1 **QUANTIFYING THE RESPONSE OF**

2 **BLAINVILLE’S BEAKED WHALES TO US NAVAL**

3 **SONAR EXERCISES IN HAWAII**

4

## 5 Eiren K. Jacobson1, E. Elizabeth Henderson2, David L. Miller1, Cornelia S.

6 **Oedekoven1, David J. Moretti3, Len Thomas1**

7 1 *Centre for Research into Ecological and Environmental Modelling, School of Mathematics and*

8 *Statistics, University of St Andrews, St Andrews, Scotland*

9 2 *Naval Information Warfare Center Pacific, San Diego, CA, USA*

10 3 *Naval Undersea Warfare Center, Newport, RI, USA*

## 11 Correspondence:

12 Eiren Jacobson

13 University of St Andrews

14 The Observatory

15 Buchanan Gardens

16 St Andrews

17 Fife

18 KY16 9LZ

19 Scotland

20 Email: [eiren.jacobson@st-andrews.ac.uk](mailto:eiren.jacobson@st-andrews.ac.uk)

21

## 22 Draft 29 April 2022

23 **Abstract**

24 Behavioral responses of beaked whales (family Ziphiidae) to naval use of mid-frequency

25 active sonar (MFAS) have been quantified for some species and regions. We describe the

26 effects of MFAS on the probability of detecting diving groups of Blainville’s beaked whales

27 on the US Navy Pacific Missile Range Facility (PMRF) in Hawaii and compare our results

28 to previously published results for the same species at the Atlantic Undersea Test and

29 Evaluation Center (AUTEC) in the Bahamas. We use passive acoustic data collected at

30 bottom-mounted hydrophones before and during six naval training exercises at PMRF along

31 with modelled sonar received levels to describe the effect of training and MFAS on foraging

32 groups of Blainville’s beaked whales. We use a multi-stage generalized additive modelling

33 approach to control for the underlying spatial distribution of vocalizations under baseline

34 conditions. At an MFAS received level of 150 dB re 1 *µ*Pa the probability of detecting groups

35 of Blainville’s beaked whales decreases by 77% (95% CI 67%-84%) compared to periods when

36 general training activity was ongoing and by 87% (95% CI 81%-91%) compared to baseline

37 conditions. Our results indicate a more pronounced response to naval training and MFAS

38 than has been previously reported. [196/200]

39 **KEYWORDS**

40 Blainville’s beaked whales, *Mesopolodon densirostris*, mid-frequency active sonar, passive

41 acoustic data, behavioral response, generalized additive model

42 **1 Introduction**

43 Beaked whales (family Ziphiidae) are a group of deep-diving cetaceans that rely on sound

44 to forage, navigate, and communicate (Aguilar de Soto et al., 2012; Johnson et al., 2004;

45 Macleod and D’Amico, 2006) and are sensitive to anthropogenic noise (Hooker et al., 2019;

46 Southall et al., 2016). Multiple mass strandings of beaked whales have been associated with

47 high-intensity anthropogenic sound sources, including naval sonar (Bernaldo de Quirós et

48 al., 2019; D’Amico et al., 2009; Simonis et al., 2020). These acute events have motivated

49 research into whether and how beaked whales respond to different types and intensities of

50 anthropogenic noise (e.g., Aguilar de Soto et al., 2006; Cholewiak et al., 2017; Stanistreet

51 et al., 2022; Tyack et al., 2011). Anthropogenic sound can disrupt the foraging dive cycles

52 of beaked whales (Falcone et al., 2017), potentially leading to cumulative sublethal impacts

53 resulting from reduced foraging opportunities (New et al., 2013; Pirotta et al., 2018), or to

54 symptoms similar to decompression sickness that can lead to injury or death (Hooker et al.,

55 2009, 2012).

56 Echolocation clicks produced by diving groups of Blainville’s beaked whales indicate foraging

57 activity and can be recorded by hydrophones (Johnson et al., 2006). Research on Blainville’s

58 beaked whales *(Mesoplodon densirostris)* using data from bottom-mounted hydrophones on a

59 U.S. Navy range in the Bahamas has shown decreases in time spent foraging and movement

60 away from naval sonar sources (Joyce et al., 2019; McCarthy et al., 2011; Tyack et al., 2011).

61 Naval sonar can be broadcast from various platforms, including vessels, helicopters, buoys,

62 submarines, and torpedoes (U.S. Department of the Navy, 2018). Most research has focused

63 on the impacts of mid-frequency active sonar [1-10 kHz bandwidth; U.S. Department of the

64 Navy (2018)] broadcast from naval vessels at the surface [hereafter referred to as MFAS;

65 Falcone et al. (2017)]. Separately, researchers have shown that, in the absence of MFAS,

66 beaked whales may alter their behavior in response to vessel noise (Aguilar de Soto et al.,

67 2006; Pirotta et al., 2012).

68 The U.S. Navy is interested in quantifying the effects of sonar on beaked whales for the

69 purpose of risk assessments and permitting associated with training activities (e.g., U.S.

70 Department of the Navy, 2017). There are different experimental and analytical ways of

71 quantifying responses to sonar (see Harris et al., 2018 for a review). Here, we focus on

72 analyses of observational data from cabled hydrophone arrays collected concurrently with

73 naval training exercises. Examples of these from previous studies include McCarthy et al.

74 (2011) who used data from the cabled hydrophone array at the U.S. Navy’s Atlantic Undersea

75 Test and Evaluation Center (AUTEC) in the Bahamas collected before, during, and after

76 naval training exercises involving MFAS. The authors used separate generalized additive

77 models (GAMs) for each period, and modelled the acoustic detection of groups of Blainville’s

78 beaked whales (group vocal periods; GVPs) as a function of location on the range and time.

79 They found that the number of GVPs was lower during the exercises than before or after.

80 Building on this work, Moretti et al. (2014) used a GAM to examine the presence or absence

81 of GVP starts within 30-min periods (i.e., whether or not a GVP started within each 30-min

82 period) on the AUTEC range as a smooth function of MFAS received level. They compared

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

84 probability of detecting animals across sonar received levels to estimate the probability of

85 disturbance. They found that the probability of detecting groups of Blainville’s beaked whales

86 was reduced by 50% at 150 dB re 1 *µ*Pa rms, which they interpreted as a 50% probability of

87 disturbance.

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

89 species on a different U.S. Navy training range in a different oceanic environment. We used

90 a spatially-referenced data set of Blainville’s beaked whale foraging dives recorded at the

91 PMRF off the island of Kauai, Hawaii (Fig. 1). Passive acoustic detections of the presence or

92 absence of GVP starts within 30-min periods were collected via a cabled hydrophone array

93 at PMRF before and during training exercises involving MFAS broadcast from navy ships.

94 Unlike AUTEC, which is situated in a deep isolated basin surrounded by steep slopes, the

95 Pacific Missile Range Facility (PMRF) in Hawaii is located on the side of an ancient volcano,

96 with a steep slope to the deep ocean floor. Previous work in this region has shown that

97 Blainville’s beaked whales are present year-round at this site, prefer sloped habitats, and that

98 acoustic detections decrease during multi-day training events involving MFAS (Henderson

99 et al., 2016; Manzano-Roth et al., 2016). As we expected the density of Blainville’s beaked

100 whales at PMRF to be low and spatially variable, our methods needed to explicitly account

101 for differences in underlying beaked whale presence across the range. An additional objective

102 was to isolate the effect of general training activity (without MFAS) from the effect of MFAS,

103 so that beaked whale response to MFAS could be quantified relative to pre-training baseline

104 periods and to periods when general training activities (without MFAS) were present on the

105 range.

106 **2 Methods**

# 107 2.1 Data Collection and Processing

## 108 2.1.1 Acoustic detection of beaked whales

109 The Pacific Missile Range Facility (PMRF) is an instrumented U.S. Navy range extending

110 70 km NW of the island of Kauai, Hawaii and encompassing 2,800 km2. The range includes

111 a cabled hydrophone array (Fig. [1)](#_bookmark0) with hydrophones at depths ranging from approximately

112 650 m to 4,700 m. We used data collected before and during six Submarine Command

113 Courses (SCCs) at PMRF. SCCs are training exercises that involve several different naval

114 platforms, occur biannually in February and August, and typically last 6–7 days. The first

115 part of the SCC involves general training activity which may include sound sources other

116 than hull-mounted MFAS from surface ships. During the second part of the training exercise,

117 hull-mounted MFAS is broadcast from naval vessels at the surface. Acoustic recordings were

118 made for a minimum of two days before each SCC as well as during the exercise. During data

119 collection, hydrophones sampled at a rate of 96 kHz. Up to 62 hydrophones were recorded

120 simultaneously by the Naval Information Warfare Center (NIWC).

121 A beaked whale echolocation detector from the Navy Acoustic Range WHale AnaLysis

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

123 first compared signal-to-noise ratio (SNR) thresholds within the expected frequency range

124 of beaked whale clicks (16–44 kHz) versus the bandwidth outside the click in a running

125 16,384-pt fast Fourier transform (FFT) spectrogram. The detected clicks were then passed

126 to a 64-pt FFT stage that measured power, bandwidth, slope, and duration characteristics

127 to classify the clicks to species. This process was followed by an automated routine in

128 MATLAB (*MATLAB*, 2017) to group detections of individual beaked whale echolocation

129 clicks into GVPs (Henderson et al., 2016). If a group of whales was detected by more than

130 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. We used half-hour periods to approximate

133 the typical vocal period of Blainville’s beaked whales during deep foraging dives (Tyack et

134 al., 2006).

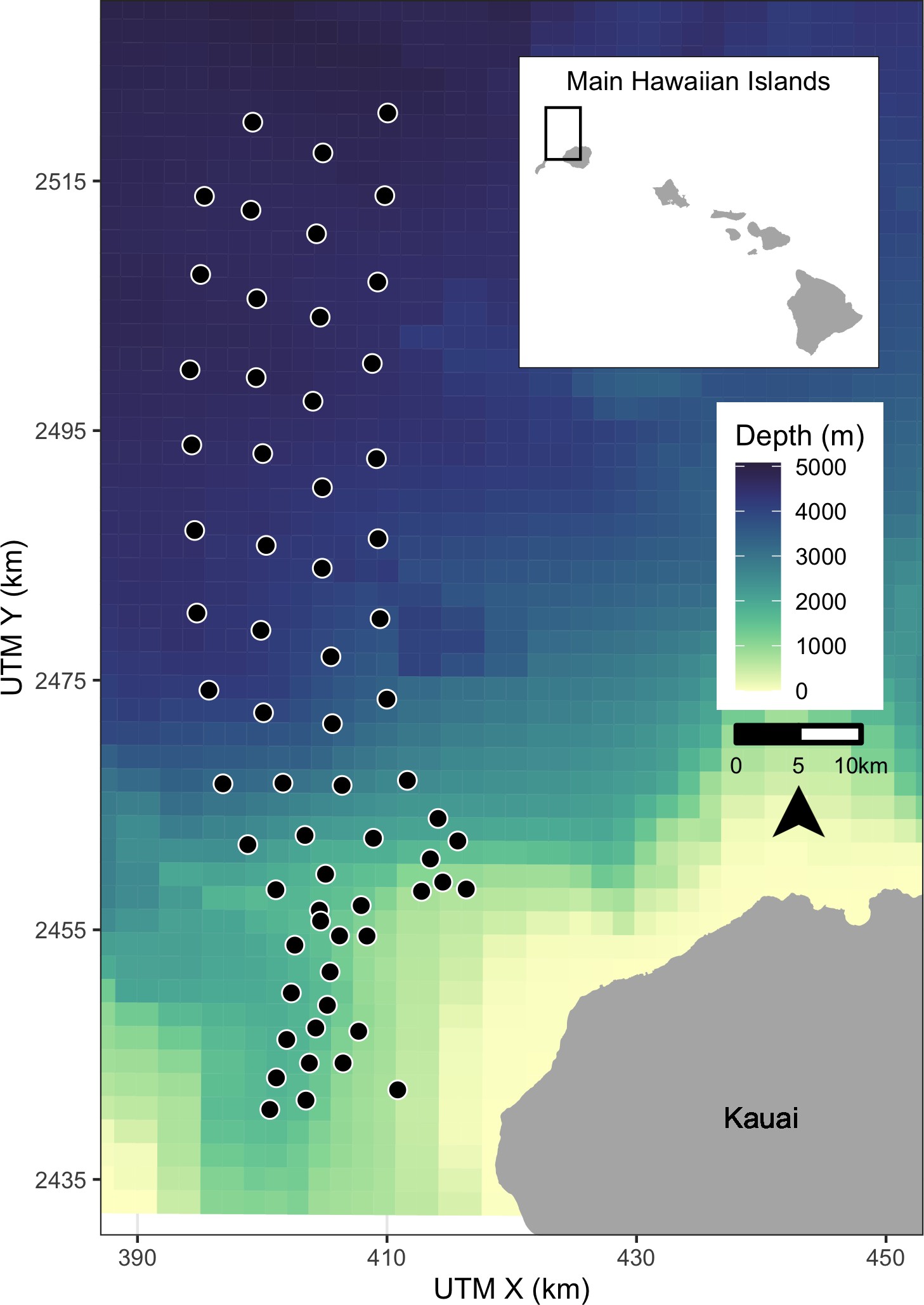


Figure 1: Map of hydrophones (black points) at the Pacific Missile Range Facility near the island of Kauai, Hawaii. For security reasons, the approximate rather than exact locations are shown here. Color scale indicates bathymetry. Inset map shows range location (black rectangle) relative to the main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

## 135 2.1.2 Modelling received levels of hull-mounted mid-frequency active sonar

136 For security reasons, classified data regarding activity that occurred on the range during each

137 SCC was passed from PMRF to one author with clearance (E.E.H.). These data indicated

138 the locations of the ships during the training periods and the start and stop times of each

139 individual training event. However, no information was provided on the start and stop

140 of sonar use; hence, periods of active sonar were determined from the range hydrophone

141 recordings by running a sonar detector from the NARWHAL algorithm suite tuned to MFAS.

142 The hydrophone recordings cannot reliably be used to determine received level when the

143 received level exceeds 140 dB re 1 *µ*Pa rms due to voltage constraints at the analog to

144 digital recorder interface. Additionally, the hydrophones are mostly 4–5 km deep, whereas

145 Blainville’s beaked whales begin clicking when they have reached depths of approximately

146 200–500 m and spend most of their foraging dive at depths of 1–1.5 km (Johnson et al., 2004,

147 2006; Madsen et al., 2013). Therefore, we used an acoustic modeling approach to estimate

148 the maximum received level of MFAS during each half-hour period around the location of

149 each hydrophone at a depth of 1,000 m.

150 First, the locations of all surface ships were noted at the start of each half-hour period and

151 the closest ship to each hydrophone was determined. MFAS propagation was modelled using

152 the parabolic equation propagation model in the program Peregrine (OASIS, Heaney and

153 Campbell, 2016). Acoustic transmission loss was estimated using a 200 Hz band around the

154 center frequency of the sonar (3.5 kHz). A nominal source level of 235 dB re 1 *µ*Pa rms @ 1 m

155 was assumed (U.S. Department of the Navy, 2018). The transmission loss was estimated

156 along the radial from the closest ship to each hydrophone from a distance of 1 km before the

157 hydrophone to 1 km past the hydrophone in 200 m increments and converted to received

158 levels based on the source level of the sonar. The maximum modeled received level along that

159 radial was determined for each hydrophone and half-hour period. However, if the distance

160 between the ship and the hydrophone was less than the depth of the water column, the

161 parabolic equation would overestimate transmission loss at that angle. In these cases, a

162 simple sonar equation was used to estimate transmission loss instead (Urick, 1983). For

163 hydrophones shallower than 1,000 m the received level was estimated at a point 20 m above

164 the sea floor with a ± 10 m buffer, while for hydrophones deeper than 1,000 m the received

165 level was estimated at a depth of 1,000 m with a ± 10 m buffer. This process resulted in

166 an estimate of received level for each hydrophone and half-hour period. Uncertainty in the

167 modeled received levels was not considered.

168 **2.2 Spatial Modelling**

169 **Summary**

170 We first used tessellation to determine the area effectively monitored by each hydrophone

171 (section 2.2.1). Then, we used a three-stage GAM approach to control for the underlying

172 spatial distribution of Blainville’s beaked whales when modelling the effects of training

173 activities and of MFAS. For the first model (M1), we used pre-activity data to create a spatial

174 model of the probability of GVPs across the range prior to the onset of naval activity (2.2.2).

175 We used the predicted values from this first model as an offset in a second model (M2) created

176 using data from when naval activity was present on the range, but MFAS was not (2.2.3).

177 We then used the predicted values from this second model as an offset in a third model (M3)

178 created using data when naval activity and MFAS were present on the range (2.2.4). Finally,

179 we used posterior simulation to calculate confidence intervals and quantified the change in

180 the probability of detecting GVPs when naval activity was present and across received levels

181 of MFAS (2.2.5). Analyses were undertaken in R (R Core Team, 2018). Code and data are

182 available at [https://github.com/eirenjacobson/MdMFASResponsePMRF.](https://github.com/eirenjacobson/MdMFASResponsePMRF)

## 183 2.2.1 Determining hydrophone effort

184 For security reasons, randomly jittered locations and depths of hydrophones at PMRF

185 were used. We projected the coordinates of each hydrophone into Universal Transverse

186 Mercator Zone 4. Because the beaked whale detection algorithm assigned GVPs to the

187 hydrophone that recorded the most echolocation clicks, and because the spatial separation of

188 the hydrophones was not uniform, effort was not the same for all hydrophones. This meant

189 that some hydrophones may have detected more GVPs because they were further away from

190 other hydrophones, not because they were located in higher-density areas. To account for

191 this, we used a Voronoi tessellation implemented in the R package deldir (Turner, 2019) to

192 define a tile for each hydrophone that contained all points on the range that were closest to

193 that hydrophone. We assumed that beaked whale groups occur within the tessellation tile

194 of the hydrophone to which the GVP is assigned, and that the area of each tessellation tile

195 influences the GVP detection rate at that hydrophone. For hydrophones on the outside of the

196 range, i.e., not surrounded by other hydrophones, we used a cutoff radius of 6.5 km to bound

197 the tessellation tiles. This distance was based on the estimated maximum detection distance

198 of individual Blainville’s beaked whale clicks at a U.S. Naval range in the Bahamas (Marques

199 et al., 2009). Due to recording capacity, a subset of PMRF hydrophones were recorded during

200 each SCC. While this subset was kept as consistent as possible, due to occasional hydrophone

201 failures slightly different combinations of hydrophones were recorded during different SCCs.

202 Therefore, we created separate tessellations for each SCC.

## 203 2.2.2 M1: Modelling the pre-activity probability of dive detection

204 In the first model, we used data collected prior to SCCs, when no naval ships were present on

205 the range and no other naval activity was known to occur, to model the spatial distribution

206 of GVP detections across the range. Because of the way that GVPs were assigned to

207 hydrophones, (see Section 2.1.1) the data were not continuous in space. To account for this,

208 we used a Markov random field (MRF) implemented in the R package mgcv (Wood, 2017)

209 to model the spatial distribution of GVP detections. Markov random fields (Rue and Held,

210 2005) model correlation in space between discrete spatial units (henceforth, “tiles”). The

211 correlation between two tiles is dictated by distance, as measured by the number of other

212 tiles one needs to pass through to travel between two tiles (“hops”); correlation is strongest

213 between a tile and its direct neighbors (those tiles it shares a border with) and decreases

214 with additional hops. This was appropriate for our data as we did not know where in each

215 tile a given GVP occurred, but we assumed that it did occur in that tile.

216 We modelled the probability of a GVP at tile *i* during SCC *s* (*µ*M1*,i,s*) as a Bernoulli random

217 variable. The linear predictor (on the logit scale) was given as:

logit (*µ*M1*,i,s*) = *β*M1*,*0 + *f* (MRF*i,s*) + *f* (Depth*i*) + log*e Ai,s* (1)

218 where *β*M1*,*0 is an intercept, *f* (MRF*i,s*) denotes the Markov random field used to smooth space

219 in the *s*th SCC, *f* (Depth*i*) is a smooth of depth at the location of each hydrophone (using a

220 thin plate spline; Wood (2003)) and log*e Ai,s* is an offset for the area (in km2) of each tile for

221 each SCC, *Ai,s*. The offset term accounts for changes in probabilities of GVP detection due

222 to the different areas monitored by each hydrophone. Because the hydrophone tessellation

223 changed between SCCs (as there were different sets of hydrophones recorded during each SSC),

224 separate MRFs were used for each SCC, but a single smoothing parameter was estimated

225 across all MRFs. This allowed for different spatial smooths for each SCC, but constrained

226 the smooths to have the same amount of wiggliness. The smooth of depth was shared across

227 SCCs. We used this model to predict the baseline probability of a GVP detection at each

228 hydrophone.

## 229 2.2.3 M2: Modelling the effect of Naval activity

230 For the second model, we used data collected prior to the onset of MFAS used during SCCs,

231 when other naval training activities occurred at PMRF. Various vessels were present on

232 the range during these periods, and other noise sources, including range tracking pingers,

233 torpedoes, and submarines, may have been present. We used data collected when training

234 activity was present on the range, but MFAS was not used, to model the effect of general

235 naval activity on beaked whale GVPs.

236 We used the predicted baseline probability of a GVP detection at each hydrophone from M1

237 as an offset to control for the underlying spatial distribution of GVPs. The model for the

238 data when naval activity was present was intercept-only, with an offset derived from M1. This

239 meant that the spatial distribution of GVPs was not allowed to change, but that we expected

240 a uniform relative change in GVPs when naval activity was present. We again modelled

241 the probability of GVP presence at tile *i* (*µ*M2*,i*) as a Bernoulli random variable, with the

242 following linear predictor:

logit (*µ*M2*,i,s*) = *β*M2*,*0 + log*e ξ*M1*,i,s,* (2)

243 where *β*M2*,*0 is an intercept and *ξ*M1*,i,s* is the prediction (on the logit scale) for tile *i* during

244 SCC *s* using model M1, included as an offset term.

## 245 2.2.4 M3: Modelling the effect of hull-mounted MFAS

246 For the third model, we used data collected when MFAS was present on the range to model

247 the effect of sonar on beaked whales. We excluded data collected during breaks in training

248 activities when sonar was not being used. The probability of a GVP when sonar was present

249 was modeled as a function of the maximum received level (modeled at each hydrophone

250 for each half-hour period; see section 2.2.1). We assumed that as the maximum received

251 level increased, the probability of dives decreased and modeled this using a monotonically

252 decreasing smooth implemented in the R package scam (Pya and Wood, 2015). To ensure

253 that the model predictions were the same at a maximum received level of 0 dB and when only

254 naval activity was present, we did not include an intercept. The probability of GVP presence

255 at tile *i* (*µ*M3*,i*) was modelled as a Bernoulli random variable where the linear predictor was:

logit (*µ*M3*,i,s*) = *f* (MaxRL*i,s*) + log*e ξ*M2*,i,s,* (3)

256 where *f* (MaxRL*i,s*) was modeled as a monotonic decreasing smooth, *ξ*M2*,i,s* denotes the prediction

257 (on the logit scale) for tile *i* during SCC *s* when naval training activities were present on the

258 range using model M2.

259 **2.2.5 Uncertainty propagation**

260 We used posterior simulation (sometimes referred to as a parametric bootstrap, Wood et al.,

261 2017) to propagate uncertainty through M1, M2, and M3. This consisted of sampling from

262 the posterior distribution of the parameters for each model in turn, calculating predictions

263 using these parameters and then refitting the subsequent model with updated offsets. We

264 generated 5,000 sets of posterior samples. Following this procedure through from M1 to M2

265 to M3 incorporated uncertainty from each model in the final predictions of the probability of

266 detecting a GVP given different combinations of covariates.

267 The prediction grid contained all possible combinations of covariates within the realized

268 covariate space; i.e., each hydrophone for each SCC with associated location, hydrophone

269 depth, and area of the tessellation tile, presence/absence of naval activity, and, if naval

270 activity was present, then either sonar absence or sonar received level. Based on the resulting

271 final posterior distribution of results (for model M3) we used 2.5%, 50%, and 97.5% quantiles

272 to obtain median predictions and credible intervals (CIs). Details of the procedure are given

273 in Appendix S1.

## 274 2.2.6 Quantifying the change in probability of GVPs

275 Finally, we calculated the expected change in the probability of detecting a GVP at each

276 hydrophone *P* (GVP) relative to either the probability of detecting a GVP during the pre-

277 activity period when no general naval training activity was present and no MFAS was present

278 (∆M3:M1), or relative to probability of detecting a GVP when general naval training activity

279 was present but no MFAS was present (∆M3:M2).

280 Using the *Nb* posterior samples, we calculated the expected *P* (GVP) under each set of

281 covariates as

*P* (GVP) = logit−1(*µ*M)*,* (1)

for each M = M1, M2, and M3. Then, we calculated the change in *P* (GVP) for each set of covariates between M3 and M1 (∆M3:M1) and between M3 and M2 (∆M3:M2) for each realization of the posterior simulation.

∆*M* 3:*M* 1

∆*M* 3:*M* 2

= *P* (GVP)M3 − *P* (GVP)M1

*P* (GVP)M1

= *P* (GVP)M3 − *P* (GVP)M2

*P* (GVP)M2

(2)

(3)

282 For each received level we calcualted the 2.5th, 50th, and 97.5th quantiles of ∆M3:M1 and

283 ∆M3:M2 to create 95% CIs of change in *P* (GVP) across possible received levels.

284 **3 Results**

# 285 3.1 Data Collection and Processing

286 Data were collected before and during six SCCs: two each in 2013, 2014, and 2017 (Table [1).](#_bookmark1)

287 The number of hydrophones for which recordings were available for each SCC varied from 49

288 to 61. A total of 190,561 30-min observations were made.

289 The exact timing of activities during these exercises varied (Fig. [2).](#_bookmark2) For most SCCs, pre-

290 activity data were available immediately preceding the onset of Naval training activity;

291 however, in February 2013 the only available pre-activity data were collected almost a month

292 prior to the onset of Naval training activity. In some SCCs, weekends or other breaks in

293 training resulted in a break in training activity on the range during the days preceding MFAS

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

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-Exercise | Phase A | Phase B |
| Feb13 | 61 | 113 | 183 | 134 |
| Aug13 | 61 | 204 | 113 | 99 |
| Feb14 | 60 | 514 | 102 | 138 |
| Aug14 | 61 | 262 | 115 | 133 |
| Feb17 | 59 | 450 | 96 | 109 |
| Aug17 | 49 | 270 | 106 | 113 |

295 Across all SCCs, hydrophones, and conditions, a total of 2458 GVPs were identified. The

296 average probability of detecting a GVP during each half-hour period was therefore 1.3%. The

297 spatial distribution of GVPs differed during the pre-activity phases of SCCs (Fig. [S2.1;](#_bookmark0) top

298 panel).

299 Modelled maximum received levels ranged from 38 to 186 dB re 1 *µ*Pa, with a median value

300 when MFAS was present of 147 dB re 1 *µ*Pa. The intensity and spatial distribution of MFAS

301 received levels varied across the range and across SCCs (Fig. [S2.2).](#_bookmark2)

302 Based on the observed data, the probability of detecting a GVP changed by -38% when

303 general naval training activity was present compared to when naval activity was absent, by

304 -61% when naval activity and MFAS were present compared to when only naval activity was

305 present, and by -76% when naval activity and MFAS were present compared to when neither

306 naval activity nor sonar were present (Fig. [S2.3).](#_bookmark3) The highest modelled received level at

307 which a GVP was observed was 164 dB re 1 *µ*Pa.

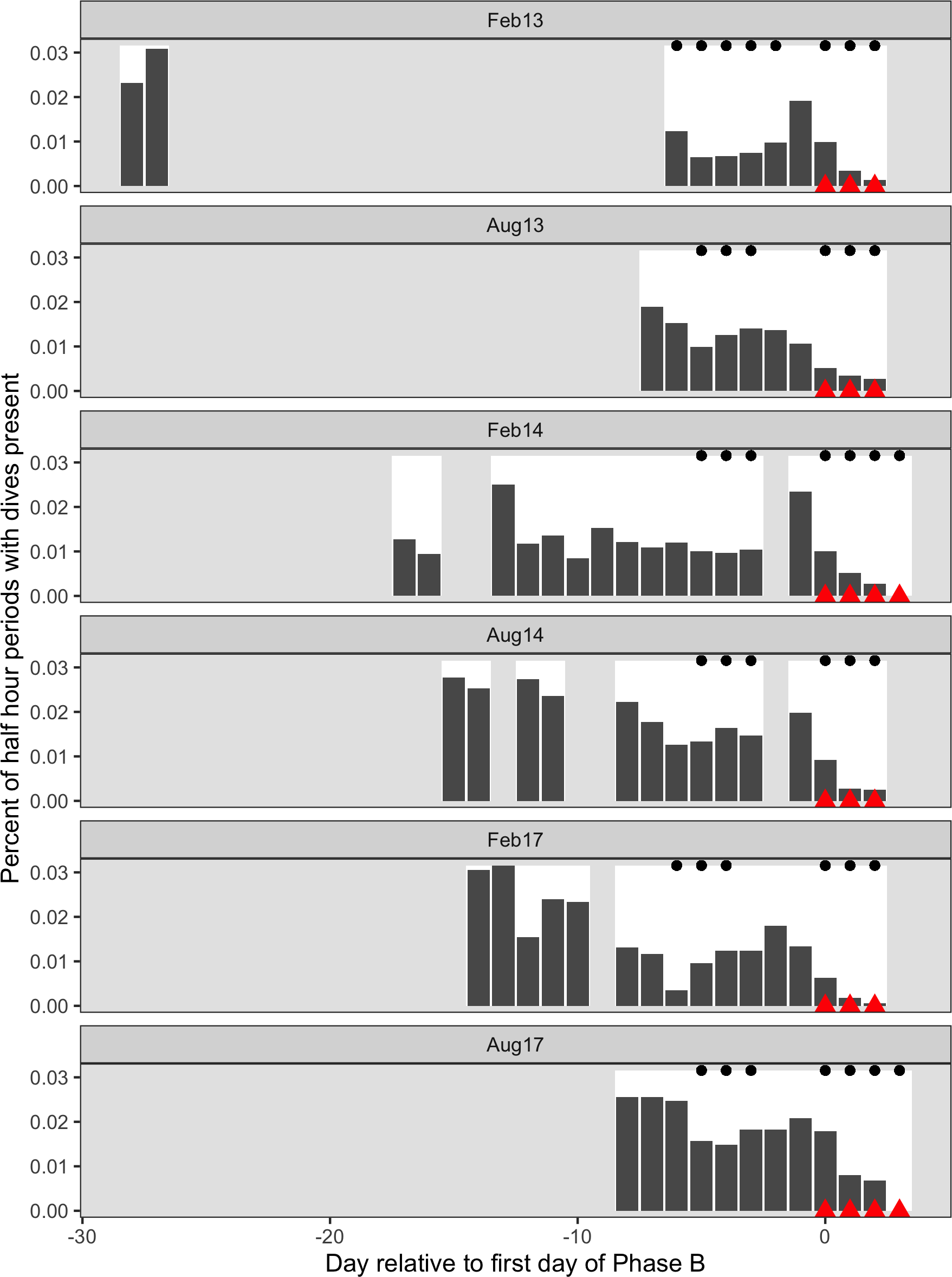


Figure 2: Time series of six recorded Naval training activities at the Pacific Missile Range Facility. The time series are aligned relative to the first day that mid-frequency active sonar (MFAS; red triangles) was used in each exercise (horizontal axis). Days with white background indicate days for 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.

308 **3.2 Spatial Modelling**

309 We created separate tessellations for each SCC (Fig. [S2.4).](#_bookmark4) In August 2017, data were

310 available from fewer hydrophones, and so in some cases the tessellated tiles, with bounding

311 radius of 6,500 m, did not completely cover the range. Hydrophone depths varied from

312 approximately 650 to 4720 m.

313 M1 fitted a spatial model of *P* (GVP) to data collected prior to the onset of naval training

314 activity. This model used a MRF smooth to account for the spatial structure of the range

315 and a spline on depth, with an offset for the log of the area effectively monitored by each

316 hydrophone. Both the MRF and spline on depth were significant at the *α* = 0.05 level

317 (*p*-value < 2E-16), indicating that GVPs varied in space. The model explained 13.5% of

318 deviance in the data set, and visual inspection of observed versus predicted values indicated

319 a good fit to the data (Fig. [S2.5).](#_bookmark5) The model M1 predicted highest *P* (GVP) at hydrophone

320 depths between 1,500 and 2,000 m (Fig. [S2.6).](#_bookmark5)

321 M2 used the predicted values from M1 as an offset and fitted a model to data when naval

322 activity was ongoing, as indicated by the presence of naval ships on the range. This model

323 was intercept-only, and *P* (GVP) when naval activity was ongoing was significantly different

324 from the baseline period at the *α* = 0.05 level (*p*-value < 2E-16). The expected *P* (GVP)

325 decreased by a median of 44% (95% CI 38% – 49%) when naval activity was present compared

326 to when it was absent.

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

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

329 modelled MFAS received level (Fig. [S2.7)](#_bookmark5) and did not include an intercept term. The smooth

330 on MFAS received level was significant at the *α* = 0.05 level (*p*-value = 2E-10) and the

331 model explained 20% of deviance in the data.

332 We did not make inference on sonar received levels below 100 dB re. 1 *µ*Pa because Blainville’s

333 beaked whales are unlikely to perceive MFAS below received levels of approximately 80

334 dB re. 1 *µ*Pa (Pacini et al., 2011) and because very little data (9 hr, or 1% of the data

335 collected when MFAS was present) were collected at received levels below 100 dB re. 1 *µ*Pa.

336 Similarly, we did not make inference on sonar received levels above 165 dB re. 1 *µ*Pa because

337 no GVPs were observed above this received level and therefore M3 predicted *P* (GVP) = 0

338 (95% CI 0–1).

339 For MFAS received levels between 100 and 165 dB re. 1 *µ*Pa, change in *P* (GVP) was calculated

340 relative to the pre-activity baseline period (∆M3:M1; Fig. 4 left panel) and to the period when

341 naval activity was present on the range (∆M3:M2; Fig. 4 right panel) using the posterior

342 samples. For illustration purposes, ∆M3:M1 and ∆M3:M2 calculated using five individual posterior

343 samples are shown in Fig. [S2.8.](#_bookmark6) At a received level of 150 dB, the posterior median of ∆M3:M1

344 was -87% (95% CI -91% – -81%) and the posterior median of ∆M3:M2 was -77% (95% CI -84%

345 – -67%). Relative to when only naval training is present, ∆M3:M2 predicts a 50% reduction in

346 *P* (GVP) at a MFAS received level of 132 dB re 1 *µ*Pa.

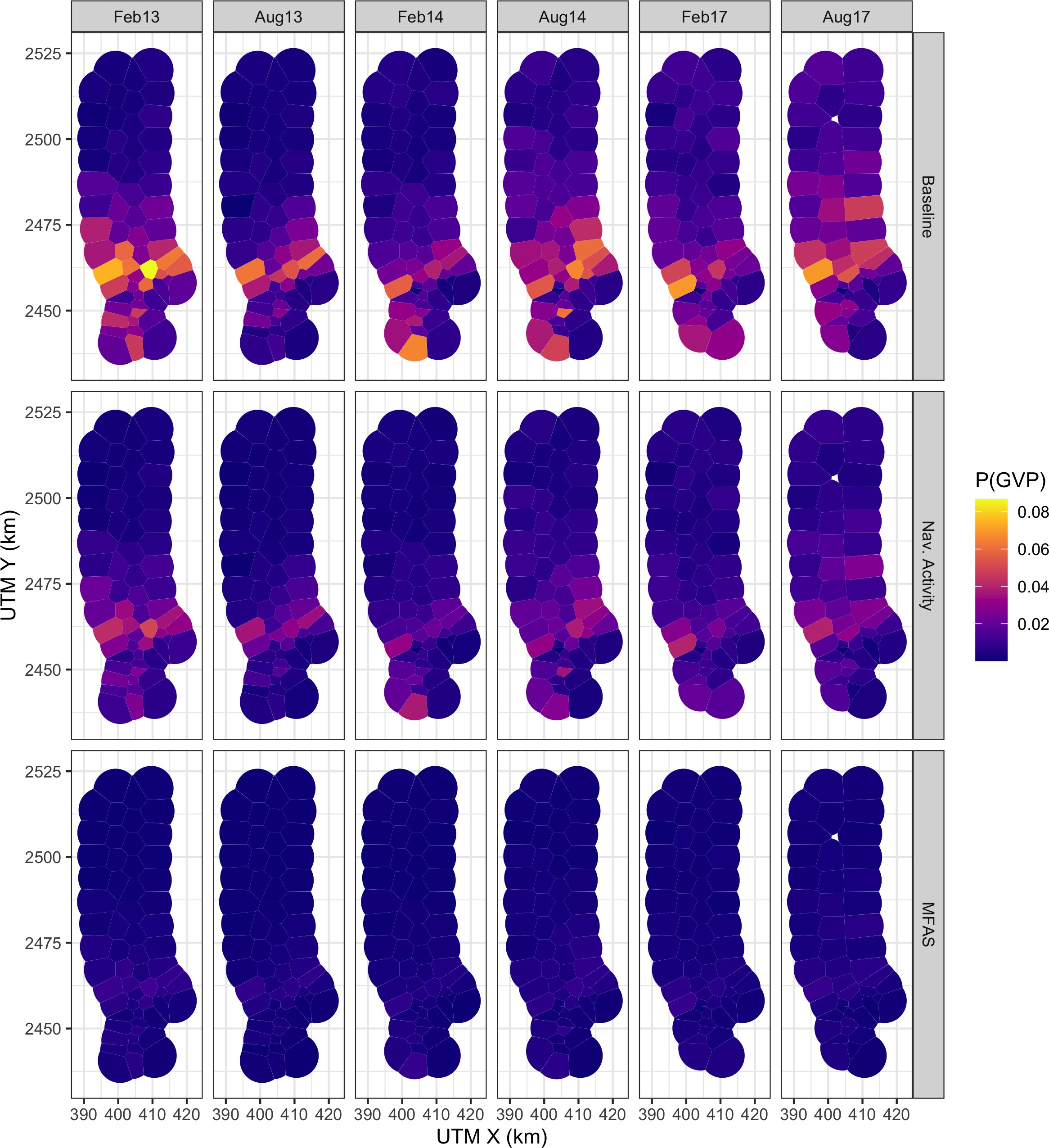


Figure 3: Map of expected probability of detecting a group vocal period (GVP; color scale) at each hydrophone during each Submarine Commander Course (SCC; columns) prior to the onset of naval training activity, during naval training activity when no mid-frequency active sonar (MFAS) was present, and during naval training activity when MFAS was present at a received level of 150 dB re 1 *µ*Pa rms (rows).

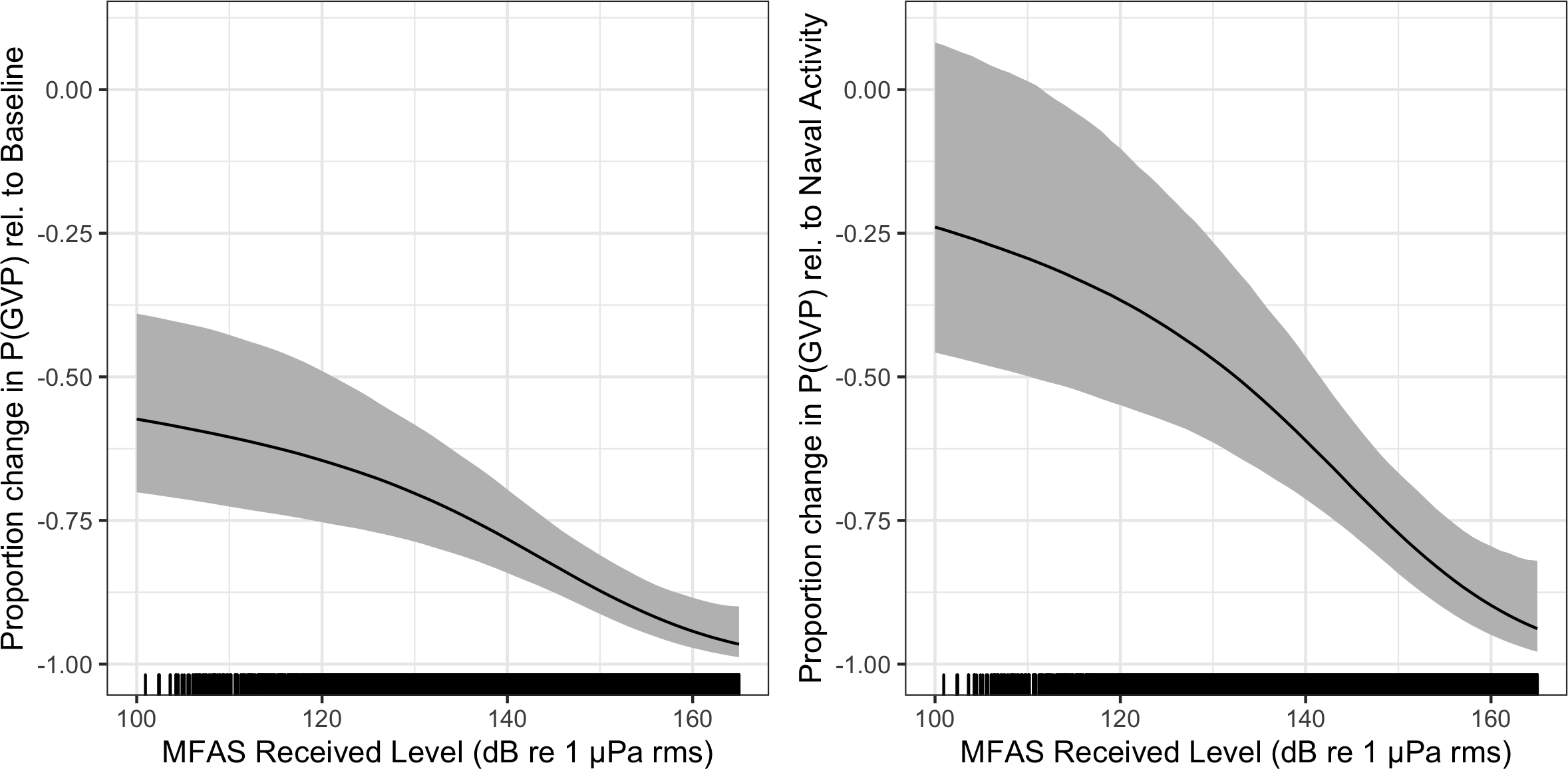


Figure 4: Median (black line) and 95% CI (gray shading) expected change in the probability of detecting a group vocal period (vertical axis) with increasing mid-frequency active sonar (MFAS) received level (horizontal axis) relative to when naval training activity but no MFAS was present on the range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

347 **4 Discussion**

348 We used a series of three linked models to quantify the response of Blainville’s beaked whales

349 to naval training exercises involving MFAS: the first model was fitted to pre-exercise baseline

350 data, the second was fitted to data collected when naval training exercises were ongoing but

351 no hull-mounted MFAS was present, and the third model was fitted to data collected during

352 naval training exercises that used hull-mounted MFAS. We found that the probability of

353 acoustic detections of Blainville’s beaked whales decreased when both naval training exercises

354 and naval training exercises using MFAS were present (Fig. [4).](#_bookmark4)

355 The methods presented here are spatially explicit and account for the spatial confounding

356 of animal distribution and naval training activity. The data used in this study are from an

357 undesigned experiment, where the spatial intensity of the treatments (naval activity and

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

359 whale foraging activity. We did not want the spatial distribution of training exercises and

360 MFAS to influence our understanding of the baseline spatial distribution of Blainville’s beaked

361 whales. Due to the spatial confounding of animal distribution and naval training activity at

362 PMRF, fitting a single model to all of the data leads to greater uncertainty in estimating the

363 impact of sonar, since changes in distribution due to MFAS could be explained as variability

364 in spatial distribution by the MRF (Appendix S3). Our three-stage modelling approach

365 addresses this issue while propagating uncertainty between the models. To our knowledge,

366 this is a novel application of GAMs.

367 The analytical approach outlined in this article could be applied to other species, regions, and

368 types of disturbance where experimental design is not possible. The use of Markov random

369 fields for the spatial term is useful for cases where exact location data are not available,

370 avoiding the inappropriate use of continuous-space smoothers. Shape-constrained smoothing

371 (in our case, monotonically decreasing smooth) is also well-suited to the kind of data we

372 modelled here, ensuring that values can only stay constant or decrease over time (or any

373 other covariate). Finally, the use of a multi-stage posterior sampling scheme for quantifying

374 uncertainty extends to other situations where multiple models are fitted and the results of

375 one part feed into another. Simulation-based approaches such as these bypass the need to

376 derive (often complex) mathematical expressions for variance (or shortcut them by assuming

377 independence).

378 The expected change in the probability of a GVP when MFAS was present included CIs

379 that reflect several sources of uncertainty (Fig. 4). The small number of GVPs when

380 MFAS was present—and therefore sparse coverage of data points across the range of received

381 levels—makes it difficult to estimate the effect of MFAS received level precisely. GVPs were

382 detected in only 1.7% of half-hour periods in the baseline data set (n = 1,831 of 105,939),

383 in 1% of periods (n = 448 of 42,049) when naval activity was present, and 0.2% (n = 50

384 of 17,593) when MFAS was present. Additional data—particularly at relatively low and

385 relatively high source levels, where uncertainty is greatest—may reduce uncertainty in the

386 expected probability of GVPs across different source levels. It is also possible that contextual

387 factors that we did not include in this analysis, such as distance to sound source (DeRuiter

388 et al., 2013; Falcone et al., 2017), may provide additional explanatory power and reduce

389 uncertainty. Finally, the observed uncertainty may reflect true individual variation in response

390 due to characteristics like age and sex (see Harris et al., 2018, sec. 2.2 for a review of relevant

391 publications).

392 The model M3, which modelled the effect of received level on *P* (GVP), was constrained to be

393 monotonically decreasing with no intercept term, so that the predicted *P* (GVP) would be

394 the same or lesser when MFAS was present compared to when only naval training activity

395 was present. However, it is possible that *P* (GVP) could be higher at relatively low MFAS

396 received levels than when only naval training is present, since animals may move away from

397 high-intensity areas, resulting in increased densities in lower-intensity areas. In our data

398 set, some hydrophones had lesser observed *P* (GVP) at low levels of MFAS and some had

399 greater (Fig. [S2.3).](#_bookmark3) Due to small sample size at low intensities, we cannot determine whether

400 observed increases in *P* (GVP) when MFAS was present at relatively low intensities was due

401 to sampling error or to avoidance of high-intensity areas. The version of the model fitted

402 as a single GAM (Appendix S3) predicted the change in *P* (GVP) to be > 0, i.e., increased

403 relative to when only naval training activity was present, at MFAS received levels below 103

404 dB re 1 *µ*Pa (Fig. [S3.1).](#_bookmark0)

405 We do not know when training activities and/or use of hull-mounted MFAS last occurred at

406 PMRF prior to the pre-activity baseline periods used in M1. If beaked whales were already

407 disturbed, then we would expect the pre-activity data to contain fewer GVPs than would

408 be expected in pristine conditions, and therefore our results would underestimate the true

409 impact of training activities and hull-mounted MFAS. Relatedly, we excluded data collected

410 between training activity within an SCC (13% of the available data) as we did not consider

411 it to be true baseline data since naval activity and/or MFAS had recently (within hours

412 or days) been present. It would be interesting to explore the complete data set, including

413 these interim periods, to investigate the timescales on which beaked whales respond to naval

414 activity (e.g., Jones-Todd et al., 2021; Joyce et al., 2019). We might expect that time since

415 training activity or MFAS could lead to recovery of *P* (GVP) towards baseline levels, perhaps

416 modulated by the cumulative exposure to training and MFAS.

417 In a regulatory context, a dose-response function as presented in Fig. [4](#_bookmark4) is often interpreted as

418 representing the proportion of a population that responds (vertical axis) to a given received

419 level (horizontal axis) (Tyack and Thomas, 2019). However, the metric used in this study—the

420 change in the probability of detecting a GVP within a 30-min period—may not directly

421 correspond to the proportion of the population that is affected. It may instead reflect a

422 change in the proportion of time that all individuals in the population spent foraging in the

423 study area. These two interpretations have different implications for understanding sublethal

424 impacts of MFAS. In the former interpretation, given exposure to a certain received level,

425 some of the population is affected and some of the population is not. In the latter, the entire

426 exposed population is affected. With our data, we cannot distinguish between these possible

427 scenarios.

428 In comparison to the risk function developed by Moretti et al. (2014) for Blainville’s beaked

429 whales at AUTEC, our risk function for PMRF predicts a more intense response to naval

430 sonar. This may be because Moretti et al. were not able to account explicitly for the effects

431 of naval training activities that did not include MFAS. Their baseline period consisted of

432 19 hr of data before the onset of MFAS; as at PMRF, it is likely that training activities

433 during this period included sound sources other than MFAS. Therefore, their risk function is

434 likely more analogous to our expected change in the probability of a detection when MFAS is

435 present relative to when naval training activity was present (Fig. [4).](#_bookmark4) Future research will

436 investigate the specific causes of changes in the probability of detecting GVPs before the

437 onset of MFAS. The reduction in detection of foraging dives could be a response to general

438 naval training activity on the range, or to specific sound sources that have not previously

439 been studied. Alternatively, it is possible that Blainville’s beaked whales are semi-resident

440 on the range and have become habituated to SCC activity; they may move off the range in

441 anticipation of MFAS (Manzano-Roth et al., 2016). Resident animals that are frequently

442 exposed to training activity and transient animals that only encounter MFAS occasionally

443 are likely to respond differently to sonar. It is not known how resident the Blainville’s beaked

444 whales are at PMRF, and offshore animals may be detected on the northern hydrophones.

445 Blainville’s beaked whales occur in multiple ocean basins and have been studied on U.S.

446 Navy training ranges in both the Atlantic (AUTEC) and the Pacific (PMRF) Oceans. The

447 AUTEC range is located in a deep basin bounded to the south, east, and west by shallow

448 waters and with maximum depths of 2,000 m. In contrast, the PMRF occurs across a steep

449 slope and into deep water, over 5,000 m in depth. Although the environments at PMRF

450 and AUTEC are different, the foraging dive behavior of Blainville’s beaked whales is similar:

451 dives occur in waters over steep slopes with gradients ranging from 3%–23%, although dives

452 occur in deeper waters (2,000–3,000 m, Henderson et al., 2016) at PMRF that at AUTEC

453 (Hazen et al., 2011; 500–1,300 m, MacLeod and Zuur, 2005). Resident Blainville’s beaked

454 whales off the island of Hawaii also occur in slightly shallower waters than at PMRF, from

455 980–1,410 m (Baird, 2011; Baird et al., 2008). It seems likely the location of the mesopelagic

456 scattering layer (indicating the presence of prey) along the slope rather than the bathymetric

457 depth drives the location of Blainville’s beaked whales; this is supported by the fact that dive

458 depths are similar across areas, occurring on average down to 1,050–1,150 m for 46–60 min

459 (Baird et al., 2008; Joyce et al., 2017; Schorr et al., 2009). Documented responses to MFAS

460 activity are comparable at both ranges, with individuals and groups moving to the periphery

461 of the range or off the range and returning 2–4 days after the cessation of the sonar (Joyce et

462 al., 2019; Manzano-Roth et al., 2016; McCarthy et al., 2011).

463 The similarities in Blainville’s beaked whale behavioral responses to navy training activity

464 across different ranges and environments at similar received levels may indicate the intrinsic

465 nature of the response. The findings presented here and in Moretti et al. (2014) may be

466 applicable to other species and regions, though species-specific dive behaviors and regional

467 differences in oceanography likely modulate the impact of MFAS. For example, existing

468 findings already demonstrate that Cuvier’s respond in a similar manner by reducing their

469 foraging dives and moving away from sonar sources (DeRuiter et al., 2013; Falcone et al.,

470 2017). Conducting a similar analysis of Cuvier’s beaked whale responses at the Southern

471 California Anti-Submarine Warfare Range (SOAR) would further support our understanding

472 of how different populations and species respond to naval sonar.

473 **Acknowledgements**

474 This study was funded by the US Navy Living Marine Resources Program (Contract

475 No. N39430-17-P-1983). We thank Petter Kvadsheim and one anonymous reviewer for

476 their helpful feedback on an earlier version of this manuscript.

477 **Authors’ Contributions**

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

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

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

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

482 Investigation: E.E.H.

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

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

485 Supervision: L.T.

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

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

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

489 **ORCID**

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

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

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

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

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

495 **References**

496 Aguilar de Soto, N., Johnson, M., Madsen, P. T., Tyack, P. L., Bocconcelli, A., &

497 Fabrizio Borsani, J. (2006). Does intense ship noise disrupt foraging in deep-diving

498 Cuvier’s beaked whales (*Ziphius cavirostris*)? *Marine Mammal Science*, *22* (3), 690–699.

499 <https://doi.org/10.1111/j.1748-7692.2006.00044.x>

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

501 & Johnson, M. (2012). No shallow talk: Cryptic strategy in the vocal communication

502 of Blainville’s beaked whales. *Marine Mammal Science*, *28* (2), E75–E92. [https:](https://doi.org/10.1111/j.1748-7692.2011.00495.x)

503 [//doi.org/10.1111/j.1748-7692.2011.00495.x](https://doi.org/10.1111/j.1748-7692.2011.00495.x)

504 Baird, R. W. (2011). Short note: Open-ocean movements of a satellite-tagged Blainville’s

505 beaked whale (*Mesoplodon densirostris*): Evidence for an offshore population in Hawai‘i?

506 *Aquatic Mammals*, *37* (4), 506–511. <https://doi.org/10.1578/AM.37.4.2011.506>

507 Baird, R. W., Webster, D. L., Schorr, G. S., McSweeney, D. J., & Barlow, J. (2008). Diel

508 variation in beaked whale diving behavior. *Marine Mammal Science*, *24* (3), 630–642.

509 <https://doi.org/10.1111/j.1748-7692.2008.00211.x>

510 Bernaldo de Quirós, Y., Fernandez, A., Baird, R. W., Brownell, R. L., N., Allen, D., Arbelo,

511 M., Arregui, M., Costidis, A., Fahlman, A., Frantzis, A., Gulland, F. M. D., Iñíguez,

512 M., Johnson, M., Komnenou, A., Koopman, H., Pabst, D. A., Roe, W. D., Sierra, E.,

513 . . . Schorr, G. (2019). Advances in research on the impacts of anti-submarine sonar

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

515 20182533. <https://doi.org/10.1098/rspb.2018.2533>

516 Cholewiak, D., DeAngelis, A. I., Palka, D., Corkeron, P. J., & Van Parijs, S. M. (2017).

517 Beaked whales demonstrate a marked acoustic response to the use of shipboard

518 echosounders. *Royal Society Open Science*, *4* (12), 170940. [https://doi.org/10.1098/](https://doi.org/10.1098/rsos.170940)

519 [rsos.170940](https://doi.org/10.1098/rsos.170940)

520 D’Amico, A., Gisiner, R., Ketten, D., Hammock, J., Johnson, C., Tyack, P., & Mead, J.

521 (2009). Beaked whale strandings and naval exercises. *Aquatic Mammals*, *35*, 452–472.

522 <https://doi.org/10.1578/AM.35.4.2009.452>

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

524 Falcone, E. A., Friedlaender, A. S., Joseph, J. E., Moretti, D., Schorr, G. S., Thomas,

525 L., & Tyack, P. L. (2013). First direct measurements of behavioural responses by

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

527 <https://doi.org/10.1098/rsbl.2013.0223>

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

529 R. D., Morrissey, R. P., & Moretti, D. J. (2017). Diving behaviour of Cuvier’s beaked

530 whales exposed to two types of military sonar. *Royal Society Open Science*, *4* (8),

531 170629. <https://doi.org/10.1098/rsos.170629>

532 Harris, C. M., Thomas, L., Falcone, E. A., Hildebrand, J., Houser, D., Kvadsheim, P. H.,

533 Lam, F.-P. A., Miller, P. J. O., Moretti, D. J., Read, A. J., Slabbekoorn, H., Southall,

534 B. L., Tyack, P. L., Wartzok, D., & Janik, V. M. (2018). Marine mammals and sonar:

535 Dose-response studies, the risk-disturbance hypothesis and the role of exposure context.

536 *Journal of Applied Ecology*, *55* (1), 396–404. <https://doi.org/10.1111/1365-2664.12955>

537 Hazen, E. L., Nowacek, D. P., St. Laurent, L., Halpin, P. N., & Moretti, D. J. (2011). The

538 relationship among oceanography, prey fields, and beaked whale foraging habitat in

539 the Tongue of the Ocean. *PLoS ONE*, *6* (4), e19269. [https://doi.org/10.1371/journal.](https://doi.org/10.1371/journal.pone.0019269)

540 [pone.0019269](https://doi.org/10.1371/journal.pone.0019269)

541 Heaney, K. D., & Campbell, R. L. (2016). Three-dimensional parabolic equation modeling

542 of mesoscale eddy deflection. *The Journal of the Acoustical Society of America*, *139* (2),

543 918–926. <https://doi.org/10.1121/1.4942112>

544 Henderson, E. E., Martin, S. W., Manzano-Roth, R., & Matsuyama, B. M. (2016). Oc-

545 currence and habitat use of foraging Blainville’s beaked whales (*Mesoplodon den-*

546 *sirostris*) on a U.S. Navy range in Hawaii. *Aquatic Mammals*, *42* (4), 549–562.

547 <https://doi.org/10.1578/AM.42.4.2016.549>

548 Hooker, S. K., Aguilar de Soto, N., Baird, R. W., Carroll, E. L., Claridge, D., Feyrer,

549 L., Miller, P. J. O., Onoufriou, A., Schorr, G., Siegal, E., & Whitehead, H. (2019).

550 Future directions in research on beaked whales. *Frontiers in Marine Science*, *5*, 514.

551 <https://doi.org/10.3389/fmars.2018.00514>

552 Hooker, S. K., Baird, R. W., & Fahlman, A. (2009). Could beaked whales get the bends?

553 Effect of diving behaviour and physiology on modelled gas exchange for three species:

554 Ziphius cavirostris, Mesoplodon densirostris and Hyperoodon ampullatus. *Respiratory*

555 *Physiology & Neurobiology*, *167* (3), 235–246. [https://doi.org/10.1016/j.resp.2009.04.](https://doi.org/10.1016/j.resp.2009.04.023)

556 [023](https://doi.org/10.1016/j.resp.2009.04.023)

557 Hooker, S. K., Fahlman, A., Moore, M. J., Soto, N. A. de, Quirós, Y. B. de, Brubakk, A.

558 O., Costa, D. P., Costidis, A. M., Dennison, S., Falke, K. J., Fernandez, A., Ferrigno,

559 M., Fitz-Clarke, J. R., Garner, M. M., Houser, D. S., Jepson, P. D., Ketten, D. R.,

560 Kvadsheim, P. H., Madsen, P. T., . . . Tyack, P. L. (2012). Deadly diving? Physiological

561 and behavioural management of decompression stress in diving mammals. *Proceedings.*

562 *Biological Sciences*, *279* (1731), 1041–1050. <https://doi.org/10.1098/rspb.2011.2088>

563 Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N., & Tyack, P. L.

564 (2004). Beaked whales echolocate on prey. *Proceedings of the Royal Society of London.*

565 *Series B: Biological Sciences*, *271*, S383–S386. <https://doi.org/10.1098/rsbl.2004.0208>

566 Johnson, M., Madsen, P. T., Zimmer, W. M. X., Soto, N. A. de, & Tyack, P. L. (2006).

567 Foraging Blainville’s beaked whales (*Mesoplodon densirostris*) produce distinct click

568 types matched to different phases of echolocation. *Journal of Experimental Biology*,

569 *209* (24), 5038–5050. <https://doi.org/10.1242/jeb.02596>

570 Jones-Todd, C. M., Pirotta, E., Durban, J. W., Claridge, D. E., Baird, R. W., Falcone, E.

571 A., Schorr, G. S., Watwood, S., & Thomas, L. (2021). Discrete-space continuous-time

572 models of marine mammal exposure to Navy sonar. *Ecological Applications*, *0* (0),

573 e02475. <https://doi.org/10.1002/eap.2475>

574 Joyce, T. W., Durban, J. W., Claridge, D. E., Dunn, C. A., Fearnbach, H., Parsons, K. M.,

575 Andrews, R. D., & Ballance, L. T. (2017). Physiological, morphological, and ecological

576 tradeoffs influence vertical habitat use of deep-diving toothed-whales in the Bahamas.

577 *PLOS ONE*, *12* (10), e0185113. <https://doi.org/10.1371/journal.pone.0185113>

578 Joyce, T. W., Durban, J. W., Claridge, D. E., Dunn, C. A., Hickmott, L. S., Fearnbach, H.,

579 Dolan, K., & Moretti, D. (2019). Behavioral responses of satellite tracked Blainville’s

580 beaked whales (*Mesoplodon densirostris*) to mid-frequency active sonar. *Marine*

581 *Mammal Science*, 1–18. <https://doi.org/10.1111/mms.12624>

582 Macleod, C. D., & D’Amico, A. (2006). A review of beaked whale behaviour and ecology

583 in relation to assessing and mitigating impacts of anthropogenic noise. *Journal of*

584 *Cetacean Research and Management*, *7* (3), 211–221.

585 MacLeod, C. D., & Zuur, A. F. (2005). Habitat utilization by Blainville’s beaked whales

586 off Great Abaco, northern Bahamas, in relation to seabed topography. *Marine Biology*,

587 *147* (1), 1–11. <https://doi.org/10.1007/s00227-004-1546-9>

588 Madsen, P. T., Aguilar de Soto, N., Arranz, P., & Johnson, M. (2013). Echolocation

589 in Blainville’s beaked whales (*Mesoplodon densirostris*). *Journal of Comparative*

590 *Physiology A*, *199* (6), 451–469. <https://doi.org/10.1007/s00359-013-0824-8>

591 Manzano-Roth, R., Henderson, E. E., Martin, S. W., Martin, C., & Matsuyama, B.

592 (2016). Impacts of U.S. Navy training events on Blainville’s beaked whale (*Mesoplodon*

593 *densirostris*) foraging dives in Hawaiian waters. *Aquatic Mammals*, *42* (4), 507–518.

594 <https://doi.org/10.1578/AM.42.4.2016.507>

595 Marques, T. A., Thomas, L., Ward, J., DiMarzio, N., & Tyack, P. L. (2009). Estimating

596 cetacean population density using fixed passive acoustic sensors: An example with

597 Blainville’s beaked whales. *The Journal of the Acoustical Society of America*, *125* (4),

598 1982–1994. <https://doi.org/10.1121/1.3089590>

599 Martin, C. R., Henderson, E. E., Martin, S. W., Helble, T. A., Manzano-Roth, R. A., 600 Matsuyama, B. M., & Alongi, G. A. (2020). *FY18 annual report on Pacific missile* 601 *range facility marine mammal monitoring*. Retrieved from Naval Information Warfare 602 Center Pacific San Diego United States website: [https://apps.dtic.mil/sti/citations/](https://apps.dtic.mil/sti/citations/AD1091141) 603 [AD1091141](https://apps.dtic.mil/sti/citations/AD1091141)

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

605 McCarthy, E., Moretti, D., Thomas, L., DiMarzio, N., Morrissey, R., Jarvis, S., Ward, 606 J., Izzi, A., & Dilley, A. (2011). Changes in spatial and temporal distribution and 607 vocal behavior of Blainville’s beaked whales (*Mesoplodon densirostris*) during multiship 608 exercises with mid-frequency sonar. *Marine Mammal Science*, *27* (3), E206–E226. 609 <https://doi.org/10.1111/j.1748-7692.2010.00457.x>

610 Moretti, D., Thomas, L., Marques, T., Harwood, J., Dilley, A., Neales, B., Shaffer, J., 611 McCarthy, E., New, L., Jarvis, S., & Morrissey, R. (2014). A risk function for behavioral 612 disruption of Blainville’s beaked whales (*Mesoplodon densirostris*) from Mid-Frequency 613 Active Sonar. *PLoS ONE*, *9* (1), e85064. <https://doi.org/10.1371/journal.pone.0085064>

614 New, L. F., Moretti, D. J., Hooker, S. K., Costa, D. P., & Simmons, S. E. (2013). Using 615 energetic models to investigate the survival and reproduction of beaked whales (family 616 Ziphiidae). *PLoS ONE*, *8* (7), e68725. <https://doi.org/10.1371/journal.pone.0068725>

617 Pacini, A. F., Nachtigall, P. E., Quintos, C. T., Schofield, T. D., Look, D. A., Levine, G. A.,

618 & Turner, J. P. (2011). Audiogram of a stranded Blainville’s beaked whale (*Mesoplodon*

619 *densirostris*) measured using auditory evoked potentials. *Journal of Experimental*

620 *Biology*, *214* (14), 2409–2415. <https://doi.org/10.1242/jeb.054338>

621 Pirotta, E., Booth, C. G., Costa, D. P., Fleishman, E., Kraus, S. D., Lusseau, D., 622 Moretti, D., New, L. F., Schick, R. S., Schwarz, L. K., Simmons, S. E., Thomas, L., 623 Tyack, P. L., Weise, M. J., Wells, R. S., & Harwood, J. (2018). Understanding the 624 population consequences of disturbance. *Ecology and Evolution*, *8* (19), 9934–9946.

625 <https://doi.org/10.1002/ece3.4458>

626 Pirotta, E., Milor, R., Quick, N., Moretti, D., Di Marzio, N., Tyack, P., Boyd, I., & 627 Hastie, G. (2012). Vessel noise affects beaked whale behavior: Results of a dedicated 628 acoustic response study. *PLoS ONE*, *7* (8), e42535. [https://doi.org/10.1371/journal.](https://doi.org/10.1371/journal.pone.0042535) 629 [pone.0042535](https://doi.org/10.1371/journal.pone.0042535)

630 Pya, N., & Wood, S. N. (2015). Shape constrained additive models. *Statistics and*

631 *Computing*, *25* (3), 543–559. <https://doi.org/10.1007/s11222-013-9448-7>

632 R Core Team. (2018). *R: A Language and Environment for Statistical Computing*.

633 Retrieved from <https://www.R-project.org/>

634 Rue, H., & Held, L. (2005). *Gaussian Markov Random fields: Theory and Applications*.

635 London: Chapman & Hall.

636 Schorr, G. S., Baird, R. W., Hanson, M. B., Webster, D. L., McSweeney, D. J., & Andrews, 637 R. D. (2009). Movements of satellite-tagged Blainville’s beaked whales off the island of 638 Hawai‘i. *Endangered Species Research*, *10*, 203–213. <https://doi.org/10.3354/esr00229>

639 Simonis, A. E., Brownell, R. L., Thayre, B. J., Trickey, J. S., Oleson, E. M., Huntington, 640 R., & Baumann-Pickering, S. (2020). Co-occurrence of beaked whale strandings and 641 naval sonar in the Mariana Islands, Western Pacific. *Proceedings of the Royal Society* 642 *B: Biological Sciences*, *287* (1921), 20200070. <https://doi.org/10.1098/rspb.2020.0070>

643 Southall, B., Nowacek, D., Miller, P., & Tyack, P. (2016). Experimental field studies to 644 measure behavioral responses of cetaceans to sonar. *Endangered Species Research*, *31*, 645 293–315. <https://doi.org/10.3354/esr00764>

646 Stanistreet, J. E., Beslin, W. A. M., Kowarski, K., Martin, S. B., Westell, A., & Moors- 647 Murphy, H. B. (2022). Changes in the acoustic activity of beaked whales and sperm 648 whales recorded during a naval training exercise off eastern Canada. *Scientific Reports*, 649 *12* (1), 1973. <https://doi.org/10.1038/s41598-022-05930-4>

650 Turner, R. (2019). *Deldir: Delaunay triangulation and Dirichlet (Voronoi) tessellation*.

651 Retrieved from [https://CRAN.R-project.org/package=deldir](https://CRAN.R-project.org/package%3Ddeldir)

652 Tyack, P. L., Johnson, M., Soto, N. A., Sturlese, A., & Madsen, P. T. (2006). Extreme

653 diving of beaked whales. *Journal of Experimental Biology*, *209* (21), 4238–4253. [https:](https://doi.org/10.1242/jeb.02505)

654 [//doi.org/10.1242/jeb.02505](https://doi.org/10.1242/jeb.02505)

655 Tyack, P. L., & Thomas, L. (2019). Using dose–response functions to improve calculations 656 of the impact of anthropogenic noise. *Aquatic Conservation: Marine and Freshwater* 657 *Ecosystems*, *29* (S1), 242–253. <https://doi.org/10.1002/aqc.3149>

658 Tyack, P. L., Zimmer, W. M. X., Moretti, D., Southall, B. L., Claridge, D. E., Durban, J. 659 W., Clark, C. W., D’Amico, A., DiMarzio, N., Jarvis, S., McCarthy, E., Morrissey, R., 660 Ward, J., & Boyd, I. L. (2011). Beaked whales respond to simulated and actual navy 661 sonar. *PLoS ONE*, *6* (3), e17009. <https://doi.org/10.1371/journal.pone.0017009>

662 U.S. Department of the Navy. (2017). *Criteria and thresholds for U.S. Navy acoustic and* 663 *explosive effects analysis (phase III)*. Retrieved from [https://www.goaeis.com/portals/](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) 664 [goaeis/files/eis/draft\_seis\_2020/supporting\_technical/Criteria\_and\_Thresholds\_](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)

665 [for\_U.S.\_Navy\_Acoustic\_and\_Explosive\_Effects\_Analysis\_June2017.pdf](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)

666 U.S. Department of the Navy. (2018). *Final environmental impact statement/overseas*

667 *environmental impact ttatement Hawaii-Southern California training and testing*. Re-

668 trieved from [https://www.hstteis.com/portals/hstteis/files/hstteis\_p3/feis/section/](https://www.hstteis.com/portals/hstteis/files/hstteis_p3/feis/section/HSTT_FEIS_3.07_Marine_Mammals_October_2018.pdf)

669 [HSTT\_FEIS\_3.07\_Marine\_Mammals\_October\_2018.pdf](https://www.hstteis.com/portals/hstteis/files/hstteis_p3/feis/section/HSTT_FEIS_3.07_Marine_Mammals_October_2018.pdf)

670 Urick, R. J. (1983). *Principles of Underwater Sound* (Third Edition, Reprint 2013). New

671 York: McGraw-Hill, Inc.

672 Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society:*

673 *Series B (Statistical Methodology)*, *65* (1), 95–114. [https://doi.org/10.1111/1467-9868.](https://doi.org/10.1111/1467-9868.00374)

674 [00374](https://doi.org/10.1111/1467-9868.00374)

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

676 Chapman; Hall/CRC.

677 Wood, S. N., Li, Z., Shaddick, G., & Augustin, N. H. (2017). Generalized additive 678 models for gigadata: Modeling the U.K. Black smoke network daily data. *Journal of* 679 *the American Statistical Association*, *112* (519), 1199–1210. [https://doi.org/10.1080/](https://doi.org/10.1080/01621459.2016.1195744)

680 [01621459.2016.1195744](https://doi.org/10.1080/01621459.2016.1195744)

681 **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, ***β***∗ ∼ MVN(***β***ˆ*,* **V*β***ˆ), where ***β***ˆ is the estimate of the model coefficients

and **V*β***ˆ is the posterior covariance matrix. Here the ***β*** 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 (**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 ***µ***∗, we calculate as follows:

***µ***∗ = *g*−1(***η***∗) = *g*−1(**X***p****β***∗ + ***ξ***)*,*

682 where *g* was the link function, ***η***∗ was the linear predictor and ***ξ*** was any offset used by 683 this prediction. Variance estimates can be obtained by taking the empirical variance of 684 the resulting predictions (Wood et al. 2017; section 7.2.6). The prediction grid contained 685 all possible combinations of covariates within the realized covariate space; i.e., each 686 hydrophone for each SCC with associated location, hydrophone depth, and area of the 687 tessellation tile, presence/absence of naval activity, and, if naval activity was present, then 688 either sonar absence or sonar received level between 35 and 190 dB in intervals of 5 dB. 689 This procedure was repeated for each model, with refitting to updated offsets from the 690 previous model.

691 An algorithm for calculating the variance from our multi-stage approach is as follows. First

692 define *Nb* as the number of samples to take (*Nb*=5,000 here), let **X***p,*M*j* for *j* = 1*,* 2*,* 3 be

693 the matrix that maps coefficients to the predictions for model M*j*. For *Nb* times:

Draw a sample from the posterior of M1:

|  |  |
| --- | --- |
| 694 | 1. |
| 695 | 2. |
| 696 | 3. |
| 697 | 4. |
| 698 | 5. |
| 699 |  |
| 700 | 6. |
| 701 | 7. |

***β***˜M1 ∼ MVN(***β***ˆM1*,* **V**M1).

Calculate a new offset for M2, ***ξ***˜M1 = **X***p,*M1***β***˜M1 + log*e* **A**. Refit M2 with ***ξ***˜M1 as the offset, to obtain M2t.

˜ ˆ

Draw a sample from the posterior of M2t:

***β***M2 ∼ MVN(***β***M2 *,* **V**M2 )

Calculate a new offset for M3,

***ξ***˜M2 = **X***p,*M2***β***˜t

+ ***ξ***˜M1 (predictions for the sonar data

locations for M2t, when no sonar was present). Refit M3 with offset ***ξ***˜M2 to obtain M3t.

M2

Predict ***µ***M1 , ***µ***M2 , and ***µ***M3 over prediction grid and store them.

702 We then calculated summary statistics (means and variances) of the *Nb* values of ***µ***M1 , ***µ***M2 , 703 and ***µ***M3 we generated. The empirical variance of the *Nb* values of ***µ***M3 gave the uncertainty, 704 incorporating components from all three models. We took appropriate pointwise quantiles 705 (e.g., 2.5th and 97.5th for a 95% interval) to form confidence bands for the functional 706 relationships between sonar received level and estimated probability of detecting GVPs.

707 **S2: Supplementary Tables and Figures**

Diagram

Description automatically generated

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 3h8ydrophone tile).

2525

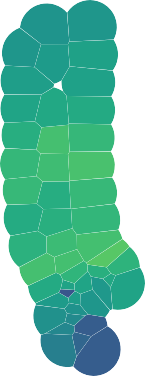
Feb13

Aug13

Feb14

Aug14

Feb17



Aug17

2500

UTM Y (km)

2475

MedRL

160



150

140

130

2450

390 400 410 420 390 400 410 420 390 400 410 420 390 400 410 420 390 400 410 420 390 400 410 420

UTM X (km)

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

Chart, box and whisker chart

Description automatically generated

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.

Diagram

Description automatically generated

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.

Chart, scatter chart

Description automatically generated

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

682

Chart, line chart

Description automatically generated

683

684

685

686

687

Figure S2.6: Spline for the relationship between P(GVP) and depth from M1 on the logit-link scale. Solid line: best fit; dashed lines: 95% CIs.

Chart

Description automatically generated

688

689

690 Figure S2.7: Spline for the relationship between P(GVP) and maximum received level from 691 M3 on the logit-link scale. Solid line: best fit; dashed lines: 95% CIs.

Chart, line chart

Description automatically generated

Figure S2.8: 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.

692 **S3: Single GAM**

693 A single GAM could be used to quantify the effect of naval sonar on Blainville’s beaked

694 whales. Here, we present such a model and compare the results to the results obtained using

695 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:

GVP*i,s,t* ∼ Bin(1*, µi,s,t*). The linear predictor on the logit scale was given as:

logit (*µi,s,t*) = *β*0 + *β*1NavTrain*t* + *f* (MRF*i,s*) + *f* (Depth*i*) + *f* (MaxRL*i, t*)Sonar*t* + log*e Ai*)

696 where *β*0 is an intercept, *β*1NavTrain*t* is the effect of naval training times an indicator

697 variable for whether naval training was present or absent at time *t*, *f* (MRF*i,s*) denotes the

698 Markov random field used to smooth space, *f* (Depth*i*) is a smooth of depth (using a thin 699 plate spline; Wood et al. 2003), *f* (MaxRL*i,t*)Sonar*t* is a smooth of sonar received level (using 700 a thin plate spline) times an indicator variable for whether sonar was present or absent at 701 time *t*, and log*e Ai* is an offset for the area (in km2) of each tile, *Ai*.

702 We fit the model to the same data used in M1, M2, and M3 (see Methods section of main

703 manuscript for details) using mgcv (Wood, 2017).

704 This single GAM (Fig. [S3.1)](#_bookmark0) predicts a 41% (95% CI 34%-46%) decrease in *P* (GVP) when 705 naval training is present compared to the baseline period, whereas the multi-stage GAM (Fig. 706 4) predicts a decrease of 44% (95% CI 38%-49%). The single GAM predicts that at a MFAS 707 received level of 150 dB re 1 *µ*Pa, *P* (GVP) decreases by 87% (95% CI 71%-95%) relative to 708 when only naval training is present, whereas the multi-stage model predicts the same

709 decrease of 87% with a narrower credible interval (95% CI 81%-92%). Relative to when only 710 naval training is present, the single GAM predicts a 50% reduction in *P* (GVP) at a MFAS 711 received level of 120 dB re 1 *µ*Pa, whereas the multi-stage model predicts a 50% reduction at 712 a MFAS received level of 132 dB re 1 *µ*Pa.

713 The major difference between this single GAM and the multi-stage model presented in the 714 main text of the manuscript is that here, the spatial smooth is constructed using data from 715 the baseline, naval training, and MFAS periods of each SCC. Therefore, the spatial

716 distribution of MFAS may influence the predicted distribution of Blainville’s beaked whales.

717 Using a single GAM leads to similar point estimates of the impact of sonar with greater

718 uncertainty than the multi-stage model.

Chart

Description automatically generated

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.