Marine Pollution Bulletin 79 (2014) 254–260

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Severity of killer whale behavioral responses to ship noise: A dose–response study

Rob Williams [a,b,*], Christine Erbe [c], Erin Ashe [a,b], Amber Beerman [a], Jodi Smith [d]

[a] Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews Fife, KY16 8LB Scotland, UK
[b] Oceans Initiative, Pearse Island, Alert Bay, BC V0N 1A0, Canada
[c] Centre for Marine Science & Technology, Curtin University, Kent Street, Bentley, 6102 Western Australia, Australia
[d] Naked Whale Research, Manchester, CA 95459, USA

ARTICLE INFO

Keywords:
Canada
Conservation
Critical habitat
Killer whale
Noise
Ship

ABSTRACT

Critical habitats of at-risk populations of northeast Pacific "resident" killer whales can be heavily trafficked by large ships, with transits occurring on average once every hour in busy shipping lanes. We modeled behavioral responses of killer whales to ship transits during 35 "natural experiments" as a dose–response function of estimated received noise levels in both broadband and audiogram-weighted terms. Interpreting effects is contingent on a subjective and seemingly arbitrary decision about severity threshold indicating a response. Subtle responses were observed around broadband received levels of 130 dB re 1 μPa (rms); more severe responses are hypothesized to occur at received levels beyond 150 dB re 1 μPa, where our study lacked data. Avoidance responses are expected to carry minor energetic costs in terms of increased energy expenditure, but future research must assess the potential for reduced prey acquisition, and potential population consequences, under these noise levels.

1. Introduction

A number of experimental and opportunistic studies have quantified the effects of small boat traffic on the fish-eating, "resident" killer whale populations in the northeastern Pacific (Erbe, 2002; Holt et al., 2008; Lusseau et al., 2009; Williams and Ashe, 2007; Williams et al., 2002a,b, 2006). These studies showed that killer whales avoid boats using stereotyped evasive tactics consistent with horizontal avoidance (i.e., changes in measures of path directness and angle between adjacent surfacings) that make the whale's path less predictable to a single boat. When exposed to repeated levels of disturbance throughout the day, the net effect is an altered activity budget, in which killer whales spend less time feeding in the presence of boats than during no-boat, control conditions (Lusseau et al., 2009; Williams et al., 2006).

A number of studies have demonstrated effects of noise from large ships on a variety of cetacean species, including Cuvier's beaked whale (Aguilar Soto et al., 2006), North Atlantic right whale (Nowacek et al., 2004; Rolland et al., 2012), beluga (Erbe and Farmer, 1998; 2000) and fin whales (Castellote et al., 2012). These studies provide a hint that ship noise can reduce a whale's foraging efficiency (Aguilar Soto et al., 2006); elevate the risk of ship strikes (Nowacek et al., 2004); and cause physiological stress that is detectable in hormone levels (Rolland et al., 2012). A combination of captive experiments and computer models (Erbe and Farmer, 2000) enabled researchers to estimate that icebreaker noise is audible to belugas and capable of eliciting behavioral responses and causing communication masking at ranges to 62 km. A temporary hearing shift was modeled to occur if a beluga stayed within 1–4 km of the icebreaker for at least 20 min. Whales have evolved in an ocean environment that becomes naturally noisy during storms and surf zones, and they have evolved some mechanisms to compensate for noise. Fin whales change their song characteristics to try to maintain communication in high levels of shipping noise (Castellote et al., 2012). There is some evidence to suggest that killer whales can compensate for increases in ambient noise by lengthening their calls (Foote et al., 2004) or increasing the source level of social calls (Holt et al., 2008). There is no evidence that killer whales can adjust their echolocation patterns to compensate for masked signals used in foraging, and no information on the upper limit to the whales' compensatory mechanisms. In many behavioral response studies, the received levels that trigger responses are rarely known (but see (Williams et al., 2002a)). A recurring theme in the literature describing marine mammals and noise is that the most rigorous behavioral studies rarely report information on the acoustic stimulus, and the best acoustic studies often have very small sample size for inferring behavioral responses (Nowacek et al., 2007).

* Corresponding author at: Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews Fife, KY16 8LB Scotland, UK. Tel.: +1 250 974 8048.

E-mail address: rmcw@st-andrews.ac.uk (R. Williams).

0025-326X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.marpolbul.2013.12.004

No studies have yet examined the responses of killer whales to presence and activities of large ships. Such studies are needed (Wright, 2008). Global shipping represents a large and growing contributor to ocean ambient soundscapes (Hildebrand, 2009), and creative solutions are needed to quantify and mitigate impacts of chronic ocean noise on sensitive marine mammals (Wright et al., 2011). A recent attempt to incorporate data on chronic ocean noise in marine spatial planning in the northeast Pacific found that large ships pass through the Haro Strait critical habitat of southern resident killer whales on average approximately once every hour, day and night, year-round (Erbe et al., 2012). Little information is available to assess the likely impact of that stressor on killer whale behavior, activity budgets, energetics or fitness, but such information would improve the conservation and management of at-risk species. Northern and southern resident killer whales have been listed under the relevant endangered species legislation of Canada and the US (Fisheries and Oceans Canada, 2011; National Marine Fisheries Service, 2008). Both countries have recognized prey depletion, contaminants and anthropogenic noise as risk factors in the whales' current conservation status and threats to be addressed to promote recovery (Fisheries and Oceans Canada, 2011; National Marine Fisheries Service, 2008).

Due to the logistical constraints and expense of experimenting on free-ranging killer whales, existing data were re-examined to assess "natural experiments" that could be used to measure the direction and magnitude of any observed behavioral responses of killer whales to large ship traffic. A long-term, land-based study (Williams et al., 2002b) has generated a large dataset that was reanalyzed to evaluate behavioral responses of northern resident killer whales (NRKW) to occasional transits by three categories of large ships: cargo vessels, cruise ships and ocean-going tugs. This archived dataset includes measurements of dive time, swimming speed, path directness, path smoothness and rates of surface-active behavior (SAB) of individually recognizable focal whales.

2. Methods

2.1. Study area

The study area for the NRKW population covered the western end of Johnstone Strait, British Columbia (BC), Canada. All data were collected from a land-based observation point on West Cracroft Island (50°30′N, 126°30′W). The study was intended to capture typical summer time conditions in important killer whale habitats. It is unknown whether killer whales should be more or less responsive to noise in winter months, or in marginal foraging habitats, but because this was a retrospective analysis of existing data (i.e., with no funding for additional field work), inference is restricted to the period during which data were collected: six years (1995–1998, 2002 and 2004), covering the months July and August. Similar data on southern resident killer whales (collected by JS) were examined for comparative analyses, but only two natural experiments were observed. The data on southern resident killer whales were not included in subsequent analyses.

2.2. Study methods

2.2.1. Theodolite tracking of individual whales

Data were collected using an electronic theodolite (Pentax ETH-10D with a precision of ±10″ of arc) connected to a laptop computer equipped with custom software (THEOPROG, (Williams et al., 2002b)). The tracking team consisted of a spotter, theodolite operator, computer operator, and video/data recorder. Killer whales entered the study area in matrifocal social units called matrilines that ranged in size from 2 to 120 individuals (Ford et al., 2000). A focal animal was selected from the group, using previously described selection criteria (Williams et al., 2002a,b) to ensure representative sampling of the population and reliability of re-sighting an individual within a tracking session. Because initial activity state can affect the probability of killer whales responding to small vessels (Williams et al., 2006), focal animals were selected during travel/forage activity, rather than resting, socializing, feeding or beach-rubbing.

Positions of surfacing animals (horizontal and vertical angle coordinates) were located using the theodolite and directly recorded into the laptop computer using THEOPROG. At each surfacing, the team recorded the focal whale's alpha-numeric ID (Ford et al., 2000), each time the whale surfaced to take a breath, and any corresponding surface active behavioral events such as breaches, pectoral fin slaps and tail (fluke) slaps. Accuracy of each whale position was confirmed by the laptop operator by viewing the positions as they were plotted in real-time. Any deviation or noticeable gap in surfacing was reviewed and confirmed by the theodolite operator.

Positions of vessels were marked with the theodolite once they entered the study area, usually while the focal whale appeared to be down on a long dive. Vessels were assigned to one of the following 10 categories:

- CAR = Self-Propelled Cargo Vessel
- CCV = Commercial Charter Vessel
- CFV = Commercial Fishing Vessel
- COL = Commercial Ocean Liner
- GPV = Government Patrol Vessel
- PRV = Professional Research Vessel
- RKG = Recreational Kayak Group
- RPV = Recreational Power Vessel
- RSV = Recreational Sailing Vessel
- TUG = Tug Boat

For the purposes of the Theodolite Threshold Study, three of these were considered large ships (COL, TUG and CAR), whereas the others were considered small vessels.

2.3. Data compilation

2.3.1. Summarizing response variables

Whale data were summarized for each track, with each track represented only once in the analyses. Five dependent whale response variables included were: inter-breath interval (dive time), speed, directness index (directness), deviation index (DEV) and surface active behavior (SAB). Refer to Table 1 for the dependent whale response variable definitions (Williams et al., 2002a,b). For completeness, we include in an appendix the R code required to calculate the directness and deviation indices from the X–Y coordinates (Appendix 1).

2.3.2. Defining natural experiments from the theodolite tracks

All tracks that included marks of large ships (cruise ships (COL), tugs (TUG) or cargo vessels (CAR)) were assessed for opportunistic natural experiments in which there was sufficient data to be able to compare and contrast behavior of the focal whale before exposure to large vessel presence and during exposure (Table 2; Appendix 2). There were a few occasions where behavior could be monitored after the ship had left the study area, but too few for a 3-way analysis. For completeness (and to facilitate inclusion of our data in future meta-analyses), we summarized whale behavior in all three segments – "Before", "During", and "After" ship encounter – even though we only used before and during comparisons in statistical analyses. For practical reasons (i.e., given the constraints of analyzing historical data, rather than conducting

Table 1
Definitions of the five killer whale behavioral variables used for scoring severity of responses.

Behavioral code	Description
Respiration (DIVE TIME)	The mean inter-breath interval is the duration (time from the onset of the first breath to the onset of the last breath) divided by the number of intervals (one less than the number of breaths). Units are seconds, representing the average time between breaths
Point-to-point speed (SPEED)	The average swimming speed (point-to-point distance covered over time) of the whale obtained by dividing the total distance traveled by the duration of tracking session, and reported in m/s
Directness index (DIRECTNESS)	Measures path predictability of a tracking session by dividing the distance between end-points of a path by the cumulative surface distance covered during all dives and multiplying by 100. Ranges from zero (a circular path) to 100 (a straight line)
Deviation index (DEVIATION)	Measures path predictability of a tracking session from one surfacing to the next. It is the mean of all angles between adjacent dives. For each surfacing in a track, this is the angle between the path taken by a dive and the straight-line path predicted by the dive before it was calculated. The deviation index is the mean of the absolute value of each of these discrepancies, in degrees, during the entire track
Surface-active behavior (SAB)	Number of surface-active behaviors (e.g. breaches, tail-slaps) counted in a track divided by the elapsed time of observation multiplied by 60 min to determine the mean rate (events per hour)

Table 2
Severity scores used in our analyses, following (Southall et al., 2007). Note that responses of severity score 1 were not possible in our study, because that category refers to brief orientation toward the sound source, and the raw data did not record this information.

Response score	Corresponding behavior
0	No change in any of the whale response variables
2	Minor change in respiration
3	Minor change in locomotion speed, direction, and/or deviation Moderate change in respiration
4	Moderate change in locomotion speed, direction, and/or deviation
5	Extensive change in locomotion speed, direction, and/or deviation

No change, 0–10%; Minor change, 10–20%; Moderate change, 20–50%; Extensive change, >50%

new field studies), the only breakpoints that we could consider involved scoring "before" or "control" segments as the time from the first mark until the last whale mark 5 min before the ship entered the study area, and the "during" or "treatment" segment as spanning the time period from 5 min before the first ship mark until 5 min after last ship mark. For completeness, we include maps of illustrative examples of what the theodolite tracks look like (Appendix 3).

For each segment of each natural experiment, the same five dependent whale response variables were calculated. Rather than conducting five statistical tests, which could result in spurious correlations, we followed recommended best practice with respect to scoring the "severity" of behavioral responses to noise exposure (Southall et al., 2007). We compared whale behavior in control and treatment segments, and based on the differences, we assigned a severity score to each natural experiment (Table 2). The decision whether to call a change "minor" or "moderate" is somewhat subjective. We defined "minor" and "moderate" changes in Table 2, based on the first author's experience conducting control-exposure experiments on killer whales since 1995. We defined a minor change as a 10–20% change in a variable, based on the 13% change in directness index observed when a single boat parallelled a male killer whale at 100 m (Williams et al., 2002b). We defined a moderate change as a 20–50% change in a variable, based on the 25% change in swimming speeds of female killer whales to a single boat parallelling the whale at 100 m (Williams et al., 2002b). We defined an extensive change as a >50% change in a variable, based on the 90% change in path smoothness when a boat leapfrogged the whale's path at 150–200 m (Williams et al., 2002a). Importantly, the severity score is meant to differentiate between minor/brief responses (0–4), those that could affect foraging, reproduction or survival (4–6), and those (7–9) that could affect vital rates (Southall et al., 2007). Although there is some degree of subjectivity in our categorization, it is important to note that (a) we are explicit and transparent about the criteria we used to assign a given response score to an experiment; (b) our decision was made by the biologists on our team, without information from the acoustician on received level; and (c) any level of subjectivity is small relative to Southall's broad categories – that is, there may be some disagreement about whether an experiment elicited a response of 2 or 3, but none of these trials elicited scores that would fall in a higher risk category (e.g., 7–9).

2.3.3. Summarizing independent variables

Candidate covariates in our analyses included natural and anthropogenic factors. For natural factors, candidate covariates included WhaleID, Year, Month, TimeOfDay, Age, and Sex. Variables describing anthropogenic factors included the following: CAR, TUG and COL (number of cargo vessels, tug boats and cruise ships ever observed during the track, respectively); Ship_Speed (speed, in m/s, of the ship that ever got closest to the whale during the track); PCA1 (point of closest approach, in meters, of the ship that ever got closest to the whale during the track); N_other_boats (number of small vessels recorded during the track); and two factors describing our best estimate of received noise level during the encounter (RL_rms and RL_weighted).

The received noise level was calculated in the following way. A representative source spectrum (rather than broadband source level) was computed for each ship. Cargo and container vessels were assumed to be 100 m long, beyond which ship length has less pronounced effects on source level than for smaller vessels (Erbe et al., 2012; McKenna et al., 2012). Using each vessel's measured speed, the source spectrum for each vessel track was computed based on the RANDI noise model (Breeding et al., 1994). For tugs, only one source level from a tethered tug (at speed v_0) was available from the database held at the Center for Marine Science & Technology. For this study, the spectrum level for tugs was adjusted for each vessel's speed (v_t) by adding 60 log (v_t/v_0) (Hamson, 1997). The source spectra of the three ship types at their mean speeds measured on site are shown in Fig. 1.

A parabolic equation (Collins et al., 1996) was used to model sound propagation based on a summer sound speed profile taken from the Global Digital Environmental Model (GDEM) database (Carnes, 2009), geoacoustic properties of clay (Hamilton, 1980), a source depth of 6 m, and a receiver depth of 5 m. Seawater absorption was also accounted for (François and Garrison, 1982a,b). This sound propagation model is described in more detail in (Erbe et al., 2012). The RL was computed in broadband (i.e., in dB re 1 μPa rms, called RL_rms) and audiogram-weighted (called RL_weighted) units. The audiogram was derived from published hearing curves (Hall and Johnson, 1972; Szymanski et al., 1999), as outlined in (Erbe, 2002).

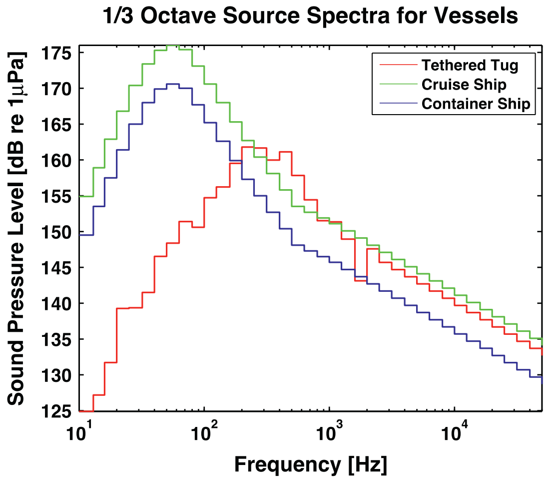

Fig. 1. One-third octave source spectra of the three vessel types at their mean speeds as measured by theodolite tracks.

2.4. Data analysis

Although the raw theodolite data were processed in THEOPROG and the behavioral responses summarized and given a severity score in Excel, all statistical analyses were conducted using generalized linear models (GLM) in R (Faraway, 2005). Ideally, one would model the response severity score itself as a function of explanatory covariates. Regrettably, there is no link function for GLMs that can cope with an ordered factor response variable (i.e., a variable in which a severity score of 6 is larger than 3, but not necessarily twice as large as 3). This statistical limitation requires that researchers, managers or regulators define a cutoff that reflects the level of impact on animals that they are willing to allow (Miller et al., 2012). Scores above that cutoff are considered a response; scores below that are considered no-response. This seemingly arbitrary decision represents a loss of information contained in the severity score itself, but does allow the causes of the response to be modeled as a binary outcome. Incidentally, statisticians are working on developing new Bayesian methods that can model the severity of a response in a GLM or generalized additive modeling framework, but they require substantial development.

We assessed 42 theodolite tracks containing ship transits to find natural experiments that could be used to model the probability of a whale responding. Of the 42 tracks considered, 35 could be considered in a before-during natural experimental framework, with sufficient information to quantify changes in whale behavior before and during a ship transit. The 7 tracks that had to be dropped contained insufficient information about whale behavior before and/or during the ship's transit to evaluate response; sparse information on the ship's track was not the limiting factor. Scoring each experiment as either a response or a non-response required using all values greater than or equal to some severity score cutoff as a somewhat arbitrary threshold. To account for the subjective nature of this step, analyses were run using severity scores of both 2 and 3 as cutoffs. There was insufficient coverage and resolution in the data to consider other levels of the Southall score as cutoffs.

We modeled the probability that a whale did (1) or did not (0) show a behavioral response to a ship transit, in a GLM framework. Candidate covariates included natural (WhaleID, Year, Month, TimeOfDay, Age, and Sex) and anthropogenic (CAR, TUG and COL; Ship_Speed; PCA1; N_other_boats; RL_rms and RL_weighted) variables. With a binomial response, one has the choice of several link functions, including logit, probit or complementary log–log. The logit link is the default for most logistic regressions. We used a probit link, because this imposes the classic sigmoidal shape thought to underlie conventional dose–response curves (Miller et al., 2012). We did not have sufficient data to be able to test alternative relationships; instead, we are assuming that killer whales will not respond to noise below some unknown, but low, received level, and that all whales would respond to noise at some unknown high level (even if that level is beyond the range of our data). In other words, the model structure assumes that if there is a dose–response relationship, it will follow a classic sigmoidal shape common to all toxicology studies, and the data are used to estimate parameters describing the curve we suspect is there. If there is no support from the data for fitting the curve, then each term will have a coefficient of zero and we will be left with an intercept-only model. We used a stepwise procedure to consider all possible combinations of candidate independent variables to choose the lowest Akaike Information Criterion (AIC; (Burnham and Anderson, 2002)). We used function *stepwise* in the "Rcmdr" library (Fox, 2005) to select the combination of terms that provided the best fit to the data, with AIC score penalizing the addition of unnecessary terms. We found that model selection was insensitive to step direction, that is, forward and backward stepwise procedures resulted in the same model being selected in all cases.

3. Results

A completely automated model selection procedure resulted in two quite different models, depending on the severity score cutoff that was used to define response. Assuming that a response is given by a score of 2 or greater on the Southall scale, the model selected by an automated stepwise procedure was (Model 1):

$$\text{Response}_2 \sim \text{Year} + \text{CAR} + \text{COL} + \text{TUG} + \text{Month} + \text{Age} + \text{RL\_rms}, \quad (\text{Model1})$$

	Estimate	Std. error	z Value	Pr(>|z|)
(Intercept)	699.74410	324.52124	2.156	0.0311*
Year	−0.34602	0.15989	−2.164	0.0305*
CAR	−10.30153	5.23157	−1.969	0.0489*
COL	−6.09617	3.02291	−2.017	0.0437*
TUG	−9.54309	4.89167	−1.951	0.0511.
Month	−3.04004	1.62113	−1.875	0.0608.
Age	0.06393	0.02682	2.383	0.0172*
RL_rms	0.18178	0.11832	1.536	0.1244

Signif. codes: 0'***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
AIC: 38.287

which means that a killer whale's response to the passage of a ship (using a score of ⩾2 as a cutoff), on average, was best explained by the number of ships in each category, year, month, the whale's age, and an increasing probability of response as received level (rms) increased. The termwise significance tests of a binomial GLM are not exact, but there is strong support for including received level (RL_rms) in Model 1 (termwise $P = 0.1244$), but AIC supported the decision to retain seemingly non-significant terms.

Binomial models are somewhat difficult to interpret with respect to explanatory power, and the usual R summaries for binomial GLMs do not contain the kind of R-squared summary statistics one normally expects in a regression. There is a tool[1] ("binomTools") to extract information from binomial models to give an idea about their explanatory power. We used function *Rsq* in

[1] http://artax.karlin.mff.cuni.cz/r-help/library/binomTools/html/Rsq.glm.html

package *binomTools* to illustrate, roughly, how much explanatory power each model had, and to assess how much additional explanatory power the various models had when including or excluding information on received level. We found that Model 1 had an R-squared value of approximately 0.58.

We reran all models with the cutoff for scoring a response set this time to ⩾3 on the Southall scale. In this case, both forward and backward stepwise model selection indicated that the preferred model was [Model 2]:

$$\text{Response}_3 \sim \text{Sex} + \text{N\_other\_boats}, \qquad \text{(Model2)}$$

which means that a killer whale's response to the passage of a ship (using a severity score of ⩾3 as a cutoff), on average, was best explained by the number of small vessels in the area and the sex of the whale. Using strictly automated procedures, Model 2 did not include information on received noise level at the whale. Because a central focus of this study is to understand whether noise was a better predictor of behavior than other variables, we compared the selected model (Model 2) to one that also contained information on received noise level. We found that

$$\text{Response}_3 \sim \text{Sex} + \text{N\_other\_boats} + \text{RL\_rms}, \qquad \text{(Model3)}$$

had similar support from the data as Model 2. The difference between Model 2 and Model 3 was ΔAIC = 1.41, which means that there is no strong statistical support for dropping noise level from the model. On the contrary, explanatory power of the model increased from *R*-squared = 0.23–0.25 when we included a term for RL. We therefore proceeded on the grounds of management interest, and used Model 3 for interpretation.

	Estimate	Std. error	*z* Value	Pr(>|z|)
(Intercept)	−8.54322	465.47010	−0.018	0.9854
SexM	−1.54243	0.62471	−2.469	0.0135*
N_other_boats	5.70421	465.45316	0.012	0.9902
RL_rms	0.02557	0.03153	0.811	0.4175

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

which means that a killer whale's response to the passage of a ship (using a severity score of ⩾3 as a cutoff), on average, was best explained by the whale's sex (with males less likely to respond than females), number of small vessels, and an increasing probability of response as received level (rms) increased. There is an equivocal case (i.e., little difference in terms of AIC or explanatory power) for dropping non-significant terms.

Fig. 2 shows the partial effect of noise, given mean values of all other terms in Model 1. Given no additional information, and ignoring all other sources of uncertainty, the best point estimate suggests that 50% of killer whales showed a response ⩾2 on the Southall severity scale at received levels of approximately 130 dB re 1 μPa rms.

The point at which half of whales showed a response ⩾3 on the Southall severity scale is likely to occur beyond the range of received levels observed in the study, i.e., >150 dB re 1 μPa rms. We do not use Model 2 or Model 3 for prediction, because the confidence intervals on RL_rms (when severity score 3 is used as the cutoff indicating a response) spanned the entire range from 0 to 1.

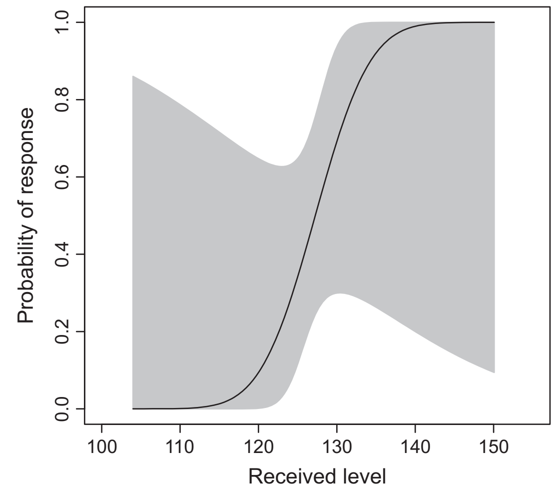

Fig. 2. The partial contribution of noise (Received level, RL_rms, in dB), as a predictor of the probability of a response ⩾2 (partial relationship in solid line; confidence intervals bounded by dark grey polygon) on the (Southall et al., 2007) severity score, conditional on mean values of all other terms in the model.

4. Discussion

Northern resident killer whales showed moderate (severity score 2–4) responses to the presence of the large ships that use Johnstone Strait in summer months, but behavioral responses were best explained by combinations of time (Year and Month), age of the animal, number of ships (CAR, COL and TUG) and the broadband noise level received by the whale (RL_rms) (Fig. 2). Evaluating the effects of ship traffic on killer whale behavior is overwhelmingly influenced by a somewhat subjective and seemingly arbitrary decision about the severity score that one uses to indicate a response. Using a cutoff of ⩾2 on the Southall severity scale, we find that whales had a 50% chance of responding to ship noise at broadband (10 Hz–50 kHz) received levels of ∼130 dB re 1 μPa root-mean-square (rms), but there is large uncertainty around that estimate (Fig. 2). Using a cutoff of ⩾3 on the Southall severity scale, we suspect that the point at which whales have 50% probability of responding to ship noise occurs beyond the range of received levels observed in our study: i.e., >150 dB re 1 μPa rms. Our models have very poor explanatory power for predicting more severe responses than those that would score a 2 on the Southall scale, because the range of traffic observed in our study never resulted in received levels higher than 150 dB, and because very few of the natural experiments we observed resulted in more severe (⩾4) behavioral responses (Appendix 2). More information is needed at both high and low received levels before one would have confidence in the shape of the dose–response curve when a threshold is set at ⩾3 on the Southall scale.

These rough estimates of sensitivity are not unexpected, given results from control-exposure studies showing subtle responses of killer whales to small vessels at received levels of 109–116 dB re 1 μPa rms (Williams et al., 2002a). Our analyses illustrate the need for a discussion about the point at which a behavioral response becomes sufficiently severe to be of conservation concern. Ultimately, it is the role of management or policy makers to decide the severity of behavioral responses that they would consider acceptable (Horowitz and Jasny, 2007). Our analyses suggest that this warrants explicit statement on the part of policy-makers, because a 130 versus 150 dB allowable harm limit would have quite different implications for real-world management. To put these thresholds in the context of real-world examples, there are many scenarios that would result in killer whales receiving a dose of 130 dB re 1 μPa (Appendix 2). This threshold can be reached from a cruise ship traveling 5.7 m/s at 700 m or a container ship traveling 5.2 m/s at 650 m. A behavioral response like the ones we describe is not in and of itself a conservation concern, but additional research is needed to model the cumulative impacts of repeated disturbance at the level of individual fitness or population dynamics.

The limitations of the study are evidenced by the wide confidence intervals shown in Fig. 1, especially at very high and very low received noise levels. Some of this uncertainty is no doubt due to real, natural variability in the whales' responsiveness to disturbance and the ecological context in which disturbance takes place (Ellison et al., 2012; Williams et al., 2006). However, lessons learned from experience elsewhere in inferring dose–response relationships to sonar and seismic surveys for many cetacean species (Miller et al., 2012) suggest that some of the variability could be reduced through increased sample size and various improvements to this study. We list proposed improvements below, in no particular order.

The dose–response curve is based on a derived parameter representing our best estimate of the noise level that the whale received. Although this is based on realistic proxy ship source levels and sound propagation models from peer-reviewed literature (Erbe et al., 2012), a dose–response curve would be improved by having better, empirical data on the actual received levels. We recently deployed 12 autonomous hydrophones in important whale habitats along the BC coast (Williams et al., 2013). It would be beneficial to conduct these control-exposure experiments while simultaneously capturing empirical data on the temporal variability in the soundscape.

The whale behavioral data are summarized over 5 min intervals, due to the temporal resolution of theodolite track data (i.e., the time of each surfacing). Telemetry data, such as DTAG deployments (Johnson and Tyack, 2003), would give finer resolution data. As the DTAG technology improves and expands to include dosimeters and calibrated hydrophones, these may give empirical values of received noise level simultaneously. Telemetry alone may not resolve this problem, though, because the flow of water over the acoustic tag may always confound our ability to measure received noise level at the whale.

In addition to the coarse resolution of the raw theodolite data, statistical limitations required us to reduce the five original behavioral variables to a single severity score, which was in turn reduced to a binary (response/no-response) categorical variable in order to conduct standard model selection exercises in a GLM framework. Investment in statistical methodological development (e.g., Bayesian methods under development for seismic and sonar; Dr. Len Thomas, University of St Andrews, pers. comm.) would allow us to extract additional information about response severity as a function of noise levels, rather than as a binary response.

Fitting a dose–response curve reliably may require a bigger sample size across a wider range of received levels (and age, sex, speed etc.) to better estimate the underlying shape and to tighten confidence intervals. Until then, we may be looking only at a relatively low and flat end of a dose–response curve. This may be particularly true because killer whales are somewhat used to noise, and because the whales have a lot of notice that the ship is coming. The ship noise will slowly increase as a ship passes, and it may be that dose–response curves will always show a better fit to sudden sounds like sonar or seismic surveys in which the sound source does not ramp up slowly. That said, the sample size in the current study is large, relative to more sophisticated and expensive control-exposure experiments on logistically challenging stressors like seismic surveys or military sonar (Miller et al., 2012, 2009). We see value in inexpensive studies like this one, especially because the land-based observation platform makes it possible to collect data under truly control (no-boat) conditions.

The response variable we measured represents current best practice in quantifying exposure and response of marine mammals to noise (Southall et al., 2007), but future studies may need to consider more ecologically relevant response variables. We did not measure vocal behavior of killer whales (echolocation or call rates, source levels etc.), and ultimately, one would want to test whether foraging efficiency or prey intake were affected by these noise levels (Williams et al., 2006). The metabolic cost of swimming in killer whales is fairly flat across the range of speeds observed in this study (Williams and Noren, 2009), so in general, these behavioral responses are expected to carry minor energetic costs in terms of increased energy expenditure, with two important caveats. First, the cost to females of having a calf swim in echelon formation is already high, at a time when lactating females may already be energetically stressed, so if female killer whales truly are more responsive than males to large ships (Model 3), then increasing their travel costs would be a conservation concern (Williams et al., 2011). Secondly, this study only looked at overt behavioral responses from surface observations. If ship noise is reducing prey acquisition through acoustic masking of echolocation signals (Clark et al., 2009), causing whales to abandon foraging opportunities (Williams et al., 2006), or by repelling fish (Slabbekoorn et al., 2010), this study would have no way of detecting those effects. The energetic cost of ship noise may be substantial in terms of reduced prey acquisition (through masking or disruption of feeding activities), even if the energetic cost of avoiding ships is relatively low. Similarly, we have not considered any physiological (i.e., hormonal) stress responses to ship noise, which have been shown to be important in other cetaceans (Rolland et al., 2012).

It is hoped that this threshold analysis can provide hypotheses to test on other datasets, such as telemetry data from DTAG deployments on killer whales around the world in the presence and absence of ships. Although the behavioral responses to ships that we documented in this study are subtle and minor, relative to some extreme responses of whales to some extreme levels of anthropogenic noise (e.g., (Fernandez et al., 2005; Jepson et al., 2003)), there are several reasons to keep ship noise on the conservation and management agenda for killer whales. In many parts of the industrialized world, ship noise is simply a more important contributor to the ocean soundscape than military sonar or seismic surveys (Croll et al., 2001; Hatch et al., 2008; McKenna et al., 2012). In critical habitat for southern resident killer whales, a large ship transits the area, on average, every hour of every day of every year, with three transits per hour observed at the busiest times (Erbe et al., 2012). There is evidence to suggest that northern and southern resident killer whales are already prey-limited, due to natural and anthropogenic stressors affecting the Chinook salmon that are the whales' preferred prey (Ford et al., 2010; Ward et al., 2009; Williams et al., 2011). If ship noise is masking (Bain and Dahlheim, 1994; Clark et al., 2009; Erbe, 2002) communication signals that killer whales use to find or share prey (Ford and Ellis, 2006), then the ubiquitous nature of global shipping traffic (Halpern et al., 2008) makes it worthwhile to evaluate whether ship noise could cause population-level consequences to whales that are already coping with multiple other natural and anthropogenic stressors. Finally, in practical terms, ship noise lends itself to mitigation much faster than the prey- and contaminant-related threats these killer whales are also facing (Leaper and Renilson, 2012).

Acknowledgements

The authors thank Christopher Clark, Phil Hammond, Patrick Miller, Brandon Southall and Len Thomas for feedback on various technical aspects of this analysis, and Marianne Gilbert, Dom Tollit, Jason Wood and an anonymous reviewer for helpful feedback on an earlier draft of the manuscript. RW collected the theodolite data with support from BC Parks and National Marine Fisheries Service, and he thanks SMRU Canada Ltd, Hemmera and Port Metro Vancouver for support for these analyses. This research was facilitated and motivated by a Marie Curie International Incoming Fellowship to RW within the 7th European Community Framework Programme (Project CONCEAL, FP7, PIIF-GA-2009-253407). CE and

RW acknowledge support from WWF-Canada for compiling ship traffic and source level data. EA and RW acknowledge ongoing support from Marisla Foundation for incorporating noise into conservation and management plans for killer whales and their critical habitat. The funders had no role in study design, data collection and analysis, decision to publish, or conclusions of the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.marpolbul.2013.12.004.

References

Aguilar Soto, N., Johnson, M., Madsen, P.T., Tyack, P.L., Bocconcelli, A., Fabrizio Borsani, J., 2006. Does intense ship noise disrupt foraging in deep-diving Cuvier's beaked whales (*Ziphius cavirostris*)? Mar. Mammal Sci. 22, 690–699.
Bain, D.E., Dahlheim, M.E., 1994. Effects of masking noise on detection thresholds of killer whales. In: Loughlin, T. (Ed.), Ma mammals and the *Exxon Valdez*. Academic Press, San Diego, pp. 243–256.
Breeding, J., Pflug, L., Bradley, M., Herbert, M., Wooten, M., 1994. RANDI 3.1 Naval Research Laboratory Technical Memorandum Naval Research Laboratory, Stennis Space Center, MS.
Burnham, K.P., Anderson, D.R., 2002. Model selection and multi-model inference: a practical information-theoretic approach. Springer Verlag.
Carnes, M.R., 2009. Description and Evaluation of GDEM-V 3.0. DTIC Document.
Castellote, M., Clark, C.W., Lammers, M.O., 2012. Acoustic and behavioural changes by fin whales (*Balaenoptera physalus*) in response to shipping and airgun noise. Biol. Conserv. 147, 115–122.
Clark, C.W., Ellison, W.T., Southall, B.L., Hatch, L., Van Parijs, S.M., Frankel, A., Ponirakis, D., 2009. Acoustic masking in marine ecosystems: intuitions, analysis, and implication. Mar. Ecol. Prog. Ser. 395, 201–222.
Collins, M.D., Cederberg, R.J., King, D.B., Chin-Bing, S.A., 1996. Comparison of algorithms for solving parabolic wave equations. J. Acoust. Soc. Am. 100, 178–182.
Croll, D.A., Clark, C.W., Calambokidis, J., Ellison, W.T., Tershy, B.R., 2001. Effect of anthropogenic low-frequency noise on the foraging ecology of *Balaenoptera* whales. Anim. Conserv. 4, 13–27.
Ellison, W., Southall, B., Clark, C., Frankel, A., 2012. A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds. Conserv. Biol. 26, 21–28.
Erbe, C., 2002. Underwater noise of whale-watching boats and its effects on killer whales (*Orcinus orca*). Mar. Mammal Sci. 18, 394–418.
Erbe, C., Farmer, D.M., 1998. Masked hearing thresholds of a beluga whales (*Delphinapterus leucas*) in icebreaker noise. Deep-Sea Res. Part II – Top. Stud Oceanogr 45, 1373–1388.
Erbe, C., Farmer, D.M., 2000. Zones of impact around icebreakers affecting beluga whales in the beaufort sea. J. Acoust. Soc. Am. 108, 1332.
Erbe, C., MacGillivray, A., Williams, R., 2012. Mapping cumulative noise from shipping to inform marine spatial planning. J. Acoust. Soc. Am. 132, EL423–EL428.
Faraway, J.J., 2005. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. CRC Press.
Fernandez, A., Edwards, J.F., Rodriguez, F., de los Monteros, A.E., Herraez, P., Castro, P., Jaber, J.R., Martin, V., Arbelo, M., 2005. "Gas and Fat Embolic Syndrome" Involving a Mass Stranding of Beaked Whales (Family Ziphiidae) Exposed to Anthropogenic Sonar Signals. Veterinary Pathology Online 42, 446-457.
Fisheries and Oceans Canada, 2011. Recovery Strategy for the Northern and Southern Resident Killer Whales (Orcinus orca) in Canada., In: Canada, F.a.O. (Ed.), Species at Risk Act Recovery Strategy Series. Fisheries and Oceans Canada, Ottawa, p. ix + 80 pp.
Foote, A.D., Osborne, R.W., Hoelzel, A.R., 2004. Environment: whale-call response to masking boat noise. Nature 428, 910–910.
Ford, J.K.B., Ellis, G.M., 2006. Selective foraging by fish-eating killer whales *Orcinus orca* in British Columbia. Mar. Ecol. Prog. Ser. 316, 185–199.
Ford, J.K.B., Ellis, G.M., Balcomb, K.C., 2000. Killer whales. In: 2nd ed. UBC Press, Vancouver.
Ford, J.K.B., Ellis, G.M., Olesiuk, P.F., Balcomb, K.C., 2010. Linking killer whale survival and prey abundance: food limitation in the oceans' apex predator? Biol Lett 6, 139–142.
Fox, J., 2005. Getting started with the R commander:a basic-statistics graphical user interface to R. J. Stat. Software 14, 1–42.
François, R.E., Garrison, G.R., 1982a. Sound absorption based on ocean measurements: Part I: Pure water and magnesium sulphate contributions. J. Acoust. Soc. Am. 72, 896–907.
François, R.E., Garrison, G.R., 1982b. Sound absorption based on ocean measurements: Part II: Boric acid contribution and equation for total absorption. J. Acoust. Soc. Am. 72, 1879–1890.
Hall, J.D., Johnson, C.S., 1972. Auditory thresholds of a killer whale *Orcinus orca* Linnaeus. J. Acoust. Soc. Am. 51, 515–517.
Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., 2008. A global map of human impact on marine ecosystems. Science 319, 948.
Hamilton, E.L., 1980. Geoacoustic modeling of the sea floor. J. Acoust. Soc. Am. 68, 1313–1340.
Hamson, R., 1997. The modelling of ambient noise due to shipping and wind sources in complex environments. Appl. Acoust. 51, 251–287.
Hatch, L., Clark, C., Merrick, R., Parijs, S., Ponirakis, D., Schwehr, K., Thompson, M., Wiley, D., 2008. Characterizing the relative contributions of large vessels to total ocean noise fields: a case study using the gerry e. studds stellwagen bank national marine sanctuary. Environ. Manag. 42, 735–752.
Hildebrand, J.A., 2009. Anthropogenic and natural sources of ambient noise in the ocean. Marine Ecology Progress Series, p. 395.
Holt, M.M., Noren, D.P., Veirs, V., Emmons, C.K., Veirs, S., 2008. Speaking up: Killer whales (Orcinus orca) increase their call amplitude in response to vessel noise. J. Acoust. Soc. Am. 125, EL27–EL32.
Horowitz, C., Jasny, M., 2007. Precautionary management of noise: lessons from the US Marine Mammal Protection Act. J. Int. Wildlife Law Policy 10, 225–232.
Jepson, P.D., Arbelo, M., Deaville, R., Patterson, I.A.P., Castro, P., Baker, J.R., Degollada, E., Ross, H.M., Herraez, P., Pocknell, A.M., Rodriguez, F., Howie, F.E., Espinosa, A., Reid, R.J., Jaber, J.R., Martin, V., Cunningham, A.A., Fernandez, A., 2003. Gas-bubble lesions in stranded cetaceans. Nature 425, 575–576.
Johnson, M.P., Tyack, P.L., 2003. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. Ocean. Eng., IEEE J. 28, 3–12.
Leaper, R., Renilson, M.R., 2012. A review of practical methods for reducing underwater noise pollution from large commercial vessels. Trans. RINA, 154.
Lusseau, D., Bain, D.E., Williams, R., Smith, J.C., 2009. Vessel traffic disrupts the foraging behavior of southern resident killer whales *Orcinus orca*. Endangered Species Res. 6, 211–221.
McKenna, M.F., Ross, D., Wiggins, S.M., Hildebrand, J.A., 2012. Underwater radiated noise from modern commercial ships. J. Acoust. Soc. Am. 131, 92–103.
Miller, P.J., Kvadsheim, P.H., Lam, F.-P.A., Wensveen, P.J., Antunes, R., Alves, A.C., Visser, F., Kleivane, L., Tyack, P.L., Sivle, L.D., 2012. The severity of behavioral changes observed during experimental exposures of killer (*Orcinus orca*), Long-Finned Pilot (*Globicephala melas*), and Sperm (*Physeter macrocephalus*) Whales to Naval Sonar. Aquat. Mammals 38, 362–401.
Miller, P.J.O., Johnson, M.P., Madsen, P.T., Biassoni, N., Quero, M., Tyack, P.L., 2009. Using at-sea experiments to study the effects of airguns on the foraging behavior of sperm whales in the Gulf of Mexico. Deep Sea Res. Part I: Oceanogr. Res. Pap. 56, 1168–1181.
National Marine Fisheries Service, 2008. Recovery plan for southern resident killer whales (*Orcinus orca*). National Marine Fisheries Service, Northwest Region, Seattle, WA, USA, pp. 251.
Nowacek, D.P., Johnson, M.P., Tyack, P.L., 2004. North Atlantic right whales (Eubalaena glacialis) ignore ships but respond to alerting stimuli. In: Proceedings of the Royal Society of London. Series B: Biological Sciences, vol. 271, pp. 227–231.
Nowacek, D.P., Thorne, L.H., Johnston, D.W., Tyack, P.L., 2007. Responses of cetaceans to anthropogenic noise. Mammal Rev. 37, 81–115.
Rolland, R.M., Parks, S.E., Hunt, K.E., Castellote, M., Corkeron, P.J., Nowacek, D.P., Wasser, S.K., Kraus, S.D., 2012. Evidence that ship noise increases stress in right whales. Proceedings of the Royal Society B: Biological Sciences.
Slabbekoorn, H., Bouton, N., van Opzeeland, I., Coers, A., ten Cate, C., Popper, A.N., 2010. A noisy spring: the impact of globally rising underwater sound levels on fish. Trends Ecol. Evol. 25, 419–427.
Southall, B., Bowles, A., Ellison, W., Finneran, J., Gentry, R., Greene, C., Kastak, D., Ketten, D., Miller, J., Nachtigall, P., 2007. Marine mammal noise exposure criteria: initial scientific recommendations. Aquat. Mammals 33, 411–521.
Szymanski, M.D., Bain, D.E., Kiehl, K., Pennington, S., Wong, S., Henry, K.R., 1999. Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms. J. Acoust. Soc. Am. 106, 1134–1141.
Ward, E.J., Holmes, E.E., Balcomb, K.C., 2009. Quantifying the effects of prey abundance on killer whale reproduction. J. Appl. Ecol. 46, 632–640.
Williams, R., Ashe, E., 2007. Killer whale evasive tactics vary with boat number. J. Zool. 272, 390–397.
Williams, R., Bain, D.E., Ford, J.K.B., Trites, A.W., 2002a. Behavioural responses of male killer whales to a 'leapfrogging' vessel. J. Cetacean Res. Manag. 4, 305–310.
Williams, R., Clark, C.W., Ponirakis, D., Ashe, E., 2013. Acoustic quality of critical habitats for three threatened whale populations. Anim. Conserv. http://dx.doi.org/10.1111/acv.12076.
Williams, R., Krkosek, M., Ashe, E., Branch, T.A., Clark, S., Hammond, P.S., Hoyt, E., Noren, D.P., Rosen, D., Winship, A., 2011. Competing conservation objectives for predators and prey: estimating killer whale prey requirements for Chinook salmon. PLoS ONE 6, e26738.
Williams, R., Lusseau, D., Hammond, P.S., 2006. Estimating relative energetic costs of human disturbance to killer whales (*Orcinus orca*). Biol. Conserv. 133, 301–311.
Williams, R., Noren, D.P., 2009. Swimming speed, respiration rate, and estimated cost of transport in adult killer whales. Mar. Mammal Sci. 25, 327–350.
Williams, R., Trites, A.W., Bain, D.E., 2002b. Behavioural responses of killer whales (*Orcinus orca*) to whale-watching boats: opportunistic observations and experimental approaches. J. Zool. 256, 255–270.
Wright, A.J., 2008. International Workshop on Shipping Noise and Marine Mammals, Hamburg, Germany, 21st-24th April 2008, In: Okeanos – Foundation for the Sea (Ed.), Hamburg, pp. 34.
Wright, A.J., Deak, T., Parsons, E., 2011. Size matters: management of stress responses and chronic stress in beaked whales and other marine mammals may require larger exclusion zones. Mar. Pollut. Bull. 63, 5–9.