SPERM WHALES EXHIBIT VARIATION IN ECHOLOCATION TACTICS WITH DEPTH AND SEA STATE BUT NOT NAVAL SONAR EXPOSURES

Author names and email addresses: Saana Isojunno[1] si66@st-andrews.ac.uk, Alexander M. von Benda-Beckmann[2] sander.vonbendabeckmann@tno.nl, Paul J. Wensveen[1,3] pjw@hi.is, Petter H. Kvadsheim[4] petter-helgevold.kvadsheim@ffi.no, Frans-Peter A. Lam[2] frans-peter.lam@tno.nl, Kalliopi C. Gkikopoulou[1] kg366@st-andrews.ac.uk, Viivi Pöyhönen[5] viivipoyhonen@hotmail.com, Peter L. Tyack[1] plt@st-andrews.ac.uk, Benjamin Benti[1] benjamin.benti@protonmail.com, Ilias Foskolos[6] ilias.foskolos@bio.au.dk, Jacqueline Bort[7] jacqueline.bort@navy.mil, Miguel Neves[8] miguelneves.reis@gmail.com, Nicoletta Biassoni[1] nicobiassoni@gmail.com, Patrick J. O. Miller[1] pm29@st-andrews.ac.uk

Author affiliations: (1) Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, UK; (2) Acoustics and Sonar, Netherlands Organization for Applied Scientific Research (TNO), The Hague, The Netherlands, (3) Faculty of Life and Environmental Sciences, University of Iceland, Reykjavik, Iceland; (4) Maritime Systems Division Sensor and Surveillance Systems, Norwegian Defense Research Establishment (FFI), Horten, Norway; 5) Department of Biology, La Rochelle University, La Rochelle, France; 6) Zoophysiology, Department of Biology, Aarhus University, Aarhus, Denmark; 7) United States Navy, Naval Facilities Engineering Command Atlantic, Norfolk, VA; 8) Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, Canada.

Correspondence: Saana Isojunno

Email: si66@st-andrews.ac.uk

Abstract

Auditory masking by anthropogenic noise may impact marine mammals relying on sound for important life functions, including echolocation. Animals have evolved antimasking strategies, but they may not be completely effective or cost-free. We formulated seven a priori hypotheses on how odontocete echolocation behavior could indicate masking. We addressed six of them using data from 15 tagged sperm whales subject to experimental exposures of pulsed and continuous active sonar (PAS and CAS). Sea state, received single-pulse sound exposure level (SEL_{sp}), whale depth and orientation towards surface and sonar were considered as candidate covariates representing different masking conditions. Echolocation behavior, including buzz duration and search range, varied strongly with depth. After controlling for depth and angle to the surface, the likelihood of buzzing following a click train decreased with sea state ($t=-7.3$, $p<0.001$). There was little evidence for changes in 10 tested variables with increasing sonar SEL_{sp}, except reduced buzzing consistent with previously reported feeding cessation ($t=-2.26$, $p=0.02$). A potential Lombard effect was detected in echolocation with sea state and SEL_{sp}, despite off-axis measurement and right-hand censoring due to acoustic clipping. The results are not conclusive on masking effects on sperm whale echolocation, highlighting challenges and opportunities for future anthropogenic masking studies.

KEYWORDS

anthropogenic noise, continuous active sonar, DTAG, auditory masking

1 | INTRODUCTION

Anthropogenic noise sources elicit changes in marine mammal behavior across a diversity of species and contexts. Behavioral responses are typically considered to be more consequential to individual fitness ("severe") if they are associated with cessation of functional behaviors, such as foraging and nursing, and are longer in duration without waning (Gomez et al., 2016; Southall et al., 2007). However, the lack of severe behavioral response to and increasing tolerance for anthropogenic activities do not necessarily equate to a lack of impact (Beale, 2007; Bejder et al., 2009). A particular concern is that vulnerable individuals tolerate anthropogenic noise because they may be less able to respond or have a higher motivation to continue activities crucial for their survival, such as foraging. Noise effects such as physiological stress, habitat degradation and auditory masking are not necessarily associated with severe behavioral responses and could continue to impact tolerant individuals. While animals have evolved to cope with fluctuations in environmental noise (Gomes et al., 2021), such mechanisms may not be cost-free and may not be completely effective in response to anthropogenic noise, providing a potential for a relatively inconspicuous yet biologically significant impact. Low-severity behavioral responses to anthropogenic noise are widely reported across marine mammal species, even at the highest received levels (Gomez et al., 2016). It is therefore important to consider the potential impacts of apparent noise tolerance when animals continue activities such as foraging in the presence of anthropogenic noise.

Auditory masking is a complex perceptual phenomenon that is "hard to detect and therefore hard to regulate" (Gisiner, 2016). Quantitative predictions can be made of when and where masking should occur (potential for masking) and how animals might compensate for it (antimasking strategies) (von Benda-Beckmann et al., 2021). The extent of auditory masking –

66 and conversely, masking release – varies with the characteristics of the signal and masker (e.g.,
67 amplitude, time-frequency overlap, and spatial overlap), sender/receiver attributes (e.g.,
68 emission/receiver directionality and auditory integration times) the acoustic environment (e.g.,
69 sound propagation) (Erbe et al., 2016; Hotchkin & Parks, 2013) and in the case of echolocation,
70 target properties (e.g., size, Nachtigall, 1980). The factors that can be controlled by the
71 sender/receiver can form part of an active antimasking strategy. For example, a calling animal
72 may be able to compensate for increased noise by increasing the call amplitude (Lombard
73 effect)(Lane & Tranel, 1971), or shift call frequency or temporal pattern relative to the masker
74 (Hotchkin & Parks, 2013). Lombard effects have been reported in social calls of several species
75 of marine mammals, ranging from partial <<1 dB to full 1 dB increase in source level for each 1
76 dB increase in ambient and/or anthropogenic noise, including pinnipeds (Fournet et al., 2021) and
77 cetaceans (killer (Holt et al., 2009), humpback (Dunlop et al., 2014; Guazzo et al., 2020), right
78 (Parks et al., 2011) and bowhead whales (Thode et al., 2020)). Echolocating dolphins and
79 porpoises are known to adjust their click source level adaptively depending on the prey distance
80 and environment (Au & Benoit-Bird, 2003; Jensen et al., 2009; Li et al., 2006). Besides acoustic
81 and auditory factors, the extent of masking also depends upon sender/receiver location and
82 orientation (Erbe et al., 2016). For example, spatial release from auditory masking can occur
83 when the listener spatially separates the signal and masker by altering its orientation (Bain &
84 Dahlheim, 1994; Holt & Schusterman, 2007; Popov et al., 2020) or increasing its distance (Erbe
85 et al., 2016) relative to the masking source. Importantly, many antimasking tactics require
86 changes in behavior (e.g., call characteristics, animal orientation, dive depth) and are therefore
87 possible to measure e.g., using passive acoustic monitoring or animal-attached sound and
88 movement recording tags. Thus, there is an opportunity to detect the occurrence of masking by

89 quantifying antimasking behaviors and their association with environmental and anthropogenic
90 noise expected to increase masking potential.

91 The introduction of increasing numbers of continuous anthropogenic noise sources in the
92 ocean raises concern over their masking potential. Intermittent noise sources such as conventional
93 pulsed active sonar (PAS) produce fluctuating noise levels that provide animals with more
94 listening opportunities during low-level periods (e.g., "within-valley listening" in bottlenose
95 dolphins (Branstetter et al., 2013), "multiple looks" in beluga whales (Erbe, 2008)) compared to
96 more continuous noise sources such as continuous active sonar (CAS) for which such temporal
97 release from masking may not be as feasible. With advancing hardware and signal processing
98 technology, continuous active sonar systems allow human users more detection opportunities,
99 and may also be used at lower source levels (Bates et al., 2018; van Vossen et al., 2011).
100 Continuous systems are subsequently becoming an operational reality for navy sonar and seismic
101 surveys ("vibroseis") around the globe (Duncan et al., 2017). This leads to a pressing need to
102 assess the impact of these new noise-emitting technologies and understand the different types of
103 impacts due to intermittent and continuous noise sources.

104 This study aimed to test indicators of auditory masking in echolocating sperm whales. We
105 formulated specific hypotheses and corresponding data indicators for 1) less successful foraging
106 and 2) potential antimasking strategies, broadly applicable to any echolocating odontocete (Table
107 1). For each hypothesis, we expected the indicator to increase as a function of masking potential,
108 such as under greater ambient noise conditions due to high sea state. Expected conditions for
109 masking were generated in different ambient noise and received sonar exposure scenarios using
110 theoretical modelling described in a recently published paper (von Benda-Beckmann et al., 2021).
111 These hypotheses are not mutually exclusive; for example, in response to high sea surface noise,

an echolocating animal may only need to increase source levels to increase echo levels above the noise floor (Hypothesis 4: amplitude masking release) when pitching up, i.e., facing the sea surface (Hypothesis 7: spatial masking release) due to hearing directionality. While a reduction in foraging effort and success could be driven by multiple underlying mechanisms (e.g., food-safety trade-off, distraction), specific antimasking strategies (e.g., Lombard effect) would provide more specific evidence that auditory masking may be driving the behavior.

To address a subset of these masking indicators (Table 1) for tagged sperm whales, we measured changes in 10 data variables (Table 2). The 10 variables were modelled as a function of candidate covariates expected to increase masking potential (sea state, received level of PAS and CAS signals) and masking release (depth, surface angle, sonar angle) while controlling for depth (ambient pressure, vertical distribution of prey) in regressions with random effects for tag deployments. We found little support for the hypothesized masking effects.

2 | METHODS

2.1 | Experimental protocols

Field data collection and experimental procedures are provided in detail in Kvadsheim et al (2021) and Isojunno et al (2020) and briefly summarized here. Sperm whales were tagged with suction-cup attached audio- and movement-recording tags (DTAG v3; (Johnson & Tyack, 2003)) north and west of Andenes, Norway in 2016-2017. At least four hours of baseline data were collected before the experimental phase, which consisted of a sequence of approaches by the source vessel (40-minute 'exposure sessions'). The sonar source was towed but not transmitting during no-sonar control approaches, which were conducted before any sonar exposure sessions. Subsequent sonar exposure sessions of one of three possible sonar types were presented in a

rotating order (Kvadsheim et al., 2021): 1) HPAS, 1 s hyperbolic upsweep from 1-2 kHz with a "high" maximum source level of 214 dB re 1 µPa m; 2) MPAS, 1 s hyperbolic upsweep from 1-2 kHz with a "medium" maximum source level of 201 dB re 1 µPa m; or 3) CAS, 19 s hyperbolic upsweep from 1-2 kHz, with a maximum source level of 201 dB (re 1 µPa m, same as MPAS) and an energy source level of 214 dB (re 1 μPa^2 m^2 s, same as HPAS). Each signal was transmitted every 20 s, resulting in 5% and 95% duty cycles for PAS and CAS, respectively.

2.2 | Data processing

The main objective of the data processing was to extract 10 data indicators to test our hypotheses (Table 1, Table 2). The indicators were calculated for each prey capture attempt ("buzz") and search ("regular") click train. Buzz and regular click trains were extracted manually by auditing stereo acoustic data (sampled at 96 kHz) both aurally and visually using spectrograms. Buzzes were defined to include a terminal echolocation periods consisting of rapid click rate (>10 Hz). Auditors judged the start and end of each buzz from both change in amplitude and pitch of the clicking sound, relative to regular click trains. The time interval between the end of a buzz and the start of the following click train defined the pause duration (s).

Movement sensor data from the tag were decimated to 5 Hz, and used to calculate depth, 3D acceleration, and body pitch angle of the whale using established methods (Johnson et al., 2009; Miller et al., 2011, 2004). Vertical velocity was calculated as the rate of change in depth. Pitch was converted to a facing angle to the surface directly above (in radians: pi/2 – pitch), hereafter termed surface angle. Following Isojunno and Miller (2018), overall dynamic body acceleration (ODBA) was calculated as the 2-norm of the whale-frame tri-axis acceleration. The acceleration values were high-pass filtered at half of the fundamental stroking rate of the animal, determined by selecting the lowest frequency peak in periodograms of acceleration (median: 0.12, range:

157 0.09-1.16 strokes per second). The 2-norm ODBA was then standardized for each tag deployment
158 by dividing by the median ODBA when the whale was at > 5 m depth. Finally, ODBA values
159 were summarized for each click train as the root mean square (rms) ODBA over the duration of
160 each train.

161 Custom detectors were used to automatically detect individual regular clicks within audited
162 regular click trains; buzz detection was also investigated but the performance of the buzz detector
163 was found to be unreliable during sonar exposures. A two-stage process was used to identify the
164 timing of regular clicks. An automated detector, using a 4 pole (Butterworth) band pass filter with
165 cut frequencies (3 – 20 kHz) and a level threshold based on the envelope click calculation was
166 used to detect clicks. The automated detections were then reviewed manually. Both processes
167 used functions from the "tagtools" Matlab toolbox developed by Mark Johnson
168 (http://www.animaltags.org/). For each regular click train, the detected clicks were used to
169 calculate the first interclick intervals (ICIs), the click rate (inverse of the median ICI across the
170 train), and overall click rate (total number of detections divided by click train duration) in
171 seconds. The median click rate was used as it was more robust to any issues in click detection.
172 The overall click rates were only used to assess the total number of detected clicks as an indicator
173 of tag performance: by definition, regular click trains have slow click rates (<5 Hz/ ICI > 0.22s,
174 (Teloni et al., 2008)) and therefore high total detection rates were used to indicate false positives
175 from sources such as nontagged whale echolocation or mechanical noise.

176 Apparent output (Madsen et al., 2005) level was measured for the first click in each regular
177 click train as the zero-to-peak sound pressure (Madsen, 2005) levels, AOL_{zp}, in a 7 ms time
178 window encompassing the click detection. A lower-gain (-12 dB) channel was used to measure
179 the AOL_{zp}, nevertheless, clipping was apparent in the recording. Therefore, unfiltered, broadband

levels were included in subsequent statistical analyses which considered the AOL_{zp} as a right-censored data variable. In other words, when AOL_{zp} was measured at the clipping level, the true AOL_{zp} was considered to be equal to or greater than the clipping level.

Beaufort sea state was determined visually in the field by experienced marine mammal observers. The modern standard Beaufort scale was used, with the following notes available to the observers at data entry: 0 – glassy, mirror-like, 0.5 – glassy & ripple patches. 1 – scale ripples, 2 – small wavelets, 2.5 – wavelets, rare white caps, 3 - whitecaps, 4 – frequent whitecaps, and 5 – many white caps/spray. Sea state was recorded at hourly intervals and every time the experimental or weather conditions changed. Sea state data were appended to each click train based on their time stamps and assigned NA values when the tagged whale was farther than 6 km distance (the longest visual range obtained during the trial) from the observation vessel.

Noise level measurements were attempted at the sperm whale echolocation band (centroid frequency 13.4-15 kHz, peak frequency 12 kHz; (Hastings & Au, 2008; Jensen et al., 2018; Møhl et al., 2003)), but they were dominated by system noise. Removal of sections with sperm whale clicks made the availability of noise measurements dependent on whale behavior – leading to a biased sample for further analyses. An alternative noise metric was therefore extracted in short windows prior to each click train. However these measurements were only weakly related to ambient noise conditions such as sea state (details provided in the Supplementary material). These results are consistent with previous work showing that animal-borne noise measurements are strongly influenced by tagged whale behavior, tag attachment, and the tag itself (von Benda-Beckmann et al., 2021, 2016). Therefore, sea state and received level of sonar were used as the primary metrics for masking potential.

Received sound exposure level integrated over the duration of the single pulse (SEL_{sp}, dB re 1 μPa^2 s) was measured in the sonar frequency band (0.98-2.24 kHz; based on the four third octave bands overlapping with 1-2 kHz sonar signal) for each sonar transmission in the DTAG acoustic recording (Miller et al., 2011). Received SEL_{sp} was used in favor of sound pressure level (SPL), as it is a better predictor of sperm whale cessation of foraging (Isojunno et al., 2020). Subsequent analyses considered the effect of SEL_{sp} during CAS and PAS separately to avoid assuming an equivalent effect.

The "sonar angle" was the angular difference between the whale's direction of travel (aka pointing angle, vector A) and direct bearing to the vessel position (vector B), calculated as $\cos^{-1}(A \cdot B / |A||B|)$. Vector B was calculated from the vessel's x-y-0 position and the whale's x-y-z position. Vector A was calculated from pitch and horizontal movement direction at the start of each click train, defined as the first 1 s of data (averaging over 5 data points at 5 Hz; circular average in the case of pitch). The function 'bearing' in R package 'geosphere' (version 1.5-10) was used to calculate the horizontal components of the whale movement direction and whale-to-vessel bearing. Two methods were used to estimate the whales' horizontal (x-y) position at 1 s resolution, due to tag magnetometer sensor failure in the 2016 sea trial. When magnetometer data was available, the track was estimated based on tag-derived high-resolution movement variables as well as visual and GPS position fixes of the whale using a Bayesian state-space model (details in Wensveen et al. (2015)). Without magnetometer data available, simple linear interpolations between positions (calculated from visual and GPS positions) were used instead.

2.3 | Statistical analysis

The objective of the statistical analysis was to estimate changes in foraging behavior and therefore any click trains occurring outside foraging states (descent, layer-restricted search, and

ascent states estimated using multivariate hidden Markov model (Isojunno et al., 2020)) were excluded from the analyzed data set. Furthermore, to ensure sufficient click detector performance, any regular click train with fewer than 5 detected clicks or total click detection rates higher than 5 Hz (ICI = 0.2s, Teloni et al., 2008) were removed.

Generalized additive mixed models (GAMMs) were applied to seven response variables (Pause, Duration, ODBA, Stroke rate, Buzz, ICI, CR; Table 2), with individual whale (tag deployment) specified as the random effect in package mgcv (version: 1.8-33) (Wood, 2008) in R (version: 4.0.3). Each model included depth as a smooth covariate, sea state as a main effect, an interaction between sea state and depth, and an interaction between sea state and surface angle (with linear assumption at the log-scale). The log-linear assumption was specified based on the expected log-relationship between ambient noise and wind speed, which scales with sea state (Ainslie, 2010). The sea state interaction terms were specified to represent spatial masking release from the sea surface, and therefore a main effect of surface angle was not included, as this would have implied an effect of posture independent of sea state. Each model also included the no-sonar control session as a presence/absence variable.

Sonar exposure was measured in terms of the maximum SEL_{sp} measured in the 20 s immediately prior to the click train (sp=single pulse), while accounting for the 3D angle to the sonar source (0 directly ahead, 180 directly behind). We considered both a combined effect (SEL_{sp}) and separate effects during CAS vs PAS exposures (SEL_{sp}_CAS, SEL_{sp}_PAS). The interaction between SEL_{sp} and sonar_angle was specified in three alternative ways: as a linear interaction (SEL_{sp}:sonar_angle), assuming no effect of SEL_{sp} at angles >90 deg (full masking release)($SEL_{sp,\ facing}$), or as a non-linear smooth interaction. These covariate combinations resulted in six alternative model structures, listed in Supplementary Table S1. The six full models were

then fitted for each seven response variables (log(Duration), log(ODBA), Pause, Stroke rate, Buzz, ICI and log(CR)).

To test whether sperm whales actively changed their orientation relative to the sea surface noise and sonar source, two more gaussian GAMMs were fitted to logit-transformed sonar and surface angles (normalized to 0-1 dividing by 180 degrees). The model for sonar angle was fitted to data collected during exposure sessions (NS, CAS, and PAS), with time since start of the session and SEL as two smooth covariates (maximum number of knots = 5). The model for source angle was fitted to baseline and postexposure data with the covariate structure state + s(depth) + SeaState + SeaState:depth, where state is a factor covariate level with the levels descent, layer-restricted search (LRS), and ascent (Table S1).

All statistical test results (see Supplementary material for details) were carefully assessed, but considering the number of tested models, the standard 5% significance level was divided by 6 to obtain a more conservative acceptance level of 0.83% for “good support” (Bland & Altman, 1995) .

A different modelling approach that allowed for right-censored data was required to model the click AOL_{zp} that included clipped values. For this, a right-censored normal regression model was built in the R package ‘nimble’ (version: 0.10.1) and fit within a Bayesian Gibbs sampling framework. All estimable parameters were assigned uninformative priors (uniform or gamma distribution). “Full” models were specified to include the hypothesized Lombard effects a-priori for sea state and sonar, while their respective null models were included opposite-direction effects. Please see the supplementary material and associated R/nimble code for further details about the model specification.

3 | RESULTS

3.1 | Analyzed data sets

A total of 131 hr of audited buzz and regular click trains were analyzed from 15 whales. Regular click trains with fewer than 5 detected clicks (n=1,118) or overall click detection rates higher than 5 Hz (n=769) were excluded from further analysis due to decreased detector performance under those conditions, with the former likely including false negatives and the latter false positives. This resulted in 1,910 buzzes and 5,963 regular click trains to be included in the analysis (Table 3).

All tagged whales produced regular clicks during sonar exposures, and the majority of sonar exposures also contained buzzes (Tables 3-4, Fig. 1). The whales clicked over a wide range of surface and sonar angles, without consistent increase in the facing angles at higher sea states and during the sonar exposures (Hypothesis 7, Table 1) (Fig. 2). During sonar exposures echolocation click trains were produced at depths of 150-1,900 m (Fig. 1). Pauses were specified as gaps in clicking > 0.76 s in duration (the median ICI during regular clicking) following an inspection of the gap duration data, which showed a clear bimodal distribution (Fig. S1).

3.2 | GAMM hypothesis tests

Random effects models (GAMMs) were used to statistically test the hypotheses (Table 1) by modelling the candidate covariates for masking potential (sea state, received level of sonar) and behavior-dependent masking release (dive depth, surface angle, sonar angle) while controlling for the effects of depth on the response variables (Table 2, Table S1).

Dive depth had a strong influence on echolocation behavior, but was not affected by exposure conditions. The smooth covariate for depth was supported at 0.83% significance level in models

292 for log(Duration), log(ODBA), Buzz, ICI and log(CR) (see Table 2 for definitions), while no
293 support was found for depth-dependency in Stroke rate or Pause occurrence (Table S3, Figs S2-
294 S6). Buzzes were longer and had a greater rms-ODBA at shallower depths (Fig. 3, Fig. S2).
295 Regular click trains were more likely to be followed by a buzz at deeper depths, and at the
296 deepest depths (>1,400 m), regular click trains also had a shorter initial ICI and faster click rate
297 (Fig. 3, Fig. 4, Fig. 5).

298 Sea state was supported as a main effect for Buzz and log(CR) at 0.83% significance level, but
299 the log(CR) model did not appear robust (Fig. 3). After controlling for the effect of depth, buzz
300 occurrence was reduced at high sea states (Table 4), especially at shallow surface angles and in
301 depths <500 m (Fig. 3, Table S2). While the sea state effect for Buzz was mediated by both
302 surface angle and depth, only the interaction term with surface angle was supported for log(CR).
303 Click rates decreased slightly with increasing sea state at shallow surface angles (Fig. S9, Table
304 S3). However, the model underestimated high click rates (see model diagnostics indicating poor
305 performance in Fig. S9) and in the raw data, there was no clear pattern with sea state after
306 controlling for depth (Table 4, Fig. 5, Table S2). Indeed, both interclick interval and click rate
307 showed more substantial variation with depth than sea state and sonar exposures (Fig. 4, Fig. 5).
308 Sea state also gained support in models for fluke stroke rate (p=0.014 in exposure models and
309 p=0.003 in nonexposure models, Table S3), with stroke rates decreasing with increasing sea state
310 (Table S2, Fig. S7).

311 Experimental exposure covariates were not supported at the 0.83% level, with the sole
312 exception of $SEL_{sp}\_CAS_{facing}$ as an explanatory covariate for ODBA. This supports a single cut-
313 off for orientation-dependent masking release from received SEL from CAS exposures when the
314 sonar angle exceeded 90˚ sonar angle. A main effect from sonar SEL was supported at 5% level

for log(ODBA) and Buzz, and in both cases, the support for the effect increased slightly with the single cut-off for masking release at ≥ 90º sonar angle ($SEL_{sp,facing}$) (Table S3). The occurrence of buzzes (following regular click trains) and ODBA were estimated to decrease with increasing sonar SEL (Table S3); a reduction in minimum specific acceleration (MSA) was also supported at 5% level in equivalent models for log(MSA). However, little reduction in ODBA was apparent in the raw data plots (Figs S10-11). SEL_{sp}_PAS and SEL_{sp}_CAS_{facing} were supported at 5%, but not 0.83%, level in the model for buzz duration (p=0.011 and p=0.026, respectively). The coefficient estimates indicated a positive effect on buzz duration during CAS and negative effect during PAS (Table S3, see also Table 4). SEL_{sp}_PAS and ti(SEL_{sp}) gained support in the full model for ICI (p=0.029 and p=0.011, respectively), but the support at 5% was lost when removing unsupported covariates from the model (i.e., when only the smooth covariate for depth was included).

No statistical support was found for changes in orientation in relation to sea state and sonar. Time since start of exposure (F=4.4, p<0.001), representing the effect of the experimental approach design and ship effect alone, rather than sonar SEL_{sp} (F=0.08, p=0.22), representing a potential masking effect, was supported as a smooth covariate for sonar angle in the model for spatial release from masking by sonar. Exposure time was also supported over SEL_{sp}_CAS and SEL_{sp}_PAS, and when excluding the 2016 data (when magnetometer data were unavailable to the horizontal track estimation). The angle to the sonar source increased as a function of time since the start of exposure, which could be partly explained by the experimental design: at the start of each exposure (including no-sonar) session, the vessel approach was set perpendicular to the focal whale expected travel direction. The relationship was perhaps more linear than expected, given the source ship was expected to reach the whale location towards the end of the session

(Fig. S12). In the model for surface angle, only the smooth covariate for depth gained support. The lack of support for sea state (t=1.2, p=0.196) was maintained including or excluding the behavior state. The diagnostics for the model are provided in Fig. S13.

3.3 | Testing Lombard effect in apparent click level (AOL_{zp})

The full model for AOL_{zp} supported both sea state (mp_1) and sonar (sp_1) effects, but it had a small coefficient of determination (R^2 = 8.6%), indicating substantial unexplained variability in the AOL_{zp} data set (n=4,045 regular click trains), of which 26% were right-censored (i.e., clipped). Of the 14 deployments contributing to the data set, 5 included no clipped data, 6 had <50% of data clipped, and 3 had more than 50% clipped data. The full model provided a slightly better fit to the data than the null model for sea state (6.0%), while there was a negligible difference to the null model for sonar (8.8%). Both the full and null model posteriors for the sea state masking potential (mp_1) parameter were centered away from zero; however, the 95% interval was narrower in the full model (4.5) than in the null model (6.4). While the lower bound of the sonar parameter (sp_1) in the null model was close to zero (posterior median 0.06, 95% 0.00 – 0.17), in the full model it was estimated to be further away from zero (median 0.42, 95% 0.05-0.79).

Apparent output from buzz clicks decreases with shorter interclick interval in sperm whales (Isojunno and Miller 2018). To check that the model result wasn't driven by behavior response that involved changes in interclick interval, the model was fitted with an additional linear covariate representing inspection range. The inspection range was calculated as $(ICI-12)/1000/2*1490ms^{-1}$ (Isojunno and Miller, 2018), providing interpretable units for the model estimates. The coefficient of determination of this model was substantially higher (24.7%). The posterior mean estimate for mp_1 was only slightly lower (median 7.58, 95% range 5.7-9.6) while

the posterior distribution for sp_1 was virtually unchanged (median 0.42, 95% range 0.08-0.78) compared to the posterior estimates in model excluding Range (Table S4) – suggesting that the effects were not driven by inspection range alone.

An exploratory analysis of AOL_{zp} showed no clear changes with sea state, sonar exposure, depth, surface angle or sonar angle (Fig. 6). With a cautionary note on the high unexplained variability, posterior model predictions for AOL_{zp} are shown in Fig. 7. The model estimates supported 2-7 dB increase in AOL_{zp} with sea state (Beaufort 1-5) when the whales faced the sea surface at relatively shallow depths (<400 m). The estimated increase in AOL_{zp} with received SEL_{sp} (from CAS and PAS combined) was relatively small and uncertain: at the highest received levels (170 dB re 1 μPa^2 s), the AOL_{zp} was estimated to increase by 4.4 dB (95% CRI: 0.6, 8.5), equivalent to the effect of sea state 3.5 at that depth (and surface angle < 90 deg).

4 | DISCUSSION

We set out to test indicators of masking in the echolocation behavior of sperm whales during different ambient masking conditions and as a function of received sonar exposures (Table 1). Most of the statistical test results showed no effect with the candidate covariates expected to mediate masking potential (sea state, received level of sonar) and masking release (depth, surface angle, sonar angle), indicating relatively stable foraging and echolocation behavior throughout different sea states and sonar exposures. Nevertheless, small reductions in buzz occurrence and locomotor activity were observed that were consistent with previously documented cessation of foraging during these controlled exposure experiments (CEEs) (Isojunno et al., 2020). While orientation-mediated effects of sea state gained some support, there was no evidence for active change in orientation in response to sonar exposures. Models incorporating orientation-dependent

masking release found a Lombard effect in echolocation due to sea surface noise, despite data limitations. The estimated effect was small compared to variation associated with changes in depth and inter-click interval. By contrast, we found strong depth-dependency in almost all the tested response variables, which could be driven by factors other than ambient noise, such as prey behavior or ambient pressure.

Theoretical modelling of masking potential predicted that regular echolocation clicks used during the search phase would be continuously masked at levels of $L_{p,1\text{-}2kHz}$ of 160 dB re 1 µPa, which corresponds roughly to the SEL = 160 dB re 1 µPa^{2}s for PAS, and the SEL = 173 dB re 1 µPa^{2}s for CAS during these experiments (von Benda-Beckmann et al., 2021). Buzz clicks were predicted less likely to be masked (in terms of detection range) due to the proximity to the prey when buzzes are produced (von Benda-Beckmann et al., 2021). These exposure levels correspond to the highest levels achieved during our experiments (n=75 and 8 regular click trains during CAS, and n=8 and 2 trains during PAS were measured ≥160 and ≥170 dB re 1 µPa^{2}s, respectively; see also Table 3). During the PAS sonar the masking for all these cases would be expected only to occur for a limited period due to the low (5%) duty cycle. Our empirical results combined with the theoretical predictions by von Benda-Beckmann et al. (2021) support that masking potential of sperm whale echolocation is limited for the exposure conditions tested in this study. That said, it is possible that we didn't have sufficient sample size for the expected masking-inducing conditions (whale facing towards the source at high SPL and ambient noise conditions) to detect the hypothesized effects. Furthermore, cessation of feeding reduced sample size for the analysis of echolocation clicks.

We couldn't find strong support for the hypothesized patterns in the data indicators, predictably perhaps given the myriad of factors that can influence echolocation masking (Luo et

al., 2015), expected discontinuous masking conditions during these experiments (von Benda-Beckmann et al., 2021), and the lack of experimental control for echolocation targets and ambient noise in observational systems such as ours. However, in-situ studies of anthropogenic masking are needed to fill the data gap on the context-dependencies of auditory masking (Hotchkin & Parks, 2013). We remain optimistic that such effects could be quantified for noise sources with sufficient masking potential and with accurate and precise measurements of 1) received levels and spectra of anthropogenic noise (the masker of interest), 2) ambient noise levels (the baseline masker), 3) animal source characteristics (the signals of interest), 4) animal behavior (potential antimasking), coupled with 5) sufficient sample size and/or experimental control across the multiple factors that influence masking. We discuss each point below, both with respect to our study and as recommendations for future research.

Medium or high received levels of anthropogenic noise are relatively straightforward to measure and estimate with the available sound recording archival tags (Johnson & Tyack, 2003). One key question is which sound metric is the most relevant for describing auditory masking potential. Clearly, metrics such as a maximum sound pressure level calculated over a long integration window ignore the finer temporal structure of received signals – and thus do not account for overlap in the time domain. However, sound exposure levels (SEL) of signals with different intermittency might not scale linearly with their masking potential (Branstetter et al., 2013; Cunningham et al., 2014; Erbe, 2008) – which is why we opted to consider both combined and separate single-pulse SEL from the CAS and PAS exposures. However, due to presence of clicks on the DTAG and system noise, it was challenging to get reliable measurements of the actual noise distribution the animal was facing. Another key consideration is the frequency band of the masker. Here we examined the effect of a relatively narrow-band and tonal masker (1-2

kHz upsweeps, though harmonics were present at high source levels; (von Benda-Beckmann et al., 2021)) on broadband echolocation clicks – the opposite of most masking studies that have considered broad-band masking sounds (e.g., white noise) on relatively narrow-band signals (Erbe et al., 2016; Hotchkin & Parks, 2013). For example, Tressler and Smotherman found different effects of broad-band and band-limited noise on bat echolocation (Tressler & Smotherman, 2009). Harmonic levels of the sonar systems can be used to extrapolate sonar levels measured in the main sonar band to higher frequencies but show substantial fluctuation over time (von Benda-Beckmann et al., 2021). Therefore, our results with respect to sonar should not be extrapolated to broadband masker signals, which may have greater masking potential on sperm whale echolocation.

Ambient noise levels in the environment fluctuate both in space and time (Hildebrand, 2009). This variation would ideally be captured by direct measurements of ambient noise. Animal-borne noise measurements (SPL_{rms}) were made in short time windows prior to each regular click train, which increased with sea state (1 dB when increasing sea state from 2.5 to 4) and decreased with dive depth (-2.7 dB between 100 and 1,000 m). Sea state and depth explained around 10% of the SPL_{rms} noise measurements (high-pass filtered >2 kHz), but also vertical speed and sonar exposures influenced the measurements (further details in Supplementary material). Measuring ambient noise on animal-born tags is complicated by many different factors, i.e., due to presence of flow-noise, noise from the nearby research vessel, body shielding, sounds produced by both the tagged whale and conspecifics, splashes near the surface and air bubbles trapped around the tags, and system noise (von Benda-Beckmann et al., 2021, 2016). We therefore opted to use the field-estimated sea state as a log-transformed proxy variable for ambient noise resulting from sea surface noise. However, with the potential for depth-dependent effects, incorporating this proxy

452 required additional interaction covariates, which complicated data visualizations and increased
453 the number of estimable parameters in the statistical models. Alternatively, noise level
454 measurements can be made using passive acoustic monitoring (PAM) devices not attached to the
455 subject animals, such as bottom-mounted acoustic recorders or towed arrays; an approach proven
456 successful in detecting Lombard effects in less directional marine mammal communication calls
457 (e.g., Holt & Schusterman, 2007; Thode et al., 2020). As well as noise measurements, at shorter
458 ranges PAM could also be used to record and track subject animals, which in combination with
459 noise modelling and animal-borne sensing could be a powerful means to obtain concurrent
460 ambient noise and animal behavior data.

461 The characteristics of the animal's signals – amplitude, spectral and temporal patterns – are
462 fundamental to understanding their potential to be masked. A disadvantage of using animal-
463 attached hydrophones to study echolocation is that the recordings are made off the main axis of
464 the source acoustic beam: for cetaceans, tags are often deployed on the back of the animal or in
465 some cases, on the head. As such, changes in signal characteristics measured off-axis could be
466 the result of changes in the source signal, or changes in the acoustic beam pattern such as greater
467 directionality (Zimmer & Tyack, 2005). Furthermore, when attached remotely (without capture),
468 unavoidable variation in tag placement can cause variation in apparent levels between
469 deployments (here analyzed as a random effect). Though similarly affected, social and
470 communication call outputs can be expected to be less directional compared to echolocation.
471 Another challenge working with echolocating species is that the echoes can be difficult to
472 characterize, even detect, further away from the source and when shielded by the animals' body,
473 as is the case for tagged sperm whales (Tønnesen et al., 2020). To address echolocation masking,
474 study systems where tag position is standardized (e.g., in capture-release programs) and where

475 received prey echoes can be detected (e.g., in beaked whales, Madsen et al., 2005) would be more
476 optimal. Such measurements could be used to proxy echo-to-noise ratios which are key to
477 quantify whether masking actually takes place (Au et al., 1988; Au & Penner, 1981; Griffin et al.,
478 1963; Luo et al., 2018).

479 Clipping of tagged-whale clicks was prevalent in our recordings (26% of the measured AOL_{zp}
480 data points). This was expected, given that sperm whale source levels can exceed 220 dB re 1
481 µPa m (Møhl et al., 2000); even lower (<-12 dB) gain hydrophones could be added, or less
482 powerful sources may provide a better model species for future work. Nevertheless, statistical
483 models that explicitly incorporate right-censored data can be useful to simultaneously quantify
484 changes in the unclipped levels, and the probability of clipping. Our right-censored regression
485 model fitted in a Bayesian framework provides a useful tool to make the most of existing clipped
486 data and in recordings where clipping is unavoidable, provided that a sufficient amount of
487 variation is present below the clipping level. The models estimated an increase in AOL_{zp} with sea
488 state and received SEL_{sp} when the whale was facing towards the sea surface (<100º) and the
489 sonar source (<30º), respectively, consistent with a Lombard effect and spatial masking release.
490 Depending on the surface angle, the sea state effect disappeared at depths exceeding 300-800 m
491 (Fig. 7). However, given the variability in the off-axis levels and the clipping issue, and the lack
492 of any visible pattern in the raw data (Fig. 6), we consider this result highly uncertain and urge
493 replication studies to further test it – ideally with more precise source level measurements and
494 greater masking levels in the main echolocation band.

495 High-resolution animal-borne sensors provide an opportunity to measure fine-scale changes in
496 acoustic and movement behavior for disturbance response studies (Johnson and Tyack, 2003). In
497 our analysis we attempted to account for animal orientation, however broad sonar angle

categories (0-45 degrees) had to be used to retain sufficient data points. This may have limited our ability to detect effects, as masking release may already be effective at smaller angles (Au & Moore, 1984). As with any bio-logging data, a key challenge is matching multivariable time series data with complex shifts in animal behavior and formulating testable hypotheses. To this end, we hope that our "masking indicators" hypotheses (Table 1) can inspire further bio-logging data analyses for echolocating odontocetes. Nevertheless, it is important to keep in mind that a change in one indicator alone does not provide conclusive evidence of masking. For example, cessation of foraging could be associated with a fear response, rather than giving up due to masking. It is the suite of effects and their association with specific conditions expected to increase masking (e.g., facing towards a source) that would provide more convincing evidence of actual masking effects.

While we were able to control sound exposure levels as part of our experimental dose-escalation design, multiple environmental factors (e.g., prey, weather) and individual and behavior context variables (e.g., age, experience, dive depth) were not within our control. For modelling of masking, prey type and associated target strength are substantial sources of uncertainty (von Benda-Beckmann et al., 2021). Many cephalopod species are relatively weak targets (Madsen et al., 2007), but could be more detectable in large schools (Tønnesen et al., 2020). Similarly, fish may or may not contain swim bladders, swim individually or in schools. Vertical stratification of prey may have contributed to the depth-dependent echolocation behavior (Isojunno & Miller, 2018). The longer and more active buzzes during shallower dives are consistent with previous analyses of sperm whale echolocation behavior (Isojunno and Miller, 2018; Teloni et al., 2008). However, ambient pressure, light and sea surface noise are also expected to vary as a function of depth and therefore confound the interpretation of one factor as

the driver of echolocation. One approach to tackle the multifactorial problem of masking is by introducing experimental control; the other is to collect more data to ensure coverage across the multivariate factors of interest, exploiting accumulating bio-logging datasets on ever larger numbers of cetacean species. Automated click detection on large numbers of existing DTAG recordings from multiple studies could be used to achieve such a sample size.

In conclusion, our study provides a useful quantitative baseline and an exhaustive evaluation of potential sonar effects on sperm whale foraging and echolocation behavior, highlighting the complexities involved in attempting to detect auditory masking on free-ranging cetaceans. Using theoretical considerations, von Benda-Beckmann et al. (2021) predicted that there is some potential for masking by CAS in sperm whales in certain conditions (when facing towards the source and in increased ambient noise, and for the highest sonar exposure levels tested here), but here we found few behavioral indications of such. Our data show that sperm whale foraging and echolocation behavior are highly depth-dependent, consistent with previous works suggesting vertical stratification of prey and pressure effects on pneumatic sound production as important drivers (Isojunno & Miller, 2018; Madsen et al., 2002). Indications of an orientation-dependent Lombard effect and reduced foraging (fluke stroke rates, buzzing) at relatively shallow depths (<500 m) with greater sea state warrant replication studies on masking in cetacean echolocation. In many cases marine mammal social sounds can be expected to be more susceptible to anthropogenic masking than echolocation, due to greater frequency band and temporal overlap for example, also inviting further study. We urge such future efforts to take advantage of marine bio-logging technology to further our understanding of anthropogenic masking of marine mammal sounds.

ACKNOWLEDGEMENTS

We thank all 3S (Sea mammals, Sonar, Safety) team members for efforts on the field data collection and access: Marije Siemensma, Lise Sivle, Rune Hansen, Lars Kleivane, Lucía Martina Martín López, Eilidh Siegal, René Dekeling, Odile Gerard, Yvonne Mather, Sander Van IJsselmuide, Mark Van Spellen, Martijn van Riet. We would like to thank 3S partners and funders for enabling this research. Many thanks to three anonymous reviewers for their time and helpful feedback on the manuscript.

FUNDING

This work was supported by the UK Defense and Science Technology Laboratory (DSTLX-1000137649), NL Ministry of Defence (Cerema-DGA #1883003901), FR Ministry of Defence, and US Navy Living Marine Resources program (N39430-17-C-1935). PLT was supported by US Office of Naval Research (ONR) grant numbers N00014-18-1-2062 and N00014-20-1-2709, as well as by the Strategic Environmental Research and Development Program (SERDP) contracts RC20-1097, RC21-3091, and RC20-7188.

DATA ACCESSIBILITY STATEMENT

All data needed to reproduce the results in this manuscript are available from https://github.com/Saana-I/MaskingIndicators, along with a readme document.

USE OF ANIMALS IN RESEARCH

Animal experiments were carried out under permits issued by the Norwegian Animal Research Authority permit no 15/223222, in compliance with ethical use of animals in experimentation.

The research protocol was approved by the University of St Andrews Animal Welfare and Ethics Committee.

REFERENCES

Ainslie, M. A. (2010). *Principles of Sonar Performance Modeling*. Berlin: Springer Verlag.

Au, W. W. L., & Benoit-Bird, K. J. (2003). Automatic gain control in the echolocation system of dolphins. *Nature*, *4744*(1997), 1997–1999. https://doi.org/10.1038/nature01727.1.

Au, W. W. L., & Moore, P. W. B. (1984). Receiving beam patterns and directivity indices of the Atlanti bottlenose dolphin Tursiops truncatus. *Journal of the Acoustical Society of America*, *75*(1), 255–262. https://doi.org/10.1121/1.390403

Au, W. W. L., Moore, P. W. B., & Pawloski, D. A. (1988). Detection of complex echoes in noise by an echolocating dolphin. *Journal of the Acoustical Society of America*, *83*(2), 662–668. https://doi.org/10.1121/1.396161

Au, W. W., & Penner, R. H. (1981). Target detection in noise by echolocating Atlantic bottlenose dolphins. *Journal of the Acoustical Society of America*, *70*(3), 687–693. https://doi.org/10.1121/1.386931

Bain, D. E., & Dahlheim, M. E. (1994). Effects of masking noise on detection thresholds of killer whales. In *Marine Mammals and Exxon Valdez* (pp. 243–255).

Bates, J. R., Grimmett, D., Canepa, G., & Tesei, A. (2018). Incoherent sub-band averaging for improved target detection and Doppler estimation in linearly frequency modulated continuous active sonar. *Proceedings of Meetings on Acoustics, Acoustical Society of America*, *055001*(055001), 55001. https://doi.org/10.1121/2.0000817

584 Beale, C. M. (2007). The Behavioral Ecology of Disturbance Responses. *International*
585 *Journal of Comparative Psychology*, *20*(20), 111–120. Retrieved from
586 http://www.comparativepsychology.org/ijcp-vol20-2-3-2007/03.Beale_PDF.pdf

587 Bejder, L., Samuels, A., Whitehead, H., Finn, H., & Allen, S. (2009). Impact assessment
588 research: use and misuse of habituation, sensitisation and tolerance in describing wildlife
589 responses to anthropogenic stimuli. *Marine Ecology Progress Series*, *395*, 177–185.
590 https://doi.org/10.3354/meps07979

591 Bland, J. M., & Altman, D. G. (1995). Statistics Notes Multiple significance tests : the
592 Bonferroni method. 310(January), 1995.

593 Branstetter, B. K., Trickey, J. S., Bakhtiari, K., Black, A., Aihara, H., & Finneran, J. J. (2013).
594 Auditory masking patterns in bottlenose dolphins (*Tursiops truncatus*) with natural,
595 anthropogenic, and synthesized noise. *The Journal of the Acoustical Society of America*, *133*(3),
596 1811–1818. https://doi.org/10.1121/1.4789939

597 Cunningham, K. A., Southall, B. L., & Reichmuth, C. (2014). Auditory sensitivity of seals and
598 sea lions in complex listening scenarios. *The Journal of the Acoustical Society of America*,
599 *136*(6), 3410–3421. https://doi.org/10.1121/1.4900568

600 Duncan, A. J., Weilgart, L. S., Leaper, R., Jasny, M., & Livermore, S. (2017). A modelling
601 comparison between received sound levels produced by a marine Vibroseis array and those from
602 an airgun array for some typical seismic survey scenarios. *Marine Pollution Bulletin*, *119*(1),
603 277–288. https://doi.org/10.1016/j.marpolbul.2017.04.001

Dunlop, R. A., Cato, D. H., & Noad, M. J. (2014). Evidence of a Lombard response in migrating humpback whales (Megaptera novaeangliae) . *The Journal of the Acoustical Society of America*, *136*(1), 430–437. https://doi.org/10.1121/1.4883598

Erbe, C. (2008). Critical ratios of beluga whales (*Delphinapterus leucas*) and masked signal duration. *The Journal of the Acoustical Society of America*, *124*(4), 2216–2223. https://doi.org/10.1121/1.2970094

Erbe, C., Reichmuth, C., Cunningham, K., Lucke, K., & Dooling, R. (2016). Communication masking in marine mammals: A review and research strategy. *Marine Pollution Bulletin*, *103*(1–2), 15–38. https://doi.org/10.1016/j.marpolbul.2015.12.007

Fournet, M. E. H., Silvestri, M., Clark, C. W., Klinck, H., & Rice, A. N. (2021). Limited vocal compensation for elevated ambient noise in bearded seals: implications for an industrializing Arctic Ocean. *Proceedings. Biological Sciences*, *288*(1945), 20202712. https://doi.org/10.1098/rspb.2020.2712

Gisiner, R. C. (2016). *The Effects of Noise on Aquatic Life II*. *875*, 355–362. https://doi.org/10.1007/978-1-4939-2981-8

Gomes, D. G. E., Francis, C. D., & Barber, J. R. (2021). Using the Past to Understand the Present: Coping with Natural and Anthropogenic Noise. *BioScience*, *XX*(X), 1–12. https://doi.org/10.1093/biosci/biaa161

Gomez, C., Lawson, J. W., Wright, A. J., Buren, A. D., Tollit, D., & Lesage, V. (2016). A systematic review on the behavioural responses of wild marine mammals to noise: the disparity between science and policy. *Canadian Journal of Zoology*, *94*(12), 801–819. https://doi.org/10.1139/cjz-2016-0098

Griffin, D. R., McCue, J. J. G., & Grinnell, A. D. (1963). The resistance of bats to jamming. *Journal of Experimental Zoology*, *152*(3), 229–250. https://doi.org/10.1002/jez.1401520303

Guazzo, R. A., Helble, T. A., Alongi, G. C., Durbach, I. N., Martin, C. R., Martin, S. W., & Henderson, E. E. (2020). The Lombard effect in singing humpback whales: Source levels increase as ambient ocean noise levels increase. *The Journal of the Acoustical Society of America*, *148*(2), 542–555. https://doi.org/10.1121/10.0001669

Hastings, M. C., & Au, W. W. L. (2008). *Principles of Marine Bioacoustics*. New Yok: Springer Verlag.

Hildebrand, J. A. (2009). Anthropogenic and natural sources of ambient noise in the ocean. *Marine Ecology Progress Series*, *395*(5), 5–20. https://doi.org/10.3354/meps08353

Holt, M. M., Noren, D. P., Veirs, V., Emmons, C. K., & Veirs, S. (2009). Speaking up: Killer whales (Orcinus orca) increase their call amplitude in response to vessel noise . *The Journal of the Acoustical Society of America*, *125*(1), EL27–EL32. https://doi.org/10.1121/1.3040028

Holt, M. M., & Schusterman, R. J. (2007). Spatial release from masking of aerial tones in pinnipeds. *The Journal of the Acoustical Society of America*, *121*(2), 1219–1225. https://doi.org/10.1121/1.2404929

Hotchkin, C., & Parks, S. (2013). The Lombard effect and other noise-induced vocal modifications: Insight from mammalian communication systems. *Biological Reviews*, *88*(4), 809–824. https://doi.org/10.1111/brv.12026

Isojunno, S., & Miller, P. J. O. (2018). Movement and biosonar behavior during prey encounters indicate that male sperm whales switch foraging strategy with depth. *Frontiers in Ecology and Evolution*. https://doi.org/10.3389/fevo.2018.00200

Isojunno, S., Wensveen, P. J., Lam, F.-P. A., Kvadsheim, P. H., von Benda-Beckmann, A. M., Martín López, L. M., Kleivane, L., Siegal, E. M., & Miller, P. J. O. (2020). When the noise goes on: received sound energy predicts sperm whale responses to both intermittent and continuous navy sonar. *The Journal of Experimental Biology*, *223*(7), jeb219741. https://doi.org/10.1242/jeb.219741

Jensen, F. H., Bejder, L., Wahlberg, M., & Madsen, P. T. (2009). Biosonar adjustments to target range of echolocating bottlenose dolphins (Tursiops sp.) in the wild. *Journal of Experimental Biology*, *212*(8), 1078–1086. https://doi.org/10.1242/jeb.025619

Jensen, F. H., Johnson, M., Ladegaard, M., Wisniewska, D. M., & Madsen, P. T. (2018). Narrow Acoustic Field of View Drives Frequency Scaling in Toothed Whale Biosonar. *Current Biology*, *28*(23), 3878-3885.e3. https://doi.org/10.1016/j.cub.2018.10.037

Johnson, M. P., Soto, N. A., & Madsen, P. T. (2009). Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: a review. *Marine Ecology Progress Series*, *395*, 55–73. https://doi.org/10.3354/meps08255

Johnson, M. P., & Tyack, P. L. (2003). A Digital Acoustic Recording Tag for Measuring the Response of Wild Marine Mammals to Sound. *IEEE Journal of Oceanic Engineering*, *28*(1), 3–12. https://doi.org/10.1109/JOE.2002.808212

Kvadsheim, P. H., Isojunno, S., Cure, C., Siemensma, M. L., Wensveen, P. J., Lam, F.-P. P. A., Hansen, R. R., Benti, B., Sivle Doksæter L., Burslem, A., Kleivane, L., & Miller, P. J. O. (2021). The 3S3 experiment data report – using operational naval sonars to study the effects of continuous active sonar, and source proximity, on sperm whales.

669 Lane, H., & Tranel, B. (1971). The Lombard Sign and the Role of Hearing in Speech. *Journal*
670 *of Speech and Hearing Research*, *14*(4), 677–709. https://doi.org/10.1044/jshr.1404.677

671 Li, S., Wang, D., Wang, K., & Akamatsu, T. (2006). Sonar gain control in echolocating
672 finless porpoises (Neophocaena phocaenoides) in an open water . *The Journal of the Acoustical*
673 *Society of America*, *120*(4), 1803–1806. https://doi.org/10.1121/1.2335674

674 Luo, J., Hage, S. R., & Moss, C. F. (2018). The Lombard Effect: From Acoustics to Neural
675 Mechanisms. *Trends in Neurosciences*, *41*(12), 938–949.
676 https://doi.org/10.1016/j.tins.2018.07.011

677 Luo, J., Siemers, B. M., & Koselj, K. (2015). How anthropogenic noise affects foraging.
678 *Global Change Biology*, *21*(9), 3278–3289. https://doi.org/10.1111/gcb.12997

679 Madsen, P. T. (2005). Marine mammals and noise: Problems with root mean square sound
680 pressure levels for transients. *The Journal of the Acoustical Society of America*, *117*(6), 1–6.
681 https://doi.org/10.1121/1.1921508

682 Madsen, P. T., Johnson, M., de Soto, N. A., Zimmer, W. M. X., & Tyack, P. (2005). Biosonar
683 performance of foraging beaked whales (*Mesoplodon densirostris*). *The Journal of Experimental*
684 *Biology*, *208*(Pt 2), 181–194. https://doi.org/10.1242/jeb.01327

685 Madsen, P. T., Payne, R., Kristiansen, N. U., Wahlberg, M., Kerr, I., & Møhl, B. (2002).
686 Sperm whale sound production studied with ultrasound time/depth-recording tags. *The Journal of*
687 *Experimental Biology*, *205*(Pt 13), 1899–1906. Retrieved from
688 http://jeb.biologists.org/content/205/13/1899.short

689 Madsen, P. T., Wilson, M., Johnson, M., Hanlon, R. T., Bocconcelli, A., Soto, N. A. De, &
690 Tyack, P. L. (2007). *Clicking for calamari : toothed whales can echolocate squid Loligo pealeii*.
691 *1*(December), 141–150. https://doi.org/10.3354/ab00014

692 Miller, P. J. O., Antunes, R., Alves, A. C., Wensveen, P. J., Kvadsheim, P. H., Kleivane, L.,
693 Nordlund, N., Lam, F-P. A., van IJsselmuide, S. P., Visser, F., Tyack, P. L. (2011). The 3S
694 experiments: studying the behavioural effects of naval sonar on killer whales (*Orcinus orca*),
695 sperm whales (*Physeter macrocephalus*), and long-finned pilot whales (*Globicephala melas*) in
696 Norwegian waters Scottish Oceans Inst. In *Scottish Oceans Institute Technical Report SOI-2011-*
697 *001*.

698 Miller, P. J. O., Johnson, M. P., Madsen, P. T., Biassoni, N., Quero, M., & Tyack, P. L.
699 (2009). Using at-sea experiments to study the effects of airguns on the foraging behavior of
700 sperm whales in the Gulf of Mexico. *Deep Sea Research I*, *56*(7), 1168–1181.
701 https://doi.org/10.1016/j.dsr.2009.02.008

702 Miller, P. J. O., Johnson, M. P., & Tyack, P. L. (2004). Sperm whale behaviour indicates the
703 use of echolocation click buzzes ‘creaks’’ in prey capture.’ *Proceedings of the Royal Society B:*
704 *Biological Sciences*, *217*(1554), 2239–2247. https://doi.org/10.1098/rspb.2004.2863

705 Møhl, B., Wahlberg, M., & Madsen, P. T. (2000). Sperm whale clicks: Directionality and
706 source level revisited. *The Journal of the Acoustical Society of America*, *107*(1), 638–648.
707 Retrieved from http://scitation.aip.org/content/asa/journal/jasa/107/1/10.1121/1.428329

708 Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., & Lund, A. (2003). The monopulsed
709 nature of sperm whale clicks. *The Journal of the Acoustical Society of America*, *114*(2), 1143–
710 1154. https://doi.org/10.1121/1.1586258

Nachtigall, P. E. (1980). Odontocete Echolocation Performance on Object Size, Shape and Material. In *Animal Sonar Systems* (pp. 71–95).

Parks, S. E., Johnson, M., Nowacek, D., & Tyack, P. L. (2011). Individual right whales call louder in increased environmental noise. *Biology Letters*, *7*(1), 33–35. https://doi.org/10.1098/rsbl.2010.0451

Popov, V. V., Supin, A. Y., Gvozdeva, A. P., Nechaev, D. I., Tarakanov, M. B., & Sysueva, E. V. (2020). Spatial release from masking in a bottlenose dolphin *Tursiops truncatus* . *The Journal of the Acoustical Society of America*, *147*(3), 1719–1726. https://doi.org/10.1121/10.0000909

Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene Jr, C. R., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., Richardson, W. J., Thomas, J. A., & Tyack, P. L. (2007). Marine mammal noise-exposure criteria: initial scientific recommendations. *Aquatic Mammals*, *33*, 411–521. Retrieved from http://www.tandfonline.com/doi/pdf/10.1080/09524622.2008.9753846

Teloni, V., Mark, J. P., Patrick, M. J. O., & Peter, M. T. (2008). Shallow food for deep divers: Dynamic foraging behavior of male sperm whales in a high latitude habitat. *Journal of Experimental Marine Biology and Ecology*, *354*(1), 119–131. https://doi.org/10.1016/j.jembe.2007.10.010

Thode, A. M., Blackwell, S. B., Conrad, A. S., Kim, K. H., Marques, T., Thomas, L., Oedekoven, C. S., Harris, D., & Bröker, K. (2020). Roaring and repetition: How bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise. *The Journal of the Acoustical Society of America*, *147*(3), 2061–2080. https://doi.org/10.1121/10.0000935

Tønnesen, P., Oliveira, C., Johnson, M., & Madsen, P. T. (2020). The long-range echo scene of the sperm whale biosonar: The sperm whale echo scene. *Biology Letters*, *16*(8). https://doi.org/10.1098/rsbl.2020.0134rsbl20200134

Tressler, J., & Smotherman, M. S. (2009). Context-dependent effects of noise on echolocation pulse characteristics in free-tailed bats. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, *195*(10), 923–934. https://doi.org/10.1007/s00359-009-0468-x

van Vossen, R., Beerens, S. P., & van de Spek, I. E. (2011). Anti-Submarine Warfare With Continuously Active Sonar. *Sea-Technology*, (November), 33–35. Retrieved from http://www.sea-technology.com/features/2011/1111/cont_active_sonar.php

von Benda-Beckmann, A. M., Isojunno, S., Zandvliet, M., Ainslie, M. A., Wensveen, P. J., Tyack, P. L., Kvadsheim, P. H., Lam, F-P. A., & Miller, P. J. O. (2021). Modeling potential masking of echolocating sperm whales exposed to continuous 1–2 kHz naval sonar. *The Journal of the Acoustical Society of America*, *149*(4), 2908–2925. https://doi.org/10.1121/10.0004769

von Benda-Beckmann, A. M., Wensveen, P. J., Samara, F. I. P., Beerens, S. P., & Miller, P. J. O. (2016). Separating underwater ambient noise from flow noise recorded on stereo acoustic tags attached to marine mammals. *Journal of Experimental Biology*, *219*(May), In review. https://doi.org/10.1242/jeb.133116

Wensveen, P. J., Thomas, L., & Miller, P. J. O. (2015). A path reconstruction method integrating dead-reckoning and position fixes applied to humpback whales. *Movement Ecology*, *3*(1), 31. https://doi.org/10.1186/s40462-015-0061-6

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *70*(3), 495–518. https://doi.org/10.1111/j.1467-9868.2007.00646.x

Zimmer, W. M. X., & Tyack, P. L. (2005). Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis. *The Journal of the Acoustical Society of America*, *117*(3), 1473–1485. https://doi.org/10.1121/1.1828501

761 Table 1. Hypotheses and behavioral indicators for auditory masking in odontocete echolocation

	Hypothesis		Mechanisms	Data indicators
Foraging effort and success	1	Reduced foraging effort	The net benefit of foraging is reduced, leading whales to give up foraging and to switch to a different behavior, or reduce locomotory effort within foraging states	Cessation of foraging behavior, support for switching from foraging to non-foraging active state published in Isojunno et al. (2020), or reduced pitching effort / fluke rates during foraging dives (Miller et al., 2009)
	2	Reduced foraging success	Prey capture success is reduced, which may also lead whales to give up during prey capture attempts and/or pursue fewer prey	Reduced buzzing within foraging states (Isojunno et al., 2020), reduced buzzing following searching behavior; fewer pauses following a buzz, shorter and less active buzzes
	3	Increased effort to capture prey	Auditory interference leads whales to increase time and locomotory effort to capture prey	Longer duration buzzes, greater locomotory activity during buzzes
Anti-masking strategies	4	Amplitude masking release	Increase echo level above masked threshold by increasing source level (Lombard effect)	Greater click source level or "apparent output level" (AOL) (Madsen et al., 2005) measured on the animal and thus measured "off-axis"
			Increase echo level above masked threshold by switching to prey type with greater target strength (e.g., from gelatinous squid to fish with swim bladders, or from individual to school targets)	Change in vertical prey layer and/or sound and movement tactics during buzzing (Madsen et al., 2005)
			Increase echo level above masked threshold by reducing transmission loss	Decrease target range for echolocation, expected to translate to shorter inter-click intervals (ICI)
	5	Spectral masking release	Reduce overlap with the masker in the frequency domain by shifting the spectral content of echolocation	Change in the frequency content of source clicks
	6	Temporal masking release	Reduce overlap with the masker in the time domain by shifting the timing of echolocation	Change in timing or rate of echolocation (click rates)
	7	Spatial masking release	Reduce overlap with the masker in space by facing away from the noise source	Increase 3D angle away from noise source (e.g., sea surface, or an underwater anthropogenic noise source)
			Reduce overlap with the masker in space by changing position away from the noise source	Horizontal or vertical avoidance to increase distance from the source

762

763

764

765 Table 2. Data variables tested in relation to each hypothesis (listed in Table 1). The data were
766 measured either for each terminal echolocation run ('buzzes'), or each search ("regular click")
767 echolocation train, separated by a pause or buzz. Hypothesis 1 was tested in Isojunno et al. (2020)
768 and Hypothesis 5 was deemed not possible to test with the present data set. "Models" column
769 lists tested model structures for each response variable (Table S1).

Hypothesis	Variable (units)	Short name	Dataset	Models
2	Pause presence following a buzz (zeros/ones)	Pause	Buzzes	1-6
2 vs 3	Buzz duration (s)	Duration		1-6
2 vs 3	Buzz ODBA rms, divided by deployment median (unitless)	ODBA		1-6
1	Fluke stroke rate (/s)	Stroke rate	Regular click (RC) trains	1-6
2	Buzz presence following a regular click train (zeros/ones)	Buzz		1-6
4	AOL of the first regular click of the train (zero-to-peak, dB)	AOLzp		9-12
4	First ICI of the regular click train (s)	ICI		1-6
6	Click rate of the regular click train (s) (inverse of median ICI)	CR		1-6
7	Facing angle to the sea surface (deg) directly above at RC start	Surface angle		7
7	3D angle to the sonar source (deg) at RC start (deg)	Sonar angle		8
7	Depth at the start of regular click train (m)	Depth		1-9,12

770

771 Table 3. Summary of collected data and sample sizes for analysis.

Tag ID	Presented sequence of exposures	Baseline-post		CAS		MPAS		HPAS		Sea state	SEL max
		RC	Buzz	RC	Buzz	RC	Buzz	RC	Buzz		
sw16_126a	CAS, MPAS, HPAS	387 (490)	67 (74)	28 (28)	3 (3)	0 (32)	0 (0)	22 (35)	0 (0)	1-3	168
sw16_130a	NS, MPAS, CAS	152 (178)	50 (54)	7 (7)	1 (1)	18 (18)	6 (6)			3-4	165
sw16_131a	NS	58 (270)	13 (39)							2.5-2.9	NA
sw16_134a	NS	67 (67)	21 (21)							2.3-2.5	NA
sw16_134b	NS, CAS, HPAS, MPAS	256 (270)	104 (110)	15 (15)	6 (6)	12 (18)	5 (7)	13 (13)	4 (4)	1-2.5	170
sw16_135a	NS, HPAS, CAS, MPAS	331 (375)	101 (113)	21 (21)	6 (6)	26 (26)	6 (6)	11 (11)	3 (3)	1-4	172
sw16_136a	NS, CAS, MPAS, HPAS	239 (282)	56 (64)	15 (15)	3 (3)	7 (13)	1 (1)	14 (14)	0 (0)	2.5-5	176
sw17_180a	NS, HPAS, MPAS, CAS	209 (324)	72 (101)	0 (23)	0 (9)	11 (11)	1 (1)	11 (14)	4 (4)	0.5-3.9	168
sw17_182a*	NS, MPAS, CAS, HPAS	42 (382)	16 (134)	0 (16)	0 (3)			0 (17)	0 (3)	1.2-3	NA
sw17_182b	NS, MPAS, CAS, HPAS	191 (227)	64 (75)	8 (14)	2 (5)	11 (14)	1 (3)	3 (4)	0 (1)	1-2	167
sw17_184a	NS, CAS, HPAS	280 (338)	85 (97)	16 (16)	3 (3)			0 (10)	0 (3)	2-4	158
sw17_186a*	NS, HPAS, CAS, MPAS	0 (593)	0 (146)	0 (22)	0 (12)	0 (0)	0 (0)	0 (43)	0 (27)		NA
sw17_186b	NS, HPAS, CAS, MPAS	338 (370)	219 (234)	10 (10)	2 (2)	13 (13)	9 (9)	34 (34)	25 (25)	2-5	157
sw17_188a	NS, MPAS, HPAS, CAS2	407 (462)	165 (185)	38 (38)	9 (9)	39 (39)	16 (16)	17 (30)	12 (19)	1-3	176
sw17_191a	NS, HPAS, MPAS	489 (489)	196 (196)			18 (18)	0 (0)	10 (10)	7 (7)	0.5-3	161

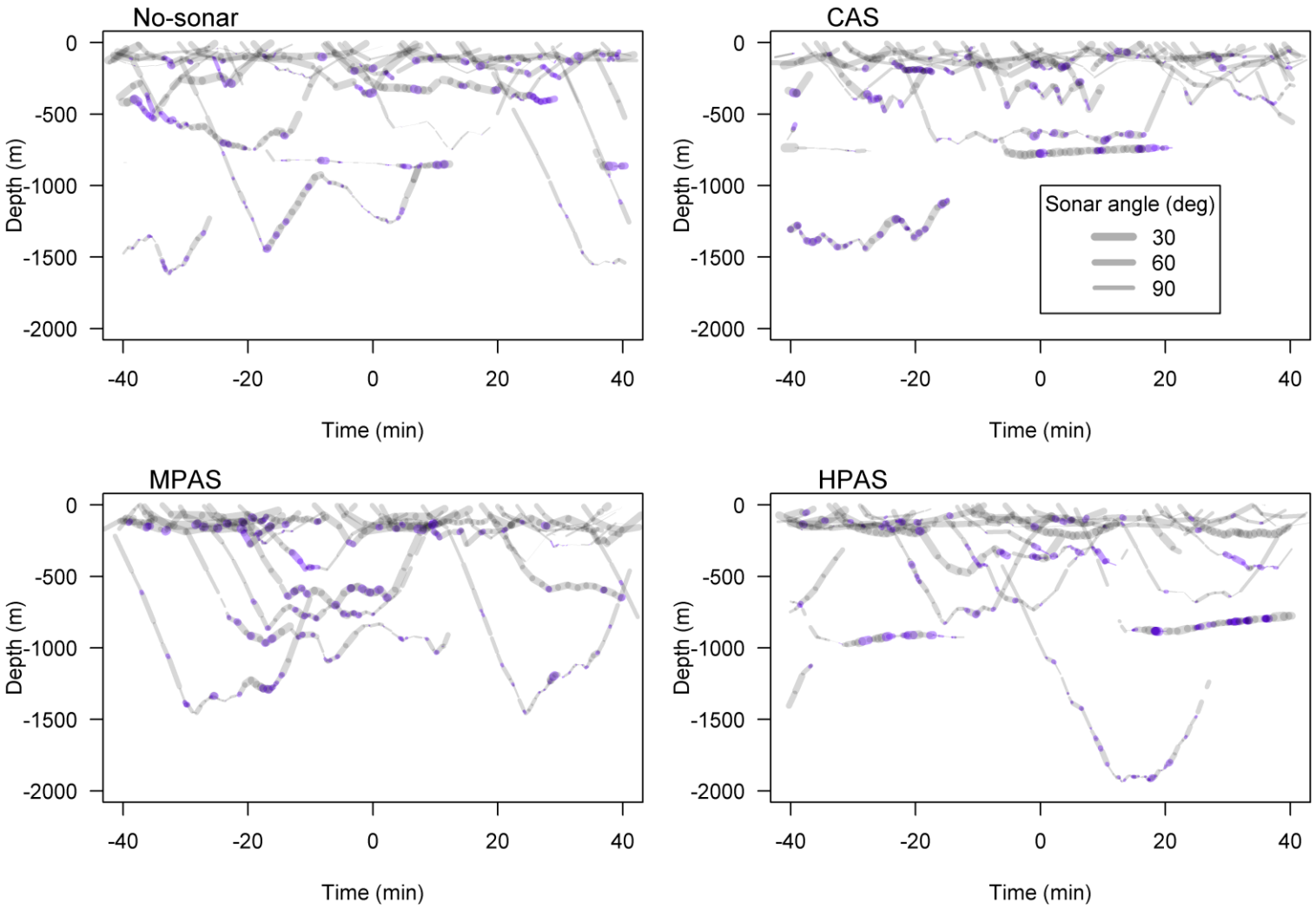
772 Note: Tag identifiers consist of the species code (sw = sperm whale), year, and Julian date. Sample sizes (number of regular click
773 trains [RC] or buzzes) are given for baseline and each exposure type (NS = no-sonar approach, CAS = continuous active sonar, CAS2
774 = two consecutive CAS exposures, MPAS = medium-level pulsed active sonar, HPAS = high-level pulsed active sonar). Brackets
775 show total sample size when including click trains with unknown sea state data (tagged whales > 6 km from the observation vessel).
776 Range of sea state (min, max) and maximum SEL values present in the *analyzed* dataset for each tag deployment are also given. Empty
777 cells indicate lack of exposure data (except for nonfocal whales marked as *, exposure data which were not included in the analysis).

Table 4. Mean and standard deviation (SD, in brackets) calculated across tag deployment means during different Beaufort sea states (SS; baseline/post-exposure data) and experimental conditions, excluding data from >500 m dive depth (to control for change in ambient noise from sea surface at depth; please see Table S2 for full data set). Note that received SEL_{sp} varied during the sonar exposures. NA = data from a single tag deployment, not possible to calculate SD.

Data	Variable	SS ≤ 2	SS 3	SS ≥ 4	No-sonar	MPAS	HPAS	CAS
Buzzes	Pause %	48.7 (32.8)	26.5 (28.9)	0 (NA)	35.4 (37)	51.6 (45.4)	42.4 (33.1)	10.7 (15.4)
	Duration (s)	21.2 (6)	19.2 (4.6)	17.3 (NA)	17.4 (7.5)	16.5 (4.8)	18.5 (9)	15.5 (4.7)
	ODBA (unitless)	3.3 (0.85)	3.7 (0.67)	3.41 (NA)	3.34 (0.72)	3.06 (0.89)	2.93 (0.59)	3.34 (1.25)
Regular click trains	Buzz %	24.8 (12.8)	21.7 (9.6)	10.5 (14.9)	23.6 (14.3)	12.1 (16.3)	18.6 (21.1)	17.4 (12.1)
	AOL * (dB)	181.8 (5.5)	178.9 (6.9)	176.2 (9.9)	181.4 (5.9)	179.3 (7.3)	179.2 (7.9)	179.2 (9.1)
	Stroke rate (/min)	4.3 (1.2)	4.0 (0.7)	4.0 (0.2)	4.4 (1.0)	4.3 (1.3)	4.3 (1.9)	4.2 (1.3)
	ICI (s)	0.88 (0.3)	0.78 (0.18)	1.09 (0.57)	0.82 (0.11)	0.82 (0.23)	0.88 (0.28)	0.89 (0.4)
	CR (/s)	1.41 (0.31)	1.74 (0.55)	2.08 (1.5)	1.61 (0.61)	1.58 (0.5)	1.49 (0.49)	1.55 (0.52)
	Surface angle (°)	109.2 (20.3)	104.2 (8.9)	119.7 (22.6)	107.6 (23)	105.7 (18.6)	104.3 (22.2)	108.9 (20.1)
	Sonar angle (°)	98.2 (15.2)	93.4 (23.6)	118 (0.6)	96.8 (32.7)	88.2 (32)	98.7 (40.9)	103.3 (31.3)
	Depth (m)	195 (78)	183 (67)	171 (26)	187 (88)	163 (46)	187 (91)	152 (88)

* Clipped values included

785 **FIGURES**



786

787 Figure 1. Start and end depths of buzzes (purple) and regular click trains (gray) as a function of exposure time (before and after start

788 of exposure at 0 min). Greater line width indicates a smaller angle towards the sonar source (Sonar angle, degrees).

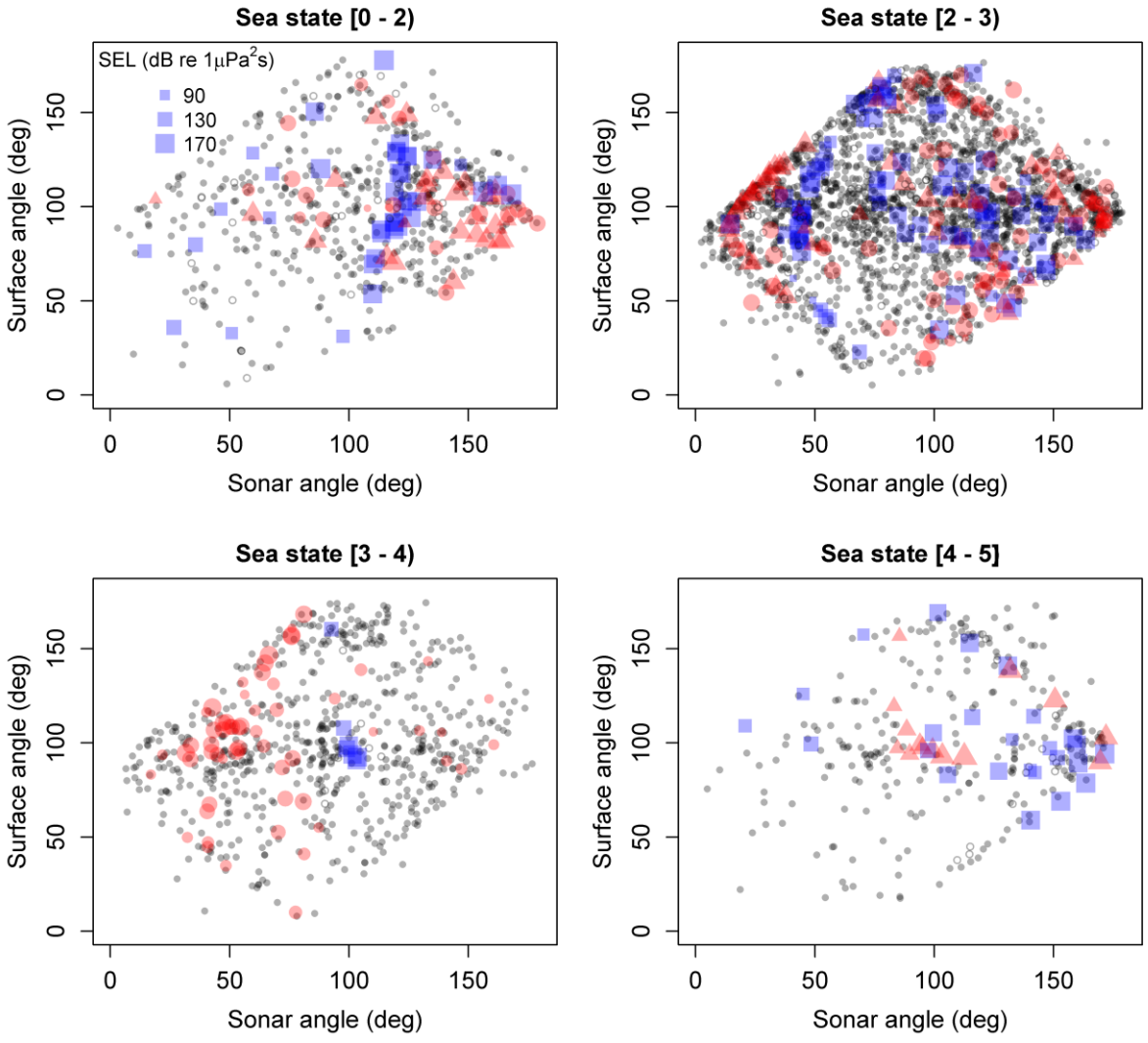
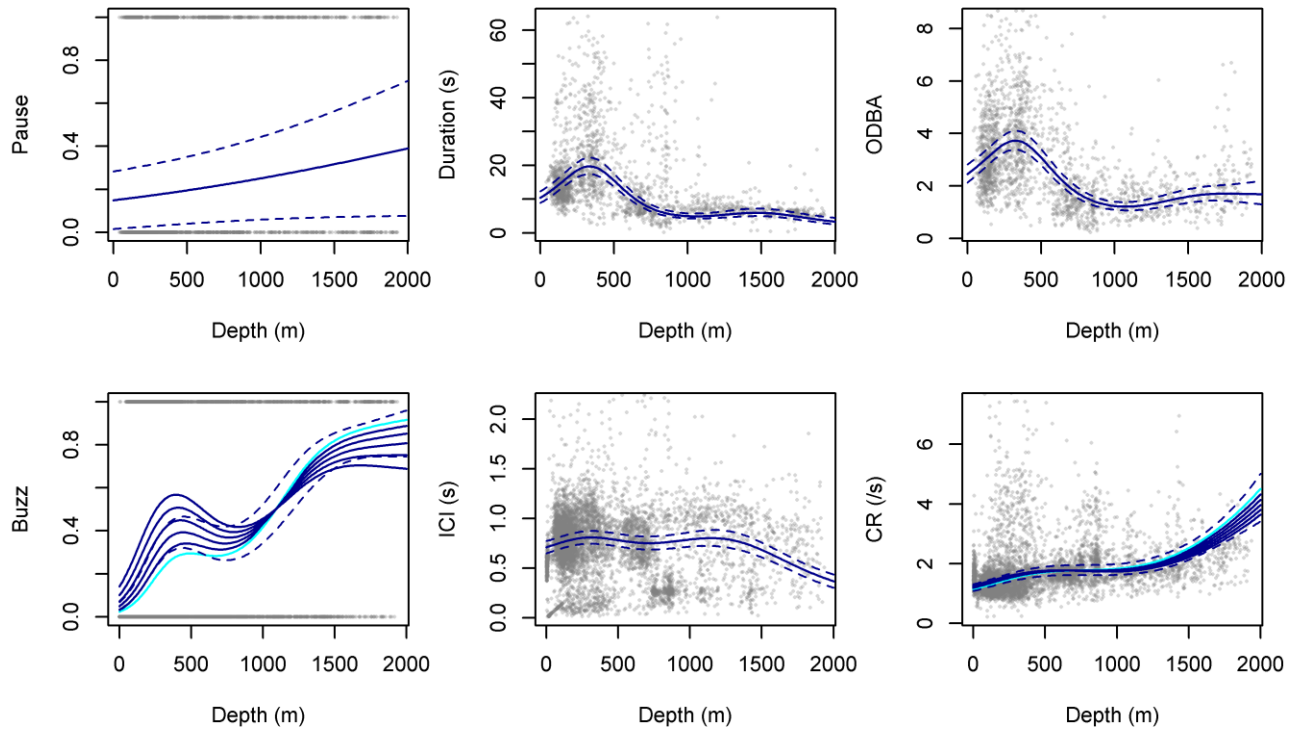


Figure 2. Surface angle (whale pitch expressed as degrees away from directly above) and sonar angle (angular difference between whale movement direction and bearing to source vessel, degrees) at the beginning of regular click trains both during and outside sonar exposures (blue = CAS continuous active sonar, red = PAS pulsed active sonar with triangles indicating HPAS and circles MPAS exposures). Solid gray circles indicate baseline/post periods and hollow circles no-sonar exposures. Note there was no apparent differences in surface and sonar angles during sonar exposure periods as expected by spatial masking release (Hypothesis 7, Table 2).



797

798 Figure 3. Observed (gray) and model-predicted values (blue) for pause occurrence following
799 buzz trains (Pause), buzz duration (Duration), buzz locomotory activity (root-mean-square
800 overall dynamic body acceleration; ODBA), buzzing occurrence following regular click trains
801 (Buzz), regular Interclick intervals (ICIs) and regular clicking rates (CR). nonexposure model
802 predictions were made as a function of depth and for those full models that supported the effects,
803 Beaufort sea state (0-5, with the highest sea sate 5 indicated in cyan), with surface angle fixed to
804 90 degrees. Dashed lines show normal confidence intervals at Beaufort sea state 3. Note the
805 relatively small differences between sea states (multiple solid lines within a panel) compared to
806 the relatively large variation in the metrics as a function of depth.

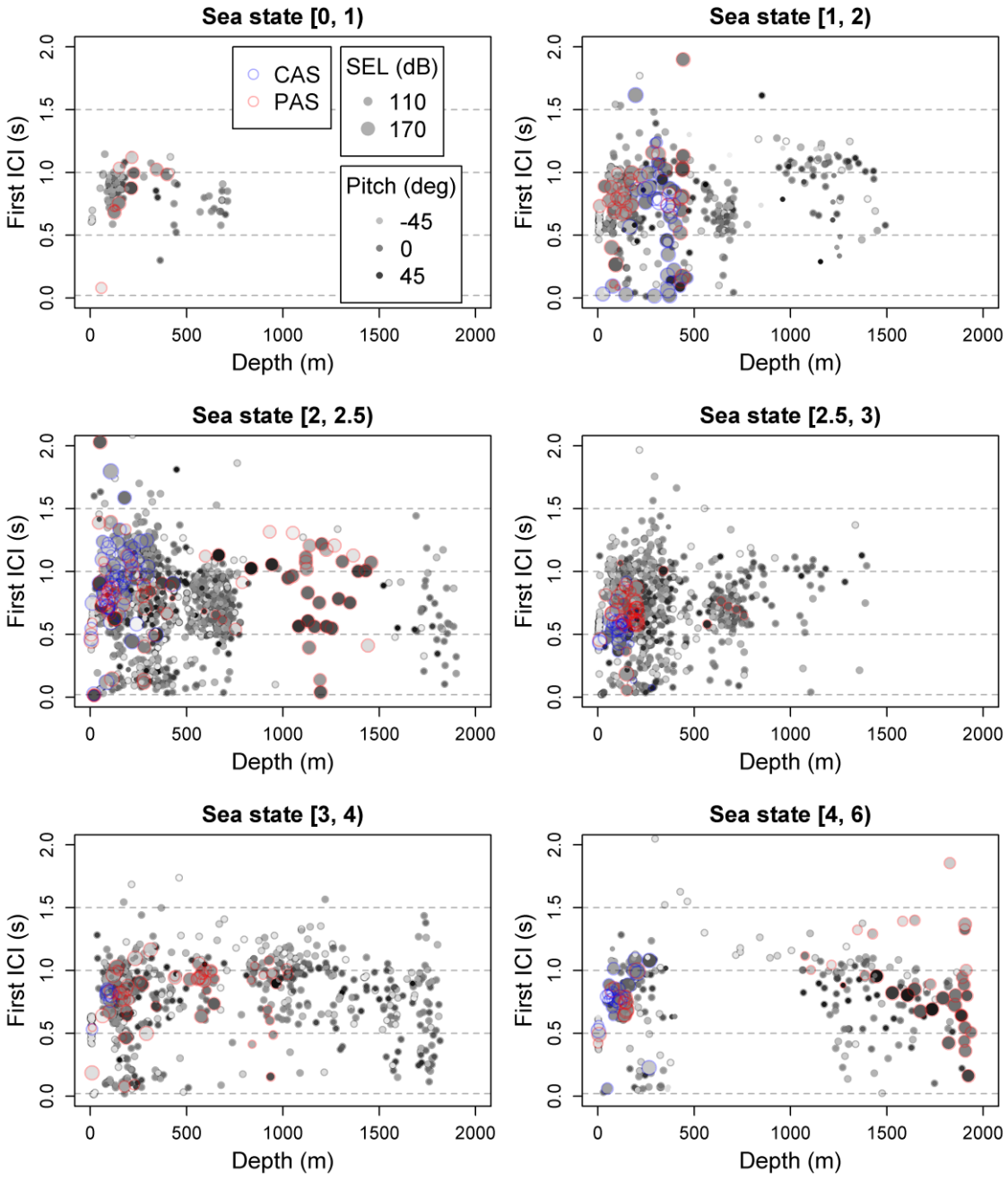


Figure 4. Interclick interval (ICI) at the start of regular click trains at different sea states and pitch angles (initial pitch of the train both during exposure and nonexposure periods (CAS = continuous active sonar, PAS = pulsed active sonar). According to Hypothesis 4 (Table 1), sperm whales were expected to reduce inspection range and therefore interclick interval in conditions with greater masking potential (during high sea states, sonar and when facing the noise source).

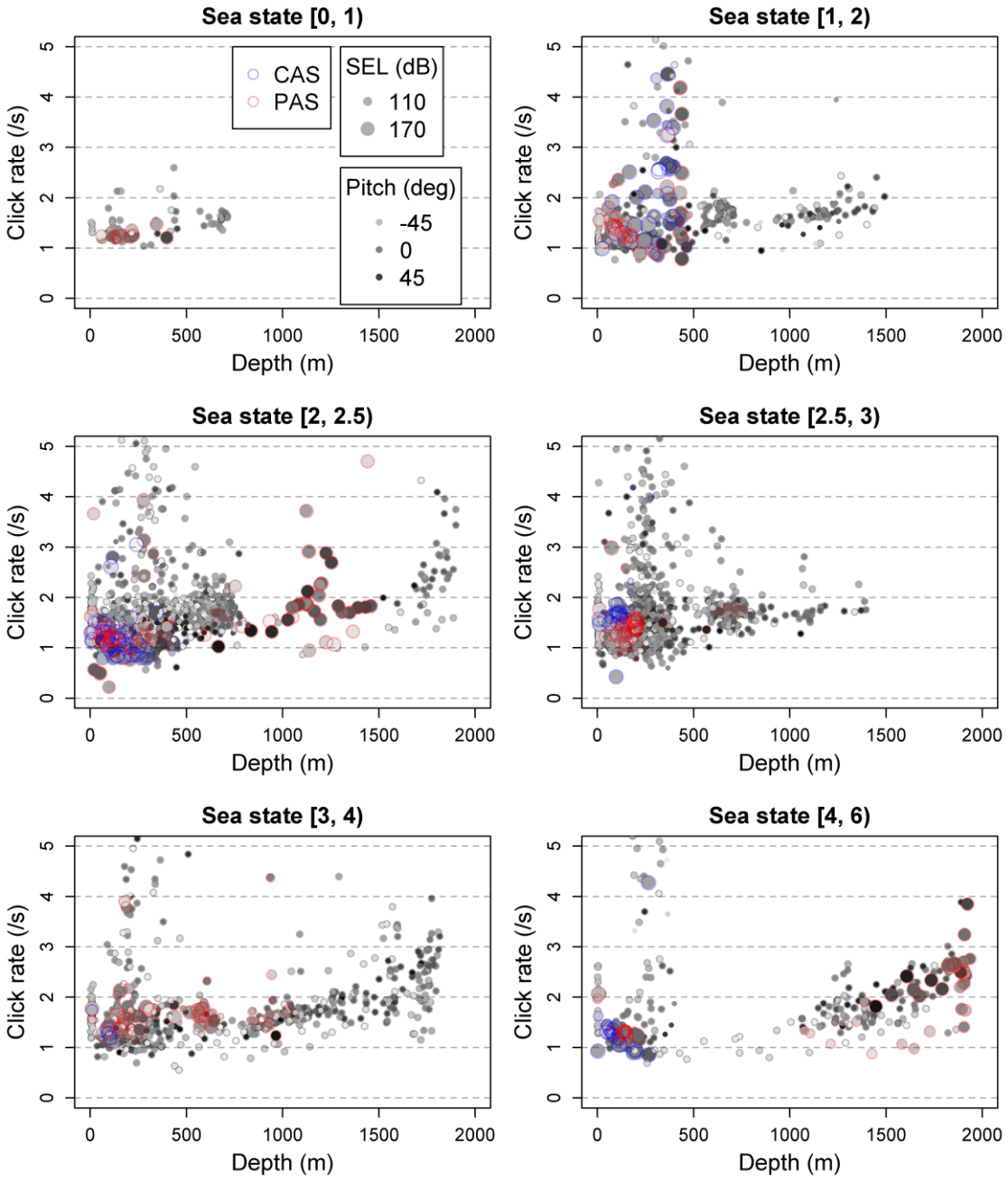


Figure 5. Regular click rates at different sea states and pitch angles both during exposure and nonexposure periods (CAS = continuous active sonar, PAS = pulsed active sonar). According to Hypothesis 6 (Table 1), sperm whales were expected to mediate the temporal pattern of their echolocation (click rates) with respect to the masking conditions.

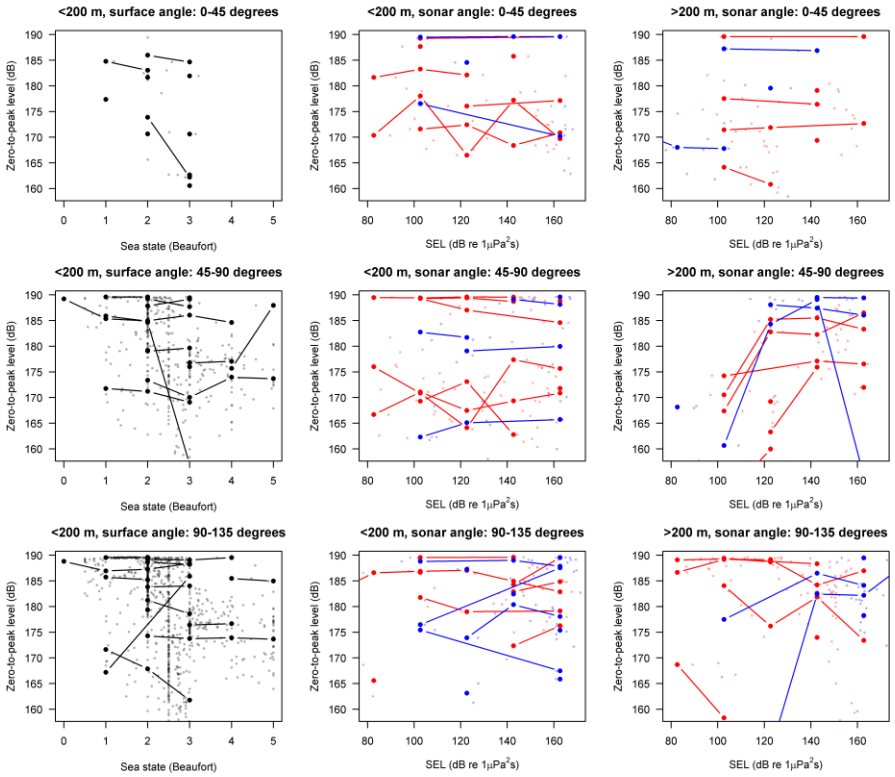
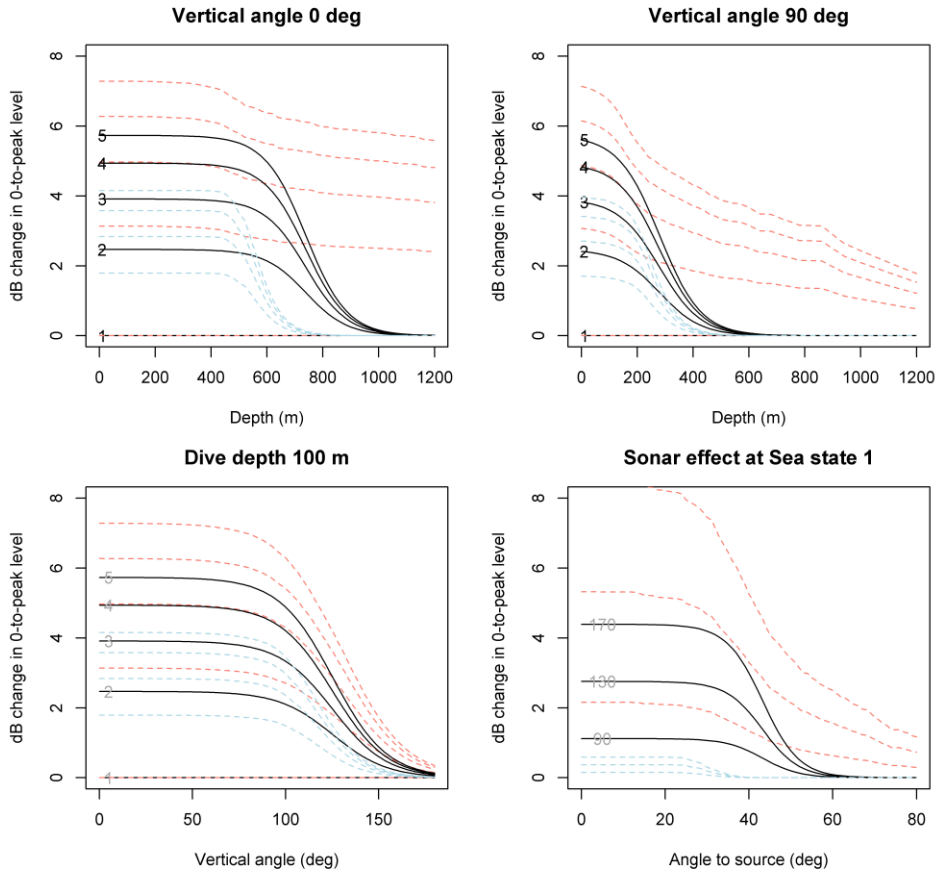


Figure 6. Apparent output level (zero-to-peak sound pressure, dB re 1 µPa) (AOL_{zp}) of regular clicks within each tag deployment (connected lines) as a function of sea state and received sound exposure level (SEL) during exposures, at different depths, surface, and sonar angles (blue = CAS continuous active sonar, red = PAS pulsed active sonar). Transparent symbols show raw data, solid symbols averages within sea state and SEL bins. Note: clipped values are included. According to the amplitude masking release Hypothesis 4 (Table 1), AOL_{zp} was expected to increase with sea state (especially when facing the sea surface at shallower depth) and with greater received level of sonar (especially when facing the sonar source).



825

826 Figure 7. Estimated change in apparent output level (zero-to-peak pressure, dB) (AOL_{zp}) of regular clicks at the start of click trains
827 as a function of sea state (numbered lines) and received sound exposure level (SEL; bottom right panel shows effect at received SEL of
828 90, 130, and 170 dB re 1 μPa^2 s), at different depths and angles to surface/sonar source. Solid lines show posterior medians, dashed
829 lines 95% credible interval (blue: lower, red: upper). The model explained 8.6% of the AOL_{zp} values.