The Challenges of Analyzing Behavioral Response Study Data: An Overview of the MOCHA

(Multi-study OCean acoustics Human effects Analysis) Project.

Catriona M. Harris[1], Len Thomas[2], Dina Sadykova[3], Stacy L. DeRuiter[4], Peter L. Tyack[5], Brandon L. Southall[6], Andrew J. Read[7], Patrick J.O. Miller[8]

1 Centre for Research into Ecological and Environmental Modelling, University of St Andrews. E.mail – catriona@mcs.st-and.ac.uk

2 Centre for Research into Ecological and Environmental Modelling, University of St Andrews. E.mail – len@mcs.st-and.ac.uk

3 Centre for Research into Ecological and Environmental Modelling, University of St Andrews. E.mail – dina@mcs.st-and.ac.uk

4 Centre for Research into Ecological and Environmental Modelling, University of St Andrews. E.mail – sldr@st-andrews.ac.uk

5 Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews. E.mail – plt@st-andrews.ac.uk

6 Southall Environmental Associates, Inc. E.mail - Brandon.Southall@sea-inc.net

7 Duke Marine Lab. E.mail – aread@duke.edu

8 Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews. E.mail – pm29@st-andrews.ac.uk

Corresponding author: Catriona M Harris, Centre for Research into Ecological and Environmental Modelling, The Observatory, Buchanan Gardens, University of St Andrews, St Andrews, KY16 9LZ, UK.

Key words: sonar; cetaceans; change-point; dose-response; state-space modelling.

Abstract

This paper describes the MOCHA project which aims to develop novel approaches for the analysis of data collected during Behavioral Response Studies (BRSs). BRSs are experiments aimed at directly quantifying the effects of controlled dosages of natural or anthropogenic stimuli (typically sound) on marine mammal behavior. These experiments typically result in low sample size, relative to variability, and so we are looking at a number of studies in combination to maximize the gain from each one. We describe a suite of analytical tools applied to BRS data on beaked whales, including a simulation study aimed at informing future experimental design.

1 Introduction

Many marine mammals rely on sound for foraging, maintaining group cohesion, navigation, finding mates and avoiding predators. Hence, they may be profoundly affected by the introduction of anthropogenic noise into the marine environment. Examples of potentially harmful noise sources include vessel noise and active acoustic devices such as naval sonar or airguns used for seismic prospecting. Potential adverse effects range from masking, i.e., reducing the range at which an animal can detect important natural signals, to longer-term behavioral disturbance, which may exclude animals from important habitat, to stranding. There is, therefore, the need to quantify the impact of noise at the level of individual animal responses and also in terms of potential population consequences.

BRSs (also sometimes called controlled exposure experiments (CEEs)) are an important approach for studying the short-term responses of animals to specific doses of potential stressors. In a BRS, focal animals are selected based upon our ability to attach tags to them and their behavior monitored using visual observations, passive acoustics, animal-borne tags or a combination of these. A sample of animals is exposed to a stimulus, such as a potentially disturbing sound, and their response monitored. Various experimental protocols have been employed, including escalating the dose until a response is observed; exposing the same animal multiple times using various stimuli; varying the context of the experiment (such as the animal's behavior before the experiment or the path of the source vessel). Care is taken to

ensure the experimental animals are not stressed more than is necessary for the experimental results, for example by ceasing the stimulus as soon as a response is detected. Various measurements are recorded before, during and after exposure, including location (in 3D) through time, acoustic behavior, orientation and movement recorded continuously by tags and behavior observed at the surface.

BRS experiments are extremely costly to undertake: a single field season may cost many hundreds of thousands of dollars. Many species of interest occur at low density, or are hard to monitor (for example if they are visually cryptic and difficult to tag), and experiments can only take place in good weather conditions and when other interfering noise sources are not present. Because of these factors, the sample sizes associated with BRS experiments are usually very low: typically fewer than ten exposures per species per field season, and sometimes substantially fewer. Despite the difficulties, BRSs are essential components of risk assessment because they offer a direct measure of the effect of sound on behavior from animals selected at random (in terms of their sensitivity to sonar). .

1.1 Data summary

Here we focus primarily on studies funded by the US Navy to investigate the effect of mid-frequency active sonar. The first such BRS was conducted in 2007 and 2008 in the Bahamas (Tyack et al., 2011) with the aim of collecting baseline data on animal behavior, and conducting CEEs to measure responses to different sound stimuli. A major focus of the Bahamas work was beaked whales, which are thought to be some of the species most vulnerable to sounds from sonar exercises as they make up the majority of animals stranded in association with naval sonar exercises (D'Amico et al, 2009). A second project, 3S (for "Sea mammals, Sonar, Safety"), which aimed to determine behavioral responses of killer, sperm and pilot whales to different sonar frequencies, commenced in 2006 with funding from The Royal Norwegian and Dutch Navies, and in 2008 additional US-Naval support allowed a significant expansion of the effort (Miller et al, 2011, 2012). Field work was conducted in Norwegian waters. A follow-up project (3S2, 2011-2013) is focusing on three different species (humpback, minke and northern bottlenose whales) and is also investigating the effectiveness of ramp-up as a mitigation strategy (Kvadsheim et al.,

2011, 2012). SOCAL-BRS is a five-year (2010-2015) study in southern Californian waters, which aims to expand on the work conducted in the Bahamas by increasing the number of species studied and by integrating with other field efforts being conducted in the same region (Southall et al., 2012). Finally, Duke University is carrying out CEEs on odontocete species off Cape Hatteras, California and Hawaii, to investigate how species-specific social structures may modulate responses to the sounds of potential predators.

All of these BRSs have relied on a suite of existing data collection technologies and methodologies that have enabled collection of a wide range of behavioral metrics. Observer-based metrics are derived from visual protocols for the collection of data on position, orientation, social configuration, surfacing behavior and swim speed of individuals and groups before, during and after a CEE (e.g., Visser et al., 2011). These data are difficult to collect for the more cryptic and deep-diving species, but animal-borne digital acoustic recording tags (DTAGs, Johnson & Tyack, 2003) have proven invaluable for this purpose. These tags are attached to cetaceans using suction cups and collect acoustic and movement data for the duration of attachment. Their deployment on unexposed animals has provided valuable baseline data on diving and foraging behavior (e.g. Tyack et al., 2006), while their use during BRS experiments has allowed the behavior of exposed individuals to be tracked in fine detail before, during and after exposure. The acoustic sensor on these tags can also provide direct measurements of received sound levels.

Whilst there has been considerable consistency in the technologies employed across all the BRSs there have also been a range of important differences that need to be understood when collating and comparing data across studies. The most obvious difference is geographical location, resulting in differences in bathymetry, distance to coastline, sound propagation properties and prey availability (i.e., motivation to remain in area). In addition, the species that have been used in CEEs have differed both within and across studies. Finally the characteristics and context of each CEE differ in terms of the behavioral context of the subjects, sound types, sound frequencies, exposure durations, position/direction/movement of source relative to animal and distance from source to animal.

1.2 Analytical challenges and the aims of MOCHA

Many diverse analyses have been applied to data from each BRS, and we do not attempt to review them all here (although see Tyack et al., 2011; Miller et al., 2012; Curé et al., 2012, DeRuiter et al., 2013 for examples). Typically the first stage has been to synthesize the observational data for each individual to determine whether it responded, gauge the magnitude of any response and relate the onset of response to exposure intensity ("dose"). Secondly, results from individual exposures may be analyzed together to estimate a (context-specific) exposure-response function. Both stages are fraught with difficulties. At the first stage, it is not clear whether or how to combine the diverse input metrics (such as position, orientation, swim speed, dive time, social configuration and surfacing behavior) all measured through time, into a single measure of response. At both stages, analyses need to account for dependencies in the data, because many measurements are taken in close succession on the same subjects. The same subjects may also be exposed multiple times. While there are many measurements on each subject, there are few exposures in total, and often even fewer individual subject animals. In such circumstances, analysis approaches based on traditional statistical hypothesis testing and estimation have restrictive assumptions and their statistical power is often very low. This has the undesirable effect that only studies on species showing the clearest responses are likely to provide statistically significant results that are easily publishable in top academic journals, where few studies are published that show "no effect" (Dwan et al., 2008). This can lead to a potential bias in published effect levels, such that some species are thought to be more sensitive on average than is actually the case and others are assumed not to respond at all.

Despite the complications, we believe that substantial progress can be made to enhance the inferences drawn from existing and future BRS data. To this end, the MOCHA project (Multi-study OCean acoustics Human effects Analysis, running from 2012 to 2015) brings together researchers undertaking BRS studies and statisticians specializing in the analysis of biological experiments of this kind in a working group format, with dedicated researcher effort between group meetings. We are primarily working on the following four areas of research:

1. Improved methods for combining diverse behavioral measures into a response metric. Consideration is being given to obtaining metrics that can be linked to biological consequences;

2. Better estimates of exposure-response functions from individual studies through the use of cutting edge (e.g. hierarchical Bayesian) statistical techniques;

3. Combine information across studies and species ("meta-analysis"), making use of expert biological opinion, to maximize the gain for each study and to make inference about taxa and contexts not yet studied. Differences in methods/protocols between studies will need to be accounted for. A component of this objective is to quantify the similarity/dissimilarity between species, placed into functional groups.

4. Sensitivity analyses to quantify where future experimental effort will most reduce uncertainty.

Effort to date has focused on deep diving odontocetes (beaked whales and sperm whales) and other odontocete species (including killer whales, Risso's dolphins and pilot whales). Our penultimate working group meeting in early 2014 will focus on baleen whales. Reports from each working group meeting are available on the project website (http://www.creem.st-and.ac.uk/mocha/).

2 Case study - analysis of BRS data from beaked whales.

Over the course of the project a number of different methodologies have been investigated, developed and applied to a variety of BRS datasets. Here we describe a suite of analyses that have been applied to one dataset to provide examples of methods that can be applied in the context of BRSs. In both 2010 and 2011 the SOCAL BRS team successfully tagged (with DTAGs) a *Ziphius cavirostris* and carried out a CEE with mid-frequency active sonar on each animal at a distance of <10km, with received levels ranging from 84-144dB re 1μPa root-mean-squared (rms). The whale tagged in 2011 was additionally incidentally exposed to MFA from a distant naval exercise (approximately 118km away) with received levels of 78-106dB re 1μPa rms. Full details of the CEEs are given by Southall et al. (2012) and an analysis of some observed responses is given by DeRuiter et al. (2013). As is the case with all CEEs that involve tagged whales, the resulting dataset included a large number of measured and derived variables.

All the data were examined by experts and changes in behavior were identified and scored according to the Southall et al. (2007) response severity scale. Miller et al. (2012) describe this as expert scoring and outline their scoring method in detail as applied to the 3S data. The ideal scenario, however, is to use quantitative methods to identify points in time where behaviour differs from baseline ("change-points"). Here we give an example of a method that can be used to identify change-points across multivariate space (e.g., DTAG dive parameters).

Mahalanobis distance is a scale-invariant measure of distance in multi-dimensional space, which takes into account correlations between parameters. It has allowed us to summarize the multivariate time-series of dive parameters as a univariate time-series, with one value per dive. Each dive in the time-series was classified as either a deep foraging dive or a shallow dive, and the Mahalanobis distance was calculated relative to the average baseline dive of the corresponding type. The time-series of Mahalanobis distances for the two exposed *Ziphius* (Figure 1) show large peaks in the distance metric around the time of the CEE exposures, corresponding to the strong avoidance response identified by expert scoring (DeRuiter et al. 2013). In contrast, the incidental naval MFA exposure did not elicit a similar response. The Mahalanobis distance was then used as a measure of response intensity and was modelled as a function of RL, source-whale range, and time since sonar exposure (DeRuiter et al. 2013) (model fit shown in Figure 1); results indicated that both RL and source-whale range influenced the strength of response.

The potentially confounding relationship between RL and range described by De Ruiter et al. (2013) posed an interesting question about experimental design. Across all BRSs the source-whale range for CEEs has been relatively low (generally less than 10km); however, the incidental exposure that occurred at a distance of 118km during the SOCAL BRS in 2011 suggests that response to similar RLs may differ depending on range. We therefore conducted a simulation study to investigate how many more CEEs would need to be carried out under different experimental designs to allow reliable estimation of range and RL parameters in the aforementioned response intensity model. The simulation consisted of 13 control whales and between 4 and 12 exposed whales. Three different experimental scenarios were

trialled (all source levels are given in dB re 1 μPa-m): 1) source level at 210dB and range randomly selected from 5, 10, 20, 40 or 80km; 2) source level at 235dB and range randomly selected as before; and 3) source level assigned as 210dB to half of the whales and 235dB to the other half and range randomly selected as before. Mahalanobis distances were simulated from the model described above, using the maximum likelihood parameter estimates, and then the model was fitted to the simulated data in exactly the same way as it was fitted to the observed data. We found that in scenarios 1-2, at least 6-10 exposure whales would be required to allow reliable parameter estimation, whereas varying both the source level and range (scenario 3) reduced the uncertainty considerably after only 4 whales were exposed (Figure 2). Differences are also apparent between scenarios 1 and 2 whereby uncertainty is lower when source level is higher (Figure2). This difference probably arises because higher source levels generate higher received levels and more intense simulated responses, which facilitates parameter estimation. This result demonstrates the utility of simulation studies in exploring the potential outcomes of different experimental designs.

Finally, the scored responses for these two individuals have been combined with expert scored responses from other studies to estimate context-specific exposure-response functions. We developed and fitted a Bayesian hierarchical model to expert-scored behavioral responses and associated RLs for 35 exposure events, including four sonar signal types and 17 cetaceans of five species. We have investigated two possible approaches which help us address slightly different questions – fitting species/species groups as a fixed effect and fitting species/species groups as a random effect. When fitted as a fixed effect we can make inference about the individual species or species group, whilst fitting as a random effect allows us to make inference about unstudied species. Thus far we have been investigating the sensitivity of the methodology to different model parameterizations and the potential for model selection methods to identify species groupings, which may help with the knowledge base for legislation.

By describing this case study we have given a brief overview of a subset of the analysis methods that have been investigated by the MOCHA team to date. We outlined above the use of Mahalanobis distance as an aid to detect change-points on a dive-by-dive basis but we are also looking at its application to fine-scale time-series data. In addition we have investigated the application of a number of quantitative univariate change-point detection methods (e.g. inflection point detection, broken stick regression, wavelet analysis) to single metrics and to the time-series of Mahalanobis distances. We have also made progress in the application of process-based time series models such as hidden Markov models (HMMs) and semi-Markov models (SMMs) to BRS data. Like Mahalanobis distance, they allow multiple metrics to be combined into one analysis, and also address the time-series nature of the data, and provide an opportunity to explore behavioral states and the probability of transitioning between these states as a function of sound exposure.

3 Concluding remarks

BRSs produce a large and complex suite of data streams, many of which could contain a signal indicating a response to sound exposure. Moreover these data relate to relatively small numbers of exposures and even smaller numbers of individuals. Therefore challenges include condensing these data into manageable and interpretable metrics for analysis, conducting analysis appropriate for small sample sizes and finally placing the results into context. The combined efforts of the individual BRS project teams and the MOCHA project are producing useful insights into a range of methods for all stages of analysis and allowing a suite of analytical tools to be developed and/or adapted for application to these data.

Acknowledgements

References

Curé C, Antunes R, Samarra F et al. (2012) Pilot whales attracted to killer whale sounds: Acoustically-mediated interspecific interactions in cetaceans. PLoS ONE 7(12): e52201.

D'Amico AD, Gisiner R, Ketten DR et al. (2009) Beaked whale strandings and naval exercises. Aquat Mamm 35:452-472.

DeRuiter SL, Southall BL, Calambokidis J et al (2013) First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active (MFA) sonar. Biol Lett 9: 20130223.

Dwan K, Altman DG, Arnaiz JA et al (2008) Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS ONE 3(8): e3081.

Johnson MP, Tyack PL (2003) A digital acoustic recording tag for measuring the response of wild marine mammals to sound. IEEE J Ocean Eng 28:3-12.

Kvadsheim P, Lam FP, Miller PJ et al (2011) Behavioural response studies of cetaceans to naval sonar signals in Norwegian waters -3S-2011 cruise report. Available via http://rapporter.ffi.no/rapporter/2011/01289.pdf. Accessed 18 June 2013

Kvadsheim P, Lam FP, Miller PJ et al (2012) Behavioural response studies of cetaceans to naval sonar signals in Norwegian waters -3S-2012 cruise report. Available via http://rapporter.ffi.no/rapporter/2012/02058.pdf. Accessed 18 June 2013

Miller PJ, Antunes R, Alves AC et al (2011) The 3S experiments: Studying the behavioural effects of naval sonar on killer whales (*Orcinus orca*), sperm whales (*Physeter macrocephalus*), and long-finned pilot whales (*Globicephala melas*) in Norwegian waters. Scottish Oceans Institute Technical Report SOI-2011- 001. Available via http://soi.st-andrews.ac.uk/documents/424.pdf. Accessed 18 June 2013.

Miller PJO, Kvadsheim PH, Lam FPA et al (2012) The severity of behavioral changes observed during experimental exposures of killer (*Orcinus orca*), long-finned pilot (*Globicephala melas*), and sperm (*Physeter macrocephalus*) whales to naval sonar. Aquat Mamm 38(4): 362-401

Southall BL et al (2007) Marine mammal noise exposure criteria: Initial scientific recommendations. Aquat Mamm 33: 411–521.

Southall BL, Moretti D, Abraham B et al (2012) Marine mammal behavioral response studies in southern California: Advances in technology and experimental methods. Mar Technol Soc J 46:48–59

Tyack PL, Johnson MP, Aguilar de Soto N et al (2006) Extreme diving of beaked whales. J Exp Biol 209:4238–4253

Tyack PL, Zimmer WMX, Moretti D et al (2011) Beaked whales respond to simulated and actual navy sonar. PLoS One 6(3): e17009

Visser F et al (2011). Group behaviour sampling protocol for behavioural response studies. In: Kvadsheim et al. (2011). Behavioural response studies of cetaceans to naval sonar signals in Norwegian waters - 3S 2011Cruise Report. Available via http://rapporter.ffi.no/rapporter/2011/01289.pdf. Accessed 18 June 2013

Figure 1: Mahalanobis distance (modelled as response intensity) for each dive in the DTAG time-series for the 2010 and 2011 whales. Each box relates to one dive cycle and the area shaded in dark grey is the period of controlled exposure and the area shaded in light grey is the incidental exposure in 2011. The dotted line relates to the fitted model.

Figure 2: Uncertainty expressed as the width of the 95% confidence intervals around each of the four parameters in the model, normalized by the parameter estimates. The three panels represent the three different experimental scenarios: source level 210dB (left panel), source level 235dB (middle panel) and source level assigned as 210dB to half of the whales and 235dB to the other half (right panel).
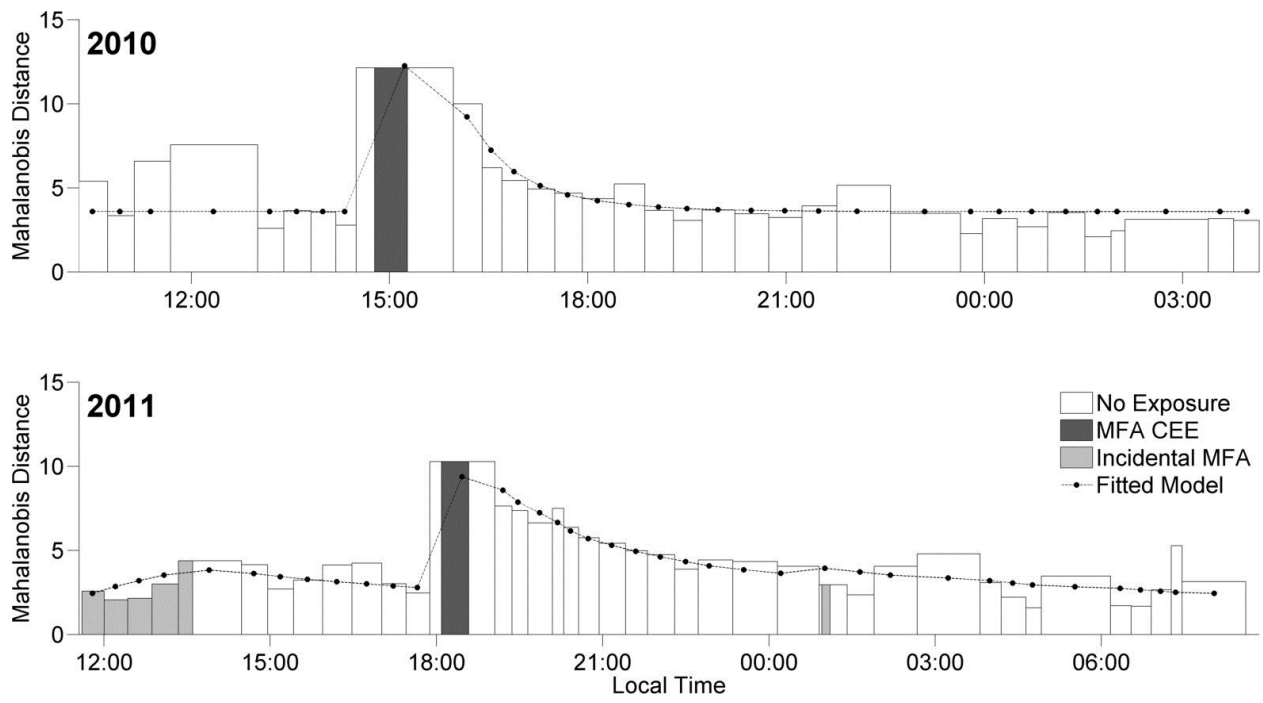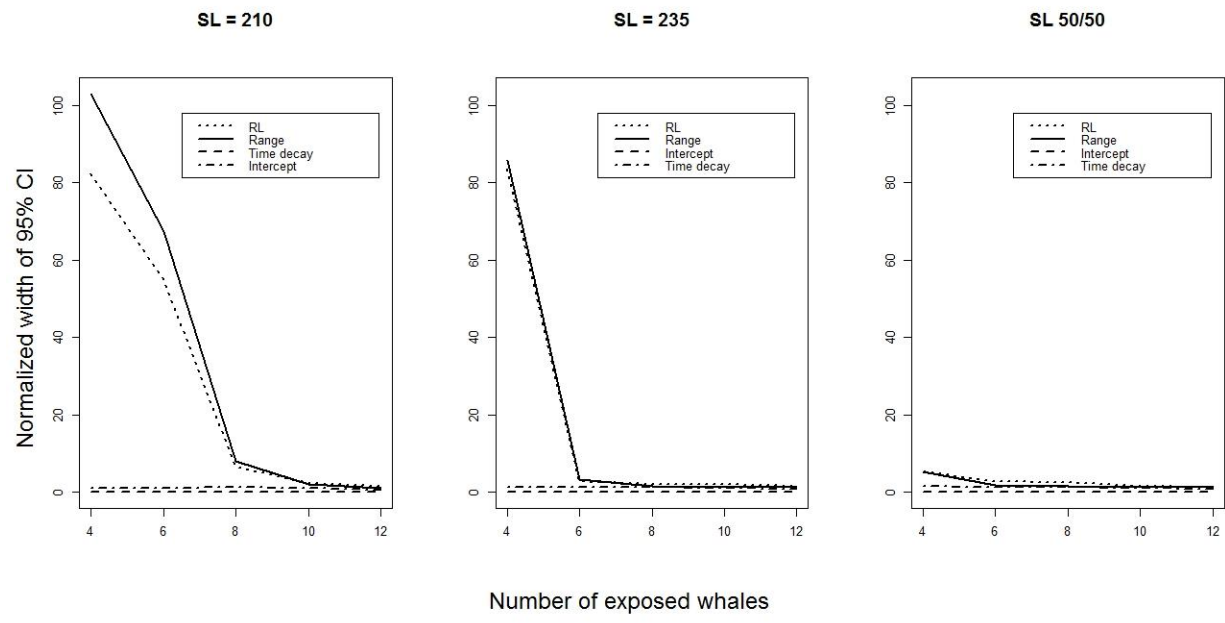
Figure 1

Figure 2