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

Done

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

Response: Done

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

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

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 (baseline) or 2 or more sources (Phase A and Phase B). Therefore, we were unable to assess the impact of multiple versus single sources.

However, as part of a follow-up study, we are looking at data from Phase A 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 Phase A 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 the detection specific to PMRF, rather than assuming that the estimate from AUTEC is applicable. However, the estimate from AUTEC is currently the best available.

The maximum distance between hydrophones is XX km. 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. I’m trying to say that we’re measuring everything relative to baseline. Add something about this to the discussion?

**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 and occasional instrument failure, not all hydrophones were recorded during all SCCs. Because different combinations of hydrophones were used during different SCCs, we created separate tessellations for each SCC.”

That is basically it. We don’t record ALL the phones at PMRF, the subset of 62 is based on the digital input cards that convert the incoming analog signal from the hydrophones to a digital signal that we record. The two card we used each had 32 slots, so we recorded 62 hydrophones plus two time-code signals. Over the years, different hydrophones would go out (or the entire string in 2017!) so we would switch which phones were recorded. Generally we tried to keep as consistent as possible. But since you created the tessellations relative to each SCC, that was accounted for. So perhaps this could be rephrased as “Due to recording capacity, a subset of PMRF hydrophones were recorded each year. While this subset was held as consistently as possible, because of occasional hydrophone failures a small number of the hydrophones that were recorded changed across years. These slight differences in hydrophone combinations during different SCCs led to the creation of separate tessellations for each SCC.” Something like that – to indicate that the majority of the phones were consistent across SCCs but just a small number changed.

**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 MFAS during our baseline and Phase A periods. We have added text to the discussion noting that if there were training exercises

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

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

**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?

I think this is rebound after the initial training exercises that did not include MFAS

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

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

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

Eh, sort of? Need to rephrase so it’s clear what we mean. Could also cite M-R here ie. Our results are consistent with previous work

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”)

Done

Reviewer 2 Comments

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.  
  
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.

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).  
  
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?