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

Eiren K. Jacobson[1], E. Elizabeth Henderson[2], David L. Miller[1], Cornelia S. Oedekoven[1], 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*

Correspondence:

Eiren Jacobson
University of St Andrews
The Observatory
Buchanan Gardens
St Andrews
Fife
KY16 9LZ
Scotland
Email: eiren.jacobson@st-andrews.ac.uk

Draft 08 December 2021

Abstract

Behavioral responses of beaked whales (family Ziphiidae) to naval use of mid-frequency active sonar (MFAS) have been quantified for some species and regions. We describe the effects of MFAS on the probability of detecting diving groups of Blainville's beaked whales on the US Navy Pacific Missile Range Facility (PMRF) in Hawaii and compare our results to previously published results for the same species at the Atlantic Undersea Test and Evaluation Center (AUTEC) in the Bahamas. We use passive acoustic data collected at bottom-mounted hydrophones before and during six naval training exercises at PMRF along with modelled sonar received levels to describe the effect of training and MFAS on foraging groups of Blainville's beaked whales. We use a multi-stage generalized additive modelling approach to control for the underlying spatial distribution of vocalizations under baseline conditions. At an MFAS received level of 150 dB re 1 μPa the probability of detecting groups of Blainville's beaked whales decreases by 61% (95% CI 56%-65%) compared to periods when general training activity was ongoing and by 88% (95% CI 87%-92%) compared to baseline conditions. Our results indicate a more pronounced response to naval training and MFAS than has been previously reported. [196/200]

KEYWORDS

Blainville's beaked whales, *Mesopolodon densirostris*, mid-frequency active sonar, passive acoustic data, behavioral response, generalized additive model

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 et al., 2004; Macleod and D'Amico, 2006) and are sensitive to anthropogenic noise (Southall et al., 2016). Multiple mass strandings of beaked whales have been associated with high-intensity anthropogenic sound sources including naval sonar (Bernaldo de Quirós et al., 2019; D'Amico et al., 2009). These acute events have motivated research into whether and how beaked whales respond to different types and intensities of anthropogenic noise (e.g., Aguilar de Soto et al., 2006; Cholewiak et al., 2017; Tyack et al., 2011). Anthropogenic sound can disrupt the foraging dive cycles of beaked whales (Falcone et al., 2017), potentially leading to cumulative sublethal impacts resulting from reduced foraging opportunities (New et al., 2013; Pirotta et al., 2018), or to symptoms similar to decompression sickness that can lead to injury or death (Hooker et al., 2009, 2012).

Echolocation clicks produced by diving groups of Blainville's beaked whales indicate foraging activity and can be recorded by hydrophones (Johnson et al., 2006). Research on Blainville's beaked whales *(Mesoplodon densirostris)* using data from bottom-mounted hydrophones on a U.S. Navy range in the Bahamas has shown decreases in time spent foraging and movement 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 (U.S. Department of the Navy, 2018). Most research has focused on the impacts of mid-frequency active sonar (MFAS) broadcast from naval vessels. Separately, researchers have shown that, in the absence of MFAS, beaked whales may alter their behavior in response to vessel noise (Aguilar de Soto et al., 2006; Pirotta et al., 2012).

The U.S. Navy is interested in quantifying the effects of sonar on beaked whales for the purpose of risk assessments and permitting associated with training activities (e.g., U.S. Department of the Navy, 2017). There are different experimental and analytical ways of quantifying responses to sonar (see Harris et al., 2018 for a review). Here, we focus on analyses of observational data from cabled hydrophone arrays collected concurrently with naval training exercises. Examples of these from previous studies include McCarthy et al. (2011) who used data from the cabled hydrophone array at the U.S. 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 or after. Building on this work, Moretti et al. (2014) used a GAM to examine the presence or absence of GVP starts within 30-min periods (i.e., whether or not a GVP started within each 30-min period) on the AUTEC range as a smooth function of MFAS received level. They 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 1 μPa, which they interpreted as a 50% probability of disturbance.

Our primary objective was to replicate the effort of Moretti et al. (2014) with the same species on a different U.S. Navy training range in a different oceanic environment. 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). Passive acoustic detections of the presence or absence of GVP starts within 30-min periods were collected via a cabled hydrophone array at PMRF before and during training exercises involving MFAS broadcast from navy ships.

Unlike AUTEC, which is situated in a deep isolated basin surrounded by steep slopes, the Pacific Missile Range Facility (PMRF) in Hawaii is located on the side of an ancient volcano, with a steep slope to the deep ocean floor. Previous work in this region has shown that

Blainville's beaked whales are present year-round at this site, prefer sloped habitats, and that acoustic detections decrease during multi-day training events involving MFAS (Henderson et al., 2016; Manzano-Roth et al., 2016). As we expected the density of Blainville's beaked whales at PMRF to be low and spatially variable, our methods needed 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.

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. Navy 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 six Submarine Command Courses (SCCs) at PMRF. SCCs are training exercises that involve several different naval platforms, occur biannually in February and August, and typically last 6-7 days. MFAS is broadcast from naval vessels during part of the training exercises. Acoustic recordings were made for a minimum of two days before each SCC as well as during the exercise. During data collection, hydrophones sampled at a rate of 96 kHz. Up to 62 hydrophones were recorded simultaneously by the Naval Information Warfare Center (NIWC).

A beaked whale echolocation detector from the Navy Acoustic Range WHale AnaLysis (NARWHAL) algorithm suite (Martin et al., 2020) was run on the recordings. This detector first compared signal-to-noise ratio (SNR) thresholds within the expected frequency range of beaked whale clicks (16-44 kHz) versus the bandwidth outside the click in a running 16,384-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 (*MATLAB*, 2017) to group detections of individual beaked whale echolocation clicks into GVPs (Henderson et al., 2016). 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. We used half-hour periods to approximate the typical vocal period of Blainville's beaked whales during deep foraging dives (Tyack et al., 2006).

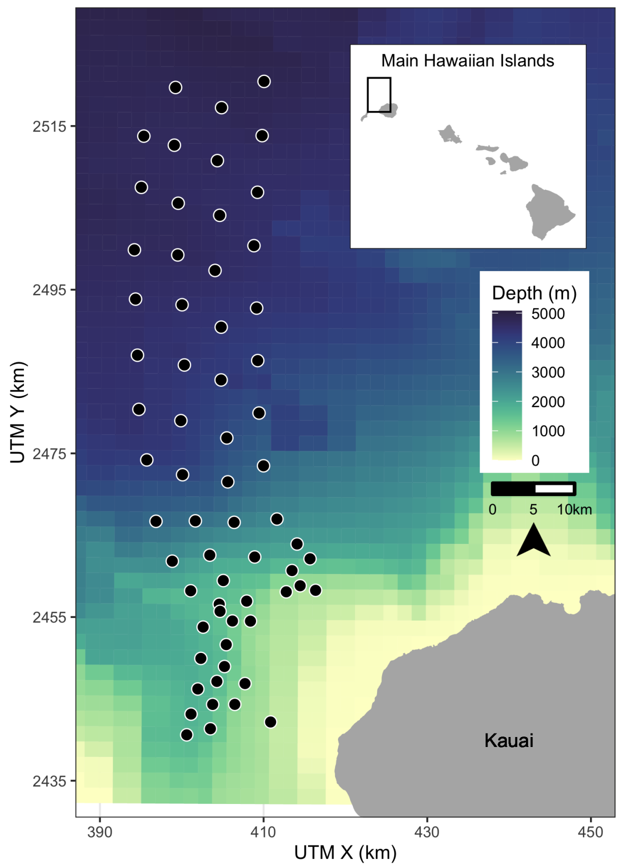


Figure 1: Map of hydrophones (black points) at the Pacific Missile Range Facility near the island of Kauai, Hawaii. For security reasons, the approximate rather than exact locations are shown here. Color scale indicates bathymetry. Inset map shows range location (black rectangle) relative to the main Hawaiian Islands.

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

For security reasons, classified data regarding activity that occurred on the range during each SCC was passed from PMRF to one author with clearance (E.E.H.). These data indicated the locations of the ships during the training periods and the start and stop times of each individual training event. However, no information was provided on the start and stop of sonar use; hence, periods of active sonar were determined from the range hydrophone recordings by running a sonar detector from the NARWHAL algorithm suite tuned to MFAS.

The hydrophone recordings cannot reliably be used to determine received level when the received level exceeds 140 dB re 1 μPa due to voltage constraints at the analog to digital recorder interface. Additionally, the hydrophones are mostly 4-5 km deep, whereas Blainville's beaked whales begin clicking when they have reached depths of approximately 200-500 m and spend most of their foraging dive at depths of 1-1.5 km (Johnson et al., 2004, 2006; Madsen et al., 2013). Therefore, we used an acoustic modeling approach to estimate the maximum received level of hull-mounted MFAS during each half-hour period around the location of each hydrophone at a depth of 1,000 m.

First, the locations of all surface ships were noted at the start of each half-hour period and the closest ship to each hydrophone was determined. MFAS propagation was modelled using the parabolic equation propagation model in the program Peregrine (OASIS, Heaney and Campbell, 2016). Acoustic transmission loss was estimated using a 200 Hz band around the center frequency of the sonar (3.5 kHz). A nominal source level of 235 dB re 1 μPa @ 1 m was assumed (U.S. Department of the Navy, 2018). 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 and converted to received levels based on the source level of the sonar. The maximum modeled received level along that radial was determined for each hydrophone and half-hour period. However, if the distance between the ship and the hydrophone was less than the depth of the water column, the parabolic equation would overestimate transmission loss at that angle. In these cases, a simple sonar equation was used to estimate transmission loss instead (Urick, 1983). For hydrophones shallower than 1,000 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 1,000 m the received level was estimated at a depth of 1,000 m with a +/- 10 m buffer. This process resulted in an estimate of received level for each hydrophone and half-hour period. Uncertainty in the modeled received levels was not considered.

2.2 Spatial Modelling

Summary

We first used tessellation to determine the area effectively monitored by each hydrophone (section 2.2.1). Then, we used a three-stage GAM approach to control for the underlying spatial distribution of Blainville's beaked whales when modelling the effects of training activities and of MFAS. For the first model (`M1`), 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 (2.2.2). We used the predicted values from this first model as an offset in a second model (`M2`) created using data from when naval activity was present on the range, but MFAS was not (2.2.3). We then used the predicted values from this second model as an offset in a third model (`M3`) created using data when naval activity and MFAS were present on the range (2.2.4). 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 (2.2.5). Analyses were undertaken in `R` (R Core Team, 2018). Code and data are available at https://github.com/eirenjacobson/MdMFASResponsePMRF.

2.2.1 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 GVPs 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. This meant that some hydrophones may have detected more GVPs because they were further away from other hydrophones, not because they were located in higher-density areas. To account for this, we used a Voronoi tessellation implemented in the `R` package `deldir` (Turner, 2019) to define a tile for each hydrophone that contained all points on the range that were closest to that hydrophone. We assumed that beaked whale groups occur within the tessellation tile of the hydrophone to which the GVP is assigned, and that the area of each tessellation tile influences the GVP detection rate at that hydrophone. For hydrophones on the outside of the range, i.e., not surrounded by other hydrophones, we used a cutoff radius of 7 km to bound the tessellation tiles. This distance was based on the estimated maximum detection distance of individual Blainville's beaked whale clicks at a U.S. Naval range in the Bahamas (Marques et al., 2009). Different combinations of hydrophones were used during different SCCs, so separate tessellations were created for each SCC.

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

In the first model, 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. Because of the way that GVPs were assigned to hydrophones (see Section 2.1.1) the data were not continuous in space. To account for this, we used a Markov random field (MRF) implemented in the `R` package `mgcv` (Wood, 2017) to model the spatial distribution of GVP detections. Markov random fields (Rue and Held, 2005) model correlation in space between discrete spatial units (henceforth, "tiles"). The correlation between two tiles is dictated by distance, as measured by the number of other tiles one needs to pass through to travel between two tiles ("hops"); correlation is strongest between a tile and its direct neighbors (those tiles it shares a border with) and decreases with additional hops. This was appropriate for our data as we did not know where in each tile a given GVP occurred, but we assumed that it did occur in that tile.

We modelled the probability of a GVP at tile i during SCC s ($\mu_{\texttt{M1},i,s}$) as a Bernoulli random variable. The linear predictor (on the logit scale) was given as:

$$\text{logit}\,(\mu_{\texttt{M1},i,s}) = \beta_{\texttt{M1},0} + f(\texttt{MRF}_{i,s}) + f(\texttt{Depth}_i) + \log_e A_{i,s} \qquad (1)$$

where $\beta_{\texttt{M1},0}$ is an intercept, $f(\texttt{MRF}_{i,s})$ denotes the Markov random field used to smooth space in the s^{th} SCC, $f(\texttt{Depth}_i)$ is a smooth of depth at the location of each hydrophone (using a thin plate spline; Wood (2003)) and $\log_e A_{i,s}$ is an offset for the area (in km^2) of each tile for each SCC, $A_{i,s}$. The offset term accounts for changes in probabilities of GVP detection due to the different areas monitored by each hydrophone. Because the hydrophone tessellation changed between SCCs (as there were different sets of hydrophones recorded during each SSC), separate MRFs were used for each SCC, but a single smoothing parameter was estimated across all MRFs. This allowed for different spatial smooths for each SCC, but constrained the smooths to have the same amount of wiggliness. The smooth of depth was shared across SCCs. We used this model to predict the baseline probability of a GVP detection at each hydrophone.

2.2.3 M2: Modelling the effect of Naval activity

For the second model, we used data collected prior to the onset of hull-mounted MFAS used during SCCs, when other naval training activities occurred at PMRF. Various vessels were present on the range during these periods, and other noise sources, including range tracking pingers, torpedoes, and submarines, may have been present. We used data collected when training activity was present on the range, but hull-mounted MFAS was not used, to model the effect of general naval activity on beaked whale GVPs.

We used the predicted baseline probability of a GVP detection at each hydrophone from `M1` as an offset to control for the underlying spatial distribution of GVPs. The model for the data when naval activity was present was intercept-only, with an offset derived from `M1`. This meant that the spatial distribution of GVPs was not allowed to change, but that we expected a uniform relative change in GVPs when naval activity was present. We again modelled the probability of GVP presence at tile i ($\mu_{\texttt{M2},i}$) as a Bernoulli random variable, with the following linear predictor:

$$\text{logit}\left(\mu_{\texttt{M2},i,s}\right) = \beta_{\texttt{M2},0} + \log_e \xi_{\texttt{M1},i,s}, \qquad (2)$$

where $\beta_{\texttt{M2},0}$ is an intercept and $\xi_{\texttt{M1},i,s}$ is the prediction (on the logit scale) for tile i during SCC s using model `M1`, included as an offset term.

2.2.4 M3: Modelling the effect of hull-mounted MFAS

For the third model, we used data collected when hull-mounted MFAS was present on the range to model the effect of sonar on beaked whales. We excluded data collected during breaks in training activities when sonar was not being used. The probability of a GVP when sonar was present was modeled as a function of the maximum received level (modeled at each hydrophone for each half-hour period; see section 2.2.1). We assumed that as the maximum received level increased, the probability of dives decreased and modeled this using a monotonically decreasing smooth implemented in the `R` package `scam` (Pya and Wood, 2015). To ensure that the model predictions were the same at a maximum received level of 0 dB and when only naval activity was present, we did not include an intercept. The probability of GVP presence at tile i ($\mu_{\texttt{M3},i}$) was modelled as a Bernoulli random variable where the linear predictor was:

$$\text{logit}\left(\mu_{\texttt{M3},i,s}\right) = f(\texttt{MaxRL}_{i,s}) + \log_e \xi_{\texttt{M2},i,s}, \qquad (3)$$

where $f(\texttt{MaxRL}_{i,s})$ was modeled as a monotonic decreasing smooth, $\xi_{\texttt{M2},i,s}$ denotes the prediction (on the logit scale) for tile i during SCC s when naval training activities were present on the range using model `M2`.

2.2.5 Uncertainty propagation

We used posterior simulation (sometimes referred to as a parametric bootstrap, Wood et al., 2017) to propagate uncertainty through M1, M2, and M3. This consisted of sampling from the posterior distribution of the parameters for each model in turn, calculating predictions using these parameters and then refitting the subsequent model with updated offsets. We generated 5,000 sets of posterior samples. Following this procedure through from M1 to M2 to M3 incorporated uncertainty from each model in the final predictions of the probability of detecting a GVP given different combinations of covariates.

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. Based on the resulting final posterior distribution of results (for model M3) we used 2.5%, 50%, and 97.5% quantiles to obtain median predictions and credible intervals (CIs). Details of the procedure are given in Appendix S1.

2.2.6 Quantifying the change in probability of GVPs

Finally, we calculated the expected change in the probability of detecting a GVP at each hydrophone $P(\text{GVP})$ relative to either the probability of detecting a GVP when no general naval training activity was present and no MFAS was present ($\Delta_{\texttt{M3:M1}}$), or relative to probability of detecting a GVP when general naval training activity was present but no MFAS was present ($\Delta_{\texttt{M3:M2}}$).

Using the N_b posterior samples, we calculated the expected $P(\text{GVP})$ under each set of covariates as

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

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

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

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

For each received level we calcualted the 2.5th, 50th, and 97.5th quantiles of $\Delta_{\texttt{M3:M1}}$ and $\Delta_{\texttt{M3:M2}}$ to create 95% CIs of change in $P(\text{GVP})$ across possible received levels.

3 Results

3.1 Data Collection and Processing

Data were collected before and during six SCCs: two each in 2013, 2014, and 2017 (Table 1). The number of hydrophones for which recordings were available for each SCC varied from 49 to 61. A total of 190,928 30-min observations were made.

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 MFAS use. MFAS was used for 3-4 days during each training event.

Across all SCCs, hydrophones, and conditions, a total of 2,312 GVPs were identified. The

Table 1: Number of hydrophones (HPs) used and number of observations made (no. 30-min periods) during each Submarine Commander Course (SCC) before the exercise began, when naval activity was present, and when naval activity and mid-frequency active (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

average probability of detecting a GVP during each half-hour period was therefore 1.2%. The spatial distribution of GVPs differed during the pre-activity phases of SCCs (Fig. S2.1; top panel).

Modelled maximum received levels ranged from 38 to 186 dB re 1 μPa, with a median value when MFAS was present of 147 dB re 1 μPa. The intensity and spatial distribution of MFAS received levels varied across the range and across SCCs (Fig. S2.2).

Based on the observed data, the probability of detecting a GVP changed by -57% when general naval training activity was present compared to when naval activity was absent, by -47% when naval activity and MFAS were present compared to when only naval activity was present, and by -77% when naval activity and MFAS were present compared to when neither naval activity nor sonar were present (Fig. S2.3). The highest modelled received level at which a GVP was observed was 164 dB re 1 μPa.

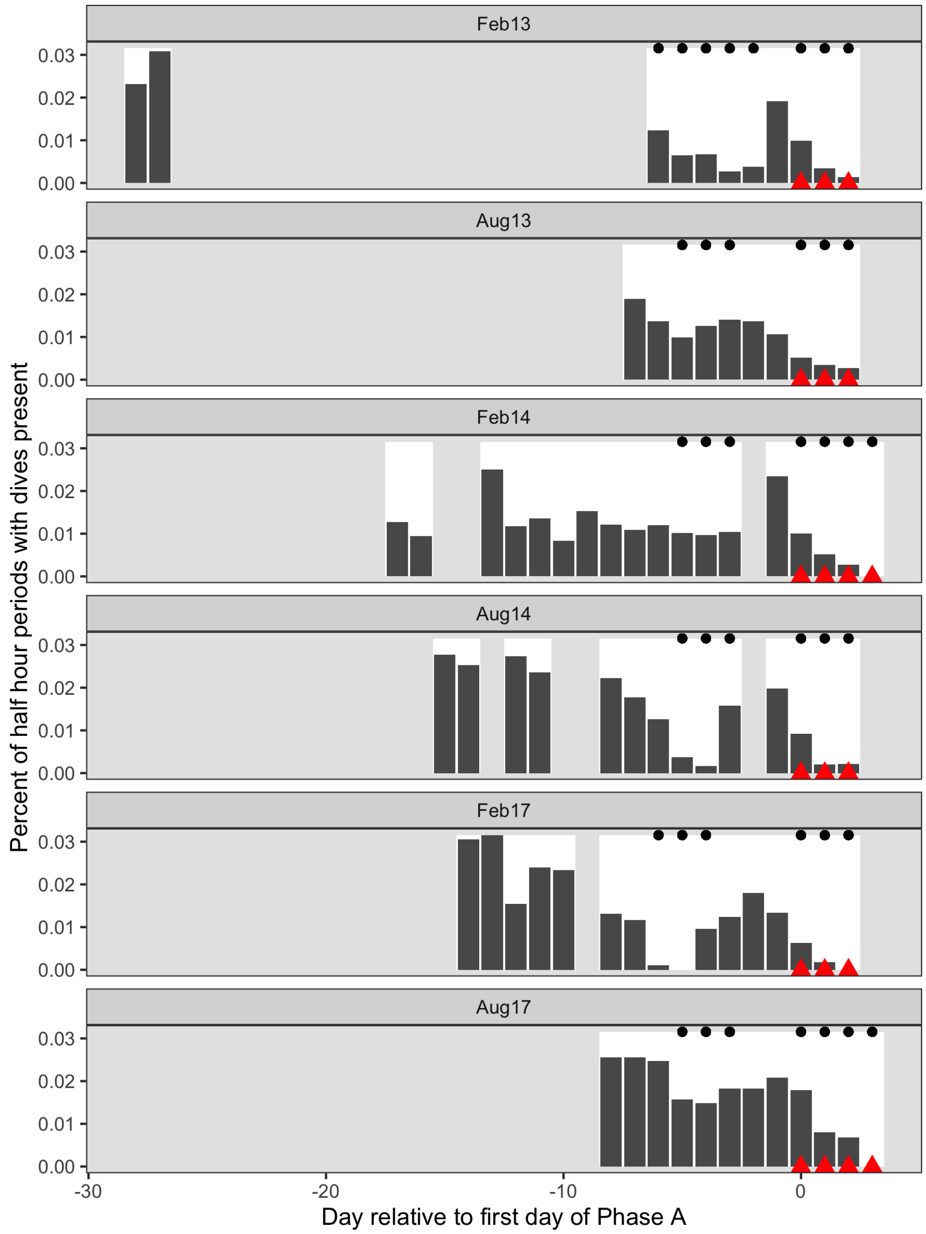


Figure 2: Time series of six recorded Naval training activities at the Pacific Missile Range Facility. The time series are aligned relative to the first day that mid-frequency active sonar (MFAS; red triangles) was used in each exercise (horizontal axis). Days with white background indicate days for which recordings and data were available. Dark gray bars indicate the proportion of 30-min periods on each day, across all hydrophones, when group vocal periods (GVPs) were detected (vertical axis). Black dots indicate days when naval training activity was present on the range.

3.2 Spatial Modelling

We created separate tessellations for each SCC (Fig. S2.4). In August 2017, data were available from fewer hydrophones, and so in some cases the tessellated tiles, with bounding radius of 6,500 m, did not completely cover the range. Hydrophone depths varied from approximately 650 to 4720 m.

`M1` fitted a spatial model of $P(\text{GVP})$ 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 at the $\alpha = 0.05$ level (p-value $<$ 2E-16), indicating that GVPs varied space. 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. S2.5). The model `M1` predicted highest $P(\text{GVP})$ at hydrophone depths between 1,500 and 2,000 m (Fig. 2).

`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 $P(\text{GVP})$ when naval activity was ongoing was significantly different from the baseline period at the $\alpha = 0.05$ level (p-value $<$ 2E-16). The expected $P(\text{GVP})$ decreased by a median of 61% (95% CI 58% - 64%) when naval 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 MFAS were present. This model used a monotonically decreasing spline on modelled MFAS received level (Fig. 2) and did not include an intercept term. The smooth on MFAS received level was significant at the $\alpha = 0.05$ level (p-value = 6.74E-10) and the model explained 12.4% of deviance in the data.

We did not make inference on sonar received levels below 100 dB re. 1 μPa because Blainville's beaked whales are unlikely to perceive MFAS below received levels of approximately 80 dB re. 1 μPa (Pacini et al., 2011) and because very little data (9 hrs, or 1% of the data collected when MFAS was present) were collected at received levels below 100 dB re. 1 μPa. Similarly, we did not make inference on sonar received levels above 165 dB re. 1 μPa because no GVPs were observed above this received level and therefore `M3` predicted $P(\text{GVP}) = 0$ (95% CI 0-1).

For MFAS received levels between 100 and 165 dB re. 1 μPa, change in $P(\text{GVP})$ was calculated relative to the pre-activity baseline period ($\Delta_{\texttt{M3:M1}}$; Fig. 4 left panel) and to the period when naval activity was present on the range ($\Delta_{\texttt{M3:M2}}$; Fig. 4 right panel) using the posterior samples. For illustration purposes, $\Delta_{\texttt{M3:M1}}$ and $\Delta_{\texttt{M3:M2}}$ calculated using five individual posterior samples are shown in Fig. S2.6. At a received level of 150 dB, the posterior median of $\Delta_{\texttt{M3:M1}}$ was -88% (95% CI -92% - -80%) and the posterior median of $\Delta_{\texttt{M3:M2}}$ was -68% (95% CI -79% - -50%). Relative to when only naval training is present, $\Delta_{\texttt{M3:M2}}$ predicts a 50% reduction in $P(\text{GVP})$ at a MFAS received level of 142 dB re 1 μPa.

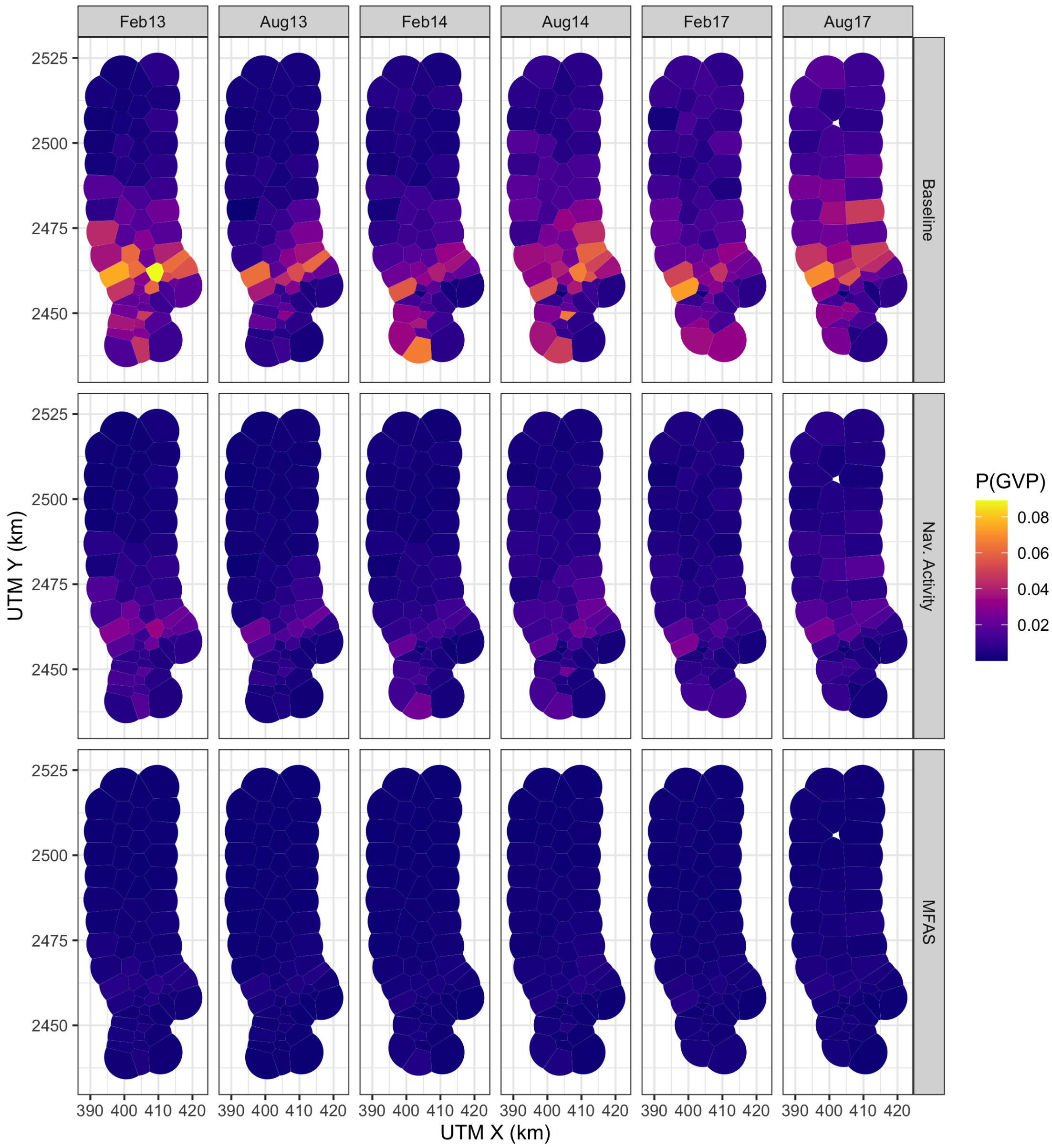


Figure 3: Map of expected probability of detecting a GVP (color scale) at each hydrophone during each SCC (columns) prior to the onset of naval training activity, during naval training activity when no MFAS was present, and during naval training activity when MFAS was present at a received level of 150 dB re 1 μPa rms (rows).

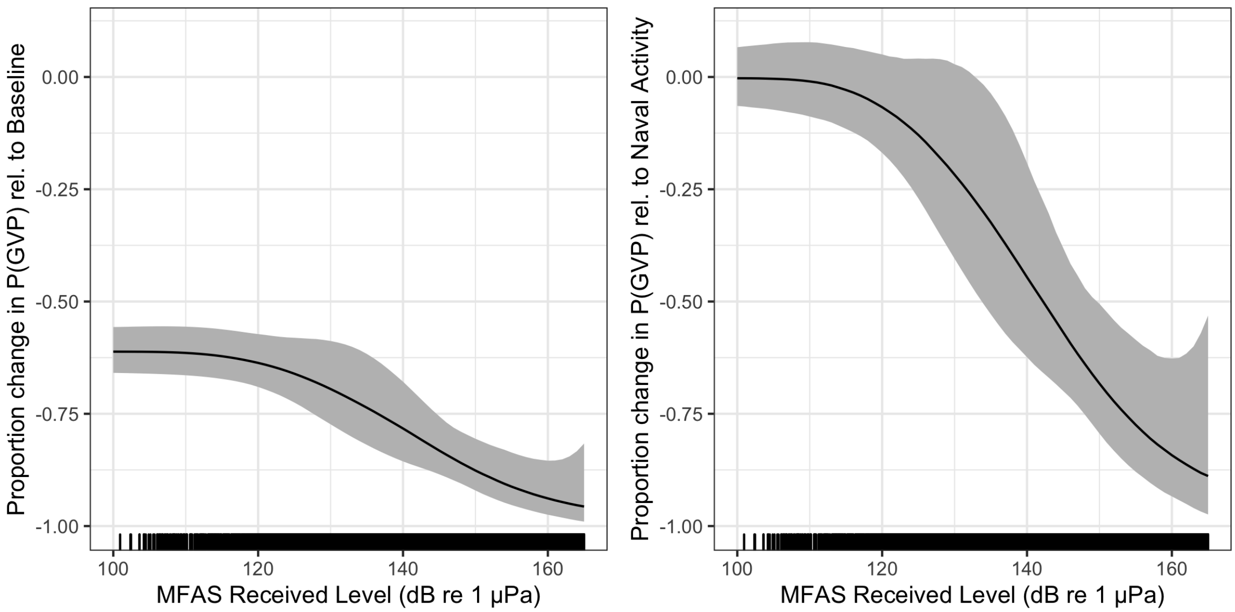


Figure 4: Median (black line) and 95% CI (gray shading) expected change in the probability of detecting a group vocal period (vertical axis) with increasing MFAS received level (horizontal axis) relative to when naval training activity but no MFAS was present on the range (left panel) and to when neither naval training activity nor MFAS were present on the range.

4 Discussion

We used a series of three linked models to quantify the response of Blainville's beaked whales to naval training exercises involving MFAS: the first model was fitted to pre-exercise baseline data, the second was fitted to data collected when naval training exercises were ongoing but no MFAS was present, and the third model was fitted to data collected during naval training exercises that used MFAS. We found that the probability of acoustic detections of Blainville's beaked whales decreased when both naval training exercises and naval training exercises using MFAS were present (Fig. 4).

The methods presented here are spatially explicit and account for the spatial confounding of animal distribution and naval training activity. The data used in this study are from an undesigned experiment, where the spatial intensity of the treatments (naval activity and MFAS) were not applied randomly with respect to either the study area or Blainville's beaked whale foraging activity. We did not want the spatial distribution of training exercises and MFAS to influence our understanding of the baseline spatial distribution of Blainville's beaked whales. Due to the spatial confounding of animal distribution and naval training activity at PMRF, fitting a single model to all of the data leads to greater uncertainty in estimating the impact of sonar, since changes in distribution due to MFAS could be explained as variability in spatial distribution by the MRF (Appendix S3). Our three-stage modelling approach addresses this issue while propagating uncertainty between the models. To our knowledge, this is a novel application of GAMs.

The analytical approach outlined in this article could be applied to other species, regions, and types of disturbance where experimental design is not possible. The use of Markov random fields for the spatial term is useful for cases where exact location data are not available, avoiding the inappropriate use of continuous-space smoothers. Shape-constrained smoothing (in our case, monotonically decreasing smooth) is also well-suited to the kind of data we modelled here, ensuring that values can only stay constant or decrease over time (or any other covariate). Finally, the use of a multi-stage posterior sampling scheme for quantifying uncertainty extends to other situations where multiple models are fitted and the results of one part feed into another. Simulation-based approaches such as these bypass the need to derive (often complex) mathematical expressions for variance (or shortcut them by assuming independence).

The expected change in the probability of a GVP when MFAS was present included CIs that reflect several sources of uncertainty (Fig. 4). The small number of GVPs when MFAS was present—and therefore sparse coverage of data points across the range of received levels—makes it difficult to estimate the effect of MFAS received level precisely. GVPs were detected in only 1.7% of half-hour periods in the baseline dataset, in 0.7% of periods (n =

323) when naval activity was present, and 0.2% (n = 48) when MFAS was present. Additional data—particularly at relatively low and relatively high source levels, where uncertainty is greatest—may reduce uncertainty in the expected probability of GVPs across different source levels. It is also possible that contextual factors that we did not include in this analysis, such as distance to sound source (DeRuiter et al., 2013; Falcone et al., 2017), may provide additional explanatory power and reduce uncertainty. Finally, the observed uncertainty may reflect true individual variation in response due to variables like age and sex (see Harris et al., 2018, sec. 2.2 for a review of relevant publications).

The model `M3`, which modelled the effect of received level on $P(\text{GVP})$, was constrained to be monotonically decreasing with no intercept term, so that the predicted $P(\text{GVP})$ would be the same or lesser when MFAS was present compared to when only naval training activity was present. However, it is possible that $P(\text{GVP})$ could be higher at relatively low MFAS received levels than when only naval training is present, since animals may move away from high-intensity areas, resulting in increased densities in lower-intensity areas. In our dataset, some hydrophones had lesser observed $P(\text{GVP})$ at low levels of MFAS and some had greater (Fig. S2.3). Due to small sample size at low intensities, we cannot determine whether observed increases in $P(\text{GVP})$ when MFAS was present at relatively low intensities was due to sampling error or to avoidance of high-intensity areas. The version of the model fitted as a single GAM (Appendix S3) predicted the change in $P(\text{GVP})$ to be > 0, i.e., increased relative to when only naval training activity was present, at MFAS received levels below 103 dB re 1 μPa (Fig. S3.1).

We excluded data collected between training activity within an SCC (13% of the available data) as we did not consider it to be true baseline data since naval activity and/or MFAS had recently (within hours or days) been present. It would be interesting to explore the complete dataset, including these interim periods, to investigate the timescales on which beaked whales respond to naval activity (e.g., Jones-Todd et al., 2021; Joyce et al., 2019). We might expect that time since training activity or MFAS could lead to recovery of $P(\mathrm{GVP})$ towards baseline levels, perhaps modulated by the cumulative exposure to training and MFAS.

In a regulatory context, a dose-response function as presented in Fig. 4 is often interpreted as representing the proportion of a population that responds (vertical axis) to a given received level (horizontal axis) (Tyack and Thomas, 2019). However, the metric used in this study—the change in the probability of detecting a GVP within a 30-min period—may not directly correspond to the proportion of the population that is affected. It may instead reflect a change in the proportion of time that all individuals in the population spent foraging in the study area. These two interpretations have different implications for understanding sub-lethal impacts of MFAS. In the former interpretation, given exposure to a certain received level, some of the population is affected and some of the population is not. In the latter, the entire exposed population is affected. With our data, we cannot distinguish between these possible scenarios.

In comparison to the risk function developed by Moretti et al. (2014) for Blainville's beaked whales at AUTEC, our risk function for PMRF predicts a more intense response to naval sonar. This may be because Moretti et al. were not able to account explicitly for the effects of naval training activities that did not include MFAS. Their baseline period consisted of 19 hours of data before the onset of MFAS; as at PMRF, it is likely that training activities during this period included sound sources other than MFAS. Therefore, their risk function is likely more analogous to our expected change in the probability of a detection when MFAS is present relative to when naval training activity was present (Fig. 4). Future research will investigate the specific causes of changes in the probability of detecting GVPs before the onset of MFAS. The reduction in detection of foraging dives could be a response to general naval training activity on the range, or to specific sound sources that have not previously been studied. Alternatively, it is possible that Blainville's beaked whales are semi-resident on the range and have become habituated to SCC activity; they may move off the range in anticipation of MFAS. Resident animals that are frequently exposed to training activity and transient animals that only encounter MFAS occasionally are likely to respond differently to sonar. It is not known how resident the Blainville's beaked whales are at PMRF, and offshore animals may be detected on the northern hydrophones.

Blainville's beaked whales occur in multiple ocean basins and have been studied on U.S. Navy training ranges in both the Atlantic (AUTEC) and the Pacific (PMRF) Oceans. The AUTEC range is located in a deep basin bounded to the south, east, and west by shallow waters and with maximum depths of 2,000 m. In contrast, the PMRF occurs across a steep slope and into deep water, over 5,000 m in depth. Although the environments at PMRF and AUTEC are different, the foraging dive behavior of Blainville's beaked whales is similar: dives occur in waters over steep slopes with gradients ranging from 3-23%, although dives occur in deeper waters (2,000-3,000 m, Henderson et al., 2016) at PMRF that at AUTEC (Hazen et al., 2011; 500-1,300 m, MacLeod and Zuur, 2005). Resident Blainville's beaked whales off the island of Hawaii also occur in slightly shallower waters than at PMRF, from 980-1,410 m (Baird, 2011; Baird et al., 2008). It seems likely the location of the mesopelagic scattering layer (indicating the presence of prey) along the slope that drives the location of Blainville's beaked whales rather than the bathymetric depth; this is supported by the fact that dive depths are similar across areas, occurring on average down to 1,050-1,150 m for 46-60 min (Baird et al., 2008; Joyce et al., 2017; Schorr et al., 2009). Documented responses to MFAS activity are comparable at both ranges, with individuals and groups moving to the periphery of the range or off the range and returning 2-4 days after the cessation of the sonar (Joyce et al., 2019; Manzano-Roth et al., 2016; McCarthy et al., 2011).

The similarities in Blainville's beaked whale behavioral responses to navy training activity across different ranges and environments at similar received levels may indicate the intrinsic nature of the response. The findings presented here and in Moretti et al. (2014) may be applicable to other species and regions, though species-specific dive behaviors and regional differences in oceanography likely modulate the impact of MFAS. For example, existing findings already demonstrate that Cuvier's respond in a similar manner by reducing their foraging dives and moving away from sonar sources (DeRuiter et al., 2013; Falcone et al., 2017). Conducting a similar analysis of Cuvier's beaked whale responses at the Southern California Anti-Submarine Warfare Range (SOAR) would further support our understanding of how different populations and species respond to naval sonar.

Acknowledgements

This study was funded by the US Navy Living Marine Resources Program (Contract No. N39430-17-P-1983).

Authors' Contributions

Conceptualization: E.E.H., D.J.M, L.T.

Data curation: E.K.J., E.E.H.

Formal analysis: E.K.J., E.E.H., C.S.O.

Funding acquisition: E.E.H., D.J.M., L.T., E.K.J.

Investigation: E.E.H.

Methodology: E.K.J., E.E.H., D.L.M., C.S.O., L.T.

Software: E.K.J., D.L.M., C.S.O.

Supervision: L.T.

Visualization: E.K.J., C.S.O.

Writing – original draft: E.K.J., E.E.H., D.L.M.

Writing – review & editing: E.K.J, E.E.H., D.L.M., C.S.O., D.J.M., L.T.

ORCID

Eiren K. Jacobson: https://orcid.org/0000-0003-0147-8367

E. Elizabeth Henderson: https://orcid.org/0000-0002-3212-1080

David L. Miller: https://orcid.org/0000-0002-9640-6755

Cornelia S. Oedekoven: https://orcid.org/0000-0002-5610-7814

Len Thomas: https://orcid.org/0000-0002-7436-067X

References

Aguilar de 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

Aguilar de Soto, N., Madsen, P. T., Tyack, P., Arranz, P., Marrero, J., Fais, A., Revelli, E., & 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

Baird, R. W. (2011). Short note: Open-ocean movements of a satellite-tagged Blainville's beaked whale (*Mesoplodon densirostris*): Evidence for an offshore population in Hawai'i? *Aquatic Mammals*, *37*(4), 506–511. https://doi.org/10.1578/AM.37.4.2011.506

Baird, R. W., Webster, D. L., Schorr, G. S., McSweeney, D. J., & Barlow, J. (2008). Diel variation in beaked whale diving behavior. *Marine Mammal Science*, *24*(3), 630–642. https://doi.org/10.1111/j.1748-7692.2008.00211.x

Bernaldo de Quirós, Y., Fernandez, A., Baird, R. W., Brownell, R. L., Aguilar de Soto, N., Allen, D., Arbelo, M., Arregui, M., Costidis, A., Fahlman, A., Frantzis, A., Gulland, F. M. D., Iñíguez, M., Johnson, M., Komnenou, A., Koopman, H., Pabst, D. A., Roe, W. D., Sierra, E., … Schorr, G. (2019). Advances in research on the impacts of anti-submarine sonar on beaked whales. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1895), 20182533. https://doi.org/10.1098/rspb.2018.2533

Cholewiak, D., DeAngelis, A. I., Palka, D., Corkeron, P. J., & Van Parijs, S. M. (2017). Beaked whales demonstrate a marked acoustic response to the use of shipboard echosounders. *Royal Society Open Science*, *4*(12), 170940. https://doi.org/10.1098/

rsos.170940

D'Amico, A., Gisiner, R., Ketten, D., Hammock, J., Johnson, C., Tyack, P., & Mead, J. (2009). Beaked whale strandings and naval exercises. *Aquatic Mammals*, *35*, 452–472. https://doi.org/10.1578/AM.35.4.2009.452

DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone, E. A., Friedlaender, A. S., Joseph, J. E., Moretti, D., Schorr, G. S., Thomas, L., & Tyack, P. L. (2013). First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. *Biology Letters*, *9*(4), 20130223. https://doi.org/10.1098/rsbl.2013.0223

Falcone, E. A., Schorr, G. S., Watwood, S. L., DeRuiter, S. L., Zerbini, A. N., Andrews, R. D., Morrissey, R. P., & 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

Harris, C. M., Thomas, L., Falcone, E. A., Hildebrand, J., Houser, D., Kvadsheim, P. H., Lam, F.-P. A., Miller, P. J. O., Moretti, D. J., Read, A. J., Slabbekoorn, H., Southall, B. L., Tyack, P. L., Wartzok, D., & Janik, V. M. (2018). Marine mammals and sonar: Dose-response studies, the risk-disturbance hypothesis and the role of exposure context. *Journal of Applied Ecology*, *55*(1), 396–404. https://doi.org/10.1111/1365-2664.12955

Hazen, E. L., Nowacek, D. P., St. Laurent, L., Halpin, P. N., & Moretti, D. J. (2011). The relationship among oceanography, prey fields, and beaked whale foraging habitat in the Tongue of the Ocean. *PLoS ONE*, *6*(4), e19269. https://doi.org/10.1371/journal.pone.0019269

Heaney, K. D., & Campbell, R. L. (2016). Three-dimensional parabolic equation modeling of mesoscale eddy deflection. *The Journal of the Acoustical Society of America*, *139*(2), 918–926. https://doi.org/10.1121/1.4942112

Henderson, E. E., Martin, S. W., Manzano-Roth, R., & Matsuyama, B. M. (2016). Occurrence and habitat use of foraging Blainville's beaked whales (*Mesoplodon densirostris*) on a U.S. Navy range in Hawaii. *Aquatic Mammals*, *42*(4), 549–562. https://doi.org/10.1578/AM.42.4.2016.549

Hooker, S. K., Baird, R. W., & Fahlman, A. (2009). Could beaked whales get the bends? Effect of diving behaviour and physiology on modelled gas exchange for three species: Ziphius cavirostris, Mesoplodon densirostris and Hyperoodon ampullatus. *Respiratory Physiology & Neurobiology*, *167*(3), 235–246. https://doi.org/10.1016/j.resp.2009.04.023

Hooker, S. K., Fahlman, A., Moore, M. J., Soto, N. A. de, Quirós, Y. B. de, Brubakk, A. O., Costa, D. P., Costidis, A. M., Dennison, S., Falke, K. J., Fernandez, A., Ferrigno, M., Fitz-Clarke, J. R., Garner, M. M., Houser, D. S., Jepson, P. D., Ketten, D. R., Kvadsheim, P. H., Madsen, P. T., ... Tyack, P. L. (2012). Deadly diving? Physiological and behavioural management of decompression stress in diving mammals. *Proceedings. Biological Sciences*, *279*(1731), 1041–1050. https://doi.org/10.1098/rspb.2011.2088

Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N., & Tyack, P. L. (2004). Beaked whales echolocate on prey. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*, S383–S386. https://doi.org/10.1098/rsbl.2004.0208

Johnson, M., Madsen, P. T., Zimmer, W. M. X., Soto, N. A. de, & Tyack, P. L. (2006). Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation. *Journal of Experimental Biology*, *209*(24), 5038–5050. https://doi.org/10.1242/jeb.02596

Jones-Todd, C. M., Pirotta, E., Durban, J. W., Claridge, D. E., Baird, R. W., Falcone, E. A., Schorr, G. S., Watwood, S., & Thomas, L. (2021). Discrete-space continuous-time models of marine mammal exposure to Navy sonar. *Ecological Applications*, *0*(0), e02475. https://doi.org/10.1002/eap.2475

Joyce, T. W., Durban, J. W., Claridge, D. E., Dunn, C. A., Fearnbach, H., Parsons, K. M., Andrews, R. D., & Ballance, L. T. (2017). Physiological, morphological, and ecological tradeoffs influence vertical habitat use of deep-diving toothed-whales in the Bahamas. *PLOS ONE*, *12*(10), e0185113. https://doi.org/10.1371/journal.pone.0185113

Joyce, T. W., Durban, J. W., Claridge, D. E., Dunn, C. A., Hickmott, L. S., Fearnbach, H., Dolan, K., & Moretti, D. (2019). Behavioral responses of satellite tracked Blainville's beaked whales (*Mesoplodon densirostris*) to mid-frequency active sonar. *Marine Mammal Science*, 1–18. https://doi.org/10.1111/mms.12624

Macleod, C. D., & D'Amico, A. (2006). A review of beaked whale behaviour and ecology in relation to assessing and mitigating impacts of anthropogenic noise. *Journal of Cetacean Research and Management*, *7*(3), 211–221.

MacLeod, C. D., & Zuur, A. F. (2005). Habitat utilization by Blainville's beaked whales off Great Abaco, northern Bahamas, in relation to seabed topography. *Marine Biology*, *147*(1), 1–11. https://doi.org/10.1007/s00227-004-1546-9

Madsen, P. T., Aguilar de Soto, N., Arranz, P., & Johnson, M. (2013). Echolocation in Blainville's beaked whales (*Mesoplodon densirostris*). *Journal of Comparative Physiology A*, *199*(6), 451–469. https://doi.org/10.1007/s00359-013-0824-8

Manzano-Roth, R., Henderson, E. E., Martin, S. W., Martin, C., & Matsuyama, B. (2016). Impacts of U.S. Navy training events on Blainville's beaked whale (*Mesoplodon densirostris*) foraging dives in Hawaiian waters. *Aquatic Mammals*, *42*(4), 507–518. https://doi.org/10.1578/AM.42.4.2016.507

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

Martin, C. R., Henderson, E. E., Martin, S. W., Helble, T. A., Manzano-Roth, R. A., Matsuyama, B. M., & Alongi, G. A. (2020). *FY18 annual report on Pacific missile range facility marine mammal monitoring.* Retrieved from Naval Information Warfare Center Pacific San Diego United States website: https://apps.dtic.mil/sti/citations/AD1091141

MATLAB. (2017). Natick, Massachusetts: The MathWorks Inc.

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

Moretti, D., Thomas, L., Marques, T., Harwood, J., Dilley, A., Neales, B., Shaffer, J., McCarthy, E., New, L., Jarvis, S., & Morrissey, R. (2014). A risk function for behavioral disruption of Blainville's beaked whales (*Mesoplodon densirostris*) from Mid-Frequency Active Sonar. *PLoS ONE*, *9*(1), e85064. https://doi.org/10.1371/journal.pone.0085064

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

Pacini, A. F., Nachtigall, P. E., Quintos, C. T., Schofield, T. D., Look, D. A., Levine, G. A., & Turner, J. P. (2011). Audiogram of a stranded Blainville's beaked whale (*Mesoplodon densirostris*) measured using auditory evoked potentials. *Journal of Experimental Biology*, *214*(14), 2409–2415. https://doi.org/10.1242/jeb.054338

Pirotta, E., Booth, C. G., Costa, D. P., Fleishman, E., Kraus, S. D., Lusseau, D., Moretti, D., New, L. F., Schick, R. S., Schwarz, L. K., Simmons, S. E., Thomas, L., Tyack, P. L., Weise, M. J., Wells, R. S., & Harwood, J. (2018). Understanding the population consequences of disturbance. *Ecology and Evolution*, *8*(19), 9934–9946. https://doi.org/10.1002/ece3.4458

Pirotta, E., Milor, R., Quick, N., Moretti, D., Di Marzio, N., Tyack, P., Boyd, I., & 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.

Schorr, G. S., Baird, R. W., Hanson, M. B., Webster, D. L., McSweeney, D. J., & Andrews, R. D. (2009). Movements of satellite-tagged Blainville's beaked whales off the island of Hawai'i. *Endangered Species Research*, *10*, 203–213. https://doi.org/10.3354/esr00229

Southall, B., Nowacek, D., Miller, P., & Tyack, P. (2016). Experimental field studies to measure behavioral responses of cetaceans to sonar. *Endangered Species Research*, *31*, 293–315. https://doi.org/10.3354/esr00764

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

Tyack, P. L., Johnson, M., Soto, N. A., Sturlese, A., & Madsen, P. T. (2006). Extreme diving of beaked whales. *Journal of Experimental Biology*, *209*(21), 4238–4253. https://doi.org/10.1242/jeb.02505

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., Clark, C. W., D'Amico, A., DiMarzio, N., Jarvis, S., McCarthy, E., Morrissey, R., Ward, J., & 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

U.S. Department of the Navy. (2017). *Criteria and thresholds for U.S. Navy acoustic and explosive effects analysis (phase III)*. Retrieved from https://www.goaeis.com/portals/goaeis/files/eis/draft_seis_2020/supporting_technical/Criteria_and_Thresholds_for_U.S._Navy_Acoustic_and_Explosive_Effects_Analysis_June2017.pdf

U.S. Department of the Navy. (2018). *Final environmental impact statement/overseas environmental impact ttatement Hawaii-Southern California training and testing*. Retrieved from https://www.hstteis.com/portals/hstteis/files/hstteis_p3/feis/section/HSTT_FEIS_3.07_Marine_Mammals_October_2018.pdf

Urick, R. J. (1983). *Principles of Underwater Sound* (Third Edition, Reprint 2013). New York: McGraw-Hill, Inc.

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. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman; Hall/CRC.

Wood, S. N., Li, Z., Shaddick, G., & Augustin, N. H. (2017). Generalized additive models for gigadata: Modeling the U.K. Black smoke network daily data. *Journal of the American Statistical Association*, *112*(519), 1199–1210. https://doi.org/10.1080/

01621459.2016.1195744

S1: Uncertainty estimation details

We used posterior simulation to propagate uncertainty through `M1`, `M2`, and `M3`. Each model was fitted via restricted maximum likelihood (REML), so the resulting estimates were empirical Bayes estimates. In this case we generated 5,000 samples from the (approximately multivariate normal) posterior of the model parameters. We generated a sample of the model parameters, $\boldsymbol{\beta}^* \sim \text{MVN}(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\hat{\beta}})$, where $\hat{\boldsymbol{\beta}}$ is the estimate of the model coefficients and $\mathbf{V}_{\hat{\beta}}$ is the posterior covariance matrix. Here the $\boldsymbol{\beta}$ for each model included the coefficients for the smooth terms in the model and fixed effects (e.g., intercept) if present. We then used the matrix that maps the model parameters to the predictions on the linear predictor scale ($\mathbf{X}_p$; Wood et al. 2017; section 7.2.6), along with the inverse link function, to generate predictions for each posterior sample. Denoting the vector of predictions $\boldsymbol{\mu}^*$, we calculate as follows:

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

where g was the link function, $\boldsymbol{\eta}^*$ was the linear predictor and $\boldsymbol{\xi}$ was any offset used by this prediction. Variance estimates can be obtained by taking the empirical variance of the resulting predictions (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 take (N_b=5,000 here), let $\mathbf{X}_{p,\mathtt{M}j}$ for $j = 1, 2, 3$ be the matrix that maps coefficients to the predictions for model $\mathtt{M}j$. For N_b times:

1. Draw a sample from the posterior of M1: $\tilde{\boldsymbol{\beta}}_{\mathtt{M1}} \sim \text{MVN}(\hat{\boldsymbol{\beta}}_{\mathtt{M1}}, \mathbf{V}_{\mathtt{M1}})$.
2. Calculate a new offset for M2, $\tilde{\boldsymbol{\xi}}_{\mathtt{M1}} = \mathbf{X}_{p,\mathtt{M1}}\tilde{\boldsymbol{\beta}}_{\mathtt{M1}} + \log_e \mathbf{A}$.
3. Refit M2 with $\tilde{\boldsymbol{\xi}}_{\mathtt{M1}}$ as the offset, to obtain M2′.
4. Draw a sample from the posterior of M2′: $\tilde{\boldsymbol{\beta}}_{\mathtt{M2'}} \sim \text{MVN}(\hat{\boldsymbol{\beta}}_{\mathtt{M2'}}, \mathbf{V}_{\mathtt{M2'}})$
5. Calculate a new offset for M3, $\tilde{\boldsymbol{\xi}}_{\mathtt{M2}} = \mathbf{X}_{p,\mathtt{M2}}\tilde{\boldsymbol{\beta}}'_{\mathtt{M2}} + \tilde{\boldsymbol{\xi}}_{\mathtt{M1}}$ (predictions for the sonar data locations for M2′, when no sonar was present).
6. Refit M3 with offset $\tilde{\boldsymbol{\xi}}_{\mathtt{M2}}$ to obtain M3′.
7. Predict $\boldsymbol{\mu}_{\mathtt{M1'}}$, $\boldsymbol{\mu}_{\mathtt{M2'}}$, and $\boldsymbol{\mu}_{\mathtt{M3'}}$ over prediction grid and store them.

We then calculated summary statistics (means and variances) of the N_b values of $\boldsymbol{\mu}_{\mathtt{M1'}}$, $\boldsymbol{\mu}_{\mathtt{M2'}}$, and $\boldsymbol{\mu}_{\mathtt{M3'}}$ we generated. The empirical variance of the N_b values of $\boldsymbol{\mu}_{\mathtt{M3'}}$ gave the uncertainty, incorporating components from all three models. We took appropriate pointwise quantiles (e.g., 2.5^{th} and 97.5^{th} for a 95% interval) to form confidence bands for the functional relationships between sonar received level and estimated probability of detecting GVPs.

681 # S2: Supplementary Tables and Figures

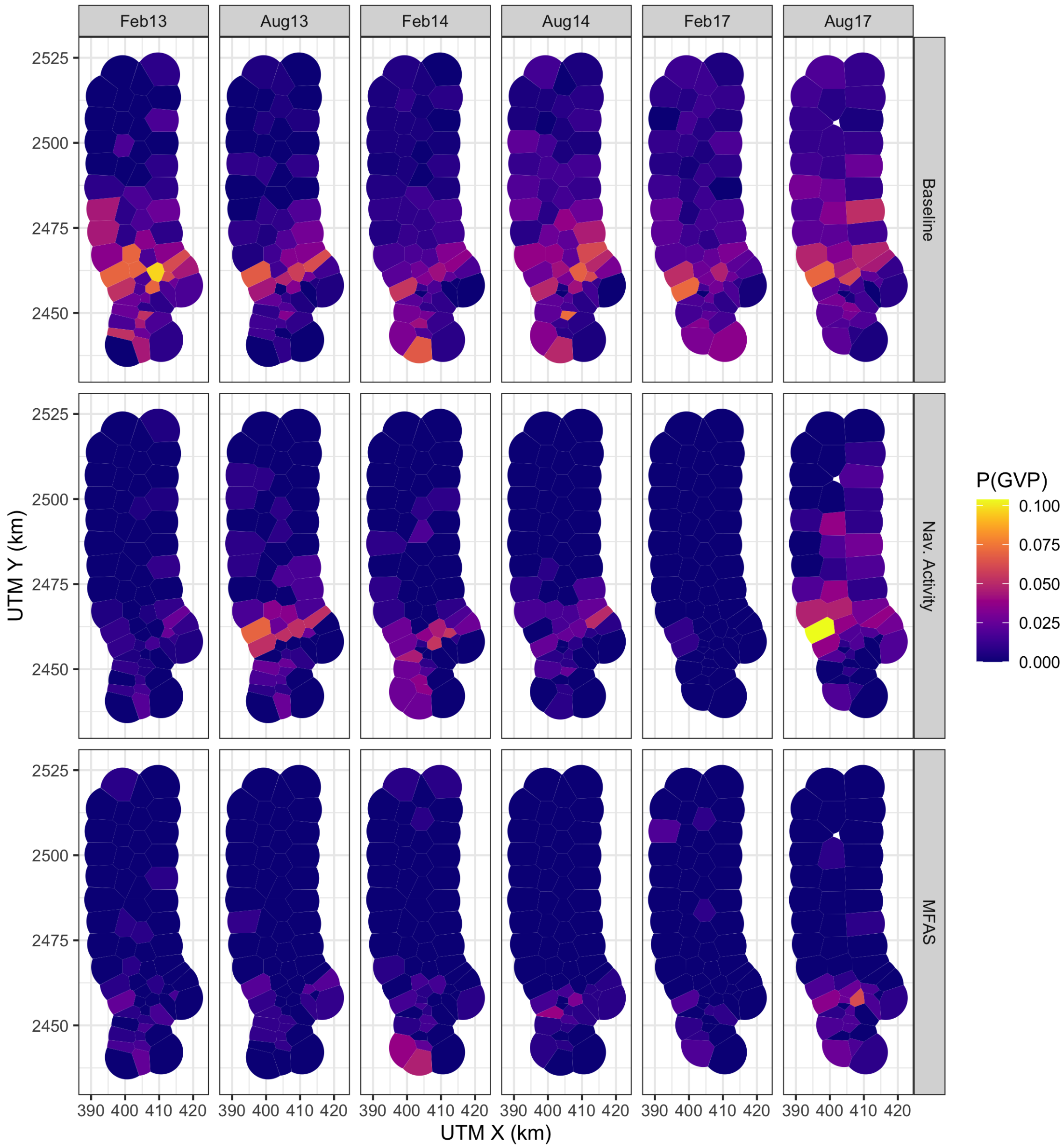


Figure S2.1: Map of observed probability of detecting a GVP at each hydrophone (color scale) during the baseline period, when naval activity was present, and when MFAS was present (rows) for each SCC (columns). Note that values of the probability of detecting a GVP are not corrected for effort (size of the hydrophone tile).

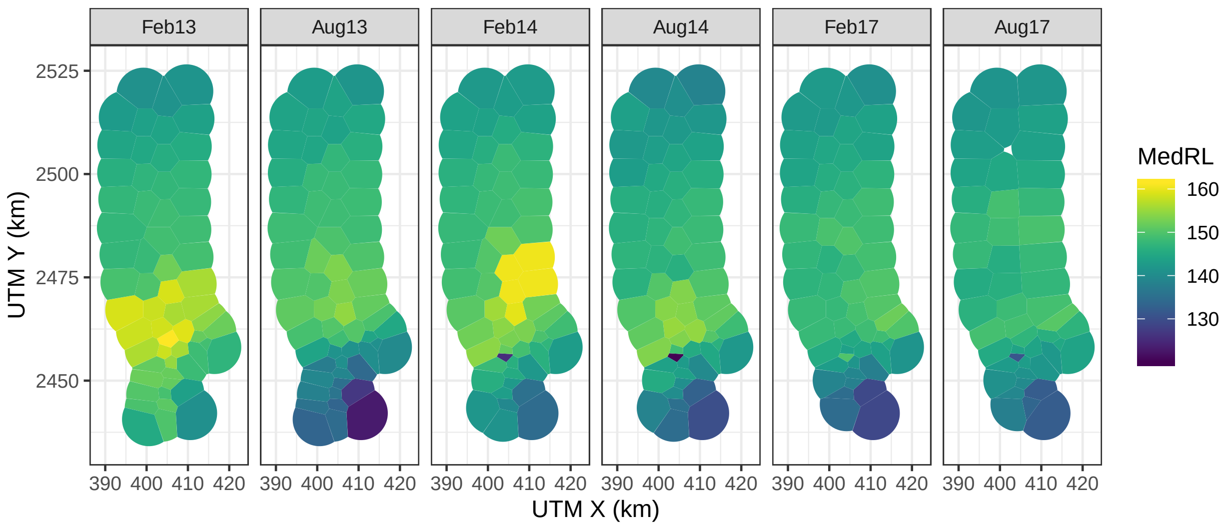


Figure S2.2: Median received level (dB re. 1 μPa) when MFAS was present (color scale) for all hydrophones and SCCs.

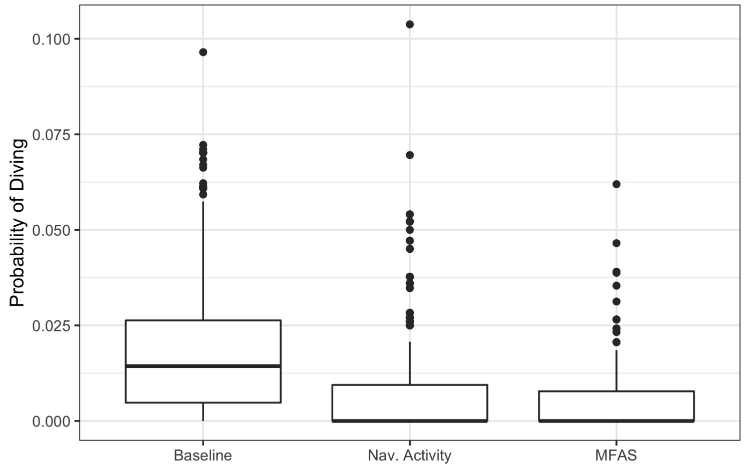


Figure S2.3: Boxplot of observed probability of a GVP for all hydrophones and SCCs (vertical axis) during baseline period, when naval activity was present, and when MFAS was present (horizontal axis). Each data point represents one hydrophone during one SCC and one phase of the training exercise.

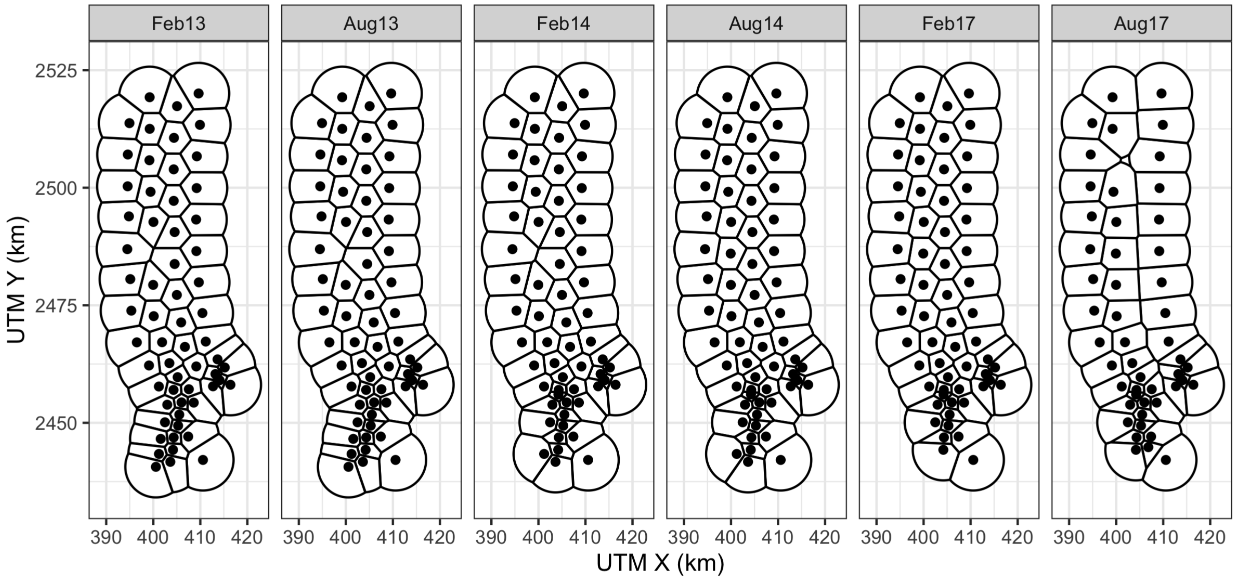


Figure S2.4: PMRF range tessellations for each of six recorded SCCs. Black lines indicate boundaries of hydrophone tiles. Black dots indicate approximate hydrophone locations.

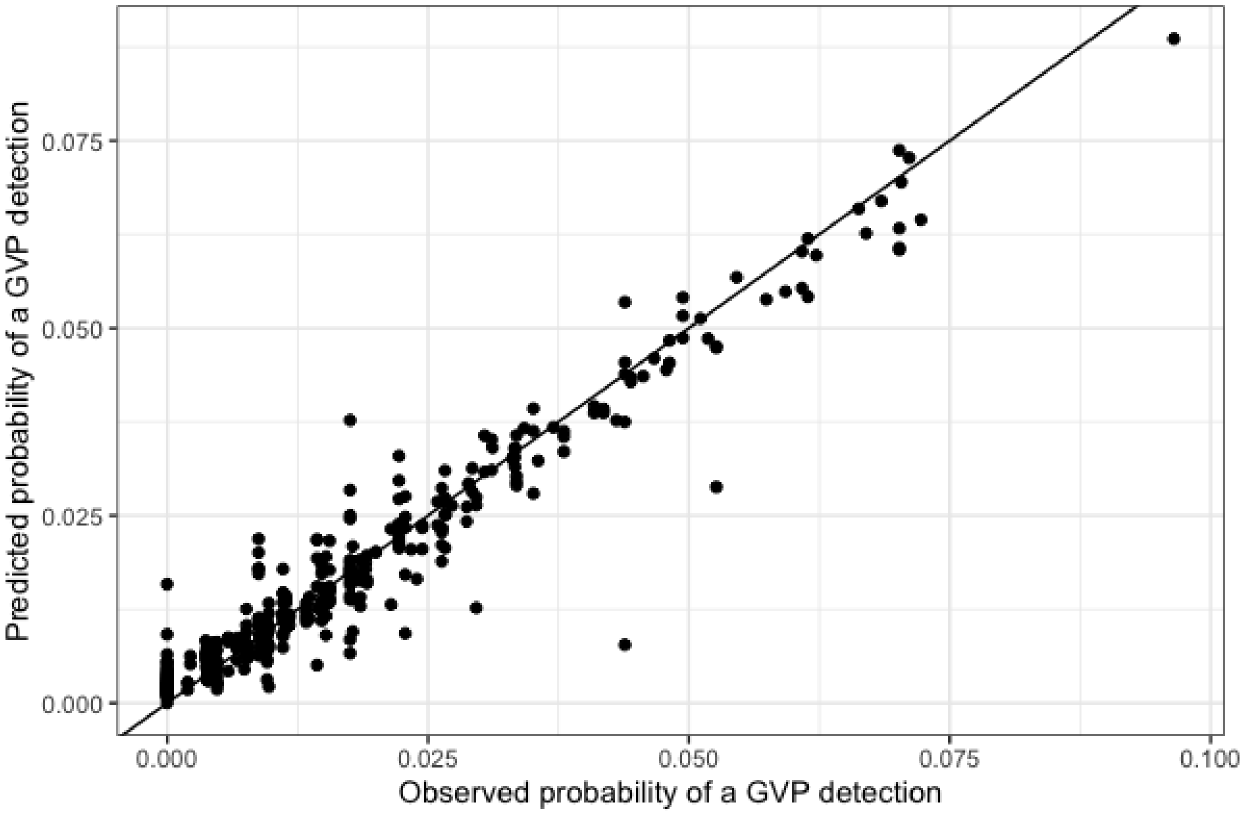
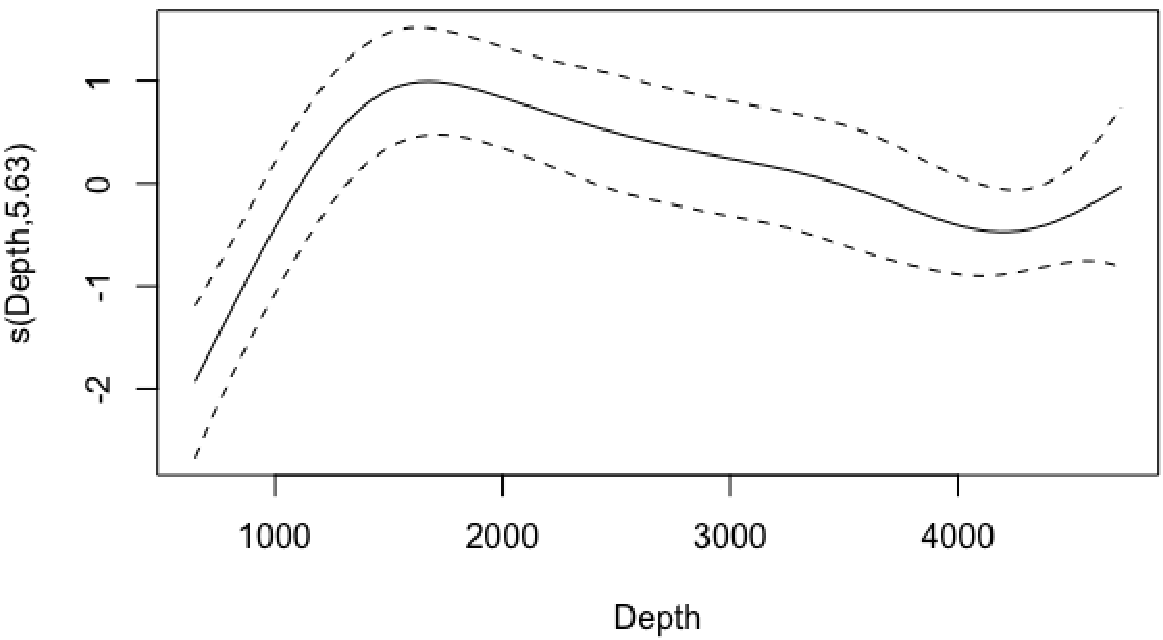


Figure S2.5: Observed (horizontal axis) versus M1 predicted (vertical axis) probability of detecting a GVP at each hydrophone during the baseline period.

\begin{figure}

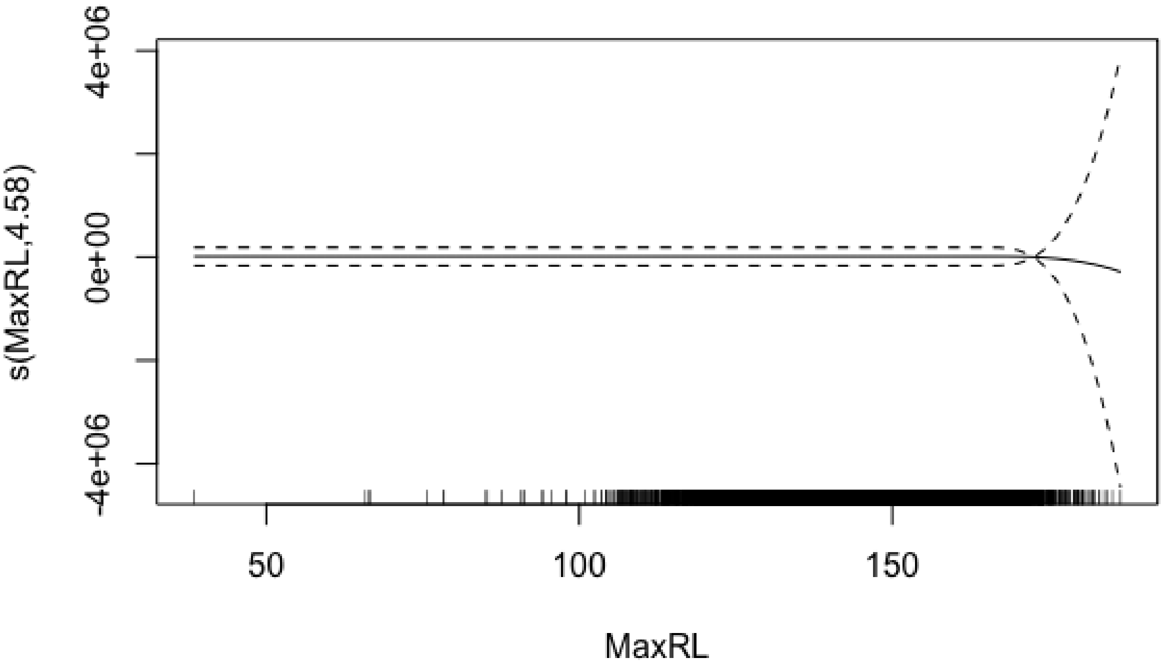


{

}

\caption{Spline for the relationship between P(GVP) and depth from M1 on the logit-link scale. Solid line: best fit; dashed lines: 95% CIs.} \end{figure}

\begin{figure}



{

}

\caption{Spline for the relationship between P(GVP) and maximum received level from M3 on the logit-link scale. Solid line: best fit; dashed lines: 95% CIs.} \end{figure}

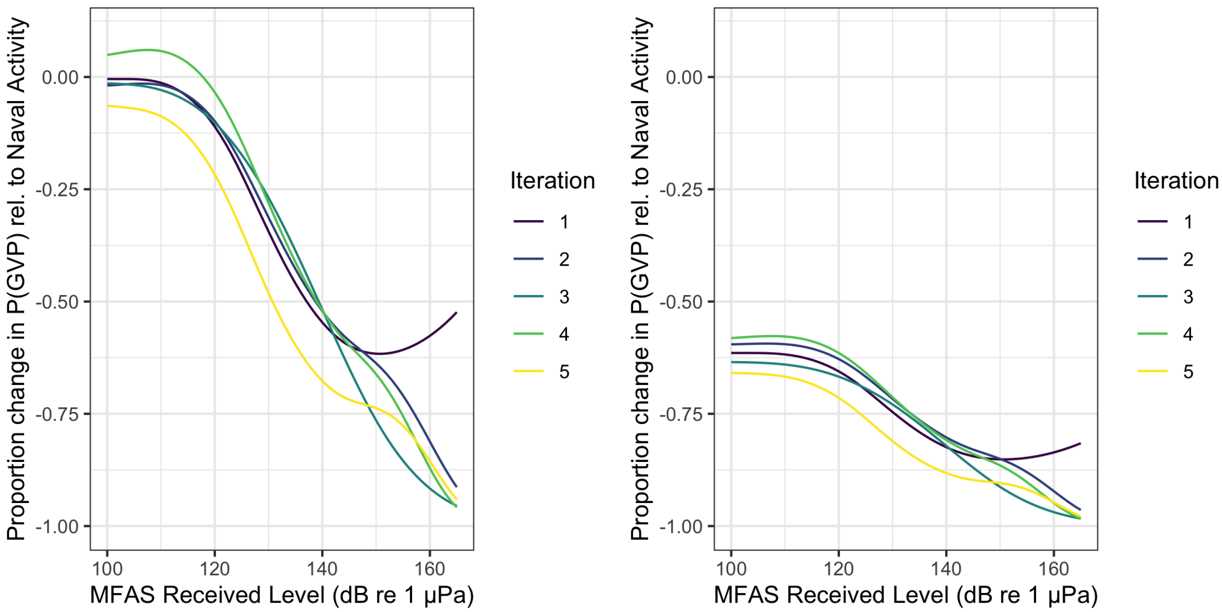


Figure S2.6: Example of five iterations (colored lines) of the 5,000 posterior samples of the expected change in the probability of detecting a group vocal period (vertical axis) with increasing MFAS received level (horizontal axis) relative to when naval training activity but no MFAS was present on the range (left panel) and to when neither naval training activity nor MFAS were present on the range.

S3: Single GAM

A single GAM could be used to quantify the effect of naval sonar on Blainville's beaked whales. Here, we present such a model and compare the results to the results obtained using the multi-stage model presented in the main text of the manuscript.

We modelled the probability of a GVP at tile i in SCC s at time t as a Bernoulli trial: $\texttt{GVP}_{i,s,t} \sim \text{Bin}(1, \mu_{i,s,t})$. The linear predictor on the logit scale was given as:

$$\text{logit}\,(\mu_{i,s,t}) = \beta_0 + \beta_1 \texttt{NavTrain}_t + f(\texttt{MRF}_{i,s}) + f(\texttt{Depth}_i) + f(\texttt{MaxRL}_i, t)\texttt{Sonar}_t + \log_e A_i)$$

where β_0 is an intercept, $\beta_1 \texttt{NavTrain}_t$ is the effect of naval training times an indicator variable for whether naval training was present or absent at time t, $f(\texttt{MRF}_{i,s})$ denotes the Markov random field used to smooth space, $f(\texttt{Depth}_i)$ is a smooth of depth (using a thin plate spline; Wood et al. 2003), $f(\texttt{MaxRL}_{i,t})\texttt{Sonar}_t$ is a smooth of sonar received level (using a thin plate spline) times an indicator variable for whether sonar was present or absent at time t, and $\log_e A_i$ is an offset for the area (in km^2) of each tile, A_i.

We fit the model to the same data used in `M1`, `M2`, and `M3` (see Methods section of main manuscript for details) using `mgcv` (Wood, 2017).

This single GAM (Fig. S3.1) predicts a 59% (95% CI 54%-64%) decrease in $P(\text{GVP})$ when naval training is present compared to the baseline period, whereas the multi-stage GAM (Fig. 4) predicts a decrease of 61% (95% CI 56%-65%). The single GAM predicts that at a MFAS received level of 150 dB re 1 μPa, $P(\text{GVP})$ decreases by 88% (95% CI 69%-96%) relative to when only naval training is present, whereas the multi-stage model predicts the same decrease of 88% with a narrower credible interval (95% CI 87%-92%). Relative to when only naval training is present, the single GAM predicts a 50% reduction in $P(\text{GVP})$ at a MFAS received level of 139 dB re 1 μPa (95% CI 118-159 dB re 1 μPa), whereas the multi-stage model predicts a 50% reduction at a MFAS received level of 142 dB re 1 μPa (95% CI

133-150 dB re 1 μPa).

The major difference between this single GAM and the multi-stage model presented in the main text of the manuscript is that here, the spatial smooth is constructed using data from the baseline, naval training, and MFAS periods of each SCC. Therefore, the spatial distribution of MFAS may influence the predicted distribution of Blainville's beaked whales. Using a single GAM leads to similar point estimates of the impact of sonar with greater uncertainty than the multi-stage model.

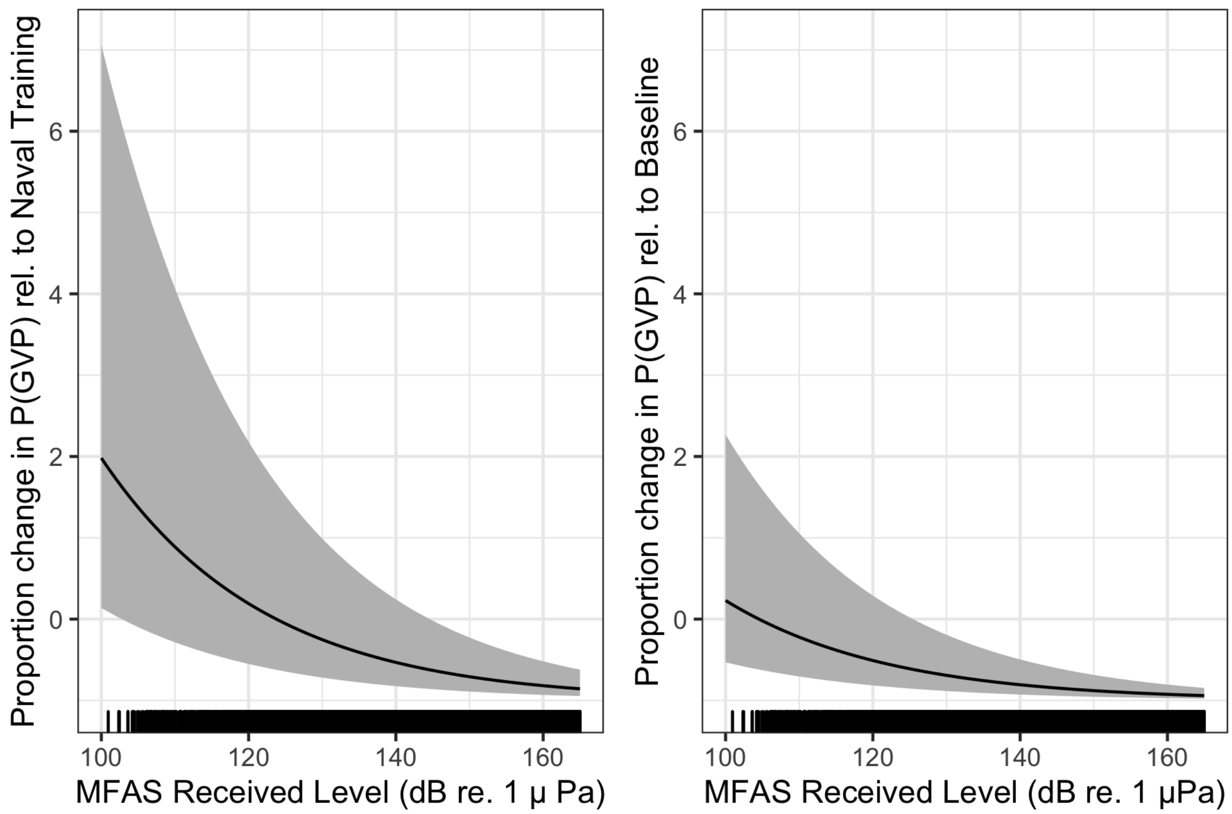


Figure S3.1: Results from a single GAM: Median (black line) and 95% CIs (gray shading) expected change in the probability of detecting a group vocal period (vertical axis) with increasing MFAS received level (horizontal axis) relative to when naval training activity but no MFAS was present on the range (left panel) and to when neither naval training activity nor MFAS were present on the range.