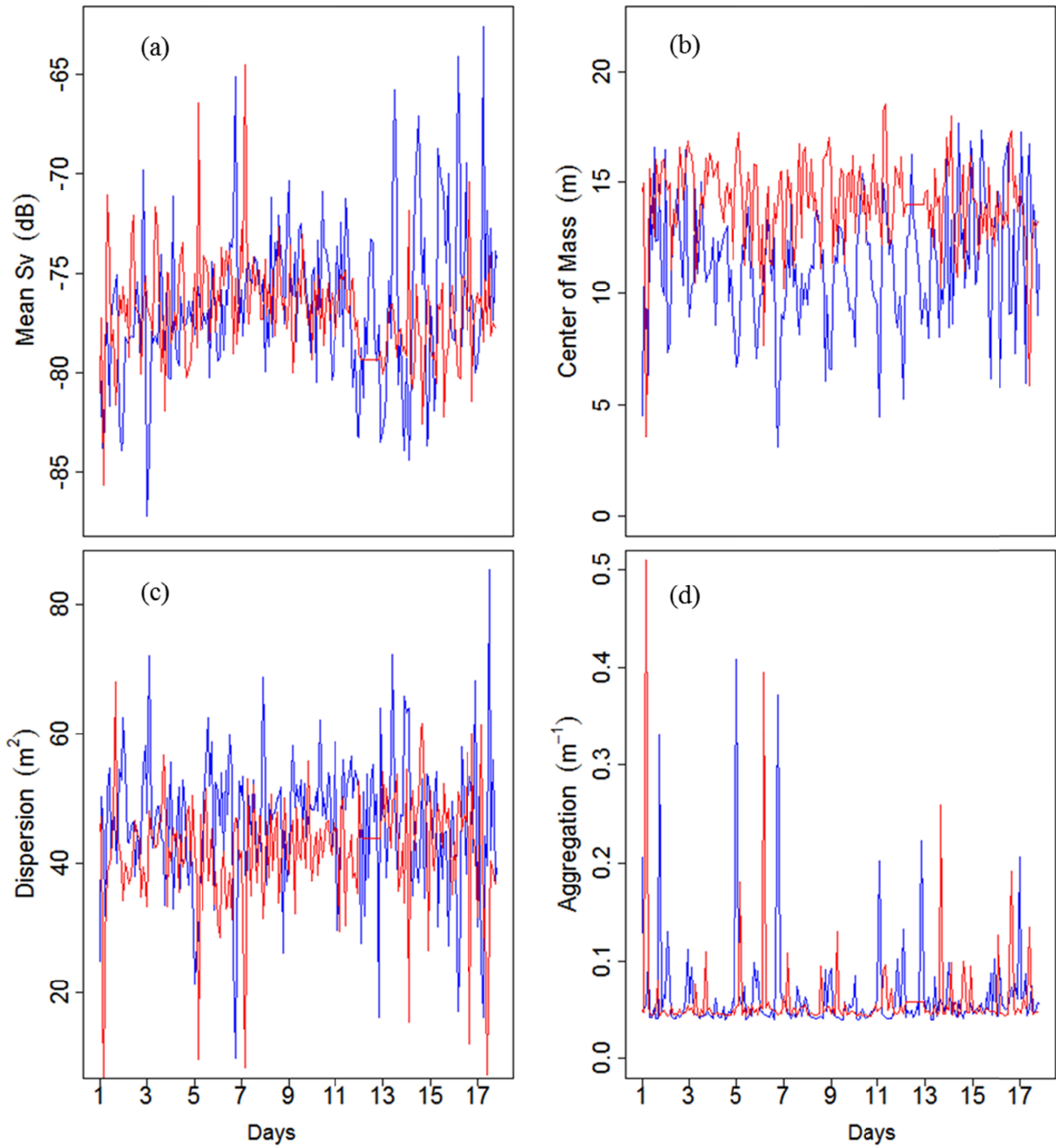


Fig. 3.4: The time series for four metrics (a) mean Sv, (b) center of mass, (c) dispersion, (d) aggregation, with Admiralty Inlet in blue and Fall of Warness in red.

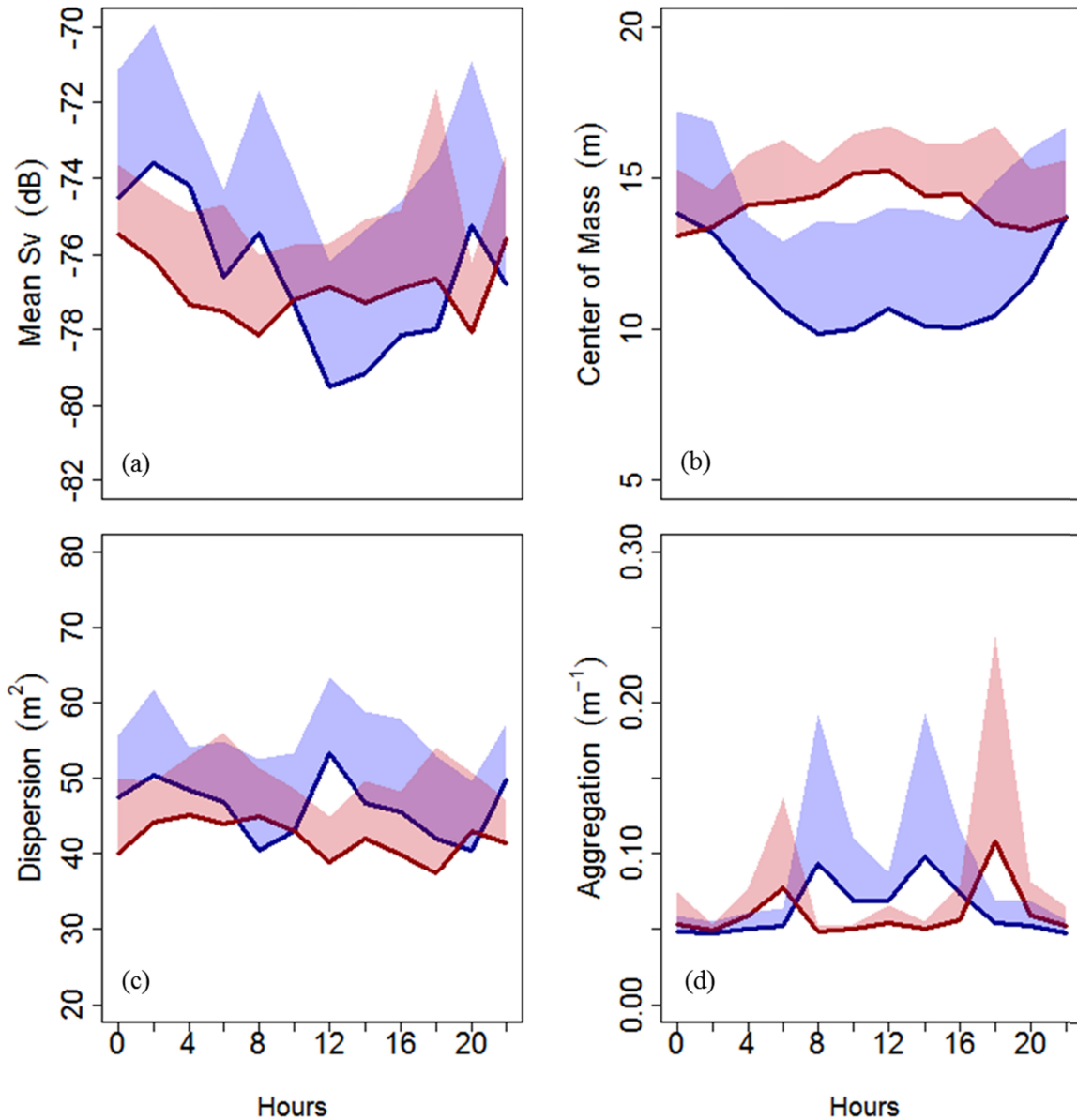


Fig. 3.5: The (a) mean Sv, (b) center of mass, (c) dispersion, (d) aggregation variability by hour for Admiralty Inlet (blue) and Fall of Warness (red). The solid lines are the means with the blocks; for clarity, only the positive standard deviation is depicted by the colored blocks.

Biomass distribution was predicted to vary with tidal speed. The tides were stronger at Fall of Warness (mean speed= $1.58 \text{ ms}^{-1}$ ), compared to Admiralty Inlet (mean speed =  $1.18 \text{ ms}^{-1}$ ). At Admiralty Inlet, the mean tidal speeds ranged between  $0.5$  and  $2 \text{ ms}^{-1}$ . The tidal speed range was greater at Fall of Warness, varying between  $0$  and  $3 \text{ ms}^{-1}$ . In Admiralty Inlet, mean Sv increased with tidal speed (slope =  $0.64$ ), but the relationship was not significant ( $p = 0.481$ ) (Fig. 3.6 a). At Fall of Warness, the biomass density

increased significantly with tidal speed, and at a steeper slope than Admiralty Inlet (slope = 1.033,  $p = 7.28E-5$ ). The center of mass decreased significantly with increased tidal speed at Admiralty Inlet (slope = -1.47,  $p = 0.0153$ ), but not at Fall of Warness (slope = 0.04,  $p = 0.947$ ). Aggregation index values did not change with increasing tidal speed at either site (slope = -0.004, slope = -0.005). Dispersion increased at Fall of Warness (slope = 2.048,  $p = 0.028$ ) but not at Admiralty Inlet, where it decreased slightly with increasing tidal speed (slope = -0.659,  $p = 0.779$ ).

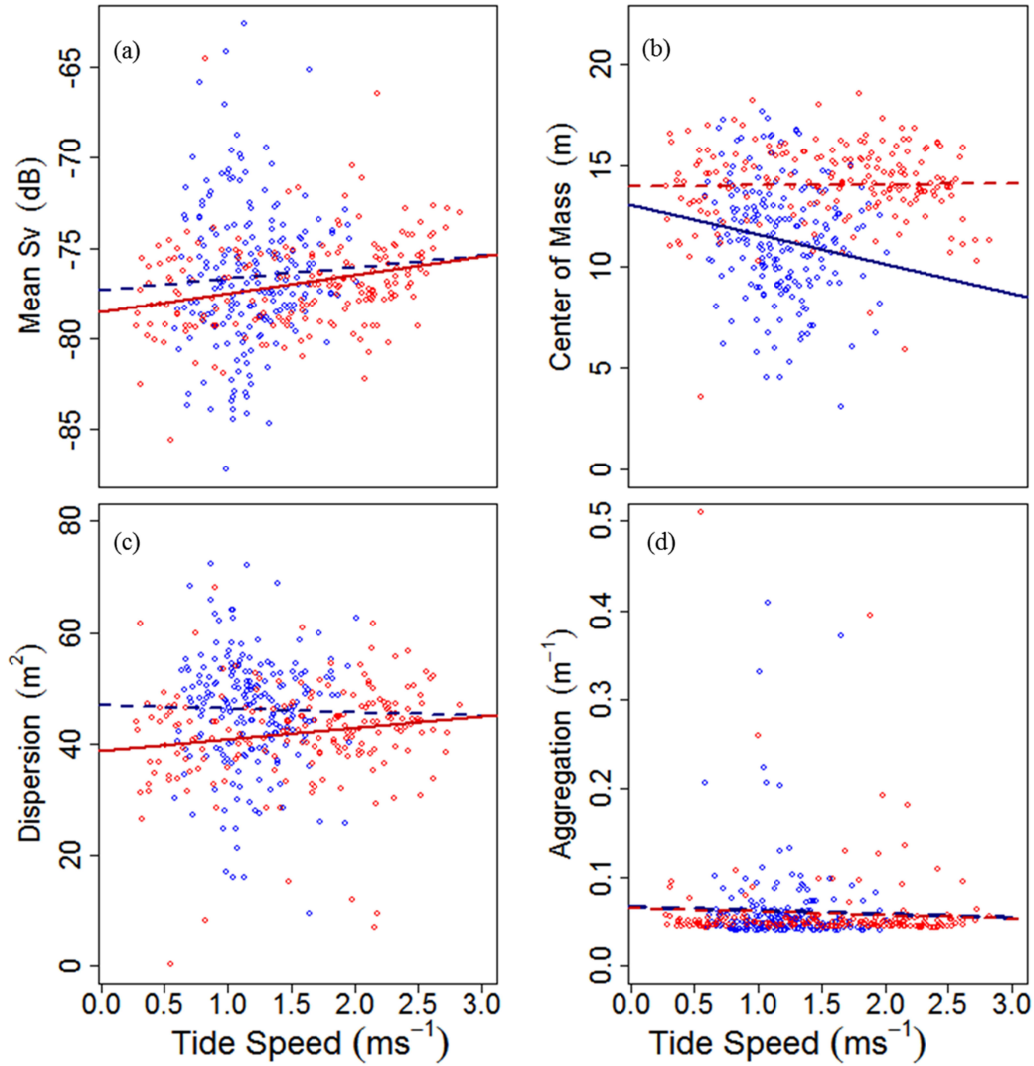


Fig. 3.6: The (a) mean Sv, (b) center of mass, (c) dispersion, (d) aggregation as a function of tidal speed for Admiralty Inlet (blue) and Fall of Warness (red). The lines of best fit from a linear regression are shown as well.

Metric values at both sites had many of the same dominant periodicities, though not always with similar amplitudes (Fig. 3.7). For mean Sv, the significant periods (compared

to the red-noise spectrum) for Admiralty Inlet were 24 hours and 4 hours (Table 3.3). Twenty four hours was by far the dominant periodicity (amplitude = 2.5), perhaps highlighting the importance of diel processes at this site. For Fall of Warness the Mean Sv significant periods were 404 hours, 12 hours, and 4 hours. Coherence between the two mean Sv metrics was the highest among all metrics, at 0.997. We discounted the 404 hour period as its significance may be due to edge effects (Chatfield, 1989). The importance of the 12 hour (amplitude = 0.9) and 4 hour frequencies (amplitude = 0.6) indicates the potential importance of tidal over diel processes for the biology of this site (Table 3.3). However, amplitudes of these frequencies are small compared to the 24 hour period amplitude at Admiralty Inlet, suggesting that there may not be as dominant a process accounting for the variability in mean Sv at Fall of Warness as there is in Admiralty Inlet.

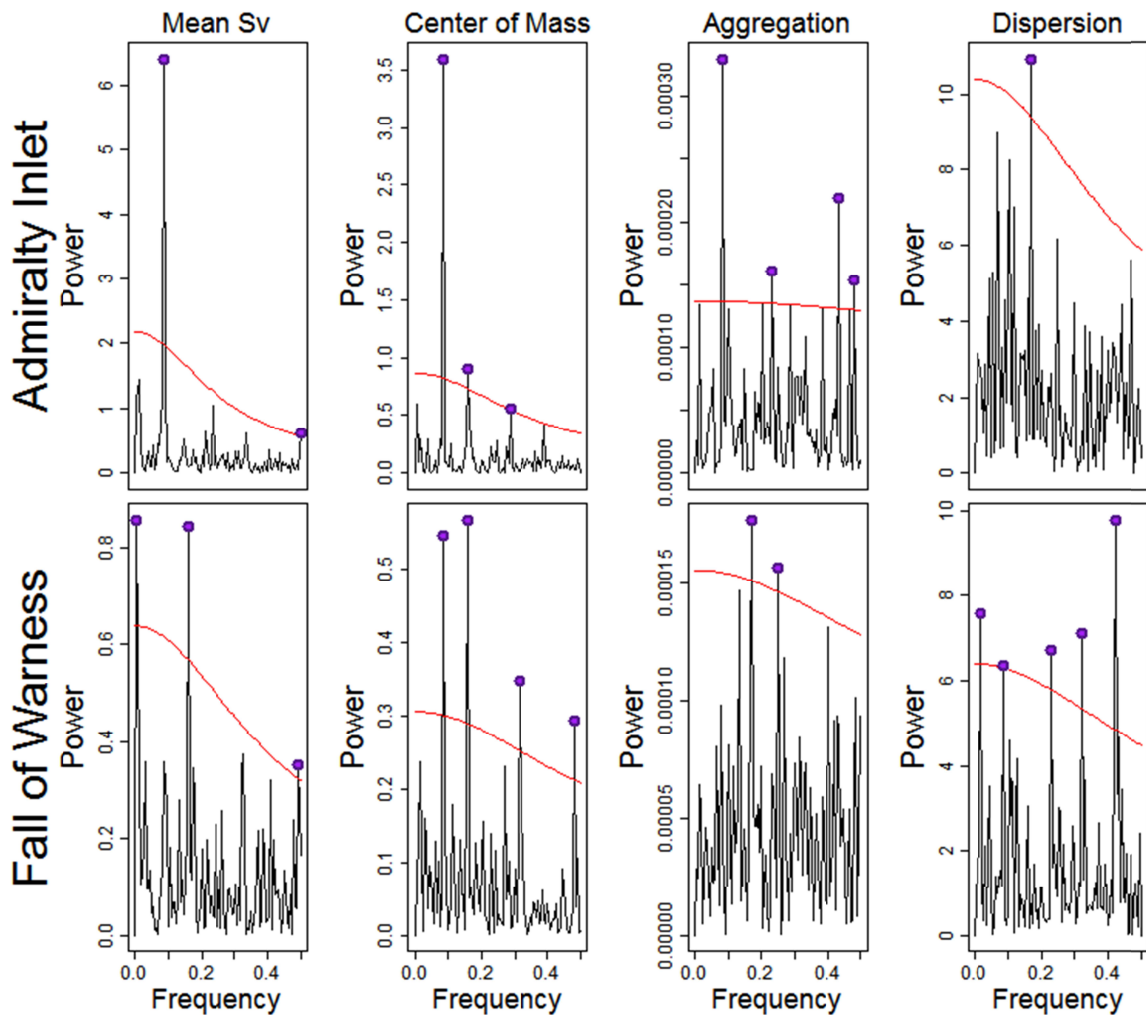


Fig. 3.7: Periodograms for the suite of metrics (mean Sv, center of mass, aggregation, dispersion) at both sites. The significant frequencies (purple dots) are above the red noise spectrum (red line).

The center of mass metrics from the two sites shared two significant frequencies (24 hour period and 12 hour period), and two similar periods (6 hours for Fall of Warness, 7 hours for Admiralty Inlet) (Table 3.3). The amplitude of the 24 hour frequency was greater at Admiralty Inlet (1.9) than at Fall of Warness (0.7), but the amplitudes for the 12 hour and 6 or 7 hour frequencies were similar (Table 3.3). Similarities in amplitudes and values of the significant frequencies at both sites, as well as the coherence between the center of mass patterns indicate that the center of mass series are very similar (0.923). The sites did not share any similar significant periods for aggregation index (Table 3.3) and the aggregation metrics at the two sites had the lowest coherence of all metric pairs (0.378). The dispersion also did not share any significant periods (Table 3.3), but the coherence for the dispersion metrics was higher (0.903) than for the aggregation index.

Table 3.3: Significant periods rounded to the nearest integer value (hour), with amplitudes in parentheses.

	Admiralty Inlet	Fall of Warness
Mean Sv	24 (2.5), 4 (0.8)	12 (0.9), 4 (0.6)
Center of Mass	24 (1.9), 12 (0.9), 7 (0.7)	12 (0.8), 24 (0.7), 6 (0.6), 4 (0.5)
Aggregation	24 (0.02), 4 (0.01), 9 (0.01)	12 (0.01), 8 (0.01)
Dispersion	12 (3.3)	4 (3.1), 135 (2.8), 6 (2.7), 9 (2.6), 24 (2.5)

Overall, the common significant frequency components were 24 hours, 12 hours, and 4 hours. These periods were used as covariates in linear regressions (Table 3.4). Amplitudes of significant frequency components had similar values between sites with the exception of the 24 hour frequency component. The greater amplitudes for the 24 hour frequency component in the mean Sv and center of mass periodograms at Admiralty Inlet indicate a greater dominance of diel processes at Admiralty Inlet than at Fall of Warness, which is supported by the pattern in hourly variability of these metrics (Fig. 3.5).

All linear regression models of the metrics, except for aggregation index, included the 24 and 12 hour periods (Table 3.4). Regression models for Admiralty Inlet and Fall of Warness mean Sv included almost the same covariates, with the exception that Admiralty Inlet model did not include the 4 hour period and tidal speed covariates. The mean Sv models for both sites were the only models that included Julian day. Besides the 24 and 12 hour periods, the center of mass model for Admiralty Inlet included tidal speed. The Fall of Warness included the 4 hour period. The aggregation model was the only model where the sites did not have any covariates in common (Table 3.4). The dispersion model was the same for Admiralty Inlet and Fall of Warness with the addition of tidal speed in

the Fall of Warness model. All model residuals formed a random pattern indicating good model fit. No VIF for any model covariates was above 5, indicating no severe multicollinearity.

Table 3.4: The covariates and corresponding p-values for linear regressions for the Admiralty Inlet and Fall of Warness. The stars indicate the strength of p-value ( $0 < *** < 0.001 < ** < 0.01 < * < 0.05$ ).

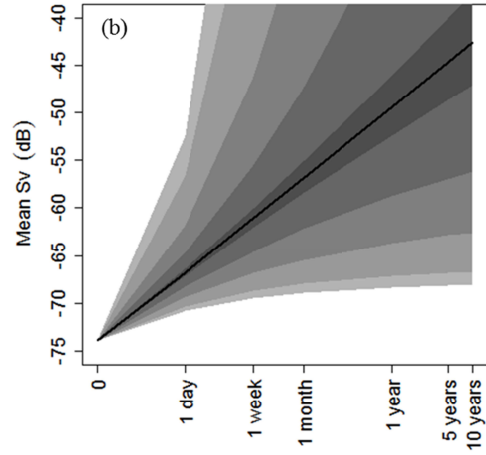
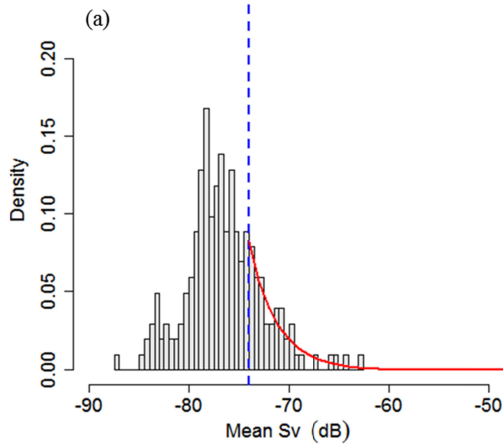
Covariate	Admiralty Inlet		Fall of Warness	
	estimate	p-value	estimate	p-value
<b>Mean Sv</b>				
Tidal speed	-	-	0.94398	0.000116 ***
Julian day	2.52E-01	1.31e-06 ***	-0.06301	0.063581
24 hour period	9.36E-01	3.37e-12 ***	1.04186	0.007225 **
12 hour period	1.053	0.051458	0.9324	0.000238 ***
4 hour period	-	-	1.04538	0.007193 **
<b>Center of Mass</b>				
Tidal speed	-2.5532	2.77e-07 ***	-	-
24 hour period	1.116	2e-16 ***	0.999	3.68e-05 ***
12 hour period	0.9933	9.07e-06 ***	0.9932	2.93e-05 ***
4 hour period	-	-	0.996	0.00233 **
<b>Aggregation</b>				
24 hour period	0.976273	3.89e-05 ***	-	-
12 hour period	-	-	0.999509	0.00413 **
4 hour period	0.976719	0.000734 ***	-	-
<b>Dispersion</b>				
Tidal speed	-	-	2.0279	0.020892 *
24 hour period	0.9753	0.03587 *	1.0065	0.002647 **
12 hour period	0.9786	0.00119 **	1.0965	0.023284 *
5 hour Period	0.976	0.04177 *	1.0125	0.000186 ***

### 3.3.3 EVA results comparison

The GPD fit did not differ greatly between sites for the mean Sv or aggregation index variables (Fig 3.8 and 3.9). Metric thresholds were similar between sites (Table 3.5), especially for the aggregation index metric. Scale parameters for both metrics were of the same order of magnitude, shape parameters differed between sites. The sums of squares between the GPD and the metric density function were used to evaluate the fit of the GPD. The mean Sv and aggregation index from Admiralty Inlet had a better GPD fit (lower sums of squares) than those of Fall of Warness (Table 3.5). It should be noted that there were small differences in the numbers of datapoints over the threshold for the

metrics between Admiralty Inlet and Fall of Warness (4 datapoints for mean Sv, 1 data point for aggregation) which may have affected the sums of squares.

### Admiralty Inlet



### Fall of Warness

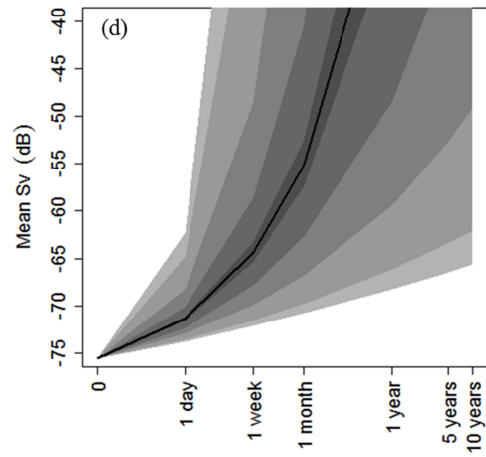
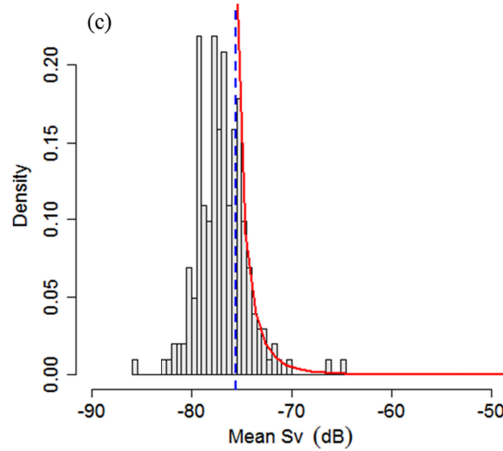
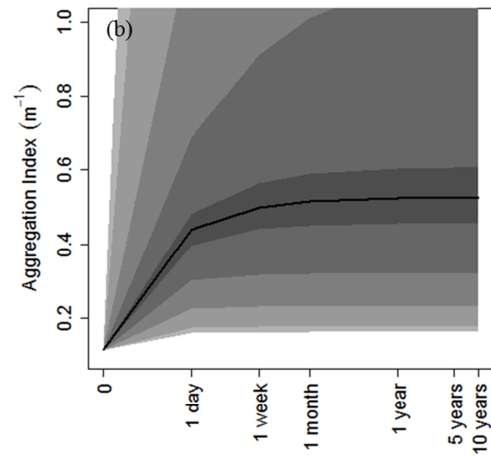
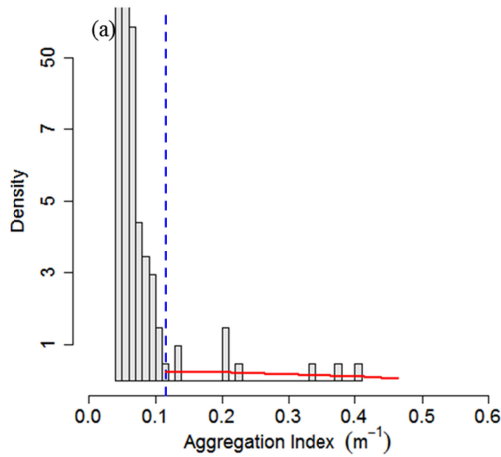


Fig. 3.8: Mean Sv (a) GPD fit and (b) return levels at Admiralty Inlet and (c) GPD fit and (d) return levels Fall of Warness. The solid lines is the best fit, and grey colors indicate credible intervals, from 10% (darkest grey) 40%, 80%, 90% lightest grey)

### Admiralty Inlet



### Fall of Warness

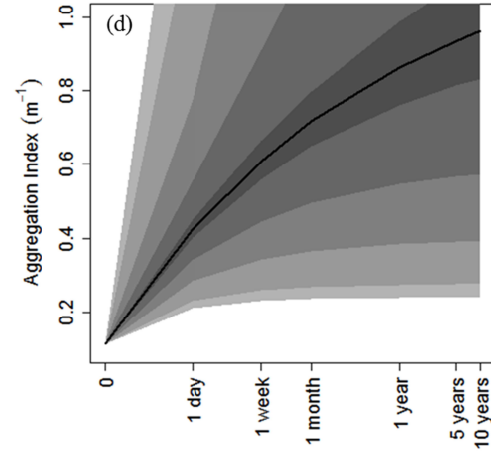
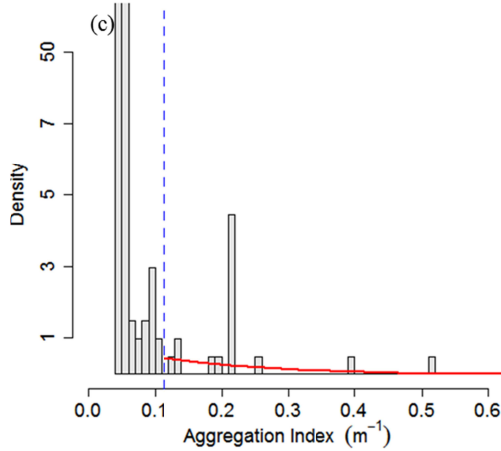


Fig. 3.9: Aggregation index (a) GPD fit and (b) return levels at Admiralty Inlet and (c) GPD fit and (d) return levels Fall of Warness. The solid lines is the best fit, and grey colors indicate credible intervals, from 10% (darkest grey) 40%, 80%, 90% lightest grey)

Table 3.5: Summary of Generalized Pareto Distribution fit for mean Sv and aggregation index metrics from both sites, with the median fits and 95% credible intervals (lower, upper) for scale and shape parameters.

	Mean Sv		Aggregation Index	
	AI	FoW	AI	FoW
Threshold	-74.0	-75.6	0.1148	0.1137
Scale	2.87 (1.80, 4.43)	1.08 (0.69, 1.64)	0.25 (0.08, 0.56)	0.15 (0.07, 0.29)
Shape	0.02 (-0.29, 0.47)	0.35 (0.05, 0.82)	-0.65 (-1.79, 0.94)	-0.13 (-0.61, 0.60)
Sums of Squares	0.000985	0.02576	0.715408	1.783951

Shapes of return level plots were similar for Mean Sv (Fig. 3.8) but not for aggregation (Fig. 3.9). The temporal difference between the data (2 weeks) and prediction (up to 10 years) explains why credible intervals spread quickly.

### 3.4 Discussion

#### 3.4.1 Tidal sites comparison

Patterns in pelagic nekton density and distribution at tidal energy sites have not been previously compared. Tidal energy sites are expected to have similar physical characteristics and are chosen because of high tidal fluxes. However, this does not dictate similarity in biological features. While there are some dissimilarities, many of the biological features at both sites are similar (Table 3.6). With the exception of aggregation, linear models of the metrics shared two or more covariates (Table 3.3). Density means and coherence values were surprisingly similar at both sites. Coherence between sites for center of mass and dispersion were similar, suggesting that temporal patterns in fish and zooplankton are in phase. Both sites had significant periodicities in all metrics that reflected tidal and diel processes (Table 3.6), but metrics at each site did not consistently identify the same process at each site. A longer dataset is necessary to examine effects of a lunar tidal cycle, and to parse lunar from diel effects.

Table 3.6: Comparison of Admiralty Inlet and Fall of Warness site ecosystem attributes, and coherence of metric pairs between sites.

	Structure		Function	
	Density	Center of Mass	Aggregation	Dispersion
Mean	AI = FoW	AI > FoW	AI < FoW	AI > FoW
Variance	AI > FoW	AI > FoW	AI = FoW	AI > FoW
Significant periodicities (hours)	same: 4 different: (AI) 24 (FoW) 12	same: 24, 12 different: (AI) 7 (FoW) 6, 4	different: (AI) 24, 9, 4 (FoW) 12, 8	different: (AI) 12 (FoW) 135, 24, 9, 6, 4
Coherence	0.997	0.923	0.378	0.903

Assessing the condition of an ecosystem is challenging. There are three primary attributes of an ecosystem: composition, structure, and function (Noss, 1990). Composition is the number and variety of elements in a system, structure is the physical organization of a system, and function includes ecological and evolutionary processes. Ecological indicators are measurable characteristics of these three attributes and changes in

indicators can be used to detect ecosystem change in response to disturbances (Niemi and McDonald, 2004). The four metrics used in this study can be used as indicators for structure (density, center of mass) and function (aggregation, dispersion). Composition is not an attribute that can be well-addressed using single-frequency, active acoustic data without trawls or another means to identify species. These metrics, while not encompassing the entire range of ecosystem attributes, present an assessment of the structure and function of the biology at the two study sites, and are appropriate indicators for comparison. This approach could be extended to include a comparison of physical metrics, such as tidal speed, or tidal range.

Species composition was not compared between sites as species were not identified during acoustic data collection. Broad taxonomical composition can be inferred from other studies. Historic studies of fauna around Fall of Warness are scarce, but some fish species that are likely to be present during the summer for spawning are mackerel (*Scromber scrombus*), herring (*Clupea harengus*), and sprat (*Sprattus sprattus*) (Aurora Environmental, 2005). Other fish that are likely to be in the vicinity are haddock (*Melanogrammus aeglefinus*), ling (*Molva molva*), saithe (*Pollachius limanda*), and cod (*Gadus morhua*). The Environmental Statement (2005) for the Fall of Warness also identifies non-commercially important butterfish (*Pholis gunnellus*) and scorpion fish (*Taurulus bubalis*). Long-term data from the Continuous Plankton Recorder show that zooplankton biomass in the North Sea fluctuates greatly on a yearly basis (Richardson and Schoeman, 2004; Pitois and Fox, 2006). Zooplankton are composed primarily of numerous copepod taxa which serve as the primary prey source for important commercial fish species such as herring (Frederiksen et al., 2006). In comparison, trawls were conducted during mobile acoustic surveys of the Admiralty Inlet site (Horne et al., 2013), consistently catching Pacific sand lance (*Ammodytes hexapterus*), northern lampfish (*Stenobranchius leucopsarus*), copper rockfish (*Sebastes caurinus*), and Pacific herring (*Clupea pallasii*). Broad zooplankton taxa are similar to those at Fall of Warness, including copepods, hydromedusae, and larval stages of fish and small pelagic crustaceans (Mackas et al., 2013). The relative abundances and distributions between the two sites are unknown. Fish at both sites (mackerel, sprat, and herring at Fall of Warness; Pacific sand lance, and Pacific herring at Admiralty Inlet) provide a prey base that support fish and apex predators in upper trophic levels (Harvey et al., 2012; Greene et al., 2015). Even though previous studies suggest that some zooplankton and fish taxa are similar between the two sites and provide similar ecosystem services, additional data on site-specific species abundance and distribution are necessary for a complete species characterization and comparison.

One primary difference between sites is that variance in the metrics was greater at Admiralty Inlet than at Fall of Warness, with the exception of the aggregation index. A

possible explanation is that water fluxes at Admiralty Inlet are more complex than at Fall of Warness. Admiralty Inlet is located near the entrance of Puget Sound at the confluence of Deception Pass, the Hood Canal Basin, and the Puget Sound main basin (Moore et al., 2008; Sutherland et al., 2011). These three water sources have different oceanographic properties (e.g. ocean water from the Strait of Juan de Fuca, fresh water from the Fraser River) and potentially carry different types of pelagic species, which may increase variability in the species composition between ebb and flood tides at Admiralty Inlet. Conversely, Fall of Warness is located on an open ocean coast, making water sources during ebb and flood tides uniform. Differences in location and physical properties of the two sites potentially explain the greater variability in metrics at Admiralty Inlet compared to Fall of Warness. An alternative explanation is that the sites' different tidal speeds (mean tidal speeds were significantly greater at Fall of Warness) could affect biomass distribution variability. Nekton motility is partially dependent on flow speed of the surrounding medium, and the ratio of nekton locometry velocity to fluid velocity increases with nekton body length (Schneider, 1994). Greater flow speeds may result in smaller nekton (especially micronekton < 5 mm (Schneider, 1994)) acting as passive particles, possibly causing metric patterns at Fall of Warness to be more uniform than at Admiralty Inlet. The significant positive relationships between tidal speed and both density and dispersion at Fall of Warness, which is not seen at Admiralty Inlet, support this hypothesis.

#### **3.4.2 Generality of EVA applicability**

This study allowed an examination of the proportion of data necessary for performing an EVA. As the proportion of data increases when determining a GPD threshold, the precision of a threshold estimate should increase because of the higher inclusion of extreme values for the threshold determination (Coles, 2001). As predicted, a higher proportion of data included in the analysis reduces the variability in the threshold estimate (i.e. threshold estimates from 10% of the data were most variable; threshold estimates from 90% of the data were least variable). However, over 500 random draws, the mean threshold estimate did not change significantly with increasing data proportion. Results from this study could be used to justify lower frequency data collection for monitoring as greater data inclusion does not change the mean threshold estimate. When sampling at low frequency (between 10 % and 70% of continuous data sampling) the standard deviation of the threshold estimate is greater than 0.1 dB over 500 draws, so the precision of the threshold estimate from a single draw is lower at lower sampling frequencies. If accuracy of the threshold estimate is a concern, then one should sample at 70% of continuous sampling or greater as the standard deviation in the threshold estimate is less than 0.1 dB over 500 draws.

While return levels and GPD threshold values differed between sites, the process of applying the peaks-over-threshold method was similar and successful in each case. As the EVA method had not been previously used for biological monitoring at MRE sites, it was important to determine whether the EVA method could be applied to another dataset. A similar proportion of data was fit to the GPD for both density and aggregation metrics at both sites. The MCMC diagnostics showed that convergence to a stationary distribution of GPD parameters was achieved at both sites for both metrics. The sums of squares results were of the same order of magnitude for both sites. The results lead to the conclusion that extreme value analysis produced comparable results at the two sites, and that this approach was an appropriate method to determine thresholds for these metrics. This study shows that EVA results may be site-specific (e.g. Admiralty Inlet mean Sv GPD threshold: -74.5 dB, Fall of Warness mean Sv GPD threshold: -75.6 dB), but that EVA can be applied as a generic biologically monitoring tool for tidal energy sites.

### **3.4.3 Standardizing MRE monitoring**

Because tidal energy is a relatively new technology, regulators are unsure of environmental impacts, and decisions on what to monitor have been largely made for site-specific concerns (e.g. Southern resident killer whales at Admiralty Inlet, sturgeon at Roosevelt Island). While regulators are unsure of generic monitoring targets, there is also uncertainty around how to monitor environmental variables. The three monitoring plans for current marine hydrokinetic energy in the US (Admiralty Inlet, Roosevelt Island, and Cobscook Bay) share broad objectives. For example, fish monitoring includes distribution, abundance, and diversity. Differences in the monitoring plans are concentrated in the choice of monitoring technologies and in the spatial and temporal scales of monitoring. Differences in monitoring methods may reflect differences in objectives but also reflect perceptions, preferences, and knowledge of developers proposing the monitoring plans. Results from monitoring of early tidal energy projects will be useful in identifying important spatiotemporal scales at which to monitor (Jacques, 2014), and the optimal sampling frequency and instrumentation to use when sampling. Standardization of tool and techniques will allow for streamlining project development, especially during a project application process, which currently is long and expensive (e.g. Verdant Power, 2010).

The goal of comparing results among tidal sites is one reason why standardization of data acquisition methods and analysis is so important to MRE monitoring. The primary justification for comparing the Fall of Warness data to the Admiralty Inlet data was that both datasets were collected with bottom-mounted echosounders. It was therefore relatively simple to subsample the Fall of Warness data to match the Admiralty Inlet sampling scheme. This comparison would not have been possible, and certainly would

have been less powerful, if data collection had not been so similar. Comparison among sites once tidal energy projects are operational will be crucial in determining whether there are generic impacts from tidal energy development. Comparisons will be difficult if monitoring methods differ among sites.

Site-specific monitoring plans are motivated by the idea that sites differ and need to have monitoring plans tailored to the biology of each site. This study suggests that not all biological characteristics of tidal energy sites are site-specific. While tidal energy is still in the developmental pilot project stage, standardization of monitoring goals and methods is a viable and necessary goal to facilitate project development and the detection of environmental impacts.

## Chapter 4:

# *SUMMARY AND SIGNIFICANCE*

The first objective of this thesis was to evaluate whether Extreme Value Analysis (EVA) could be used for impact characterization. The peaks-over-threshold (POT) method, where extreme values are defined as values above a threshold and then fit to a Generalized Pareto Distribution (GPD), was successfully applied to Admiralty Inlet mean Sv and aggregation metrics using a Bayesian MCMC simulation. A derivative method to objectively determine the GPD threshold was developed. The univariate POT was deemed more appropriate for this dataset than the bivariate POT. The MCMC simulation converged for both metrics and return level plots were generated using the resulting MCMC posteriors for the GPD parameter estimates values. Return level plots had wide confidence intervals for return level estimates, indicating uncertainty in predictions for periods many times larger than the length of the baseline dataset.

EVA has many applications for MRE biological monitoring, and defining extreme value thresholds for each site will help MRE regulators assess the risk of impacts, as well as establish a baseline for expected extreme value periodicity. Evaluating periodicities of extreme events will allow both site developers and regulators to predict the risk of damage to tidal technology or other MRE technologies, and the risk of MRE technologies impacting the environment. Techniques used in this study should decrease the time needed to develop environmental monitoring plans, reduce costs of permitting, and help identify relevant monitoring variables. Using EVA, regulators will be able to set variable or metric thresholds and model the frequency of rare values that may be associated with biological impacts. Regulators could use this approach and results from threshold analyses to compare conditions among tidal energy sites to set more generic monitoring guidelines. These methods are also applicable to other forms of environmental monitoring around most anthropogenic disturbances.

The second objective of the study was to compare the distribution and abundance of fish and macrozooplankton at two tidal energy sites. This comparison enabled evaluation of the similarities of biological characteristics, and to determine whether EVA could be used as a generic approach for biological monitoring at tidal energy sites. The two datasets were collected using active acoustics and were processed to maximize comparability. Metrics describing variability in vertical biomass distribution in the water column over space and through time were used to compare the two sites. Results show that sites had similar biological patterns. EVA was successfully applied to mean Sv and aggregation

index series at both sites. The main difference between the two sites was that the variance in metrics was greater at Admiralty Inlet than at Fall of Warness, which could be explained by physical differences between the sites. These analyses lead to the conclusions that tidal energy sites have similar biological characteristics, and that similar monitoring methods could be used at both sites.

The amount of data necessary to perform an EVA was also evaluated using Fall of Warness mean Sv data. While it is possible to obtain a threshold estimate using 10% of the available data, accuracy is increased as the amount of data used increased. Evaluating the amount of data needed could be used to address a problem frequently encountered when designing a monitoring plan, which is the temporal and spatial resolution at which to monitor. While the scope of the study can be determined by monitoring objectives and targets, sampling frequency and scope of study can also be determined by resource availability (e.g. financial, data storage, power capability) particularly for autonomous instrument packages. While collecting as much data as possible is a typical goal, there can be a point of diminishing return. A power analysis or autocorrelation study may show that increasing the resolution of data does not result in a corresponding increase in information. Collecting additional data may be costly and may be the only option in lieu of adequate sample design (e.g. Verdant Power). Determining the appropriate sample design to detect impacts is an important problem that is needed to optimize monitoring plan formulation and implementation.

This study was based on the analysis of acoustic backscatter data. No effort was made to proportion acoustic backscatter into species or functional groups. While species identification is possible, it requires direct sampling (e.g. trawls), which are constrained in high flow areas. Acoustic backscatter classification was not required for this study as it was tangential to the primary objectives, but without classification comments on community and composition are not possible. Using acoustic backscatter as a common currency is an advantage when comparing sites because it enabled direct comparison of temporal and spatial organism distribution patterns. As much of environmental regulation is species specific, a natural extension of this project would be to investigate whether thresholds from EVA could be applied to datasets that include taxonomical composition, and see if differences in thresholds are proportional to species abundances and residence times. This would aid proposing generic monitoring guidelines for more specific monitoring targets (e.g. special status marine mammal and fish species).

One primary issue that this study addressed was the definition and detection of biological impacts. While it is possible to define thresholds based on statistical significance, determining biological significance is more challenging. Specific knowledge and studies will be essential to characterize biological significance, and consensus on defining

biological impact, even in case-specific settings, is expected to be difficult to reach. It will be interesting to see how thresholds generated from EVA will be used to inform regulatory decisions, and if they will differ from those that are deemed biologically significant.

Investigating the generality of monitoring methods was another important component of this study. The case is made that standardized monitoring is advantageous for site comparisons. Site comparisons are important for determining general impacts of human activities. Environmental monitoring around human disturbances also requires observed changes to be attributed to one or more disturbances. This is the grand challenge of monitoring. Changes may be detected, but whether observed changes are caused by human activity or simply natural fluctuations is difficult to determine. Control sites can be used to address this issue, but this approach has been criticized due to the difficulty of finding appropriate control sites for a treatment site (Underwood, 1994). An alternative to using control sites would be to compare monitoring data from several impact sites (e.g. operational tidal energy projects) and to quantify change in a common set of monitoring variables to see if similar changes can be detected. To be effective, it is critical that consistent methodology be used at all sites. The comparison of biological characteristics at two sites suggests that standardization of methods is a realistic regulatory objective.

## References

- Agarwal, A., Venugopal, V., and Harrison, G.P. (2013). The assessment of extreme wave analysis methods applied to potential marine energy sites using numerical model data. *Renewable and Sustainable Energy Reviews* 27, 244–257.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika* 52, 317–332.
- Aurora Environmental (2005). Environmental Statement (Stromness, Orkney).
- Bartlett, M.S. (1947). The use of transformations. *Biometrics* 3, 39–52.
- Behrens, C.N., Lopes, H.F., and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling* 4, 227–244.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of extremes* (Chichester, England: John Wiley & Sons).
- Belsley, D.A., Kuh, E., and Welsch, R.E. (1980). Detecting and Assessing Collinearity. In *Regression Diagnostics*, (John Wiley & Sons, Inc.), pp. 85–191.
- Bijleveld, A.I., Van Gils, J.A., Van der Meer, J., Dekinga, A., Kraan, C., Van der Veer, H.W., and Piersma, T. (2012). Designing a benthic monitoring programme with multiple conflicting objectives. *Methods in Ecology and Evolution* 3, 526–536.
- Boehlert, G.W., and Gill, A.B. (2010). Environmental and Ecological Effects of Ocean Renewable Energy Development: A Current Synthesis. *Oceanography* 23, 68–81.
- Box, G.E., and Jenkins, G.M. (1976). *Time series analysis: forecasting and control* (Holden-Day).
- Burgos, J.M., and Horne, J.K. (2007). Sensitivity analysis and parameter selection for detecting aggregations in acoustic data. *ICES J. Mar. Sci.* 64, 160–168.
- Cada, G., Ahlgrimm, J., Bahleda, M., Bigford, T., Stavrakas, S.D., Hall, D., Moursund, R., and Sale, M. (2007). Potential impacts of hydrokinetic and wave energy conversion technologies on aquatic environments. *Fisheries* 32, 174–181.
- Chatfield, C. (1989). *The Analysis of Time Series: An Introduction* (Chapman and Hall).
- Coetzee, J. (2000). Use of a shoal analysis and patch estimation system (SHAPES) to characterise sardine schools. *Aquatic Living Resources* 13, 1–10.

- Coles, S. (2001). *An introduction to statistical modeling of extreme values* (London: Springer).
- Copping, A., Battey, H., Brown-Saracino, J., Massaua, M., and Smith, C. (2014). An international assessment of the environmental effects of marine energy development. *Ocean & Coastal Management* 99, 3–13.
- Copping, A., Hanna, L., Whiting, J., Geerlofs, S., Grear, M., Blake, K., Coffey, A., Massaua, M., Brown-Saracino, J., and Battey, H. (2013). *Environmental Effects of Marine Energy Development Around the World. Annex Iv Final Report* (Richland, WA: Pacific Northwest National Laboratory).
- Copping, A.E., and Geerlofs, S.H. (2011). The Contribution of Environmental Siting and Permitting Requirements to the Cost of Energy for Marine and Hydrokinetic Devices. PNNL--20963 21.
- Crome, F.H.J., Thomas, M.R., and Moore, L.A. (1996). A Novel Bayesian Approach to Assessing Impacts of Rain Forest Logging. *Ecological Applications* 6, 1104–1123.
- Doelle, M. (2009). Role of Strategic Environmental Assessments in Energy Governance: A Case Study of Tidal Energy in Nova Scotia's Bay of Fundy. *J. Energy & Nat. Resources L.* 27, 112.
- Duinker, P.N., Burbidge, E.L., Boardley, S.R., and Greig, L.A. (2013). Scientific dimensions of cumulative effects assessment: toward improvements in guidance for practice. *Environmental Reviews* 21, 40–52.
- Dupuis, D.J. (1999). Exceedances over high thresholds: A guide to threshold selection. *Extremes* 1, 251–261.
- Federal Energy Regulatory Commission (2004). *Handbook for Hydroelectric Project Licensing and 5 MW Exemptions from Licensing* (Federal Energy Regulatory Commission).
- Federal Energy Regulatory Commission (2008). *Licensing Hydrokinetic Pilot Projects*.
- Frederiksen, M., Edwards, M., Richardson, A.J., Halliday, N.C., and Wanless, S. (2006). From plankton to top predators: bottom-up control of a marine food web across four trophic levels. *Journal of Animal Ecology* 75, 1259–1268.
- Frid, C., Andonegi, E., Depestele, J., Judd, A., Rihan, D., Rogers, S.I., and Kenchington, E. (2012). The environmental interactions of tidal and wave energy generation devices. *Environmental Impact Assessment Review* 32, 133–139.

- Garthwaite, P.H., and O'Hagan, A. (2000). Quantifying Expert Opinion in the UK Water Industry: An Experimental Study. *Journal of the Royal Statistical Society: Series D (The Statistician)* 49, 455–477.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013). *Bayesian Data Analysis, Third Edition* (CRC Press).
- Gelman, A., and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 7, 457–472.
- Germano, J.D. (2000). Ecology, statistics, and the art of misdiagnosis: The need for a paradigm shift. *Environ. Rev.* 7, 167–190.
- Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In *Bayesian Statistics*, (University Press), pp. 169–193.
- Gray, T., Haggett, C., and Bell, D. (2005). Offshore wind farms and commercial fisheries in the UK: A study in Stakeholder Consultation. *Ethics, Place & Environment: A Journal of Philosophy & Geography* 8, 127–140.
- Greene, C., Kuehne, L., Rice, C., Fresh, K., and Penttila, D. (2015). Forty years of change in forage fish and jellyfish abundance across greater Puget Sound, Washington (USA): anthropogenic and climate associations. *Marine Ecology Progress Series* 525, 153–170.
- Handegard, N.O., Buisson, L. du, Brehmer, P., Chalmers, S.J., De Robertis, A., Huse, G., Kloser, R., Macaulay, G., Maury, O., Ressler, P.H., et al. (2013). Towards an acoustic-based coupled observation and modelling system for monitoring and predicting ecosystem dynamics of the open ocean. *Fish and Fisheries* 14, 605–615.
- Harvey, C.J., Williams, G.D., and Levin, P.S. (2012). Food web structure and trophic control in central Puget Sound. *Estuaries and Coasts* 35, 821–838.
- Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- Hewitt, J.E., Thrush, S.E., and Cummings, V.J. (2001). Assessing environmental impacts: effects of spatial and temporal variability at likely impact scales. *Ecological Applications* 11, 1502–1516.
- Hobbs, N.T., and Hilborn, R. (2006). Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. *Ecological Applications* 16, 5–19.

- Horne, J.K. (2000). Acoustic approaches to remote species identification: a review. *Fisheries Oceanography* 9, 356–371.
- Horne, J.K., Jacques, D.A., Parker-Stetter, S.L., Linder, H.L., and Nomura, J.M. (2013). Evaluating acoustic technologies to monitor aquatic organisms at renewable energy sites (U.S. Department of the Interior, Bureau of Ocean Energy Management).
- Inger, R., Attrill, M.J., Bearhop, S., Broderick, A.C., James Grecian, W., Hodgson, D.J., Mills, C., Sheehan, E., Votier, S.C., Witt, M.J., et al. (2009). Marine renewable energy: potential benefits to biodiversity? An urgent call for research. *Journal of Applied Ecology* 46, 1145–1153.
- Jacques, D. (2014). Describing and Comparing Variability of Fish and Macrozooplankton Density at Marine Hydrokinetic Energy Sites. University of Washington.
- Jonathan, P., and Ewans, K. (2013). Statistical modelling of extreme ocean environments for marine design: A review. *Ocean Engineering* 62, 91–109.
- Jones, R.N. (2001). An environmental risk assessment/management framework for climate change impact assessments. *Natural Hazards* 23, 197–230.
- Lovell, D.P. (2013). Biological importance and statistical significance. *J. Agric. Food Chem.* 61, 8340–8348.
- Lundquist, C., Thrush, S., Coco, G., and Hewitt, J. (2010). Interactions between disturbance and dispersal reduce persistence thresholds in a benthic community. *Marine Ecology Progress Series* 413, 217–228.
- Mackas, D., Galbraith, M., Faust, D., Masson, D., Young, K., Shaw, W., Romaine, S., Trudel, M., Dower, J., Campbell, R., et al. (2013). Zooplankton time series from the Strait of Georgia: Results from year-round sampling at deep water locations, 1990–2010. *Progress in Oceanography* 115, 129–159.
- Maclennan, D.N., Fernandes, P.G., and Dalen, J. (2002). A consistent approach to definitions and symbols in fisheries acoustics. *ICES Journal of Marine Science: Journal Du Conseil* 59, 365–369.
- Mapstone, B.D. (1995). Scalable Decision Rules for Environmental Impact Studies: Effect Size, Type I, and Type II Errors. *Ecological Applications* 5, 401.
- Martin, T.G., Kuhnert, P.M., Mengersen, K., and Possingham, H.P. (2005). The power of expert opinion in ecological models using bayesian methods: impact of grazing on birds. *Ecological Applications* 15, 266–280.

- Martínez-Abraín, A. (2008). Statistical significance and biological relevance: A call for a more cautious interpretation of results in ecology. *Acta Oecologica* 34, 9–11.
- Massey Jr, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association* 46, 68–78.
- Mazas, F., and Hamm, L. (2011). A multi-distribution approach to POT methods for determining extreme wave heights. *Coastal Engineering* 58, 385–394.
- Moore, S.K., Mantua, N.J., Newton, J.A., Kawase, M., Warner, M.J., and Kellogg, J.P. (2008). A descriptive analysis of temporal and spatial patterns of variability in Puget Sound oceanographic properties. *Estuarine, Coastal and Shelf Science* 80, 545–554.
- National Ocean Service (2011). Tide Tables 2011 – West Coast of North and South America including the Hawaiian Islands.
- National Ocean Service (2013). Tide Tables 2013 – Europe and West Coast of Africa including the Mediterranean Sea.
- Nerzic, R., Frelin, C., Prevosto, M., and Quiniou-Ramus, V. (2007). Joint Distributions of Wind/Waves/Current In West Africa And Derivation of Multivariate Extreme I-FORM Contours. (International Society of Offshore and Polar Engineers),.
- Niemi, G.J., and McDonald, M.E. (2004). Application of ecological indicators. *Annual Review of Ecology, Evolution, and Systematics* 35, 89–111.
- Northrop, P.J., and Jonathan, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics* 22, 799–809.
- Noss, R.F. (1990). Indicators for monitoring biodiversity: a hierarchical approach. *Conservation Biology* 4, 355–364.
- Nunnallee, E.P. (1990). An alternative to thresholding during echo-integration data collection. *Rapports Et Procès-verbaux Des Réunions / Conseil Permanent International Pour L'exploration De La Mer* 189, 92–94.
- Ocean Renewable Power Company (2011). Comments on the Draft Pilot Project License Application Cobscook Bay Tidal Energy Project.
- Osenberg, C.W., Schmitt, R.J., Holbrook, S.J., Abu-Saba, K.E., and Flegal, A.R. (1994). Detection of Environmental Impacts: Natural Variability, Effect Size, and Power Analysis. *Ecological Applications* 4, 16–30.

- Pelc, R., and Fujita, R.M. (2002). Renewable energy from the ocean. *Marine Policy* 26, 471–479.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* 119–131.
- Pitois, S.G., and Fox, C.J. (2006). Long-term changes in zooplankton biomass concentration and mean size over the Northwest European shelf inferred from Continuous Plankton Recorder data. *ICES J. Mar. Sci.* 63, 785–798.
- Polagye, B., Van Cleve, B., Copping, A., and Kirkendall, K. (2011). Environmental Effects of Tidal Energy Development. NOAA Technical Memorandum NMFS F/SPO-116 186.
- Public Utility District No. 1 of Snohomish County (2012). Admiralty Inlet Tidal Project Final Monitoring and Mitigation Plans.
- Richardson, A.J., and Schoeman, D.S. (2004). Climate impact on plankton ecosystems in the Northeast Atlantic. *Science* 305, 1609–1612.
- Scarrott, C., and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal* 10, 33–60.
- Schmitt, R.J., and Osenberg, C.W. (1996). *Detecting Ecological Impacts: Concepts and Applications in Coastal Habitats* (Academic Press).
- Schneider, D.C. (1994). *Scale-dependent patterns and species interactions in marine nekton* (Blackwell Scientific: Oxford).
- Shumway, D.R.H., and Stoffer, P.D.S. (2011). Spectral Analysis and Filtering. In *Time Series Analysis and Its Applications*, (Springer New York), pp. 173–265.
- Simmonds, J., and MacLennan, D. (2005). *Fisheries acoustics: Theory and practice* (Oxford, UK: Blackwell Science Ltd).
- Stephenson, A.G. (2002). evd: Extreme Value Distributions. *R News* 2, 31–32.
- Stewart-Oaten, A., and Bence, J.R. (2001). Temporal and spatial variation in environmental impact assessment. *Ecological Monographs* 71, 305–339.
- Sutherland, D.A., MacCready, P., Banas, N.S., and Smedstad, L.F. (2011). A model study of the Salish Sea estuarine circulation. *Journal of Physical Oceanography* 41, 1125–1143.
- Thompson, P., Cai, Y., Reeve, D., and Stander, J. (2009). Automated threshold selection methods for extreme wave analysis. *Coastal Engineering* 56, 1013–1021.

Underwood, A.J. (1994). On Beyond BACI: Sampling Designs that Might Reliably Detect Environmental Disturbances. *Ecological Applications* 4, 3–15.

Underwood, A.J. (1996). Detection, interpretation, prediction and management of environmental disturbances: some roles for experimental marine ecology. *Journal of Experimental Marine Biology and Ecology* 200, 1–27.

Urmy, S.S., Horne, J.K., and Barbee, D.H. (2012). Measuring the vertical distributional variability of pelagic fauna in Monterey Bay. *ICES Journal of Marine Science* 69, 184–196.

Verdant Power (2010). Final Kinetic Hydropower Pilot License Application - Volume 4.

Viehman, H. (2012). Fish in a tidall dynamic region in Maine: hydroacoustic assessments in relation to tidal power development. University of Maine.

Wand, M.P., and Jones, M.C. (1994). Kernel Smoothing (CRC Press).

Watkins, J.L., and Brierley, A.S. (1996). A post-processing technique to remove background noise from echo integration data. *ICES Journal of Marine Science: Journal Du Conseil* 53, 339–344.

Wilkie, D.S., and Carpenter, J.F. (1999). Bushmeat hunting in the Congo Basin: an assessment of impacts and options for mitigation. *Biodiversity and Conservation* 8, 927–955.

Williamson, B.J., Blondel, P., Armstrong, E., Bell, P.S., Hall, C., Waggitt, J.J., and Scott, B.E. (2015). A self-contained subsea platform for acoustic monitoring of the environment around marine renewable energy devices—field deployments at wave and tidal energy sites in Orkney, Scotland.