**MMSCI-5285 Response to Reviewers**

**R1 Comment:** Line 44-45: can also cite Hooker et al. 2019

Response: Done.

**R1 Comment:** Line 49-50: add Stanistreet et al. 2022 reference

Response: Done.

**R1 Comment:** Line 59: Add McCarthy et al. 2011 reference

Response: Done.

**R1 Comment:** Line 61-62: There was also Erin Falcone’s work measuring Cuvier’s response to helo-dipping sonars (Falcone et al. 2017) which should be mentioned here

Response: Done.

**R1 Comment:** Line 144: what about the instances where multiple ships were broadcasting (cumulative exposure)? Did you have a covariate for multiple ships to see if there was any interaction/effect by including whether there were multiple sources?

Response: Due to the sensitive nature of the naval training activities, we were not given information about the exact number of ships (or other sources) on the range or using sonar at any given time. For each time period, we were given a covariate that was 0 if no sources were present, 1 if 1 source was present, and 2 if 2 or more sources were present. In practice, all data used in this study had either 0 sources (no activity) or 2 or more sources (when general training or hull-mounted MFAS were present). Therefore, we were unable to assess the impact of multiple versus single sources.

We did experiment with using a metric of cumulative exposure (SEL), but dropped the idea because it required the assumption that whales did not move around the range over the course of a SCC (i.e., the SEL calculated at a single point in space is not what a whale actually experiences).

However, as part of a follow-up study, we are looking at data from the general training period in more detail. We now have data on distance and sonar transmission from the closest surface ships, surface recovery boats, torpedoes, and submarines to each hydrophone. We are investigating whether the decrease in probability of detecting group vocal periods during this period can be attributed to one or more of these sources.

**R1 Comment:** Line 150: “center hydrophone” right? Or did you do all hydrophones in the group? Need to clarify

Response: Modified text to clarify that the transmission loss was estimated from the closest ship to each individual hydrophone.

**R1 Comment:** Line 190: can we assume that the detection range (6.5 km) will be the same at PMRF as it is at AUTEC? As you mentioned earlier, the two ranges have different bathymetries and depths. Has it been estimated what the detection range is at PMRF?

Response: We agree that it would be preferable to use an estimate of detectability specific to PMRF (and ideally, specific to each hydrophone, since detection range is unlikely to be uniform around the range) rather than assuming that the estimate from AUTEC is applicable. However, the estimate from AUTEC is currently the best available.

We chose presence of group vocal periods (GVPs) during half hour periods as our metric of interest because we thought it would be the most robust to potential differences in detectability around the range. I.e., we are not confident that we detect all echolocation clicks, but we are confident that we detect all diving groups.

In the southern portion of the range (where the bathymetry is most complex) the phones are spaced closer than 6 km so we have high confidence we aren’t missing groups on the range. In this area groups are usually detected on more than one phone. In the northern part of the range where the hydrophones are spaced further apart, the bathymetry is deep and generally less complex, so propagation becomes largely spherical. Therefore, we still feel confident we aren’t missing groups, although they are more likely to be detected on a single phone rather than multiple phones.

The estimate of a detection range of 6.5 km affects the hydrophones at the edge of the range the most, since we have bounded the tessellations based on this radius. If the estimate of 6.5 km is incorrect, the offset for area (used as effort) would be incorrect for the edge hydrophones. However, since the effort is being held constant across phases of the training exercise, we would not expect this to affect our results on the scale of the PMRF. I.e., if the estimate of effort is incorrect for some hydrophones, then beaked whale occurrence on those hydrophones would be incorrect, but relative changes during the training exercise should be the same.

For edge hydrophones, one potential issue with our method is that beaked whales at the edge of edge hydrophone tiles are quite far (6.5 km) from the hydrophone, and therefore are not actually experiencing the RL modelled at the hydrophone. We would expect this to lead to an underestimate of the effect of hull-mounted MFAS.

Previously, we included a covariate for edge hydrophones in the baseline model, to account for possible error in the estimated effective detection area. The model with the covariate for edge hydrophones did not perform as well as the model without (according to AIC) so we dropped it from the model.

**R1 Comment:** Line 193: what do you mean “different combinations of hydrophones were used”? Are you saying that the range operators only had certain hydrophones on/recording for one SCC and then a different hydrophone set for another SCC? Please clarify. Does this also mean you generated different probability models pre-SCC for each SCC as opposed to combining them all (so one M1 per SCC)? – I see the answer to my question in section 2.2.2. Maybe introduce on line 193 a little more clearly (ie: “different combinations of hydrophones were used due to...”)?

Response: We have modified the text to explain that “Due to recording capacity, a subset of PMRF hydrophones were recorded during each SCC. While this subset was kept as consistent as possible, due to occasional hydrophone failures different combinations of hydrophones were recorded during different SCCs. Therefore, we created separate tessellations for each SCC.”

**R1 Comment:** Line 197: please include how long it had been since the last exercise ie “no other naval activity was known to occur for at least 1 week prior” as we saw with McCarthy et al. 2011 and Stanistreet et al. 2022 that it takes time for baseline foraging to resume (sometimes as much as 1 week)

Response: We don’t have access to the full range schedules and therefore cannot be certain when MFAS was last used on the range. We can only be sure that there was no hull-mounted MFAS during our baseline and general training periods. If there were training exercises preceding the baseline period, the effects of training exercises and hull-mounted MFAS reported here could be underestimated.

We have added the following to the discussion: “We do not know when training activities and/or use of hull-mounted MFAS last occurred at PMRF prior to the pre-activity baseline periods used in M1. If beaked whales were already disturbed, then we would expect the pre-activity data to contain fewer GVPs than would be expected in pristine conditions, and therefore our results would underestimate the true impact of training activities and hull-mounted MFAS.”

**R1 Comment:** Line 270: I think there’s a typo here? “no general naval training activity present and MFAS was present”? Same with the M2:M3 text

Response: This is correct. We have modified the sentence to clarify that “Finally, we calculated the expected change in the probability of detecting a GVP at each hydrophone relative to either the probability of detecting a GVP during the pre-activity period when no general naval training activity was present and no MFAS was present or relative to probability of detecting a GVP when general naval training activity was present but no MFAS was present.”

**R1 Comment:** Line 299: where on the range was the highest RL? Center? Bottom left?

Response: Figure S2.2 shows median received levels at each hydrophone during each SCC. We reference this figure in the preceding paragraph: “The intensity and spatial distribution of MFAS received levels varied across the range and across SCCs (Fig. S2.2).”

**R1 Comment:**Interesting that in Figure 2 there was an increase in the % half hour periods between days when ships are out but no MFA used. (I’m looking specifically at Feb 13, Feb 14 and Aug14 where the percentage is noticeably larger the day before the sonar). What are the authors’ thoughts on this? Do you think Blainville’s are predicting a sonar event and trying to get more foraging dives in before they are displaced to lesser quality habitats?

Response: We think this may be a rebound effect after the initial training exercises that did not include hull-mounted MFAS. However, we did not include this data in our study, as it was neither “training” nor “baseline”, but something in between. In future, it would be interesting to look at fine-scale temporal changes in GVP detections before, during, and after training exercises. (See also discussion paragraph beginning at Line 398).

**R1 Comment:** Line 375: include the n value for # half hour periods in the baseline dataset (the readers will see the 0 inflatedness that way)

Response: Done.

**R1 Comment:** Lines 379-380: here is where you can address my question from line 144

Response: See above.

**R1 Comment:** Lines 427-428: I thought that is what Manzano-Roth found so wouldn’t this sentence switch from a conditional to a statement?

Response: We’ve added Manzano-Roth as a reference here but kept it conditional, since neither study can definitively attribute observed changes in detection of GVPs to animal movement.

**R1 Comment:** Lines 442-444: “location of the mesopelagic scattering layer (indicating the presence of prey) along the slope that drives the location of Blainville’s beaked whales rather than the bathymetric” (remove “that”)

Response: Done.

**R2 Comment:** The use of the term “MFAS” (mid frequency active sonar) should be reconsidered. At least it should be defined but my preference would be to replace it with the frequency band meant. Mid frequency means different things between different navies and if you ask a fishery acoustician it means something very different. Please define or replace MFAS.

Response: We have added a definition and citation for a U.S. Navy report the first time this term is used.

**R2 Comment:** I propose to replace the term “general navy activity” with “navy activity without 1-10 kHz sonar” (or whatever frequency range MFAS means here) to increase readability. Or at least it should be defined the first time it is mentioned.

Response: We have added an explanation that general training activity refers to periods without hull-mounted MFAS broadcast from surface ships.

**R2 Comment:** The dB metrics used for received and source level are not clearly defined. I assume it is rms levels. Add proper definition of even better reference the ISO standard (ISO 18405:2017).

Response: Yes, everything is reported in rms following the ISO standard. We have added this to the text.   
  
**R2 Comment:** In cases were estimated max received level at the hydrophones were lower than 140dB my understanding is that the phones have enough dynamic range to record it. Did you do any comparison between the measured and estimated levels for those cases?

Response: Early on in this analysis, we did compare measured and modelled received levels. However, in addition to the problem of hydrophones clipping at 140 dB, the hydrophones are very close to the sea floor, while the modelled received levels were modelled at a depth of 1000 m, where diving beaked whales would be experiencing sonar. The measured received levels therefore tended to be lower than the modelled levels, since sonar was propagating from ships at the surface. In future, it would be good to conduct a separate study to compare measured and modelled received levels at depth across the range.