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 K. Jacobson
The Observatory
Buchanan Gardens
University of St Andrews
St Andrews
Fife
KY16 9LZ
Scotland

Email: eiren.jacobson@st-andrews.ac.uk

Draft 09 March 2021

1

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 using risk functions, which give the probability of a response as a function of covariates such as received level. We develop a novel risk function for Blainville's beaked whales (BBWs) on the US Navy Pacific Missile Range Facility (PMRF) in Hawaii and compare our risk function to one 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 along with modelled sonar received levels to describe the effect of training and MFAS on foraging groups of BBWs. We used a multi-stage generalized additive modelling (~~GAM~~) approach to control for the underlying spatial distribution of vocalizations under baseline conditions. At a MFAS received level of 150 dB re 1μPa the probability of detecting groups of BBWs decreased by 78% (95% CI 62%-100%) when compared to periods when general training activity was ongoing and by 92% (95% CI 87%-100%) when compared to baseline conditions. Our results indicate a more pronounced response to naval training and MFAS than has been previously reported. [200/200]

KEYWORDS

1 Introduction

[handwritten margin note: reference format? — author initials?? — too many authors instead of "et al"?]

Beaked whales (family Ziphiidae) are a group of deep-diving cetaceans that rely on sound to forage, navigate, and communicate (N. 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 foraging dive cycles of beaked whales (Falcone et al., 2017), potentially leading to cumulative sublethal impacts resulting from reduced foraging opportunities (L. F. New, Moretti, Hooker, Costa, and Simmons, 2013), or to symptoms similar to decompression sickness that can lead to injury or death (Bernaldo de Quirós et al., 2019). 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 naval vessels. Separately, researchers have shown that, in the absence of MFAS, beaked whales may alter their behavior in response to vessel noise (N. Aguilar de 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 (e.g., Navy, 2017). 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

68 before, during, and after naval training exercises involving MFAS. The authors used separate

69 generalized additive models (GAMs) for each period, and modelled the acoustic detection of

70 groups of Blainville's beaked whales (group vocal periods; GVPs) as a function of location on

71 the range and time. They found that the number of GVPs was lower during the exercises

72 than before or after.

73 Building on this work, Moretti et al. (2014) used a GAM to model the presence of acoustic

74 detections of groups of Blainville's beaked whales on the AUTEC range as a smooth function

75 of MFAS received level. They then compared the expected probability of detecting animals

76 when no sonar was present to the expected probability of detecting animals across sonar

77 received levels to estimate the probability of disturbance. They found that the probability of

78 detecting groups of Blainville's beaked whales was reduced by 50% at 150 dB re 1μPa, which

79 they interpreted as a 50% probability of disturbance.

80 Our primary objective was to replicate the effort of Moretti et al. (2014) with the same

81 species on a different US Navy training range in a different oceanic environment. Unlike

82 AUTEC, which occurs in a deep isolated basin surrounded by steep slopes, the Pacific Missile

83 Range Facility (PMRF) range in Hawaii is located on the side of an ancient volcano, with a

84 steep slope down to the deep ocean floor. Density of Blainville's beaked whales at PMRF is

85 lower and more variable than at AUTEC, so we wanted to explicitly account for differences

86 in underlying beaked whale presence across the range.

87 An additional objective was to isolate the effect of general training activity from the effect of

88 MFAS, so that beaked whale response to MFAS could be quantified relative to pre-training

89 baseline periods and to periods when general training activities were present on the range.

90 We used a spatially-referenced dataset of Blainville's beaked whale foraging dives recorded at

91 the PMRF off the island of Kauai, Hawaii (Fig. 1). Acoustic detections of Blainville's beaked

92 whales were collected via a cabled hydrophone array at PMRF before and during training

93 exercises involving MFAS broadcast from navy ships. Previous work in this region has shown

4

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 linked models were fitted. The first was fitted to data collected before the

training exercises began in order to estimate the underlying spatial distribution of acoustic

detections. The expected values from this model were used as offsets in the second model,

which was fitted to data collected when training exercises were ongoing by no MFAS was

broadcast. The expected values from this second model were then used as offsets in the

third model, which was fitted to data collected when training exercises that did use MFAS

were ongoing. Uncertainty was propagated through all models using posterior simulation.

With 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 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 six Submarine Command

Courses (SCCs) at PMRF. SCCs are training exercises that occur biannually in February

and August and typically last 6-7 days. Acoustic recordings were made for a minimum of two

5

117 days before each SCC as well as during the exercise. During data collection, hydrophones

118 sampled at a rate of 96 kHz, with the high pass filter on each hydrophone set at either 50 Hz,

119 100 Hz, or 10 kHz. Up to 62 of the range hydrophones were recorded simultaneously by the

120 Naval Information Warfare Center (NIWC).

121 A beaked whale detector from the Navy Acoustic Range WHale AnaLysis (NARWHAL)

122 algorithm suite (C. R. Martin et al., 2020) was run on the recordings. This detector first

123 compares signal-to-noise (SNR) thresholds within the expected beaked whale click frequency

124 range (16-44 kHz) versus the bandwidth outside the click in a running 16,384-pt fast Fourier

125 transform (FFT) spectrogram. The detected clicks were then passed to a 64-pt FFT stage

126 that measured power, bandwidth, slope, and duration characteristics to classify the clicks

127 to species. This process was followed by an automated routine in Matlab (*MATLAB*, 2019)

128 to group detections of individual beaked whale echolocation clicks into GVPs (Henderson,

129 Aschettino, Deakos, Alongi, and Leota, 2019). If a group of whales was detected by more

130 than one hydrophone, the GVP was assigned to the hydrophone that recorded the most clicks.

131 The data were then aggregated to indicate the presence or absence of the start of a GVP for

132 each hydrophone within each half-hour period.
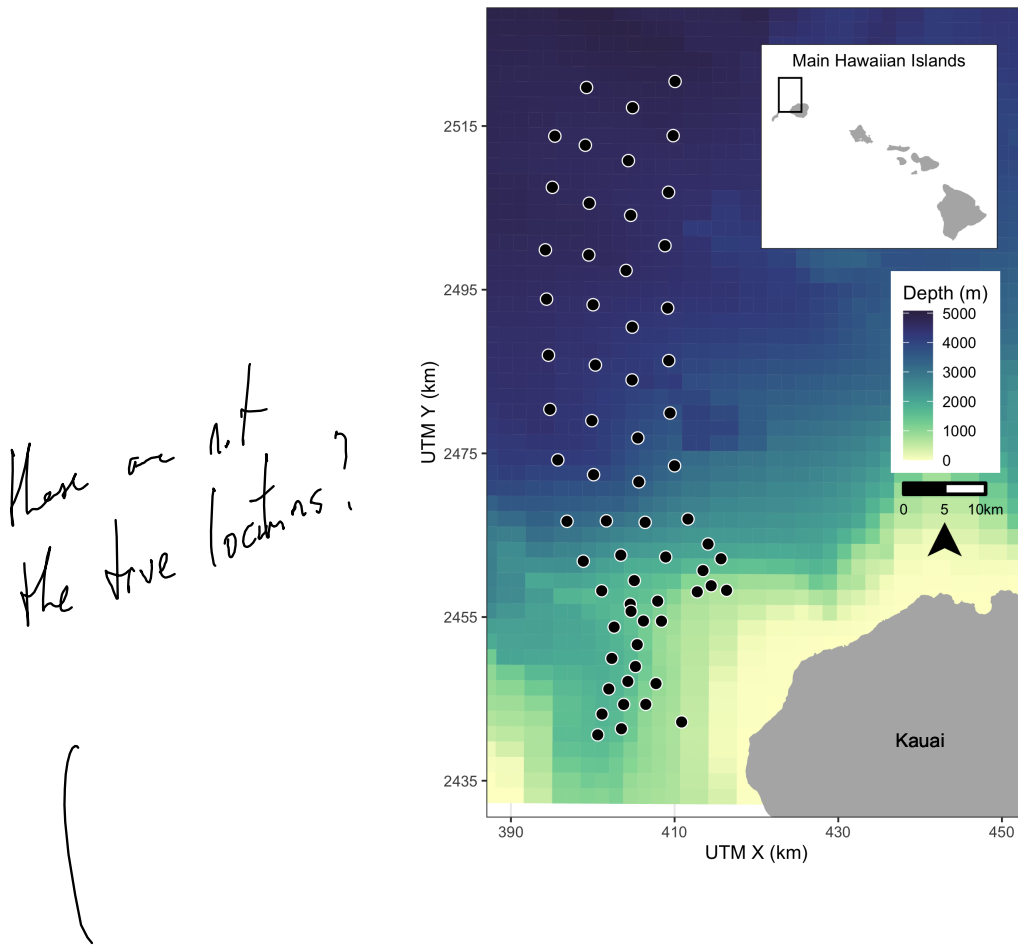
Here are not the true locations.?

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.

? r

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 be reliably used to determine received level when the received level exceeds XXX dB re. 1 μ Pa [LIZ Q: HOW DO WE SAY THIS? I WANT TO EXPLAIN WHY WE DIDN'T USE THE RECORDED LEVELS.]. Therefore, we used propagation modelling to estimate transmission loss and calculate the expected maximum received level of hull-mounted MFAS around the location of each hydrophone. First, the locations of all surface ships were noted for 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). Transmission loss was estimated using a 200 Hz band around the center frequency of the sonar (3.5 kHz). 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. [LIZ Q: WAS THE SOURCE LEVEL KNOWN OR ASSUMED? CAN WE SAY WHAT IT WAS?] 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. Transmission loss was estimated at depth since Blainville's beaked whales do not begin clicking until they have reached approximately 200-500 m depth of their foraging dive and spend most of their foraging dive at around 1,000-1,500 m (Johnson et al., 2004, Johnson, Madsen, Zimmer, Soto, and Tyack (2006), Madsen, Aguilar de Soto, Arranz, and Johnson (2013)). 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

We used a multi-stage generalized additive modelling (GAM; S. N. Wood, 2017) approach to control for the underlying spatial distribution of Blainville's beaked whales when modelling the effects of training activities and of MFAS. 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. We then used the predicted values from this second model as an offset in a third 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.

2.2.1 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. We account for this by determining the area effectively monitored by each hydrophone. ~~To do this,~~ 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

9

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 6,500 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, Thomas, Ward, DiMarzio, and Tyack, 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

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) 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 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 is 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 as a Bernoulli trial: $\texttt{GVP}_i \sim \text{Bin}(1, \mu_{\texttt{M1},i})$. The linear predictor ~~for the model was modelling~~ on the logit scale: *was given as:*

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

where $\beta_{\texttt{M1},0}$ is an intercept, $f(\texttt{MRF}_{i,s})$ denotes the Markov random field used to smooth space in SCC s, $f(\texttt{Depth}_i)$ is a smooth of depth (using a thin plate spline; Simon N Wood (2003))

10

213 and $\log_e A_i$ is an offset for the area (in km^2) of each tile, A_i. The offset term accounts changes

214 in probabilities of GVP detection due to the differing area monitored by each hydrophone.

215 Because the hydrophone tessellation changed between SCCs (as there were different sets

216 of hydrophones recorded during each SSC), separate MRFs were used for each SCC, but a

217 single smoothing parameter was estimated across all MRFs. This allows for different spatial

218 smooths for each SCC, but constrains the smooths to have the same amount of wiggliness.

219 The smooth of depth was shared across SCCs.

2.2.3 M2: Modelling the effect of Naval activity

221 For a few days prior to the onset of hull-mounted MFA sonar used during SCCs, other naval

222 training activities occurred at PMRF. Various vessels were present on the range during this

223 period and other noise sources, including torpedoes and submarines, may have been present.

224 We used data collected when training activity was present on the range, but hull-mounted

225 MFAS was not used, to model the effect of general naval activity on beaked whale GVPs.

226 Initially, we tried to use low-frequency noise levels in the 10-999 Hz range measured on range

227 hydrophones as a covariate in this model, but found that the measured noise levels were not

228 consistent with known locations of naval training activities.

229 We used the predicted baseline probability of a GVP detection from M1 as an offset to control

230 for the underlying spatial distribution of GVPs. The model for the data when ships were

231 present was intercept-only, with an offset derived from M1. We again modelled GVP presence

232 at tile i as $\texttt{GVP}_i \sim \text{Bin}(1, \mu_{\texttt{M2},i})$, with a linear predictor on the logit scale:

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

233 where $\beta_{\texttt{M2},0}$ is an intercept and $\xi_{\texttt{M1},i}$ is the prediction (on the logit scale) for tile i using model

234 M1, included as an offset term.

Say something about how this assumes constant level change (no spatial effects from naval activity)?

11

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

We used data collected when hull-mounted MFAS was present on the range to model the effect of sonar on beaked whales. 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). We assumed that as the maximum received level increased, the probability of dives decreased and modeled this using a monotonically decreasing smooth ~~so that the relationship held for all possible realizations of the smoo~~th (Pya and Wood, 2015). To ensure that the model predictions were the same at a maximum received level of 0 dB and when ships were not present, we did not include an intercept. GVP presence at tile i was modelled as a Bernoulli trial $\mathtt{GVP}_i \sim \mathrm{Bin}(1, \mu_{\mathtt{M3},i})$ where the linear predictor on the logit scale was:

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

where $f(\mathtt{MaxRL}_i)$ was modeled as a monotonic decreasing smooth, $\xi_{\mathtt{M2},i}$ denotes the prediction (on the logit scale) for tile i when Naval training activities were present on the range using model $\mathtt{M2}$.

2.2.5 Uncertainty propagation

We used posterior simulation (sometimes referred to as a parametric bootstrap; Wood, Li, Shaddick, and Augustin (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. Following this procedure through from M1 to M2 to M3 incorporated uncertainty from each model in the final results.

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.

Based on the resulting final posterior distribution of results (for model M3) we can use appropriate quantiles to obtain average predictions and intervals. Mathematical 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 $\mathbb{P}(\text{GVP})$ 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 posterior samples from the model, we calculated the expected $\mathbb{P}(\text{GVP})$ under each set of covariates as

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

$M' =$

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

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

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

13

274 For each received level we calcualted the 2.5th, 50th, and 97.5th quantiles of $\Delta_{M3':M1'}$ and

275 $\Delta_{M3':M2'}$ to create 95% CIs of change in $\mathbb{P}(\text{GVP})$ across possible received levels. We consider

276 that the probability of disturbance is equal to 1 wherever the 95% CI does not include 0, and

277 0 otherwise.

2.2.7 Implementation

279 NOTE: this section may be moved to an appendix if we have space constraints, but I do

280 prefer to acknowledge the contributions of package developers in the main text if possible.

281 Statistical analyses presented in this manuscript were conducted in R (v. 3.5.2; R Core Team,

282 2018). Code and data are available at [CITE zenodo repo]. Data import and manipulation

283 was accomplished using tidyverse (Wickham et al., 2019) packages. `mgcv` (S. N. Wood,

284 2017) and `scam` (Pya and Wood, 2015) were used to formulate and fit models. Map creation

285 was facilitated by the `fields` (Nychka, Douglas, Furrer, Reinhard, Paige, John, and Sain,

286 Stephan, 2017), `ggsn` (Santos Baquero, 2019), `marmap` (Pante and Simon-Bouhet, 2013),

287 `rgdal` (R. Bivand, Keitt, and Rowlingson, 2019), `sf` (Pebesma, 2018), and `sp` (R. S. Bivand,

288 Pebesma, and Gomez-Rubio, 2013) packages. All graphics were produced using `ggplot2`

289 (Wickham, 2016), with color palettes from the `viridis` (Garnier, 2018) and `cmocean` (Thyng,

290 2019) packages. The manuscript was written in `rmarkdown` (Xie, Allaire, and Grolemund,

291 2018).

3   Results

3.1   Results of Data Collection and Processing

294 Data were collected before and during six SCCs: two each in in 2013, 2014, and 2017 (Table

295 1). The number of hydrophones for which recordings were available varied from 49 to 61. A

14

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 MFAS 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

total of 190928 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.
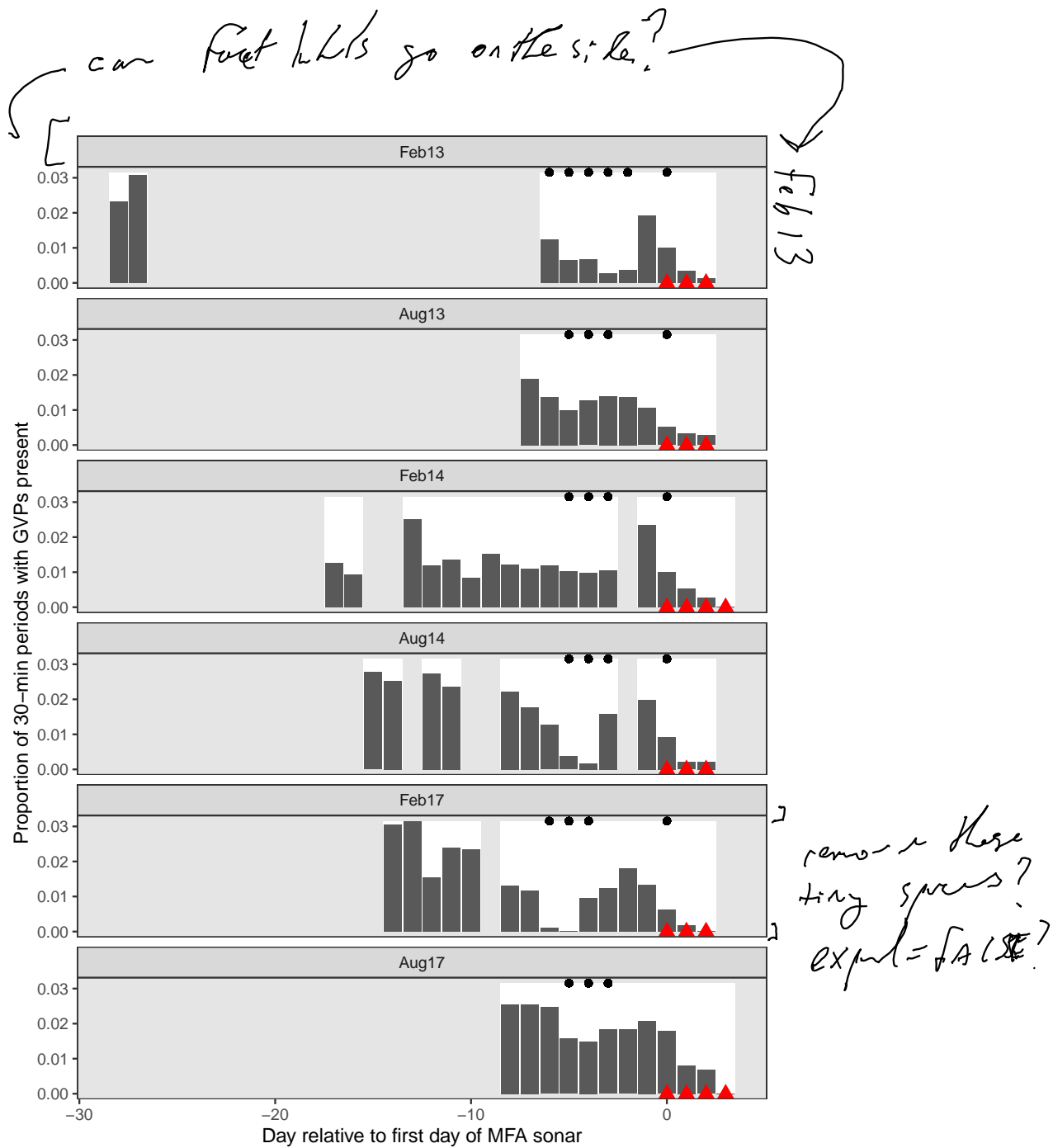
Figure 2: Timeseries of six recorded Naval training activities at PMRF. The timeseries are aligned relative to the first day that MFAS (red triangles) was used in each exercise (x-axis). Days with white background indicate days for which recordings and data were available. Gray bars indicate the proportion 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.

303 Across all SCCs, hydrophones, and conditions, a total of 2312 GVPs were identified. The

16

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. S3; top panel).

Modelled maximum received levels ranged from 38 to 186 dB re. 1 μ Pa, with a median value when MFAS 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. S4).

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 MFAS were present compared to when only ships were present, and by -77% when naval activity and MFAS were present compared to when neither ships nor sonar were present (Fig. S2).

3.2   Results of spatial modelling

We created separate tessellations for each SCC (Fig. S1). 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 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 1,500 and 2,000 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

17

you should probably state the level you were doing the F/t-test on the smooths/MFS over if the p-value is < 10^16

₃₂₈ 64% (95% CI 59% - 68%) when naval training activity was present compared to when it was

₃₂₉ absent.

₃₃₀ `M3` used the predicted values from `M2` as an offset and fitted a model to data when naval

₃₃₁ activity and MFAS were present. This model used a monotonically decreasing spline on

₃₃₂ modelled MFAS received level (Fig. SX) and did not include an intercept term. The smooth

₃₃₃ on MFAS received level had significant explanatory power (p-value = 6.74E-10) and the

₃₃₄ model explained 12.4% of deviance in the data.

₃₃₅ For MFAS received levels above 100 dB, change in $\mathbb{P}(GVP)$ was calculated relative to the

₃₃₆ pre-activity baseline period ($\Delta_{M3':M1'}$) and to the period when naval activity was present

₃₃₇ on the range ($\Delta_{M3':M2'}$; Fig. 4 & Fig. 5). At a received level of 150 dB, $\Delta_{M3':M1'}$ was -92%

₃₃₈ (95% CI -100% - -87%) and $\Delta_{M3':M2'}$ was -78% (95% CI -100% - -62%).
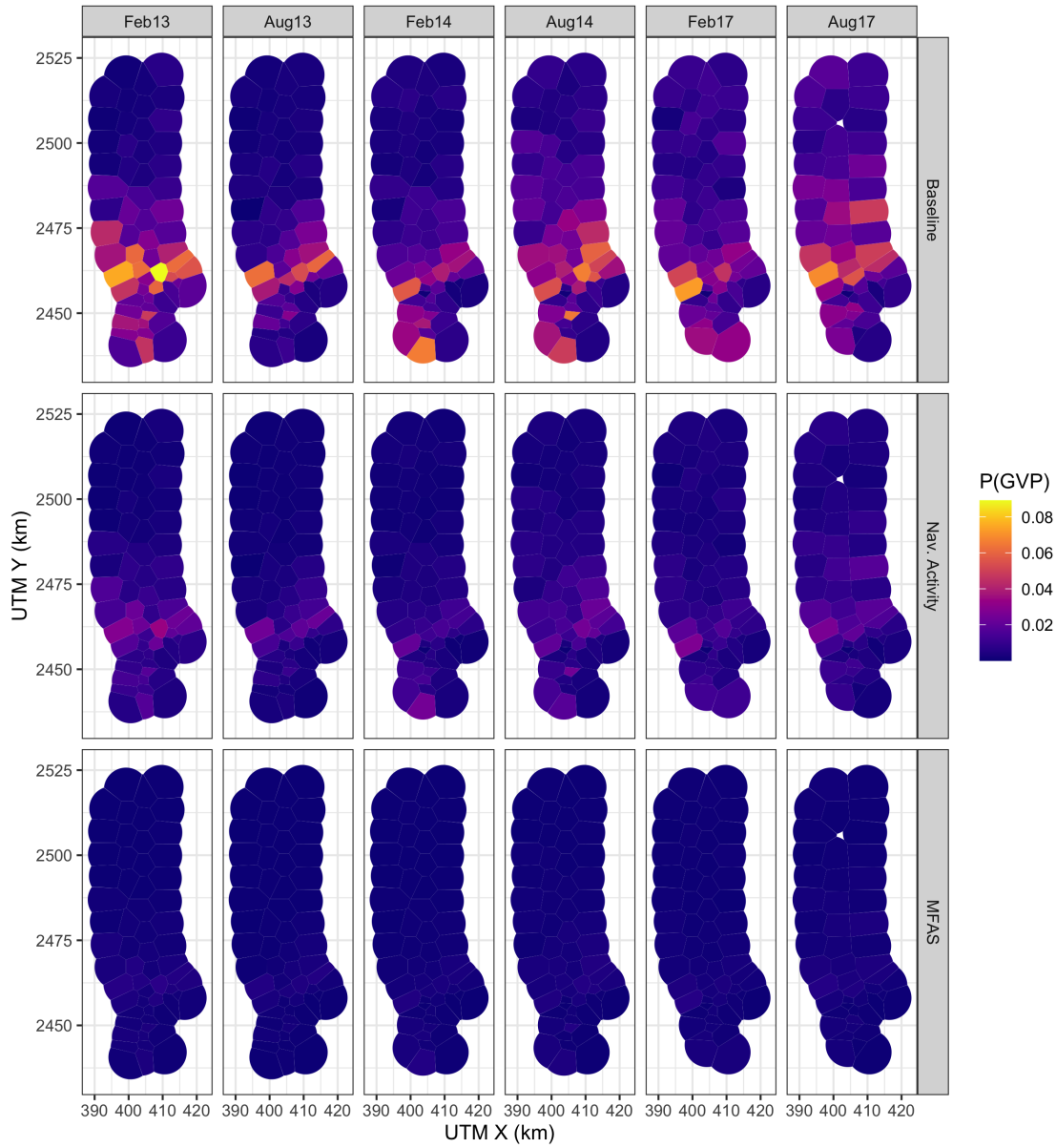
Figure 3: Map of expected probability of diving (color scale) at each hydrophone during each SCC (columns) prior to the onset of Naval training activity, during naval training activity when no MFAS was present, and during naval training activity when MFAS was present at a level of 150 dB re. 1 uPa rms (rows).
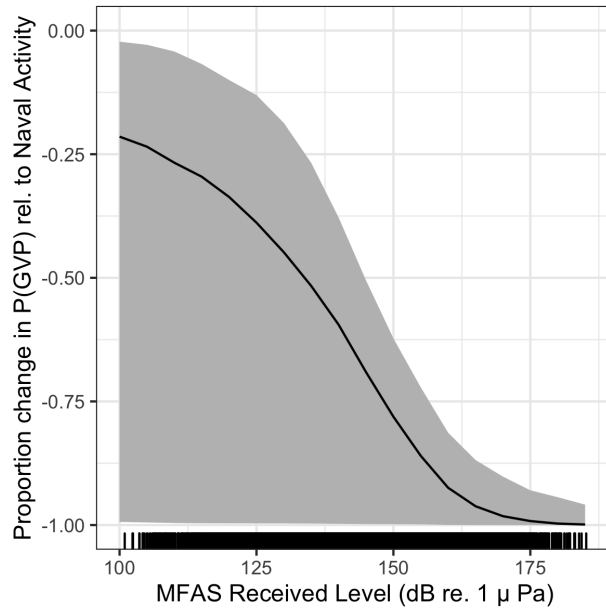
Figure 4: Median (black line) and 95% CIs (gray shading) expected change in the probability of detecting a group vocal period (y-axis) with increasing MFAS received level (x-axis) relative to when Naval training activity but no MFAS is present on the range.
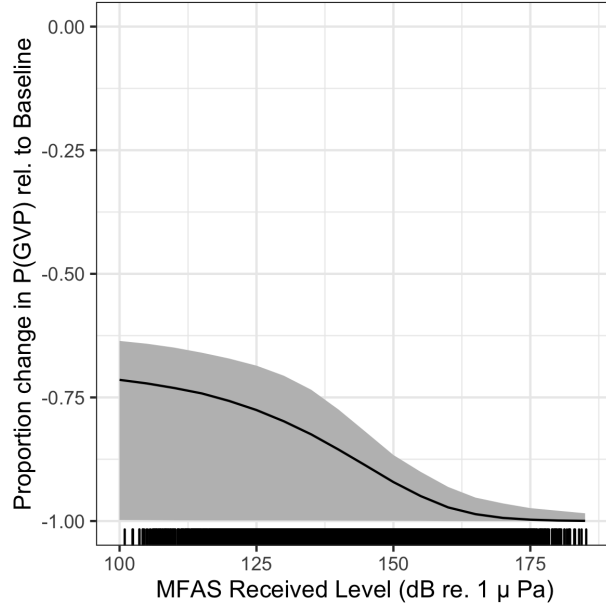
Figure 5: Median (black line) and 95% CIs (gray shading) expected change in the probability of detecting a group vocal period (y-axis) with increasing MFAS received level (x-axis) relative to when neither Naval training activity nor MFAS is present on the range.

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 and 5).

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

21

MFAS) were not applied randomly with respect to either the study area or Blainville's beaked

whale presence. 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 would lead to underestimates of the impact of sonar,

since changes in distribution due to MFAS could be explained as spatial changes by the MRF

(Appendix S3). Our three-stage modelling approach addresses this issue while propagating

uncertainty between the models.

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 distance data is not available, avoiding

the use of continuous smoothers when true location data is not available. Shape-constrained

smoothing is also well-suited to the kind of data we model 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 extends to any situation 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) expressions (or shortcut them by assuming independence).

We provide example code as Supplementary Material so that other researchers may apply

and/or expand on these methods.

In a regulatory context, a dose-response function as presented in Figs. 4 and 5 is often

interpreted as representing the proportion of a population that responds (y-axis) to a given

received level (x-axis) (Tyack and Thomas, 2019). However, the metric used in this study

– the change in the probability of detecting a GVP within a half-hour period – does not

directly correspond to the proportion of the population that is affected. Rather, it may

reflect a change in the proportion of time spent foraging, which affects the entire population.

These two interpretations have different implications for understanding sub-lethal impacts of

22

376 MFAS. In the traditional interpretation, given exposure to a certain received level, some of
377 the population is affected and some of the population is not. In our interpretation, the entire
378 exposed population is affected.

379 In comparison to the risk function developed by Moretti et al. (2014) for Blainville's beaked
380 whales at AUTEC, our risk function predicts a more intense response to naval sonar. This
381 may be because Moretti et al. were not able to explicitly account for the effects of naval
382 training activities that did not include MFAS. Their baseline period consisted of 19 hours of
383 data before the onset of MFAS; as at PMRF, it is likely that training activities during this
384 period included sound sources other than MFAS. Therefore, their risk function is probably
385 more analagous to our expected change in the probability of a detection when MFAS is
386 present relative to when naval training activity was present (Fig. 4).

387 In future, we would like to further investigate the specific causes of changes in the probability
388 of detecting GVPs before the onset of MFAS. The reduction in detection of foraging dives
389 could be a response to general Naval training activity on the range, or to specific sound
390 sources that have not previously been studied. Alternatively, it is possible that Blainville's
391 beaked whales are semi-resident on the range and have become habituated to SCC activity;
392 they may move off the range in anticipation of MFAS.

393 The AUTEC range is located in the Tongue of the Ocean (TOGO) off the island of Andros
394 in The Bahamas, in a deep basin bounded to the south, east, and west by shallow waters
395 and with maximum depths of 2000 m. In contrast, PMRF is located to the northwest of the
396 island of Kauai and occurs across a steep slope and into deep water, over 5000 m in depth.
397 Although the environments at PMRF and AUTEC are different, the foraging dive behavior
398 of Blainville's beaked whales is similar at AUTEC and PMRF; dives occur in deeper waters
399 over steep slopes with gradients ranging from 3-23%, although dives occur in deeper waters
400 (2000-3000 m; Henderson et al., 2016) at PMRF that at AUTEC (500-1300 m; MacLeod and
401 Zuur, 2005, Hazen, Nowacek, St. Laurent, Halpin, and Moretti (2011)). Resident Blainville's

beaked whales off the Big Island also occur in slightly shallower waters than at PMRF, from 980-1410 m (R. Baird, 2011; R. W. Baird, Webster, Schorr, McSweeney, and Barlow, 2008). Therefore it is likely the location of the mesopelagic scattering layer 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 1050-1150 m for 46-60 min [R. W. Baird et al. (2008); Schorr et al. (2009); Joyce et al. (2017)).

Similarly, responses to MFAS activity are similar 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). Tagged individuals that were to the north of AUTEC at the onset of sonar moved 27 – 43 km away from the edge of the range; the individual that was south of the range initially moved south but may have been trapped by the shallow water at the edge of the basin and so turned north towards the range, then south again (Joyce et al., 2019). In contrast, Blainville's beaked whales at PMRF can easily move away from the range in any direction without impediment. While no Blainville's beaked whales have been tagged at PMRF, they have been tagged off the Big Island and the resident, island-associated animals appear to remain relatively close to shore, with median distances of 4-27 km (Schorr et al., 2009). These resident animals may also travel to other islands as well (R. W. Baird, 2019), unlike Blainville's beaked whales in the Bahamas where no animals have been photo-identifed at both AUTEC and the nearby island of Abaco (Claridge, 2013). However, two Blainville's beaked whales have been tagged off the Big Island that moved away into deep water; these may represent offshore animals that are also found in Hawaiian waters (R. Baird, 2011; R. W. Baird, 2019; R. W. Baird et al., 2008). 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 there may be offshore animals as well found on the northern hydrophones. Regardless, the similarities

24

429 in Blainville's beaked whale behavioral responses to Navy training activity across different
430 ranges and environments and at similar received levels may indicate the intrinsic nature of the
431 response. Conducting a similar analysis of Cuvier's beaked whale responses at the Southern
432 California Anti-Submarine Warfare Range (SOAR) would further support this assessment;
433 existing findings already demonstrate that Cuvier's respond in a similar manner by reducing
434 their foraging dives and moving away from the ensonified area (DeRuiter et al., 2013; Falcone
435 et al., 2017).

Acknowledgements

Authors' Contributions

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

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

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

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

444 Investigation: E.E.H.

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

446 Software: E.K.J., D.L.M.

447 Supervision: L.T.

448 Visualization: E.K.J.

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

450 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

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., . . . 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. (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. (2019). Behavior and Ecology of Not-So-Social Odontocetes: Cuvier's and Blainville's Beaked Whales. In B. Würsig (Ed.), *Ethology and Behavioral Ecology of Odontocetes* (pp. 305–329). https://doi.org/10.1007/978-3-030-16663-2_14

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.

473 https://doi.org/10.1111/j.1748-7692.2008.00211.x

474 Bernaldo de Quirós, Y., Fernandez, A., Baird, R. W., Brownell, R. L., Aguilar de Soto, N.,

475 Allen, D., . . . Schorr, G. (2019). Advances in research on the impacts of anti-submarine

476 sonar on beaked whales. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1895),

477 20182533. https://doi.org/10.1098/rspb.2018.2533

478 Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R,*

479 *Second edition.* Retrieved from https://asdar-book.org/

480 Bivand, R., Keitt, T., & Rowlingson, B. (2019). *Rgdal: Bindings for the 'geospatial' data*

481 *abstraction library* [Manual]. Retrieved from https://CRAN.R-project.org/package=rgdal

482 Claridge, D. E. (2013). *Population ecology of Blainville's beaked whales (Mesoplodon den-*

483 *sirostris)* (Thesis, University of St Andrews; pp.). Retrieved from https://research-

484 repository.st-andrews.ac.uk/handle/10023/3741

485 Cox, T., Ragen, T., Read, A., Vos, E., Baird, R., Balcomb, K., . . . others. (2006).

486 Understanding the impacts of anthropogenic sound on beaked whales1. *Journal of*

487 *Cetacean Research and Management*, *7*(3), 177–187.

488 DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone,

489 E. A., . . . Tyack, P. L. (2013). First direct measurements of behavioural responses by

490 Cuvier's beaked whales to mid-frequency active sonar. *Biology Letters*, *9*(4), 20130223.

491 https://doi.org/10.1098/rsbl.2013.0223

492 Falcone, E. A., Schorr, G. S., Watwood, S. L., DeRuiter, S. L., Zerbini, A. N., Andrews, R.

493 D., . . . Moretti, D. J. (2017). Diving behaviour of Cuvier's beaked whales exposed to two

494 types of military sonar. *Royal Society Open Science*, *4*(8), 170629. https://doi.org/10.

495 1098/rsos.170629

496 Garnier, S. (2018). *Viridis: Default color maps from 'matplotlib'* [Manual]. Retrieved from

27

https://CRAN.R-project.org/package=viridis

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

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., Aschettino, J., Deakos, M., Alongi, G., & Leota, T. (2019). Quantifying the Behavior of Humpback Whales (Megaptera novaeangliae) and Potential Responses to Sonar. *Aquatic Mammals*, *45*(6), 612–631. https://doi.org/10.1578/AM.45.6.2019.612

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

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),

28

5038–5050. https://doi.org/10.1242/jeb.02596

Joyce, T. W., Durban, J. W., Claridge, D. E., Dunn, C. A., Fearnbach, H., Parsons, K. M., . . . 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., . . . Moretti, D. (2019). Behavioral responses of satellite tracked Blainville's beaked whales (*Mesoplodon densirostris*) to mid-frequency active sonar. *Marine Mammal Science*, mms.12624. 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.* 11.

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. (2019). Natick, Massachusetts: The MathWorks Inc.

McCarthy, E., Moretti, D., Thomas, L., DiMarzio, N., Morrissey, R., Jarvis, S., . . . 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., . . . 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

Navy, U. D. of the. (2017). Criteria and thresholds for US navy acoustic and explosive effects analysis (phase III). *Space and Naval Warfare Systems Command, US Navy, Department of Defence, San Diego, California.*

Navy, U. D. of the. (2018). *Final Environmental Impact Statement/Overseas Environmental Impact Statement 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

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

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

Pante, E., & Simon-Bouhet, B. (2013). Marmap: A package for importing, plotting and analyzing bathymetric and topographic data in r. *PLoS ONE*, *8*(9), e73051.

Pebesma, E. (2018). Simple features for r: Standardized support for spatial vector data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

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.

Santos Baquero, O. (2019). *Ggsn: North Symbols and Scale Bars for Maps Created with 'ggplot2' or 'ggmap'* [Manual]. Retrieved from https://CRAN.R-project.org/package=ggsn

Schorr, Baird, Hanson, Webster, McSweeney, & Andrews, R. (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

Thyng, K. (2019). *Cmocean: Beautiful colour maps for oceanography.* Retrieved from

31

https://CRAN.R-project.org/package=cmocean

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. (2016). *Ggplot2: Elegant graphics for data analysis.* Retrieved from https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

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

Xie, Y., Allaire, J., & Grolemund, G. (2018). *R markdown: The definitive guide.* Retrieved

32

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 results are empirical Bayes estimates. In this case we can generate samples from the (multivariate normal) posterior of the model parameters. After generating a sample, $\boldsymbol{\beta}^* \sim \text{MVN}(\hat{\boldsymbol{\beta}}, \mathbf{V}_\beta)$, 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 et al. (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 $\mathbf{X}_{p,\text{M}j}$ for $j = 1, 2, 3$ be the $\mathbf{L}_p$ matrix that maps coefficients to the predictions for model $\text{M}j$. For N_b times:

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

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

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

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

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

644   locations for M2$'$).

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

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

647 We can then calculate summary statistics (means and variances) of the N_b values of $\boldsymbol{\mu}_{\text{M1}'}$,

648 $\boldsymbol{\mu}_{\text{M2}'}$, and $\boldsymbol{\mu}_{\text{M3}'}$ that we have generated. The empirical variance of the N_b values of $\boldsymbol{\mu}_{\text{M3}'}$ will give

649 the uncertainty incorporating components from all three stages. We can take appropriate

650 pointwise quantiles to form confidence bands for the functional relationships between sonar

651 received level and estimated probability of detecting GVPs. (eg Fig 4/5)?

S2: Supplementary Tables and Figures

To add:

- Fig. of obs v exp for M1

- Fig. of spline on depth from M1
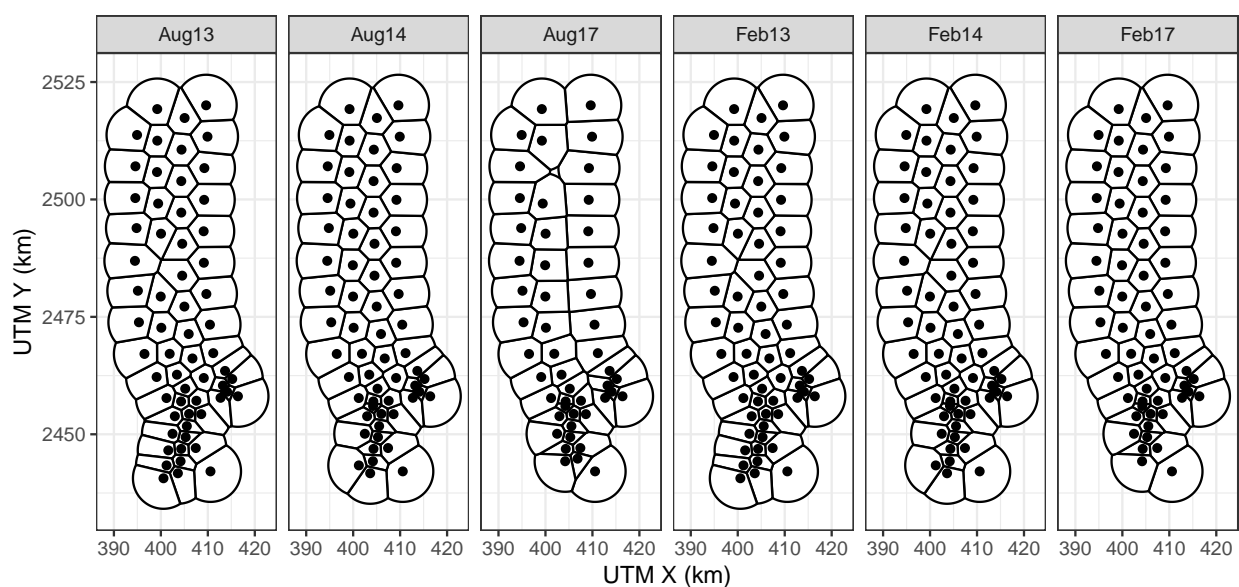
- Fig. of spline on MFAS from M3

Figure S1: PMRF range tessellations for each of 6 recorded SCCs. Black lines indicate boundaries of hydrophone tiles. Black dots indicate approximate hydrophone locations.
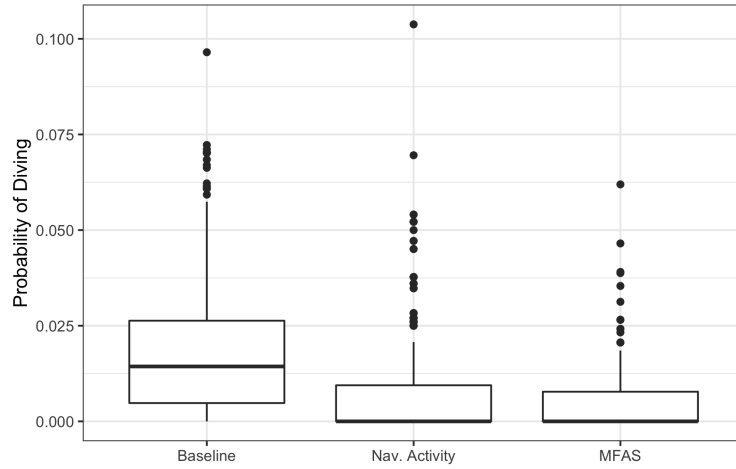
Figure S2: Boxplot of observed probability of a GVP across all hydrophones and SCCs (y-axis) during baseline period, when naval activity was present, and when MFAS was present (x-axis).
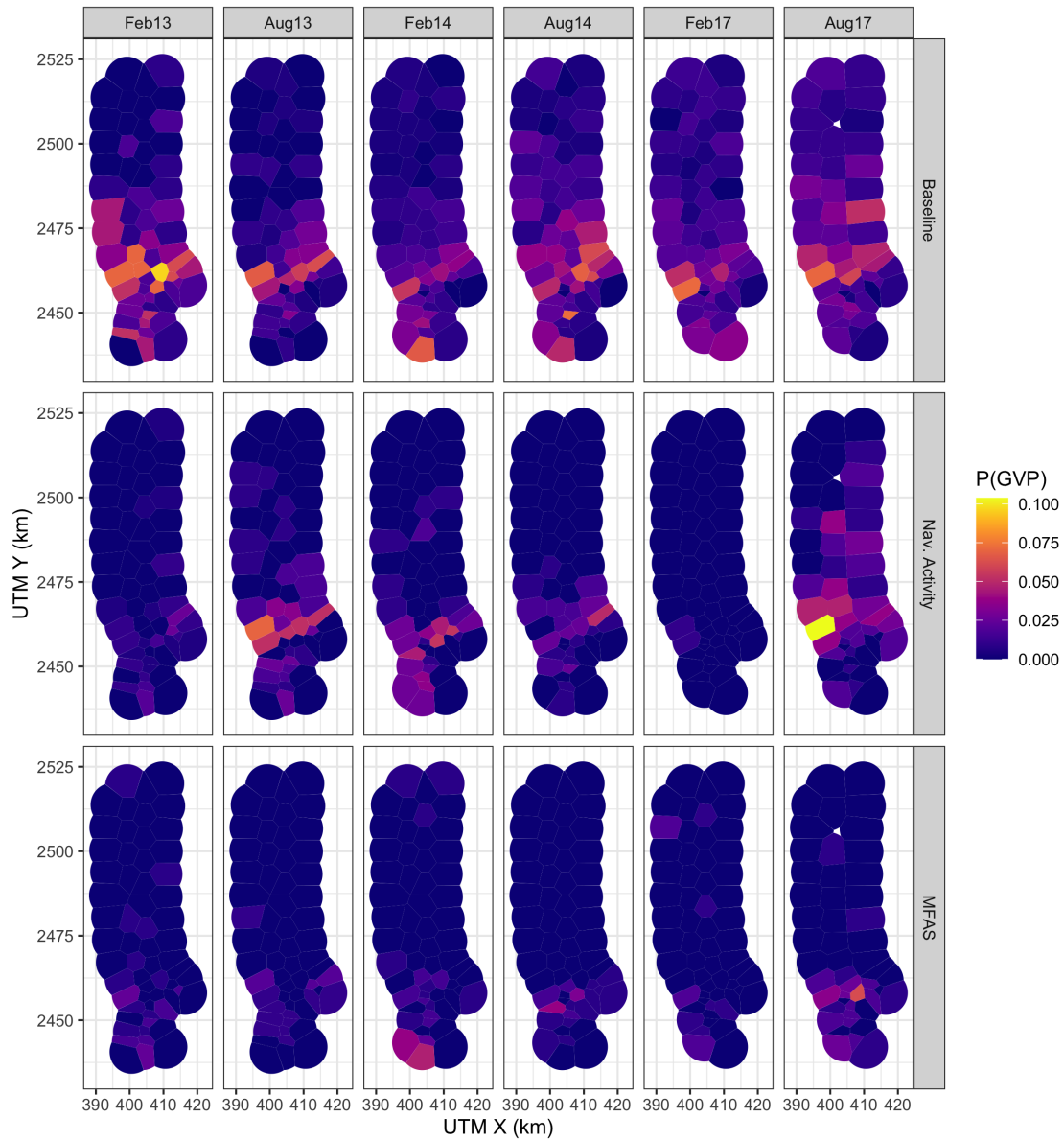
Figure S3: Map of observed probability of a GVP at each hydrophone (color scale) during the baseline period, when naval activity was present, and when MFAS was present (rows) for each SCC (columns). Note that values of PDive are not corrected for effort (size of the hydrophone tile).

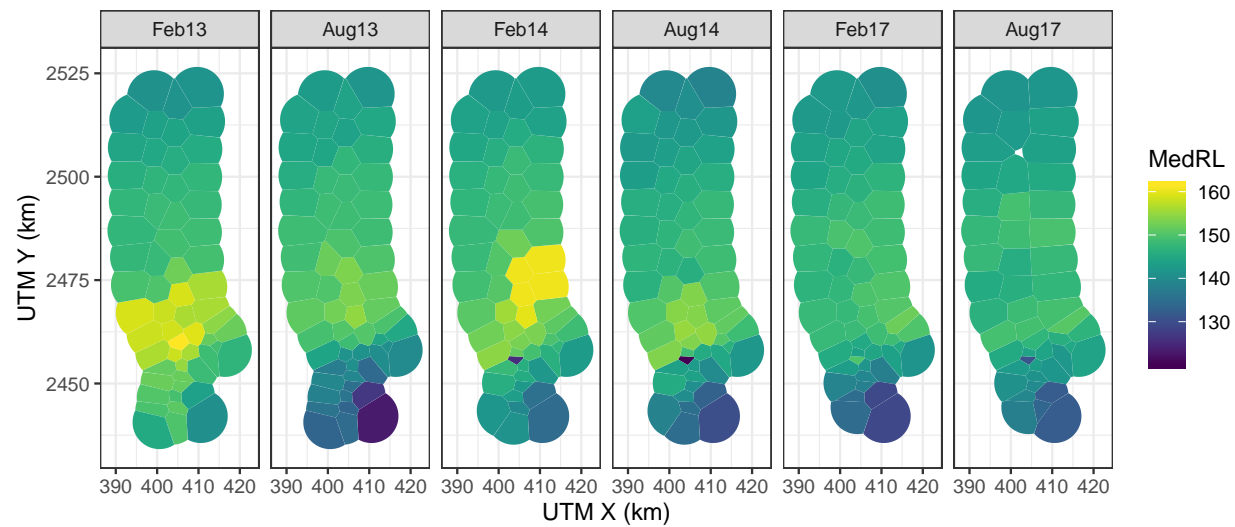Figure S4: Median received level when MFAS was present (color scale) for all hydrophones and SCCs.

S3: Single GAM

Results from a single giant GAM will be presented here.