Supplementary Material

# Supplementary Figures and Tables

## Supplementary Figures



**Supplementary Figure 1.** Regression fit to three empirical behaviour studies on southern and northern killer whales. The data plotted from each study is identified by symbol in the legend. The y-axis is a relative severity score (SS) of an observed whale’s behavioural response to noise of increasing broadband energy (Southall 2007). If no response was observed then this corresponded to SS=0. The width of the confidence intervals around the regression line inform the uncertainty around the dose-response function pictured in Figure 2 of the manuscript.

**Supplementary Methods**

This supplementary material describes the analyses that were required to generate acoustic behavioural dose-response functions. As well, we provide the pseudo-code details around the assumptions of the SRKW noise-exposure model.

To develop the behavioural response thresholds for two severity responses, ‘low’ and ‘moderate’ severity behavioural responses, the following data from three studies on resident killer whales in the vicinity of the Salish Sea.

Study 1: Theodolite

There were 35 incidences of northern resident killer whales (NRKW) exposed to commercial vessel traffic in Johnstone Strait, off northern Vancouver Island. The (Southall) severity scores of behaviour observed in the NRKW was matched to estimates of the received noise levels (RL) in the Appendix B of the Theodolite Study of Killer Whale Response to Ship Traffic (Williams et al. 2014).

Study 2: DTAG

Six RL and Southall severity scores were taken from Table 8 of the DTAG Study of NRKW Responses to Ship Traffic (Wright et al. 2017). The DTAG study used randomisation tests to determine when behaviours were statistically different than baseline behaviours. The percent change in behaviour was measured, and a single DTAG deployment (oo11\_245a) resulted in a 100% change in rate of social sound production that was assumed to equate to a Southall severity score of 6.

Study 3: PAM

The Passive Acoustic Monitoring Study of Killer Whale Responses to Ship Traffic found clear evidence of the Lombard effect, whereby SRKW increase the source level (SL) of their calls as RL increases. Southall et al. (2007) suggest that the Lombard effect, with duration less than the exposure sound, should be classified as a behavioural response severity score of 3. Southall et al. (2007) created their severity scores in order to rank the severity of behavioural responses to anthropogenic noise, however, the evolution of the Lombard effect long pre-dates anthropogenic noise (Brumm and Zollinger 2011) and SRKW clearly exhibit the Lombard effect to both anthropogenic and non-anthropogenic noise. To use the Lombard results (a continuous response) from the PAM study to generate combined thresholds in this study, a threshold (a single discreet point) was selected for deciding when the Lombard effect has occcurred in response to anthropogenic noise.

Wood et al. (2013) provided cumulative probability plots of sound levels at Lime Kiln split into four sound types; ships, boats, depth sounders, and ‘remainder’ (which is assumed to be mostly non-anthropogenic sounds). This study was focused on setting behavioural response thresholds to commercial vessel noise, therefore cumulative probability curve from ships was used to determine a conservative threshold above which an anthropogenic Lombard effect is deemed to occur. This study chose the 10% cumulative probability of ship noise level of 110 dB re 1µPa as 90% of ship noise is higher amplitude than this level at Lime Kiln. This threshold of 110 dB re 1µPa was deemed to provide a conservative estimate of the cut-off between ship noise driven Lombard effects and non-anthropogenic noise Lombard effects.

Once the ship-related noise threshold above which a Lombard effect was determined to occur, it was necessary to develop a rationale to convert the Lombard effect to Southall severity scores. In other words, RL thresholds were needed that would result in specific Southall severity scores. This was done for Southall severity scores 3 through 6, as 3 is the lowest severity score that Southall et al. (2007) considered for a Lombard effect and 6 is the highest Southall severity response to ships recorded in the theodolite and DTAG studies on NRKW. Like the theodolite and DTAG studies, attributing a Southall severity score to a change in behaviour requires setting thresholds for change that equate to specific Southall severity scores. This was done for the PAM data using similar percent change thresholds that were used in the theodolite study for Southall severity 3 to 5. When a 100% change in rate of social sound production was observed, this was assumed to equate to a Southall severity score of 6. Using these percent changes in SRKW RL above the Lombard effect ship threshold (110 dB re 1µPa), the RL thresholds corresponding to each Southall severity score were calculated.

Creating Combined Thresholds from All Three Studies

The RL and resulting Southall severity scores from the three studies are consolidated in Supplementary Figure 1. There is a large amount of variation in correlating these 45 individual estimates of RL, and even a large amount of variability is apparent within each Southall severity score. This variation may be due to the RL being only one of several variables that may contribute to a behavioural response, and these reasons are explored in our discussion, but we acknowledge here that the behaviour state and context of the animal exposed to noise, as well as its location in relation to the noise, and its previous experience with that noise may all influence whether a response occurs and its severity (Ellison et al. 2012).

Two thresholds of behavioural response were developed and correspond to a low severity and moderate severity response. Behavioural responses were categorised as either ‘low’ or ‘moderate’, corresponding to Southall severity scores of 2-3 and 4-6 (Southall et al. (2007). Low behavioural responses are relatively minor and brief, while moderate behavioural responses have a higher likelihood of affecting foraging, reproduction and subsequent long-term vital rates of individual whales.

While the regression line in Supplementary Figure 1 could be used to predict the average RL for low and moderate behavioural response (129 dB re 1 µPa for low and 137 dB re 1 µPa for moderate) and therefore as a single threshold for estimates of SRKW noise overlap and population modelling, this would ignore the large variance in RL seen in these studies. Dose-response curves plot the relationship between the probabilities of an effect occurring at a given level of noise exposure. Dose-response curves have also been applied to estimates of marine mammal injury and behavioural response due to anthropogenic noise, specifically by the US Navy (US Navy 2008, 2012). This study uses the dose-response formula defined by the US Navy (2008) but uses the inputs derived from these three SRKW combined thresholds, and . The US Navy (2008) dose-response formula is as follows:

where: R = risk from 0 to 1 (i.e. probability of a behavioural response)

 L = Received Level (RL) in dB

 B = basement RL in dB

 K = the RL increment above basement in dB at which there is 50% risk

 A = risk transition sharpness parameter

The basement RL (B) is the RL below which likely risk is so low that it does not warrant calculation. The US Navy (2008) dose-response curve was based on large behavioural responses (i.e. Southall severity scores ≥6) to Navy sonar and B was set at 120 dB re 1 µPa. Given the quieter noise sources, lower levels of behavioural response being modelled in this study, and that a behavioural response was documented at 114 dB re 1 µPa, in the Theodolite study, B was set at 90 and 100 dB re 1 µPa for low and moderate behavioural response respectively, and it is assumed that the probability of a low or moderate response is zero below this level.

Based on the regression in Supplementary Figure 1, the 50% probability (and 95% confidence intervals; 95%CI) for low (Southall severity 2.5) and moderate (Southall severity 5) responses was set at RLs of 129.5 (95%CI: 126.9, 132) and 137.2 (95%CI: 133.5, 140.9) dB re 1 µPa, respectively. Consequently, K was set at 39 (129.5-90=39.5) and 37 (137.2-100=137.2) for low and moderate dose-response curves.

The US Navy (2008) used risk transition sharpness parameters (A) of 10 for odontocetes and pinnipeds, and 8 for mysticetes. A dose-response curve with A=8 has a lower slope than a curve with A=10. The lower slope was used for mysticetes in recognition that animals were responding at RL below the 50% risk level. In a sensitivity analysis in our context, the slope parameter (A) was increased to 10 and decreased to 6 to compare with A=8 used in this study. A decrease in ‘A’ generates higher estimates of behavioural response at lower RL, but decreases the estimates of the number of animals responding at higher RL. An increase in ‘A’ does the opposite. Given the large variance in RL of killer whale behavioural responses highlighted in the current study and to be more precautionary, a more conservative A=8 slope was adopted for the SRKW dose-response curves in this study.

# Pseudo-code for Southern Resident Killer Whale noise-exposure model

We now proceed by providing the pseudo-code for incorporating the dose response function into the SRKW Noise Exposure Model, and how a behavioural response to RL translates into ‘potential lost foraging time’. The numbers in the following pseudo-code correspond to the numbers in Figure 3 of the manuscript.

Estimate presence and distribution of SRKW pods in the Slowdown region. Each of the three SRKW pods (J, K and L) was determined present or absent in the Slowdown study region based on monthly probabilities determined from an 11 year dataset (2001-2011) of effort corrected sightings maintained by the Vancouver Aquarium (BC Cetacean Sighting Network), and The Whale Museum (OrcaMaster; dataset described in Olsen et al. 2018). If the pod was present, the location of the 78 individuals (based on January 1, 2017 population estimates from the Center for Whale Research) belonging to that pod were distributed as per relative density predictions by pod and month (August through October) probability of occurrence. The number of unique centroids was determined based on the probability of pods being sighted with one another vs on their own. The centroids are then generated from a multivariate copula distribution based on the marginal distributions of each pod.

Determine distribution of individual SRKW at predicted pod locations. Individual whales in each pod were distributed over a 5 km radius according to the SRKW spatial density within the slowdown trial study area, and assigned to the appropriate 200 m grid cell.

Determination of received broadband Sound Pressure Level and Power Spectral Density at 50 kHz. This step involves matching the whale‘s location to correct grid of noise data. The predictive regional acoustic model developed by JASCO Applied Sciences estimates on a scale of minutes a) the broadband received Sound Pressure Levels (SPLs) and b) the Power Spectrum Density (PSDs) at 50 kHz. These are condensed to sequential 5-minute increments of noise maxima.

Model dual behavioural responses based on SPL dose. The number of low and moderate severity behavioural responses (BRs) to the broadband regional acoustic model for each day of the trial was counted based on SRKW-specific dose-response relationships discussed above. A low severity responses was assumed to last 5 minutes, or the time it takes for a ship travelling at 18 knots to traverse a circle of 1.4 km radius. Moderate severity responses were assumed to last 5 times as long with a duration of 25 minutes during which no low severity BRs were permitted. Therefore, although moderate severity BRs had a lower chance of occurring, their net effect on ‘potential lost foraging time’ is greater than low severity BRs.

Model residual high frequency click masking based on 50 kHz PSD. At lower broadband noise levels, when no behavioural response was predicted, the model estimated the degree of additional or residual high frequency masking using a precautionary maximum click detection range (threshold) of 250 m and calculating a proportional detection distance range reduction due to 50 kHz noise levels. A 1-dimensional loss function was used to simply translate this proportional loss into proportion of minutes within each 5-minute increment that residual click masking occurred.

Accumulate BRs and click masking over a twenty-four hour period. The above process was repeated for each 5-minute time increment of the day, and the number of low and moderate behavioural responses and degree of masking were accumulated (summed) for each individual whale across the 288 5-minute periods per day. ‘Potential lost foraging time’ was accumulated by summing all 25-minute moderate severity BRs, all 5-minute low severity BRs and all residual click masking minutes.

Accumulate BRs and click masking over the vessel slowdown trial period. The twenty-four hour period totals were accumulated for slowdown trial period (August 7th to October 6th) for each individual whale assigned to each of pod J, K, and L. A number of metrics could then be derived to summarize ‘time lost’ including time lost per individual over the slowdown period, per pod, or across all SRKWs.

Calculate 95% confidence intervals. The entire model simulation was run 500 times to generate the 95% quantiles or confidence intervals (CIs) for all outputs of the noise-exposure simulation.

Re-run steps 1 through 8 of this model for each of the six model scenarios in Table 2 for details of the traffic, vessel speed and participation level of each scenario.

The following table shows how vessel type codes from the Marine Traffic AIS dataset (left) were assigned to the vessel categories in the regional vessel noise model (right). Sailing vessels were excluded from the Recreational vessel category and were not included in the model (i.e., they were assumed not to be under power).

|  |  |
| --- | --- |
| MarineTraffic AIS Vessel Type | Noise Model Category |
| Cargo/Containership | Containership |
| Container Ship | Containership |
| Ro-Ro/Passenger Ship | Ferry (Ro-ro Passenger) |
| Factory Trawler | Fishing |
| Fish Carrier | Fishing |
| Fish Factory | Fishing |
| Fishing | Fishing |
| Fishing Vessel | Fishing |
| Trawler | Fishing |
| Buoy-Laying Vessel | Government |
| Fishery Patrol Vessel | Government |
| Fishery Research Vessel | Government |
| Law Enforce | Government |
| Logistics Naval Vessel | Government |
| Military Ops | Government |
| Patrol Vessel | Government |
| Replenishment Vessel | Government |
| Research/Survey Vessel | Government |
| Bulk Carrier | Bulker/Gen. Cargo |
| Cargo | Bulker/Gen. Cargo |
| Cargo - Hazard A (Major) | Bulker/Gen. Cargo |
| Chemical Tanker | Bulker/Gen. Cargo |
| General Cargo | Bulker/Gen. Cargo |
| LPG Tanker | Bulker/Gen. Cargo |
| Rail/Vehicles Carrier | Bulker/Gen. Cargo |
| Reefer | Bulker/Gen. Cargo |
| Self Discharging Bulk Carrier | Bulker/Gen. Cargo |
| Timber Carrier | Bulker/Gen. Cargo |
| Wood Chips Carrier | Bulker/Gen. Cargo |
| Anti-Pollution | Other |
| Cable Layer | Other |
| Dive Vessel | Other |
| Drill Ship | Other |
| Heavy Lift Vessel | Other |
| High Speed Craft | Other |
| Hopper Dredger | Other |
| Local Vessel | Other |
| Other | Other |
| Pilot Vessel | Other |
| Port Tender | Other |
| Reserved | Other |
| SAR | Other |
| Tender | Other |
| Unspecified | Other |
| Wing In Grnd | Other |
| Passenger | Passenger (under 100 m), Cruise (over 100 m) |
| Passengers Ship | Passenger (under 100 m), Cruise (over 100 m) |
| Pleasure Craft | Recreational |
| Yacht | Recreational |
| Crude Oil Tanker | Tanker |
| Oil Products Tanker | Tanker |
| Oil/Chemical Tanker | Tanker |
| Tanker | Tanker |
| Anchor Handling Vessel | Tug |
| Fire Fighting Vessel | Tug |
| Multi Purpose Offshore Vessel | Tug |
| Offshore Supply Ship | Tug |
| Pollution Control Vessel | Tug |
| Pusher Tug | Tug |
| Towing Vessel | Tug |
| Tug | Tug |
| Vehicles Carrier | Vehicles Carrier |
| Ro-Ro/Container Carrier | Vehicles Carrier or Ferry (Ro-ro Cargo)\* |
| Ro-Ro Cargo | Vehicles Carrier or Ferry (Ro-ro Cargo)\* |

\* Ferry (Ro-ro Cargo) vessels in this category were manually assigned based on MMSI.