QUANTIFYING THE RESPONSE OF BLAINVILLE'S BEAKED WHALES TO US NAVAL SONAR EXERCISES IN HAWAII

Eiren K. Jacobson[1]*, E. Elizabeth Henderson[2], Cornelia S. Oedekoven[1], David L. Miller[1], Stephanie L. Watwood[3], David J. Moretti[3], Len Thomas[1]

[1] *Centre for Research into Ecological and Environmental Modelling, School of Mathematics and Statistics, University of St Andrews, St Andrews, Scotland*

[2] *Naval Information Warfare Center Pacific, San Diego, CA, USA*

[3] *Naval Undersea Warfare Center, Newport, RI, USA*

Corresponding author (email: eiren.jacobson@st-andrews.ac.uk)

Draft 29 April 2020

Abstract

Naval use of mid-frequency active (MFA) sonar has been associated with injury and death of multiple species of marine mammals. Deep-diving beaked whales (family Ziphiidae) are particularly susceptible to naval sonar. The US Navy operates multiple training and testing facilities where MFA sonar is used regularly, and where cumulative sublethal impacts of exposure to MFA sonar could have negative effects on beaked whale populations. Responses of beaked whales to MFA sonar have been quantified in the form of risk functions for some species and regions. Here we develop a risk function for Blainville's beaked whales *(Mesoplodon densirostris)* on the Pacific Missile Range Facility (PMRF) in Hawaii and compare our risk function to another developed for the same species in a different ocean basin. We used passive acoustic data collected at bottom-mounted hydrophones before and during six naval training exercises at PMRF in conjunction with modelled sonar received levels to describe the effect of Naval training and MFA sonar on foraging groups of Blainville's beaked whales. We used a multi-stage generalized additive modelling (GAM) approach to control for the underlying spatial distribution of Blainville's beaked whales when modelling the effects of training activities and of MFA sonar. Our results show that at a MFA sonar received level of 150 dB re 1μPa the probability of detecting groups of Blainville's beaked whales decreased by 78% (95% CI 62%-100%) when compared to periods when general Naval training activity was ongoing and by 92% (95% CI 87%-100%) when compared to pre-training baseline periods. These results indicate a more pronounced response to Naval training and MFA sonar than has been previously reported. In future, we would like to apply these methods to data from other Navy ranges to better understand how beaked whale responses to Naval training and sonar differ across species and locations.

1 Introduction

Beaked whales (family Ziphiidae) are a group of deep-diving cetaceans that rely on sound to forage, navigate, and communicate (Aguilar de Soto et al., 2012; Johnson, Madsen, Zimmer, Aguilar de Soto, and Tyack, 2004; Macleod and D'Amico, 2006). Multiple mass strandings of beaked whales have been associated with high-intensity anthropogenic sound sources. These acute events have motivated research into whether and how beaked whales respond to different types and intensities of anthropogenic noise (Cox et al., 2006).

Anthropogenic sound can disrupt the patterned dive cycles of beaked whales (Falcone et al., 2017), potentially leading to death (Jepson et al., 2003) or to cumulative sublethal impacts (L. F. New, Moretti, Hooker, Costa, and Simmons, 2013). For example, research on Blainville's beaked whales *Mesoplodon densirostris* on a US Navy range in the Bahamas has shown that animals may stop foraging and/or move away from naval sonar sources (Joyce et al., 2019; Tyack et al., 2011).

Naval sonar can be broadcast from various platforms, including vessels, helicopters, buoys, submarines, and torpedoes (Harris et al., 2019; Navy, 2018). Most research has focused on the impacts of mid-frequency active sonar (MFAS) broadcast from US Naval vessels. Separately, researchers have shown that, in the absence of MFAS, beaked whales may alter their behavior in response to vessel noise (Aguilar Soto et al., 2006; Pirotta et al., 2012).

The US Navy is interested in quantifying the effects of sonar on beaked whales for the purpose of risk assessments and permitting associated with training activities [CITE?]. There are different experimental and analytical ways of quantifying responses to sonar. Here, we focus on analyses of data from cabled hydrophone arrays.

For example, (McCarthy et al., 2011) used data from the cabled hydrophone array at the US Navy's Atlantic Undersea Test and Evaluation Center (AUTEC) in the Bahamas collected before, during, and after naval training exercises involving MFAS. The authors used separate generalized additive models (GAMs) for each period, and modelled the acoustic detection of groups of Blainville's beaked whales (group vocal periods; GVPs) as a function of location on the range and time. They found that the number of GVPs was lower during the exercises than before, and also lower during an exercise than after.

Building on this work, (Moretti et al., 2014) used a GAM to model the presence of acoustic detections of groups of Blainville's beaked whales on the AUTEC range as a smooth function of MFAS received level. They then compared the expected probability of detecting animals when no sonar was present to the expected probability of detecting animals across sonar received levels to estimate the probability of disturbance. They found that the probability of detecting groups of Blainville's beaked whales was reduced by 50% at 150 dB re 1μPa, which they interpreted as a 50% probability of disturbance.

In the present study, our primary objective was to replicate the effort of Moretti et al. with the same species on a different US Navy training range in a different oceanic environment. Unlike AUTEC, which occurs in a deep isolated basin surrounded by steep slopes, the Pacific Missile Range Facility (PMRF) range in Hawaii is located on the side of an ancient volcano,

with a steep slope down to the deep ocean floor. Density of Blainville's beaked whales at PMRF is lower and more variable than at AUTEC, so we wanted to explicitly account for differences in underlying beaked whale presence across the range.

An additional objective was to isolate the effect of general training activity from the effect of MFAS, so that beaked whale response to MFAS could be quantified relative to pre-training baseline periods and to periods when general training activities were present on the range.

To accomplish these objectives, we used a spatially referenced dataset of Blainville's beaked whale foraging dives recorded at the PMRF off the island of Kauai, Hawaii (Fig. 1). Acoustic detections of Blainville's beaked whales were collected via a cabled hydrophone array at PMRF before and during naval training exercises involving MFAS. Previous work in this region has shown that Blainville's beaked whales are present year-round at this site, prefer slope habitats, and that acoustic detections decrease during multi-day training events involving MFAS (Henderson, Martin, Manzano-Roth, and Matsuyama, 2016; Manzano-Roth, Henderson, Martin, Martin, and Matsuyama, 2016).

A series of three models were fitted to data collected before the training exercises began, data collected when training exercises were ongoing but no MFAS was present, and to data collected when training exercises including MFAS were present. The expected values from each model were used as an offset in the next model, and uncertainty was propagated through all models using posterior simulation. Using this set of model results, we quantified the expected decrease in detection of GVPs across increasing sonar received levels relative to both the pre-training baseline period and the period when training activities were ongoing but no hull-mounted MFAS was present.

2 Methods

2.1 Data Collection and Processing

2.1.1 Acoustic detection of beaked whales

The Pacific Missile Range Facility (PMRF) is an instrumented U.S. Naval range extending 70 km NW of the island of Kauai, Hawaii and encompassing 2,800 km^2. The range includes a cabled hydrophone array (Fig. 1) with hydrophones at depths ranging from approximately 650 m to 4,700 m. We used data collected before and during Submarine Commander Courses (SCCs) at the PMRF. SCCs are training exercises that occur biannually in February and August. SCCs typically last 6-7 days, and the period before the onset of the SCC is recorded for a minimum of 2 days. During data collection, hydrophones sampled at a rate of 96 kHz, with the high pass filter on each phone set at either 50 Hz, 100 Hz, or 10 kHz. Up to 62 of the range hydrophones were recorded simultaneously by the Naval Information Warfare Center (NIWC).

A beaked whale detector from the Navy Acoustic Range WHale AnaLysis (NARWHAL) algorithm suite [CITE] was run on recordings from PMRF. This detector first compares

3

signal-to-noise (SNR) thresholds within the expected beaked whale click frequency range (16 - 44 kHz) versus the bandwidth outside the click in a running 16384-pt fast Fourier transform (FFT) spectrogram. The detected clicks were then passed to a 64-pt FFT stage that measured power, bandwidth, slope, and duration characteristics to classify the clicks to species. This process was followed by an automated routine in Matlab (CITE Mathworks 2019) to group detections of individual beaked whale echolocation clicks into GVPs (CITE). If a group of whales was detected by more than one hydrophone, the GVP was assigned to the hydrophone that recorded the most clicks. The data were then aggregated to indicate the presence or absence of the start of a GVP for each hydrophone within each half-hour period.
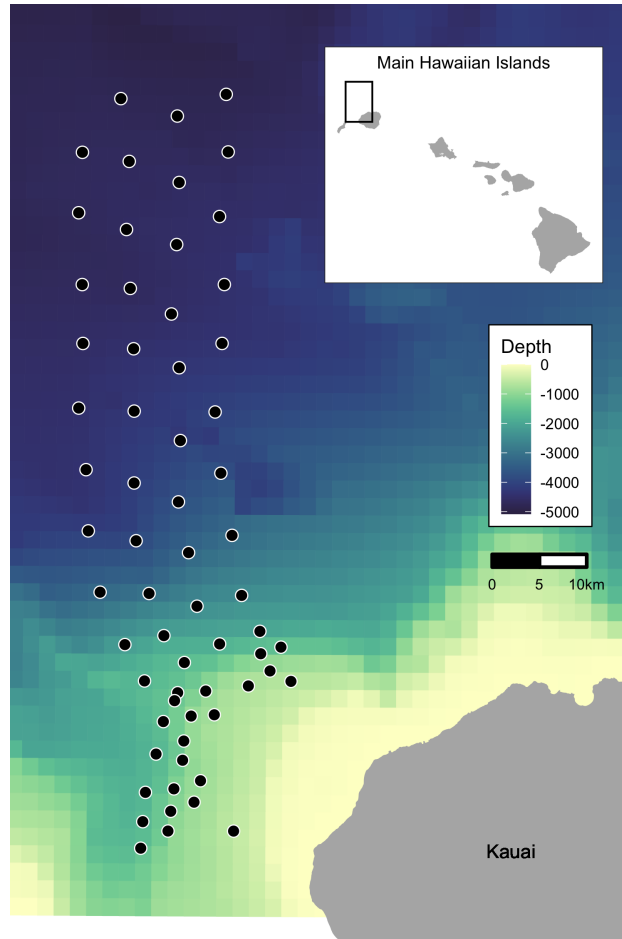

Figure 1: Map of approximate locations of hydrophones (black points) at the Pacific Missile Range Facility near the island of Kauai, Hawaii. Color scale indicates bathymetry. Inset map shows range location relative to the Main Hawaiian Islands.

2.1.2 Modelling received levels of hull-mounted mid-frequency active sonar

Classified ship positional data and other activity that occured on the range during each SCC were provided by PMRF. These data indicated the locations of the ships during the training periods and the start and stop times of each individual training event, but no information

was provided on the start and stop of sonar use. Periods of active sonar were determined by running a sonar detector from the NARWHAL algorithm suite tuned to MFAS. The locations of all surface ships were noted for each half-hour period and the closest ship to each hydrophone was determined. Propagation modelling was used to calculate the expected received level of hull-mounted mid-frequency active sonar at the location of each hydrophone from the closest ship during each half-hour period of each SCC.

The propagation modelling was done using the parabolic equation propagation model in the program Peregrine (OASIS; Heaney and Campbell, 2016) to estimate the transmission loss between the ship and the hydrophone; this was then converted to a received level at the hydrophone location based on the source level of the sonar. However, if the distance between the ship and the hydrophone was less than the depth of the water column, the parabolic equation overestimates transmission loss at that angle and so a simple sonar equation was used to estimate transmission loss instead. Transmission loss was estimated using a 200 Hz band around the center frequency of the sonar type (here, 3.5 kHz). Transmission loss was estimated at depth since Blainville's beaked whales don't begin clicking until they have reached approximately XX m depth of their foraging dive and spend most of their foraging dive at around 1000 m (CITE).

For hydrophones shallower than 1000 m the received level was estimated at a point 20 m above the sea floor with a +/- 10 m buffer, while for hydrophones deeper than 1000 m the received level was estimated at a depth of 1000 m with a +/- 10 m buffer. The location of the beaked whale foraging group was assumed to be within 4-6 km of the hydrophone with the most click detections, since beaked whale echolocation clicks attenuate beyond that distance (T. A. Marques, Thomas, Ward, DiMarzio, and Tyack, 2009; McCarthy et al., 2011). Therefore the transmission loss was estimated along the radial from the ship to the hydrophone from a distance of 1 km before the hydrophone to 1 km past the hydrophone in 200 m increments. The maximum modeled received level along that radial was determined for each hydrophone and half-hour period and aggregated with the data on beaked whale group detections. Uncertainty in the modeled received level was not considered.

2.2 Spatial Modelling

We used a multi-stage generalized additive modelling (GAM) approach to control for the underlying spatial distribution of Blainville's beaked whales when modelling the effects of training activities and of MFA sonar. We first used a tessellation to determine the area effectively monitored by each hydrophone. Then, we used pre-activity data to create a spatial model of the probability of GVPs across the range prior to the onset of naval activity. We used the predicted values from this model as an offset in a model created using data from when naval activity was present on the range, but MFAS was not. Again, we used the predicted values from this model as an offset in a model created using data when naval activity and MFAS were present on the range. Finally, we used posterior simulation to calculate confidence intervals and quantified the change in the probability of detecting GVPs when Naval activity was present and across received levels of MFAS.

2.3.1 Determining hydrophone effort

For security reasons, randomly jittered locations and depths of hydrophones at PMRF were used. We projected the coordinates of each hydrophone into Universal Transverse Mercator Zone 4.

Because the beaked whale detection algorithm assigned groups of whales to the hydrophone that recorded the most echolocation clicks, and because the spatial separation of the hydrophones was not uniform, effort was not the same for all hydrophones. To determine the area effectively monitored by each hydrophone, we used a Voronoi tessellation implemented in the R (R Core Team, 2018) package `deldir` (Turner, 2019) to define a tile for each hydrophone that contained all points on the range that were closest to that hydrophone. The area of each tile corresponded to the effective area monitored. We assumed that beaked whale groups occur within the tessellation tile of the hydrophone to which the GVP is assigned. For hydrophones on the outside of the range, i.e., not surrounded by other hydrophones, we used a cutoff radius of 6500 m to bound the tessellation tile. This distance is based on the maximum detection distance of individual Blainville's beaked whale clicks at a U.S. Naval range in the Bahamas (T. A. Marques et al., 2009). Different combinations of hydrophones were used during different SCCs, so separate tessellations were created for each SCC.

2.3.2 M1: Modelling the pre-activity probability of dive detection

We used data collected prior to SCCs, when no Naval ships were present on the range and no other Naval activity was known to occur, to model the spatial distribution of GVP detections across the range. The exact locations of beaked whale groups was not known; rather, detections of beaked whale groups were "snapped" to hydrophone locations depending on which hydrophone detected the most echolocation clicks. Therefore, the data were not continuous in space. To account for this, we used a Markov random field to model the spatial distribution of GVP detections. A Markov random field (Rue and Held, 2005) is a method for modelling correlation in space between discrete spatial units. Each unit is correlated more strongly with its neighbours (those units which touch) than those that are more hops away. This gives a graph structure, where distance between tiles is measured as the number of hops required. This is appropriate for our data as we did not know where in each tile a given GVP occured, but we assumed that it did occur in that tile.

The R package `mgcv` (S. N. Wood, 2017) was used to formulate the model on the tessellation described in the previous section. The linear predictor for the model was:

$$\text{logit}\left(\mu_{\text{M1},i}\right) = \beta_{\text{M1},0} + f(\text{MRF}_i) + f(\text{Depth}_i) + \log_e A_i, \qquad \text{(M1)}$$

where $\text{DivePresent}_i \sim \text{Bin}(1, _{\text{M1},i})$. The spatial smooth MRF is given by $f(\text{MRF}_i)$, $f(\text{Depth}_i)$ is a smooth of depth (using a thin plate spline; Simon N Wood (2003)) and $\log_e A_i$ is an offset for the area (in km^2) of each tile, A_i. The offset term accounts changes in probabilities of GVP detection due to the differing area monitored by each hydrophone. Because the

6

²⁰³ hydrophone tessellation change between SCCs, separate MRFs were used for each SCC, but
²⁰⁴ a single smoothing parameter was estimated across all MRFs. Therefore different spatial
²⁰⁵ patterns could occur, but with the same amount of variation. The smooth of depth was
²⁰⁶ shared across SCCs.

²⁰⁷ NOTE: f(MRF) could be indexed by SCC to indicate that the smooth function is different
²⁰⁸ for each.

2.3.3 M2: Modelling the effect of Naval activity

²¹⁰ For a few days prior to the onset of hull-mounted MFA sonar used during SCCs, other
²¹¹ Naval training activities occurred at the PMRF. Various vessels were present on the range
²¹² during this period and other noise sources, including torpedoes and submarines, may have
²¹³ been present. We used data collected when training activity was present on the range, but
²¹⁴ hull-mounted MFA sonar was not used, to model the effect of general Naval activity on
²¹⁵ beaked whale GVPs. Initially, we tried to use low-frequency noise levels in the 10-999 Hz
²¹⁶ range measured on range hydrophones as a covariate in this model, but found that the
²¹⁷ measured noise levels were not consistent with known locations of Naval training activities
²¹⁸ (see Appendix B for details).

²¹⁹ We used the predicted baseline probability of a GVP detection from Model 1 as an offset to
²²⁰ control for the underlying spatial distribution of GVPs. The model for the data when ships
²²¹ were present was intercept-only, with an offset derived from `M1`. This model was simply:

$$\text{logit}\left(\mu_{\texttt{M2},i}\right) = \beta_{\texttt{M2},0} + \log_e \xi_{\texttt{M1},i}, \qquad \text{(M2)}$$

²²² where $\texttt{DivePresent}_i \sim \text{Bin}(1, \mu_{\texttt{M2},i})$. $\xi_{\texttt{M1},i}$ denotes the prediction (on the logit scale) for tile i
²²³ using model `M1`. This was again modeled in the R package `mgcv`.

2.3.4 M3: Modelling the effect of hull-mounted MFA sonar

²²⁵ We used data collected when hull-mounted MFA sonar was present on the range to model
²²⁶ the effect of sonar on beaked whales. The probability of a dive when sonar was present was
²²⁷ modeled as a function of the maximum received level (modeled at each hydrophone; see
²²⁸ section 2.2). We assumed that as the maximum received level increased, the probability of
²²⁹ dives decreased and modeled this using a shape constrained smooth so that the relationship
²³⁰ held for all possible realizations of the smooth. To ensure that the model predictions were
²³¹ the same at a maximum received level of 0 dB and when ships were not present, we did not
²³² include an intercept. This model was written as:

$$\text{logit}\left(\mu_{\texttt{M3},i}\right) = f(\texttt{MaxRL}_i) + \log_e \xi_{\texttt{M2},i}, \qquad \text{(M3)}$$

where $\texttt{DivePresent}_i \sim \mathrm{Bin}(1, \mu_{\texttt{M3},i})$. $f(\texttt{MaxRL}_i)$ was modeled as a monotonic decreasing smooth using the R package $\texttt{scam}$ (Pya and Wood, 2015). $\xi_{\texttt{M2},i}$ denotes the prediction (on the logit scale) for tile i when Naval training activites were present on the range using model $\texttt{M2}$.

2.3.5 Uncertainty propagation

We used posterior simulation to propagate uncertainty through M1, M2, and M3. Each model was fitted via restricted maximum likelihood (REML; Wood (2008)), so the results are empirical Bayes estimates. In this case we can generate samples from the (multivariate normal) posterior of the model parameters, using mvtnorm?. After generating a sample, $\boldsymbol{\beta}^* \sim \mathrm{MVN}(\hat{\boldsymbol{\beta}}, \mathbf{V}_\beta)$, we can use the matrix that maps the model parameters to the predictions on the linear predictor scale (often referred to as the $\mathbf{L}_p$ matrix or $\mathbf{X}_p$ matrix; Wood, Li, Shaddick, and Augustin (2017); section 7.2.6), along with the inverse link function to generate predictions for each posterior sample. Here the $\boldsymbol{\beta}$ for each model includes the coefficients for the smooth terms in the model and fixed effects (e.g., intercept) if present. Predictions, $\boldsymbol{\mu}^*$, can be written as:

$$\boldsymbol{\mu}^* = g^{-1}(\boldsymbol{\eta}^*) = g^{-1}(\mathbf{X}_p \boldsymbol{\beta}^* + \boldsymbol{\xi}),$$

where g is the link function, $\boldsymbol{\eta}^*$ is the linear predictor and $\boldsymbol{\xi}$ is any offset used by this prediction. By sampling from the posterior of $\hat{\boldsymbol{\beta}}$, and then taking the variance of the resulting predictions we can obtain variance estimates (Wood et al. (2017); section 7.2.6). The prediction grid contained all possible combinations of covariates within the realized covariate space; i.e., each hydrophone for each SCC with associated location, hydrophone depth, and area of the tessellation tile, presence/absence of Naval activity, and, if Naval activity was present, then either sonar absence or sonar received level between 35 and 190 dB in intervals of 5 dB. This procedure was repeated for each model, with refitting to updated offsets from the previous model.

An algorithm for calculating the variance from our multi-stage approach is as follows. First define N_b as the number of samples to make, let $\mathbf{X}_{p,\texttt{M}j}$ for $j = 1, 2, 3$ be the $\mathbf{L}_p$ matrix that maps coefficients to the predictions for model $\texttt{M}j$. For N_b times:

1. Draw a sample from the posterior of $\texttt{M1}$: $\tilde{\boldsymbol{\beta}}_{\texttt{M1}} \sim \mathrm{MVN}(\hat{\boldsymbol{\beta}}_{\texttt{M1}}, \mathbf{V}_{\texttt{M1}})$.

2. Calculate a new offset for $\texttt{M2}$, $\tilde{\boldsymbol{\xi}}_{\texttt{M1}} = \mathbf{X}_{p,\texttt{M1}} \tilde{\boldsymbol{\beta}}_{\texttt{M1}} + \log_e \mathbf{A}$.

3. Refit $\texttt{M2}$ with $\tilde{\boldsymbol{\xi}}_{\texttt{M1}}$ as the offset, to obtain $\texttt{M2}'$.

4. Draw a sample from the posterior of $\texttt{M2}'$: $\tilde{\boldsymbol{\beta}}_{\texttt{M2}'} \sim \mathrm{MVN}(\hat{\boldsymbol{\beta}}_{\texttt{M2}'}, \mathbf{V}_{\texttt{M2}'})$

5. Calculate a new offset for $\texttt{M3}$, $\tilde{\boldsymbol{\xi}}_{\texttt{M2}} = \mathbf{X}_{p,\texttt{M2}} \tilde{\boldsymbol{\beta}}'_{\texttt{M2}} + \tilde{\boldsymbol{\xi}}_{\texttt{M1}}$ (predictions for the sonar data locations for $\texttt{M2}'$).

6. Refit $\texttt{M3}$ with offset $\tilde{\boldsymbol{\xi}}_{\texttt{M2}}$ to obtain $\texttt{M3}'$.

7. Predict $\boldsymbol{\mu}_{\texttt{M1}'}$, $\boldsymbol{\mu}_{\texttt{M2}'}$, and $\boldsymbol{\mu}_{\texttt{M3}'}$ over prediction grid and store them.

We can then calculate summary statistics (means and variances) of the N_b values of $\boldsymbol{\mu}_{\texttt{M1}'}$, $\boldsymbol{\mu}_{\texttt{M2}'}$, and $\boldsymbol{\mu}_{\texttt{M3}'}$ we have generated. The empirical variance of the N_b values of $\boldsymbol{\mu}_{\texttt{M3}'}$ will give the uncertainty, incorporating components from all three models. We can take appropriate pointwise quantiles to form confidence bands for the functional relationships between sonar received level and estimated probability of detecting GVPs.

2.3.6 Quantifying the change in probability of GVPs

Finally, we calculated the expected change in $\mathbb{P}(\text{GVP})$ relative to either the distribution of GVPs when no general Naval training activity was present and no MFA sonar was present ($\Delta_{M3':M1'}$), or relative to the distribution of GVPs when general Naval training activity was present but no MFA sonar was present ($\Delta_{M3':M2'}$).

Using the N_b bootstrapped model realizations we calculated the expected $\mathbb{P}(\text{GVP})$ under each set of covariates as

$$\mathbb{P}(\text{GVP}) = \text{logit}^{-1}(\mu_{\texttt{M}'}), \tag{1}$$

for each $\texttt{M1}'$, $\texttt{M2}'$, and $\texttt{M3}'$. Then, we calculated the change in $\mathbb{P}(\text{GVP})$ for each set of covariates $\texttt{M3}'$ and $\texttt{M1}'$ ($\Delta_{M3':M1'}$) and between $\texttt{M3}'$ and $\texttt{M2}'$ ($\Delta_{M3':M2'}$) for each realization of the posterior simulation.

$$\Delta_{M3':M1'} = \frac{\mathbb{P}(\text{GVP})_{\texttt{M3}'} - \mathbb{P}(\text{GVP})_{\texttt{M1}'}}{\mathbb{P}(\text{GVP})_{\texttt{M1}'}} \tag{2}$$

$$\Delta_{M3':M2'} = \frac{\mathbb{P}(\text{GVP})_{\texttt{M3}'} - \mathbb{P}(\text{GVP})_{\texttt{M2}'}}{\mathbb{P}(\text{GVP})_{\texttt{M2}'}} \tag{3}$$

For each received level we calcualted the 2.5th, 50th, and 97.5th quantiles of $\Delta_{M3':M1'}$ and $\Delta_{M3':M2'}$ to create 95% CIs of change in $\mathbb{P}(\text{GVP})$ across possible received levels. We consider that the probability of disturbance is equal to 1 wherever the 95% CI does not include 0, and 0 otherwise.

2.3.7 Implementation

Statistical analyses presented in this manuscript were conducted in R (v. 3.5.2; R Core Team, 2018). Data import and manipulation was accomplished using the packages `dplyr` (Wickham, François, Henry, and Müller, 2020), `lubridate` (Grolemund and Wickham, 2011), `readr` (Wickham, Hester, and Francois, 2018), `readxl` (Wickham and Bryan, 2019), and `tidyr` (Wickham and Henry, 2020). Map creation was facilitated by the `fields` (Douglas Nychka, Reinhard Furrer, John Paige, and Stephan Sain, 2017), `ggsn`, `marmap`, `rgdal`, `sf`, and `sp` packages. All graphics were produced using `ggplot2`, with color palettes from the `viridis` and `cmocean` packages. The manuscript was written in `rmarkdown` (Xie, Allaire, and Grolemund, 2018).

3 Results

3.1 Results of Data Collection and Processing

Data were collected before and during six SCCs; two each in in 2013, 2014, and 2017 (Table 1). The number of hydrophones for which recordings were available varied from 49 to 61. A total of 190928 30-min observations were made.

Table 1: No. of hydrophones used and number of observations made (no. 30-min periods) for each SCC before the exercise began, when Naval activity was present, and when Naval activity and MFA sonar were present.

SCC	HPs	Pre-Activity	Nav. Activity	MFA Sonar
Feb13	61	114	193	124
Aug13	61	209	115	97
Feb14	60	513	111	129
Aug14	61	263	120	128
Feb17	59	450	97	108
Aug17	49	270	106	113

The exact timing of activities during these exercises varied (Fig. 2). For most SCCs, pre-activity data were available immediately preceding the onset of Naval training activity; however, in February 2013 the only available pre-activity data were collected almost a month prior to the onset of Naval training activity. In some SCCs, weekends or other breaks in training resulted in a break in training activity on the range during the days preceding MFA sonar use. MFA sonar was used for 3-4 days during each training event.

Across all SCCs, hydrophones, and conditions, a total of 2312 GVPs were identified. The average $\mathbb{P}(\text{GVP})$ in the dataset was therefore 1%. The spatial distribution of GVPs differed during the pre-activity phases of SCCs (Fig. SX; top panel).

Modelled maximum received levels ranged from 38 to 186 dB re. 1 μ Pa, with a median value when MFA sonar was present of 147 dB re. 1 μ Pa. The intensity and spatial distribution of MFA received levels varied across the range and across SCCs (Fig. SX).

Based on the observed data, $\mathbb{P}(\text{GVP})$ changed by -57% when Naval activity was present compared to when Naval activity was absent, by -47% when Naval activity and MFA sonar were present compared to when only ships were present, and by -77% when Naval activity and MFA sonar were present compared to when neither ships nor sonar were present (Fig. SX).
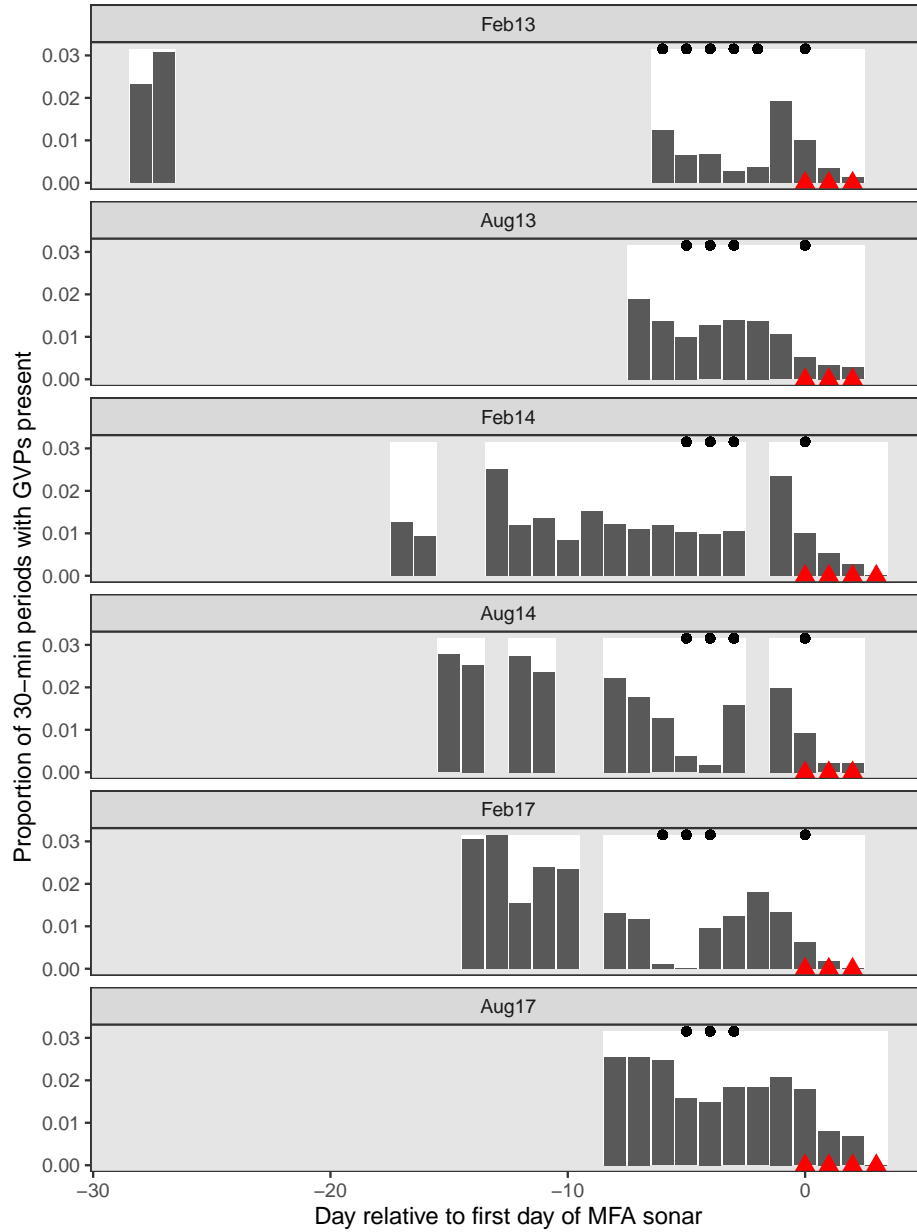
10

Figure 2: Timeseries of six recorded Naval training activities at PMRF. The timeseries are aligned relative to the first day that MFA sonar (red triangles) was used in each exercise (x-axis). Days with white background indicate days for which recordings and data were available. Gray bars indicate the proportion of 30-min periods on each day, across all hydrophones, when GVPs were detected. Black dots indicate days when Naval activity was present on the range.

3.2 Results of spatial modelling

We created separate tesselations for each SCC (Fig. SX). In August 2017, data were available from fewer hydrophones, and so in some cases the tesselated tiles, with bounding radius of 6500 m, did not completely cover the range. Hydrophone depths varied from 648 to 4716 m.

`M1` fitted a spatial model of $\mathbb{P}(\text{GVP})$ to data collected prior to the onset of Naval training activity. This model used a MRF smooth to account for the spatial structure of the range and a spline on depth, with an offset for the log of the area effectively monitored by each hydrophone. Both the MRF and spline on depth were significant (p-value < 2E-16). The model explained 14.1% of deviance in the dataset, and visual inspection of observed versus predicted values indicated a good fit to the data (Fig. SX). The model `M1` predicted highest $\mathbb{P}(\text{GVP})$ at hydrophone depths between 1500 and 2000 m (Fig. SX).

`M2` used the predicted values from `M1` as an offset and fitted a model of to data when Naval activity was ongoing, as indicated by the presence of Naval ships on the range. This model was intercept-only, and $\mathbb{P}(\text{GVP})$ when Naval training was ongoing was significantly different from the baseline period (p-value < 2E-16). The expected $\mathbb{P}(\text{GVP})$ decreased by a median of 64% (95% CI 59% - 68%) when Naval training activity was present compared to when it was absent.

`M3` used the predicted values from `M2` as an offset and fitted a model to data when Naval activity and MFA sonar were present. This model used a monotonically decreasing spline on modelled MFA sonar received level (Fig. SX) and did not include an intercept term. The smooth on MFA sonar received level had significant explanatory power (p-value = 6.74E-10) and the model explained 12.4% of deviance in the data.

For MFA sonar received levels above 100 dB, change in $\mathbb{P}(\text{GVP})$ was calculated relative to the pre-activity baseline period ($\Delta_{M3':M1'}$) and to the period when Naval activity was present on the range ($\Delta_{M3':M2'}$; Fig. 4 & Fig. 5). At a received level of 150 dB, $\Delta_{M3':M1'}$ was -92% (95% CI -100% - -87%) and $\Delta_{M3':M2'}$ was -78% (95% CI -100% - -62%).

4 Discussion

- Describe why we didn't use a single giant GAM – didn't want contamination of the baseline period by the spatial distribution of sonar, would lead to underestimates of the impact of sonar. Could present the single giant GAM in an appendix.

- Emphasize novelty – more sophisticated method

- Discuss unusual timeline of Feb13

- Discuss what "Naval activity" could mean

- GVPs appear to decrease over the course of MFA sonar; this is something we could investigate with a spatio-temporal model in teh future (hour since onset of MFA? SEL?)
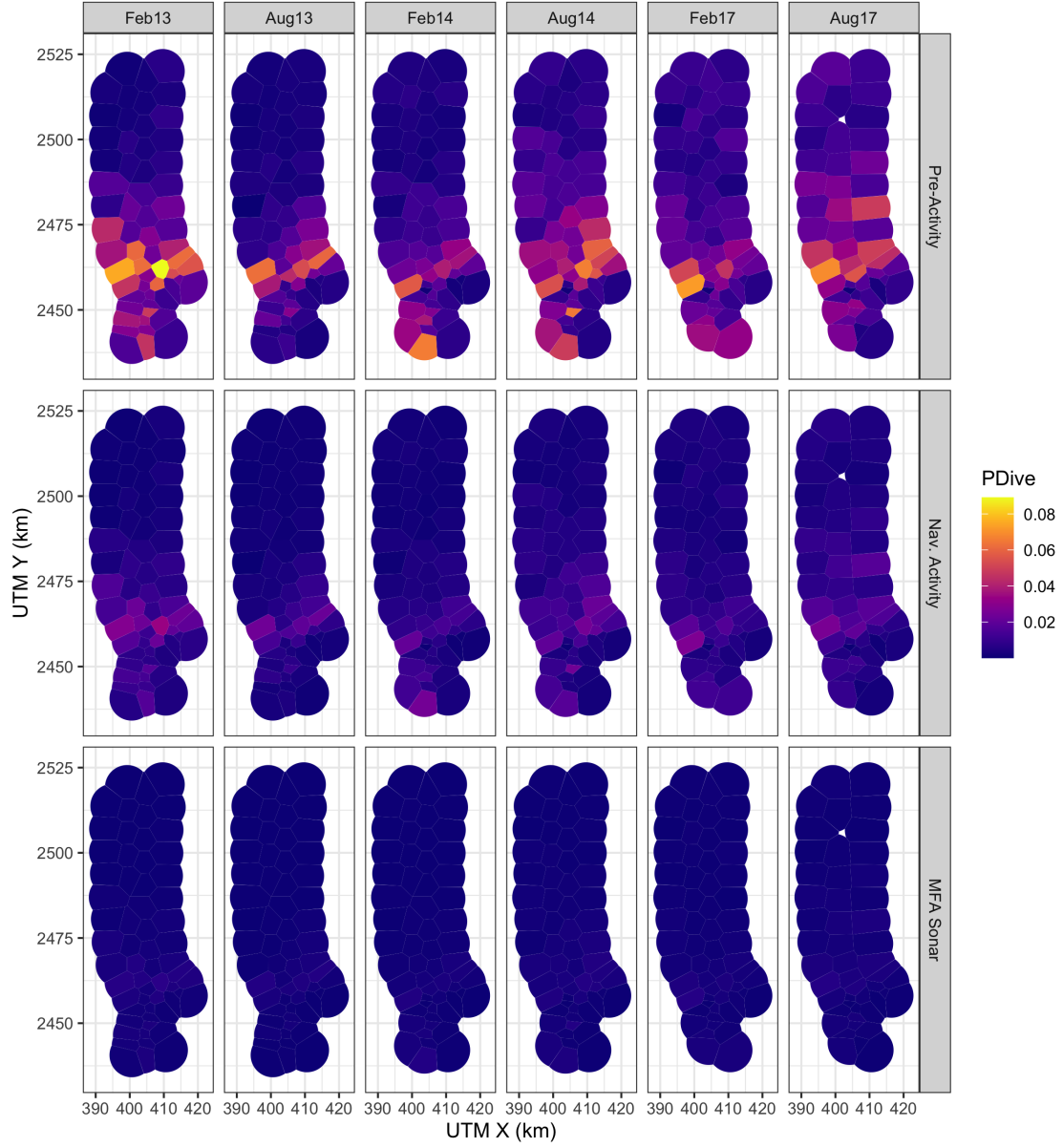
12

Figure 3: Map of expected probability of diving (color scale) at each hydrophone during each SCC (columns) prior to the onset of Naval training activity, during Naval training activity when no MFA sonar was present, and during Naval training activity when MFA sonar was present at a level of 150 dB re. 1 uPa rms (rows).
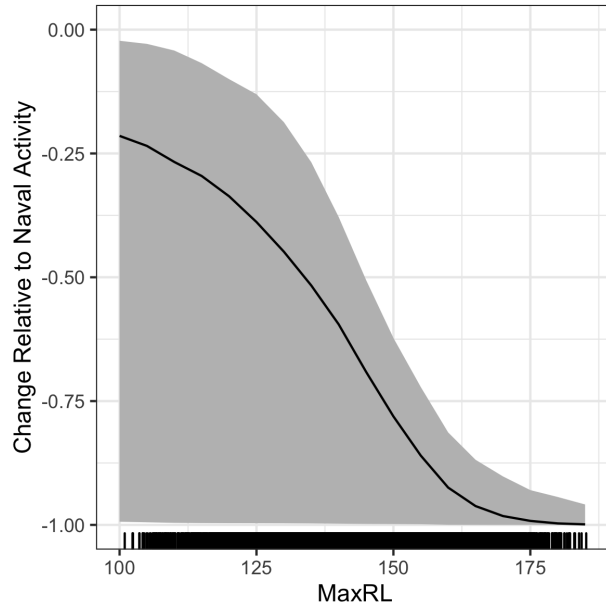
Figure 4: Median (black line) and 95% CIs (gray shading) expected change in the probability of detecting a group vocal period (y-axis) with increasing MFA sonar received level (x-axis) relative to when Naval training activity but no MFA sonar is present on the range.
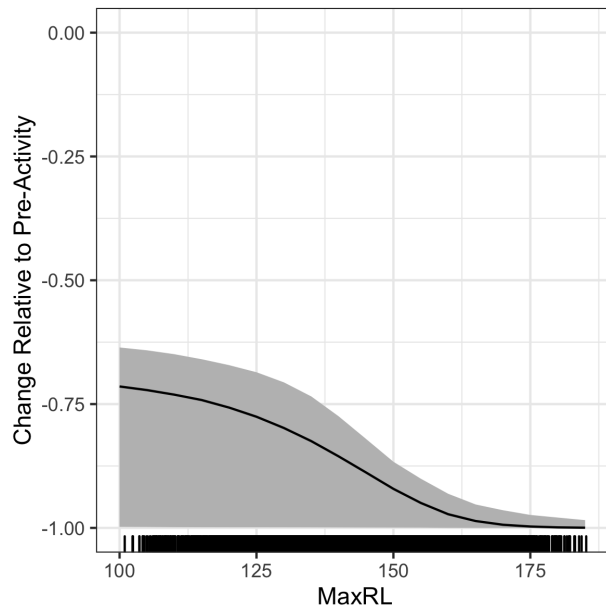

Figure 5: Median (black line) and 95% CIs (gray shading) expected change in the probability of detecting a group vocal period (y-axis) with increasing MFA sonar received level (x-axis) relative to when neither Naval training activity nor MFA sonar is present on the range.

- Discuss dose-response and p(disturbance) in context of (Tyack and Thomas, 2019)

- Compare results to Moretti et al 2014; in particular the fact that their "before" was likely actually similar to our "training without sonar"period as it was only 19 hours of data before the onset of sonar and it was the same training scenario as an SCC at PMRF. Therefore our risk function results of the decrease in dives from training without sonar to training with sonar are actually quite similar. Then we can discuss the fact that environment/habitat (e.g. deep basin with shallow slopes all around vs deep open ocean) doesn't seem to play much of a role in Blainville's response, and the response seems to be more of an intrinsic characterisitic. Also can mention here the same effort at SCORE with Cuvier's – in light of these results we expect similar results there even though different species but similar habitat to AUTEC.

References

Aguilar de Soto, N., Madsen, P. T., Tyack, P., Arranz, P., Marrero, J., Fais, A., . . . Johnson, M. (2012). No shallow talk: Cryptic strategy in the vocal communication of Blainville's beaked whales. *Marine Mammal Science*, *28*(2), E75–E92. https://doi.org/10.1111/j.1748-7692.2011.00495.x

Aguilar Soto, N., Johnson, M., Madsen, P. T., Tyack, P. L., Bocconcelli, A., & Fabrizio Borsani, J. (2006). Does intense ship noise disrupt foraging in deep-diving Cuvier's beaked whales (Ziphius cavirostris)? *Marine Mammal Science*, *22*(3), 690–699. https://doi.org/10.1111/j.1748-7692.2006.00044.x

Cox, T., Ragen, T., Read, A., Vos, E., Baird, R., Balcomb, K., . . . others. (2006). Understanding the impacts of anthropogenic sound on beaked whales1. *J. CETACEAN RES. MANAGE*, *7*(3), 177–187.

Douglas Nychka, Reinhard Furrer, John Paige, & Stephan Sain. (2017). *Fields: Tools for spatial data.* https://doi.org/10.5065/D6W957CT

Falcone, E. A., Schorr, G. S., Watwood, S. L., DeRuiter, S. L., Zerbini, A. N., Andrews, R. D., . . . Moretti, D. J. (2017). Diving behaviour of Cuvier's beaked whales exposed to two types of military sonar. *Royal Society Open Science*, *4*(8), 170629. https://doi.org/10.1098/rsos.170629

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. Retrieved from http://www.jstatsoft.org/v40/i03/

Harris, C. M., Martin, S. W., Martin, C., Helble, T. A., Henderson, E. E., Paxton, C. G. M., & Thomas, L. (2019). Changes in the Spatial Distribution of Acoustically Derived Minke Whale (Balaenoptera acutorostrata) Tracks in Response to Navy Training. *Aquatic*

388   *Mammals*, *45*(6), 661–674. https://doi.org/10.1578/AM.45.6.2019.661

389   Henderson, E. E., Martin, S. W., Manzano-Roth, R., & Matsuyama, B. M. (2016). Occurrence
390   and Habitat Use of Foraging Blainville's Beaked Whales (Mesoplodon densirostris) on a
391   U.S. Navy Range in Hawaii. *Aquatic Mammals*, *42*(4), 549–562. https://doi.org/10.1578/
392   AM.42.4.2016.549

393   Jepson, P. D., Arbelo, M., Deaville, R., Patterson, I. A. P., Castro, P., Baker, J. R., . . .
394   Fernández, A. (2003). Gas-bubble lesions in stranded cetaceans. *Nature*, *425*(6958),
395   575–576. https://doi.org/10.1038/425575a

396   Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N., & Tyack, P. L. (2004).
397   Beaked whales echolocate on prey. *Proceedings of the Royal Society of London. Series B:*
398   *Biological Sciences*, *271*(suppl_6). https://doi.org/10.1098/rsbl.2004.0208

399   Joyce, T. W., Durban, J. W., Claridge, D. E., Dunn, C. A., Hickmott, L. S., Fearnbach, H.,
400   . . . Moretti, D. (2019). Behavioral responses of satellite tracked Blainville's beaked whales
401   (*Mesoplodon densirostris*) to mid-frequency active sonar. *Marine Mammal Science*,
402   mms.12624. https://doi.org/10.1111/mms.12624

403   Macleod, C. D., & D'Amico, A. (2006). *A review of beaked whale behaviour and ecology in*
404   *relation to assessing and mitigating impacts of anthropogenic noise.* 11.

405   Manzano-Roth, R., Henderson, E. E., Martin, S. W., Martin, C., & Matsuyama, B.
406   (2016). Impacts of U.S. Navy Training Events on Blainville's Beaked Whale (Mesoplodon
407   densirostris) Foraging Dives in Hawaiian Waters. *Aquatic Mammals*, *42*(4), 507–518.
408   https://doi.org/10.1578/AM.42.4.2016.507

409   Marques, T. A., Thomas, L., Ward, J., DiMarzio, N., & Tyack, P. L. (2009). Estimat-
410   ing cetacean population density using fixed passive acoustic sensors: An example with
411   Blainville's beaked whales. *The Journal of the Acoustical Society of America*, *125*(4),
412   1982–1994. https://doi.org/10.1121/1.3089590

413   McCarthy, E., Moretti, D., Thomas, L., DiMarzio, N., Morrissey, R., Jarvis, S., . . . Dilley, A.
414   (2011). Changes in spatial and temporal distribution and vocal behavior of Blainville's
415   beaked whales (Mesoplodon densirostris) during multiship exercises with mid-frequency
416   sonar. *Marine Mammal Science*, *27*(3), E206–E226. https://doi.org/10.1111/j.1748-
417   7692.2010.00457.x

418   Moretti, D., Thomas, L., Marques, T., Harwood, J., Dilley, A., Neales, B., . . . Morrissey,
419   R. (2014). A Risk Function for Behavioral Disruption of Blainville's Beaked Whales
420   (Mesoplodon densirostris) from Mid-Frequency Active Sonar. *PLoS ONE*, *9*(1), e85064.
421   https://doi.org/10.1371/journal.pone.0085064

422   Navy, U. D. of the. (2018). *Final Environmental Impact Statement/Overseas Environmental*
423   *Impact Statement Hawaii-Southern California Training and Testing.* Retrieved from
424   https://www.hstteis.com/portals/hstteis/files/hstteis_p3/feis/section/HSTT_FEIS_3.

07_Marine_Mammals_October_2018.pdf

New, L. F., Moretti, D. J., Hooker, S. K., Costa, D. P., & Simmons, S. E. (2013). Using Energetic Models to Investigate the Survival and Reproduction of Beaked Whales (family Ziphiidae). *PLoS ONE*, *8*(7), e68725. https://doi.org/10.1371/journal.pone.0068725

Pirotta, E., Milor, R., Quick, N., Moretti, D., Di Marzio, N., Tyack, P., . . . Hastie, G. (2012). Vessel Noise Affects Beaked Whale Behavior: Results of a Dedicated Acoustic Response Study. *PLoS ONE*, *7*(8), e42535. https://doi.org/10.1371/journal.pone.0042535

Pya, N., & Wood, S. N. (2015). Shape constrained additive models. *Statistics and Computing*, *25*(3), 543–559. https://doi.org/10.1007/s11222-013-9448-7

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.R-project.org/

Rue, H., & Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman & Hall.

Turner, R. (2019). *Deldir: Delaunay Triangulation and Dirichlet (Voronoi) Tessellation*. Retrieved from https://CRAN.R-project.org/package=deldir

Tyack, P. L., & Thomas, L. (2019). Using dose–response functions to improve calculations of the impact of anthropogenic noise. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *29*(S1), 242–253. https://doi.org/10.1002/aqc.3149

Tyack, P. L., Zimmer, W. M. X., Moretti, D., Southall, B. L., Claridge, D. E., Durban, J. W., . . . Boyd, I. L. (2011). Beaked Whales Respond to Simulated and Actual Navy Sonar. *PLoS ONE*, *6*(3), e17009. https://doi.org/10.1371/journal.pone.0017009

Wickham, H., & Bryan, J. (2019). *Readxl: Read excel files*. Retrieved from https://CRAN.R-project.org/package=readxl

Wickham, H., & Henry, L. (2020). *Tidyr: Tidy messy data*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Hester, J., & Francois, R. (2018). *Readr: Read rectangular text data*. Retrieved from https://CRAN.R-project.org/package=readr

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95–114. https://doi.org/10.1111/1467-9868.00374

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(3), 495–518.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.).

461   Chapman; Hall/CRC.

462   Wood, S. N., Li, Z., Shaddick, G., & Augustin, N. H. (2017). Generalized Additive Models
463   for Gigadata: Modeling the U.K. Black Smoke Network Daily Data. *Journal of the*
464   *American Statistical Association, 112*(519), 1199–1210. https://doi.org/10.1080/01621459.
465   2016.1195744

466   Xie, Y., Allaire, J., & Grolemund, G. (2018). *R markdown: The definitive guide.* Retrieved
467   from https://bookdown.org/yihui/rmarkdown