

NORTHERN RESIDENT KILLER WHALE RESPONSES TO VESSELS VARIED WITH NUMBER OF BOATS

Final report to NOAA Fisheries for Contract AB133F04SE0736

Rob Williams¹ and Erin Ashe²

¹ Pearse Island, Box 193 Alert Bay BC V0N 1A0 Canada

² 2103 N 54th St. Seattle, WA 98103 USA

21 January 2006

Abstract

Vessel traffic has been implicated as a potential contributing factor to the at-risk status of two killer whale populations in western Canada and the US. Whalewatching guidelines can help mitigate this potential threat, especially when these are developed using experimental impact assessments that allow animal response to inform vessel management. Two published experimental studies on one of these populations documented stereotyped avoidance responses. Opportunistic observations in these studies in the mid-1990s suggested an inflection point in avoidance behavior when approximately 3 boats approached whales to within 1000m. Our experiment was designed to test whether whales responded differently to approach by few (1-3) versus many (>3) vessels. Data were collected in summer 2004, in Johnstone Strait, British Columbia (BC), Canada, using a theodolite to track positions of boats and individually identifiable whales. Experimental trials included 20-minute “no boat” and 20-minute “boat” phases (with local whalewatching vessels volunteering to act as experimental treatments), during which data were collected continuously on the focal whale. Responses of the 16 adult male killer whales tracked differed significantly between treatment levels (Wilcoxon’s test $P=0.0148$). Swimming path became less direct when approached by few boats, and whales increased directness when approached by many boats. Pooling both treatments, reducing the comparison to presence vs. absence, would have masked these significant responses with strong statistical confidence (Wilcoxon’s test $P>0.999$). Consistent with previous experiments, inter-breath interval, swimming speed, angle between successive dives, and rate of surface active behavior did not differ significantly. This apparent distinction between “few” and “many” boats, though, was supported by 140 opportunistic observations on 26 whales from the population of 216. Generalized Additive Models were used to control for effects of potentially confounding variables, and confirmed a non-linear relationship between the number of boats approaching within 1000m and a whales’ swimming path directness, with an inflection point around 3 boats. One objective measure of this inflection point (refitting the GAMs, but placing knots manually from 1-16 boats, and minimizing AIC score) showed that whale behavior was best described as a non-linear function of boat number, with a knot placed at three boats, which suggests that the experimental track treatment level of few (1-3) versus many (>3) boats was appropriate. The avoidance response by northern resident killer whales differed when there were few versus many boats, and this result has important implications for the design and interpretation of other vessel impact assessments. Pooling all traffic conditions into categories of absence versus presence of boats could allow researchers to falsely conclude that there was no impact of boat traffic on whale behavior when, in fact, there was more than one response. We recommend that presence/absence analyses be supplemented by more detailed analyses when possible.

Introduction

A growing number of studies examine the impacts of human activity on the behavior, energetics and reproductive success of a variety of wildlife taxa. While short-term responses of animals to anthropogenic actions are relatively easy to demonstrate, it has proven conceptually difficult to identify whether these responses could impact fitness of targeted animals or populations.

Despite the difficulty in conducting impact assessments on free-ranging marine species, a number of studies have quantified the impacts of vessel traffic on cetacean behavior (e.g., Kruse 1991, Lusseau In Press, Williams *et al.* 2002ab, Scheidat *et al.* 2004), activity and energetic budgets (Williams 2003), and possibly reproductive success (Bejder 2005). One feature common

to these studies is that they report small, subtle behavioral cues that cetaceans displayed when approached by boats. The studies were conducted from land, or used sophisticated analysis techniques to account for the effect of the research platform on the effects of additional boats that they were trying to measure. Taken as a whole, the literature in this field presents overwhelming evidence that at least behavioral responses of whales to boats are generally small, but real.

In Johnstone Strait, British Columbia (BC), Canada, the research and whalewatch communities work closely together to draft common-sense guidelines, to test their efficacy experimentally, and update them in an iterative process. Commercial whalewatch operators recommended that boaters approach whales slowly, from the side, and to parallel the animals' path at distances no closer than 100m (JSKWC 1996). In 1995, the whalewatch community began cooperating with land-based researchers to provide experimental control periods to quantify the effects of their industry on whales.

One experimental study that was conducted in direct response to the relevance of the 100m guideline began in 1995 (Williams *et al.* 2002a). Erratic swimming paths during experimental approaches were interpreted using a predator-prey analogy, where the killer whale (prey) evades an approaching predator (boat) using a simple turning gambit (Williams 1999). Subtle adoption of a relatively irregular path is consistent with a "horizontal avoidance" tactic. Northern resident killer whales, on average, increased their distance to boats by surfacing unpredictably. Opportunistic observations revealed that the magnitude of this horizontal avoidance pattern increased as boats approached closer than 100m.

The next experimental study in this sequence measured how whales responded to one boat violating local whalewatching guidelines by 'leapfrogging' – this activity involves speeding up and placing one's boat in the predicted path of the whale. It generally results in a less than 100m point of closest approach which is in violation of the 100m guideline. That study reported that whales responded by varying the directness of their swimming paths (Williams *et al.* 2002b) in the same way that they responded to one boat following whalewatching guidelines (Williams *et al.* 2002a).

The consistency between these experimental and opportunistic studies is important. A body of evidence is mounting to suggest that northern resident killer whales, on average, display stereotyped responses to evade boats. Summary statistics on whale swimming paths can be considered a response variable that may be used as a proxy representing a whale's tolerance for a variety of vessel activities.

The opportunistic component of the first study (Williams *et al.* 2002a) revealed an apparent discrepancy between whale response to boat number and proximity. The findings suggested that killer whales responded differently to a few boats than to many. The authors speculated that:

"An irregular path may be a useful avoidance tactic with a single boat, but ineffective with more than one. In a multiple-vessel scenario, a dive that takes a whale farther from one boat may bring it closer to another. ... This compromise deserves further attention in the form of a multiple-vessel experiment."

We decided to test whether whales responded differently to experimental approach by few boats than many.

Both populations of resident, fish-eating killer whales are listed under Canada's Species at Risk Act (SARA) and vessel traffic has been implicated in the current status of both killer whale populations (Baird 2001). Canada is developing new legislation for viewing marine with the stated goal: "To set in place a contemporary management framework designed to address the non-consumptive use and protection of marine mammals in Canada. Current management proposals include: approach distances; vessel operation and vessel numbers around marine mammals."

For killer whales, some of these proposals carry a sound scientific rationale. However, the hypothesis that whales showed different responses to few boats approaching closely than to many boats (Williams *et al.* 2002a) has not yet been tested experimentally. Here, we report on the results of the first experiments to test how killer whales reacted to experimental approach by varying numbers of boats.

Boats are of particular concern for the depleted status of the southern resident (SRKW) population (Baird 2001), but assessing the role of boat traffic on SRKWs is problematic, primarily because there are too many boats to allow much experimental control over studies¹. Efforts to control traffic via time-area closures along the west side of San Juan Island during the 2003 and 2004 field seasons had limited success and primarily provide relatively brief “no-boat” situations only during morning hours. In addition, the unpredictability of recreational boaters unaware of the voluntary “no-go” zone significantly limits opportunities for controlled experiments with the southern resident population. We believe that working with the allopatric northern resident (NRKW) population allowed us to conduct an experimental impact assessment by proxy.

The primary goal of this study, then, was to test whether resident killer whales responded differently to experimental approach by few boats than to many boats. A loess curve fitted to scatterplots of opportunistic data (collected in 1995 and 1996) that correlated path directness with boat number (Williams 2002a) suggested that whale behavior changed when more than two or three boats approached whales within 1000m. We used that observation to define a ‘node’ separating what appeared to look like ‘few’ versus ‘many’ boats to a killer whale in the summer months of 1995-6. The apparent discontinuity in whale behavior at two to three boats was used to design our current experiment. Our secondary goal was to describe whale behavior opportunistically – across a wider range of traffic conditions, accounting for potentially confounding effects, and with a larger sample of individuals than could be obtained practically using experimental approaches.

Methods

Field methods

Data were collected from 1 July through 11 September 2004, from a land-based cliff observation point on the south shore of West Cracroft Island in Johnstone Strait (50°30'N, 126°30'W; Figure 1). The cliff observation point offered an unobstructed view of the Robson Bight (Michael Bigg) Ecological Reserve (RBMBER, which is largely free from whale-oriented vessel traffic) and the waters immediately adjacent to the Reserve where whalewatching vessels often congregate, while waiting for whales to leave the Reserve.

Northern resident killer whales use this deep, narrow channel heavily and reliably. Data on killer whale and boat positions were collected using an electronic theodolite (Pentax ETH-10D with a precision of a 10” arc) connected to a laptop computer equipped with custom software (Theoprog, available from Dr. D.E. Bain, dbain@u.washington.edu). The observation cliff height was determined by stretching a rope of known length at the water’s edge on the beach below the cliff and obtaining horizontal and vertical angle coordinates for both ends using the theodolite (Williams 1999). A detailed description of cliff height measurement procedure can be found in Appendix 2. Height was calculated using trigonometric relationships described by Davis *et al.* (1981) and Würsig *et al.* (1991).

¹ Fisheries and Oceans Canada noted that “[i]n 2001, the M3 program observed an average of 18 vessels (commercial and private) around whales at any time in the Victoria/Haro Strait area from dawn to dusk” and that “[u]p to 50 vessels actively viewing whales have been observed in the Victoria/Haro Strait area at any one time. (DFO 2002)”

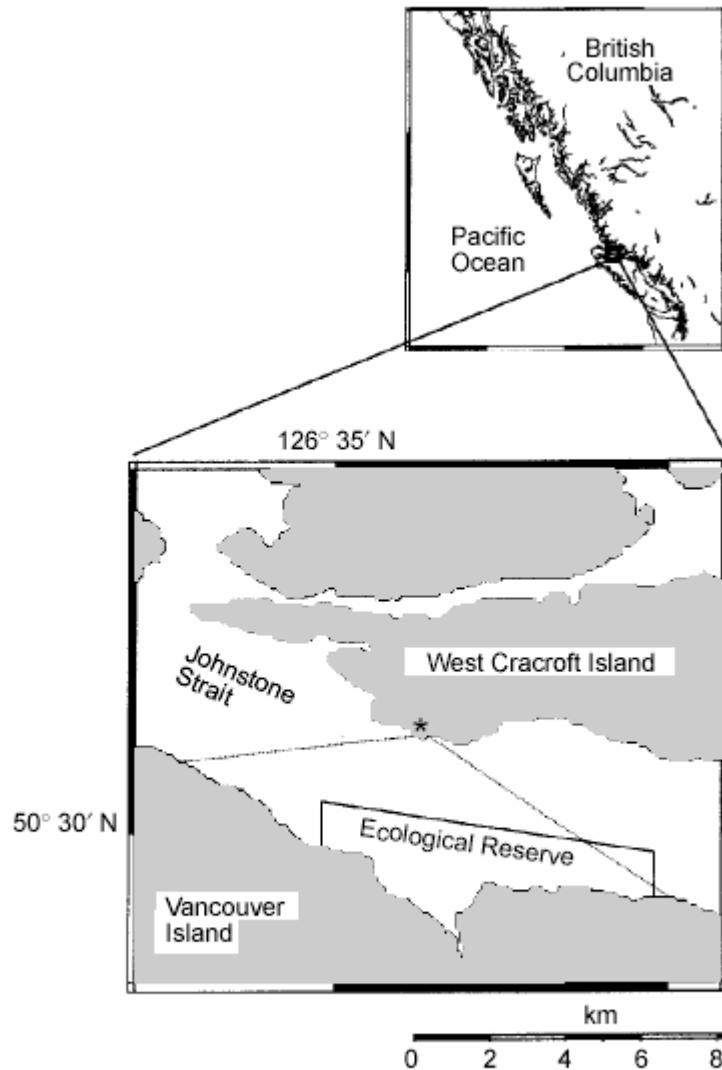


Fig. 1. Study area in Johnstone Strait, BC, Canada, showing lines of sight (- - -), position of theodolite (*), and boundaries of Robson Bight–Michael Bigg Ecological Reserve.

Choosing focal animals:

As whales entered the field of view from a study site, individuals were identified from a combination of factors. Focal animals were identified using careful examination of natural markings from long-term photo-identification (Bigg *et al.* 1990, Hammond *et al.* 1990, Ford *et al.* 2000) studies.

Advanced warning from local researchers regarding matriline identity allowed selection of focal animals to be drawn as evenly as practicable from all matriline. Males were preferentially chosen for experimental observation, because we knew from previous experience that sex-based difference in avoidance responses are possible (Williams *et al.* 2002a), and that males are easier to identify. For opportunistic observations, individuals were chosen to provide as representative a sample as practicable of all matriline that use the area in summer.

Focal animals were chosen for experimental trials based on: ease of repeat identification; position mid-strait that would allow continuous observation for at least 40 minutes and the greatest distance from boats. Statistical independence in the experimental component of the study is addressed by using each whale as its own control. Vessel presence or absence played no role in choosing focal animals for opportunistic observations.

Measuring behavior of focal animals and activity of vessel:

Once an individual was chosen as the focal whale, the position of the focal whale was 'marked' using the theodolite at each surfacing (*i.e.*, the time and position of each breath was recorded), and the coordinates simultaneously and automatically entered into a laptop computer attached to the theodolite. The theodolite operator recorded vessel positions as frequently as possible, giving each boat a unique identifier to plot its trajectory with respect to that of the focal animal. This computer-theodolite apparatus reduced the likelihood of transcription errors, and allowed real-time estimation of distances between boats and whales.

In addition to recording positions of boats and whales, Theoprog was used to record activity states, behavioral events (*e.g.*, respirations and surface active behaviors such as breaches or tail-slaps) and other notes such as sea state, weather condition, and general comments on boat type and activity (Williams *et al.* 2002ab). Each record of at least 20-minutes of observation of a focal animal may be referred to subsequently as a sample or a track. Boat and whale data were summarized for each track, such that each was represented only once in the analyses. At each surfacing, the total number of boats within 100, 400 and 1000m was counted, and then, the maximum, inclusive value for each of these radii was recorded for each track. These are referred to subsequently as MAX100, MAX400 and MAX1000. Similarly, the minimum distance ever observed between the focal whale and any boat was recorded. This is referred to subsequently as the point of closest approach (**PCA**). Together, these allowed compilation for each track, a parameter for each of a suite of *candidate* covariates or predictors.

The five dependent (*i.e.*, whale response) variables were calculated using the same methods used in earlier studies on this population (Williams *et al.* 2002ab) and an ongoing NMFS-funded study on southern resident killer whales led by David Bain. The dependent variables included:

1. Inter-breath interval [**RESP**]: A mean time between breaths was calculated (in seconds) for each track. If an animal breathed twice in a row at the same location, each breath was recorded, however this was exceedingly rare, so we also refer to the average inter-breath interval as the average dive time.
2. Swimming Speed [**SPEED**]: The average swimming speed of the whale during the track was obtained by dividing the total distance traveled by the duration of the tracking session and reported in km/h. Note that this represents total surface distance covered over time, rather than the crow's flight, or progressive, distance.

Two measures of path predictability were calculated: a *directness index* and a *deviation index* (Figure 2).

3. Directness Index [**DI**]: The directness index measures how straight a path is, and is a proxy for path predictability on the scale of a tracking session. It is generated by dividing the distance between end-points of a path (*i.e.*, crow's flight distance) by the cumulative surface distance covered during all dives. The directness index can be thought of as the ratio of the diameter of a path to its perimeter, and ranges from zero (a circular path) to 100 (a straight line). For some analyses, this index is presented as a proportion (*i.e.*, divided by 100, such that it ranges from 0 to 1).
4. Deviation Index [**DEV**]: The deviation index is a proxy for path predictability from one surfacing to the next. It is the mean of all angles between adjacent dives, and can be

considered an inverse measure of a path's smoothness – *i.e.*, a deviation of 0 is a straight line, and higher numbers represent more jagged paths. For each surfacing in a track, we calculated the angle between the path taken by a dive and the straight-line path predicted by the dive before it. If an animal breathed twice in a row at the same location, the direction of travel was undefined. However, we replaced this undefined value with 0 representing the change in direction for the purpose of calculating average deviation. The deviation index is the mean of the absolute value of each of these discrepancies, in degrees (potentially ranging from 0 to 180), during the entire track.

5. **Surface-active Behavior [SAB]:** We recorded each time that surface-active events such as spy-hopping, tail-slapping or breaching occurred. Note that we used the time of each exhalation observed to define a surfacing, and scored a bout of surface-active behavior as one event if, for example, more than one tail-slap or pectoral fin-slap occurred on a surfacing. However, if an animal breached or spy-hopped repeatedly, each breach or spy-hop would be scored as a bout, because a whale was assumed to breathe each time it breached or spy-hopped – *i.e.*, we used each surfacing, each time the animal's blowhole cleared the surface of the water as an independent event. We report the result as the rate of surface-active events per hour in order to standardize this measure in order to account for unequal track duration.

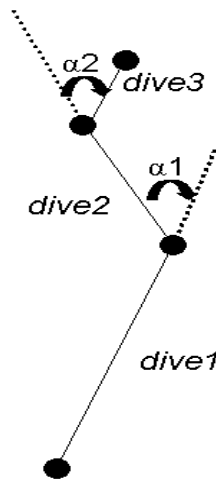


Figure 2. A segment of a hypothetical track, with four surfacings (•), and therefore three dives (–). The directness index is calculated by dividing the “crow’s flight” distance covered during the track (*i.e.*, from the first to the fourth surfacing) by the cumulative distance covered (*i.e.*, the sum of the lengths of dives 1, 2 and 3). Each angle (α) represents the difference (in degrees) between the position of a surfacing and that predicted (····) from the preceding one. The deviation index is the mean of these angular deviations from all pairs of succeeding and preceding dives.

Calculation of these candidate explanatory variables is described in greater detail in previous work (Williams 1999, Williams *et al.* 2002ab, Williams 2003). However, note that the vessel traffic variables represent minima or maxima for each observation (*e.g.*, point of closest approach, maximum number of boats within 100m, etc.). Response variables, on the other hand, represent average whale behavior observed during the tracking session.

Experimental trials

The local whalewatching community had agreed prior to the beginning of the whalewatch season to participate in experiments. For a detailed description of the experimental tracks refer to Appendix 3. At the outset of an experimental treatment, we requested their cooperation avoiding the focal whale for approximately 20 minutes, and then asked the whalewatching vessel to approach the focal whale for a typical whalewatch for 20 minutes. No attempt was made to direct

traffic other than commercial whalewatch boats during the experimental trial. However, all vessel traffic within 1000m of the focal was recorded by the theodolite at least twice. Boat traffic variables were summarized for every experimental trial for *post-hoc* categorization of treatment level. Each experimental trial was divided into two components, a control and a treatment phase. The control section, in all cases, referred to the data collected over a continuous 20-minute period during which no boat approached the focal whale within 1000m. The treatment section referred to the 20-minute minimum period immediately following this control period in which at least one primary experimental vessel approached within 1000m of the focal for the duration of the control treatment. Each animal served as its own control and the control section always preceded the treatment section.

Trials were classified *post-hoc* into one of two treatment levels based on the number of boats that actually approached during the experimental section: either *few* or *many*. We then compared the response using the observed distribution of the number of boats approaching to define 'few' and 'many' *a posteriori*. Recall that our hypothesis was that a node existed that between which whale response to boats differed. We suspected that that node existed at approximately 2-3 boats (Williams *et al.* 2002a), but we had limited opportunities to collect sufficient data to allow a multi-level treatment that would assess whether such a node was placed more precisely on average at two or three or some other number. While a balanced sample size is preferred, the opportunities to conduct experiments on free-ranging whales at all were rare due to commercial fishing openings, cruise ship traffic, and other recreational and commercial vessels transiting Johnstone Strait. Similarly, while local commercial whalewatchers were remarkably cooperative with this study, they were also engaged primarily in other activities, had additional schedules to follow, and preferentially watched groups of whales in locations far from the no-entry Reserve when given the option. Therefore, while we could control to some extent whether a boat approached, we had no control over how many boats approached.

Opportunistic observations

On most occasions, behavioral data were collected under conditions that offered no ability to manipulate traffic around the focal whale. We called these "opportunistic observations." In such cases, whales either entered the study area already accompanied by boats, or were observed in the absence of boats on occasions when no boat was available to serve as the experimental treatment. During these occasions, focal animals were selected and whale behavior was measured in the same way as it was during experimental trials. Boat activity was also measured using the same methods as described for the experimental trials.

Analysis of experimental trials with two treatment levels: few vs. many boats

We chose to test for evidence of significant behavioral responses of whales to our experimental treatments using non-parametric tests. For each trial, we calculated the whale's *response* to experimental approach. The response data were then placed in two categories: a sample of responses of whales exposed to approach by few (1-3) boats, and a second sample of responses of whales to many (>3) boats approaching. We used, the Mann-Whitney U test (Zar 1996) to assess whether the treatment variable influenced the ranking of each observation of whale behavioral response more than one would expect by chance alone.

Analysis of experimental trials with one treatment level: boat presence

As a secondary check of our distinction between 'few' and 'many' boats, we reran the analyses but ignored this grouping variable, thereby pooling data from all trials. For each trial, we contrasted the behavior of the focal whale during the control phase with its subsequent behavior, irrespective of the number of boats that approached. If these control data and treatment data were normally distributed and had equal variance, a paired t-test could be used to ask whether the average response to boats differed significantly from zero. We chose to use the non-parametric equivalent of the paired t-test, the Wilcoxon matched-pairs signed-ranks test.

Analysis of opportunistic data

The advantages of our experimental approach to addressing our primary research question were manifold: each whale served as its own control, which eliminated most of the potentially confounding variables, such as individual, age, sex, month and time of day. A shortcoming of our approach was logistical difficulty, and thus restricted sample size. Inclusion of opportunistic data was chosen in order to increase sample size, and to allow us to assess whether the results predicted from the experimental trials could be observed in larger datasets, in observations from more whales (especially females and young animals), and in data spanning a wider range of traffic conditions. Note then, that opportunistic observations were a minimum of 20 min in length, but observations longer than 20 minutes were not split into smaller, 20-min tracks, which might have led to spurious variance estimates through pseudoreplication. The only samples to be split were the experimental trials, which were split into adjacent, paired control and treatment samples.

We analyzed all opportunistic observations using multivariate descriptive models rather than strict statistical hypothesis testing, because the secondary goal of our study was to describe whether whale behavior varied across a range of traffic conditions in a manner consistent with the results from our stricter experimental tests. When linking anthropogenic activity and natural variability to whale behavior, the need to specify a mathematical function is problematic. Many biological and ecological interactions do not have a functional relationship that can be easily defined or predicted even from a good understanding of the biology of the system being described. Generalized additive models (GAMs) offer a powerful, robust means of addressing all of these problems concurrently. Generalized additive models permit the response distribution to follow any member of the exponential family: No mathematical relationship need be specified *a priori*: the GAM chooses linear and non-linear combinations of candidate explanatory variables that best describe the response.

We chose to describe variation in the five response variables (whale behavior), as linear or non-linear functions of candidate predictors, using a set of objective criteria for adding or dropping terms. The decision to link a predictor to a response using linear or non-linear terms was automated by *mgcv*, which determines the shape of the relationships between predictors and response using thin-plate regression splines (Wood 2003). The optimal degree of flexibility that can be justified by the data is estimated in a maximum likelihood framework, while penalizing the model for overparameterization or oversmoothing (Wood 2000). The higher the estimated degrees of freedom (edf) given to any term in the model, the more flexible that relationship is estimated to be. This distinction in approaches to model specification and model selection in the two software packages is an important one, and is outlined in greater detail in Appendix 4.

Potential explanatory variables considered for inclusion in the model were *Month*, *Time*, *Age*, *Sex*, *Point of Closest Approach (PCA)*, *Number of boats within 100m (MAX100)*, *Number of boats within 400m (MAX400)*, and *Number of boats within 1000m (MAX1000)*. Factor variables were entered as linear or grouping terms. Continuous variables were entered as candidates for smoothing ($s(x)$) by *mgcv*.

The following summarizes our backward stepwise model specification procedure adopted for each of the five response variables, y , during this study, using the framework proposed by Wood (2001):

1. A fully saturated model was fitted to the data: $\{y \sim Month + Time + s(Age) + Sex + s(PCA) + s(MAX100) + s(MAX400) + s(MAX1000)\}$ with the default degree of smoothing (10 knots, 9 df).
2. Model fit was assessed using the *summary.gam* and *plot.gam* functions in *mgcv*, which showed coefficients, GCV score, explanatory power (deviance explained) and fit (residual plots).
3. For each linear term, the parameter coefficient (slope) was examined to see if it was near 0 and the significance term to see if it was near 1. If so, the term was removed to see if

the GCV score decreased and the explanatory power of the model increased. If so, the term was dropped from the model. If no marked improvement was detected by removing the term, then it remained in the model. Term-wise P-values for each coefficient were reported, and give an approximate measure of confidence in the decision to retain the term, given that the other terms are in the model.

4. For each smooth model term, the estimated number of degrees of freedom was examined to see if it was near 1. The 95% confidence intervals for that term were examined to see if they included zero across the range of observations. If so, the term was dropped temporarily, to see if the GCV score dropped and the explanatory power of the model increased.
5. A term was dropped from the final model if it satisfied all three of the conditions in step 4 (*i.e.*, $\text{edf} \approx 1$; 95% CI's include zero across range of x ; and dropping the term decreased the GCV score and increased the explanatory). If the first criterion was met ($\text{edf} \approx 1$), but not the other two, then the smooth term was replaced by a linear term. If replacing the smooth term with a linear one reduced the model's explanatory power, then it was retained as a smooth term with $\text{edf} \approx 1$.

One downside of our decision to use GAMs in `mgcv` was that it demanded a backward stepwise approach for model specification (that is, start with a saturated model, and drop terms when appropriate). A forward stepwise approach (adding terms sequentially to a skeleton model) would have been nice in terms of parsimony, but would have presented problems beyond the scope of this study to resolve. Forward stepwise procedures suffer from the problem that the likelihood of a term entering the model is influenced by the order in which the term is presented. Given that we have both boat-related and boat-unrelated terms in the model, it would be somewhat subjective to decide whether to include boat-related terms before or after individual (*e.g.*, age and sex) or environmental (*e.g.*, day) terms. We chose a backward stepwise procedure then for two reasons: first, it is the approach advocated by the author of the GAM-fitting procedure in R (Wood 2001); and secondly, it objectively sorts out the problem of inadvertently biasing the model specification procedure by deciding to present boat-related variables before or after the other terms.

Finally, we conducted one additional analysis of a particular relationship of interest, namely the one between the animals' path directness and the number of boats. This is of interest both in terms of the suitability of our decision to use three boats as the cutoff between few and many boats in our experimental data, and in terms of designing studies on other populations (such as southern resident killer whales). This problem is a common one in interpolation. Recall that the GAMs use splines for interpolating between observed data. That is to say that "knots" are placed at observed data, and a variety of mathematical functions are used to interpolate between these points. B-Splines (more commonly, `bsplines`) are an advanced curve form that allows for the order to be set by the user (that is, the knots are placed manually) instead of being defined by the number of control points. To that end, we conditioned on the selected model describing variation in path directness index, but placed knots manually along the x -axis describing variation in boat number. We used AIC to determine objectively where a node ought to be placed, *i.e.*, we used the variability in whale behavior to determine what constituted few versus many boats.

Results

Effort and sample size

The observer team was available for data collection on 72 of the 73 days between 1 July and 11 September 2004. Whales were present for at least part of the day on 60 of these 73 days. Fog and rain prevented data collection on a total of 11 days. In total, our study represents 792 hours of search effort. The no-boat sections of experimental tracks were interrupted by fishing boats, ferries, cruise ships, recreational boaters, and kayakers transiting the study area and coming within 1000m of the focal animal. While the focal whale may be within the Reserve, a vessel transiting on the Reserve boundary can easily approach within 1000m. In addition, a 10-day commercial fishing opening in which hundreds of commercial fishing boats at a time were within

the study area (including the Reserve) limited opportunities for experimental tracks. Experimental track sample size was also constrained by commercial whalewatch vessel availability. Whalewatch operators preferentially target groups of whales well away from the Reserve when multiple groups of whales provide the option. Note that opportunities to conduct experimental approaches were so rare that we decided to restrict our trials to males. The one experimental approach of a female was excluded from the analyses. A total of 140 tracks (both opportunistic and experimental) were recorded on 26 focal whales. Details on sample size by age, sex, and individual identity are provided in Appendix 1. In total, our study represents 73.8 hours of continuous observation (*i.e.*, of tracks >20 minutes in length) of focal animals using the theodolite.

Post-hoc definition of treatment levels: few versus many boats

Of the 16 experimental trials, focal animals were approached within 1000m by 1-3 boats on eight occasions, and 4-17 on the remaining eight (Table 1). We used these categories in to define a node in subsequent analyses to separate our two treatment levels: *few* (1-3) versus *many* (4-17) boats.

Number of boats	Number of trials
1	1
2	4
3	3
4	1
5	1
6	1
8	1
9	1
11	1
12	1
17	1

Table 1. Number of times that a given number of boats approached the focal whale within 1000m during 16 experimental trials. Note that during 8 of the 16 trials, the focal animal was approached by 1-3 boats within 1000m. In the remaining 8 trials, the focal whale was approached by 4-17 boats, once each.

Responses of whales to experimental approach by few versus many boats

Figure 3 shows the mean values and standard error bars for each of the five response variables during the control and treatment phases of the “few” and “many” trials. While this figure shows the distribution of the observed values, it should be interpreted with caution, because all subsequent analyses used non-parametric statistics based on ranks, rather than raw data. Consequently, apparent differences suggested by non-overlapping confidence intervals may not be significant ones in non-parametric analyses, and vice versa. The distribution of the raw data are shown for illustrative purposes only.

1. Inter-breath interval [RESP]

Whales approached within 1000 m by few boats had mean dive times approximately 5.0s longer than during the preceding, control conditions. Those approached within 1000 m by many boats had mean dive times 3.8s longer than during control conditions. These responses did not differ significantly by treatment level ($p = 0.9591$). Thus, any apparent change in dive time did not vary with respect to number of boats approaching within 1000m.

2. Speed [SPEED]

Whales approached within 1000 m by few boats had mean swimming speed approximately 0.89km/h faster than that observed during the preceding, control conditions. Those approached within 1000 m by many boats had mean swimming speeds approximately 0.75km/h faster than during control conditions. These responses did not differ significantly by treatment level ($p>0.9999$). Thus, any apparent increase in swimming speed did not vary with respect to number of boats approaching within 1000m.

3. Directness index [DI]

On average, whales approached within 1000 m by few boats decreased their index of path directness by 14.0 points over that observed during preceding, control conditions. Those approached within 1000 m by many boats actually increased the directness of their paths on average by 12.9 points than during control conditions. These responses did differ significantly by treatment level ($p=0.0148$). Thus, whales showed significantly different responses to experimental approach by few versus many boats. Those approached by few boats adopted paths that were less direct than paths observed previously. Those approached by many boats adopted paths that were significantly more direct than observed previously.

We calculated the mean response of focal animals to approach by few boats, in order to allow comparison with previously published studies. In the control phases of our experiment, the average directness index was 82. When approached by 1-3 boats, the mean directness index declined to 67.9. Whales approached by 4-17 boats actually increased their mean directness index from 70.9 to 83.8. We used these data and the formula for calculating directness index given in the methods to estimate the distance a whale would have to travel along a circuitous route to cover 100m of straight-line distance. Recall that directness index is calculated by dividing the progressive distance by the surface distance covered in a track. The surface distance required to cover 100m in a given track, then, is simply 100m divided by the path directness index expressed as a percentage.

In the absence of boats (*i.e.*, a whale displaying a path directness index of 82), a whale would have to swim 122m (*i.e.*, $100/0.82$) along a circuitous path to cover the straight-line distance of 100m. The same whales adopting a mean swimming path directness index of 67.9 would have to swim 147m along a circuitous path to cover the 100m straight-line distance. The total distance a whale would have to travel in the presence of 1-3 boats represents a 20.7% increase over the total distance it would swim to cover the same ground in the absence of boats. Whales approached by 4-17 boats would actually increase the efficiency of their paths: from 141 (control) to 119m (treatment) required to make 100m of headway.

4. Deviation index [DEV]

Whales approached by few boats showed an average angular deviation between successive dives that was 0.3° greater than that observed during preceding, control conditions. Those approached by many boats showed an average angular deviation between successive dives that was 6.5° less than that observed during preceding, control conditions. These responses did not differ significantly by treatment level ($p=0.4418$). Thus, any apparent difference in the way that whales responded by altering their deviation index did not differ with respect to number of boats approaching within 1000m.

5. Surface Active Behavior [SAB]

Surface active behavior was observed only twice during the 16 experimental trials (*i.e.*, the 32 experimental segments). Both bouts occurred during the no-boat sections preceding exposure to many boats. However, given that no surface active events were observed in the no-boat sections preceding exposure to few boats, or during any of the 16 treatment sections, rigorous statistical analysis is problematic. We interpret this as an interesting anecdote that in both cases where many boats approached surface active animals within 1000m, surface activity ceased. This may be due entirely to chance.

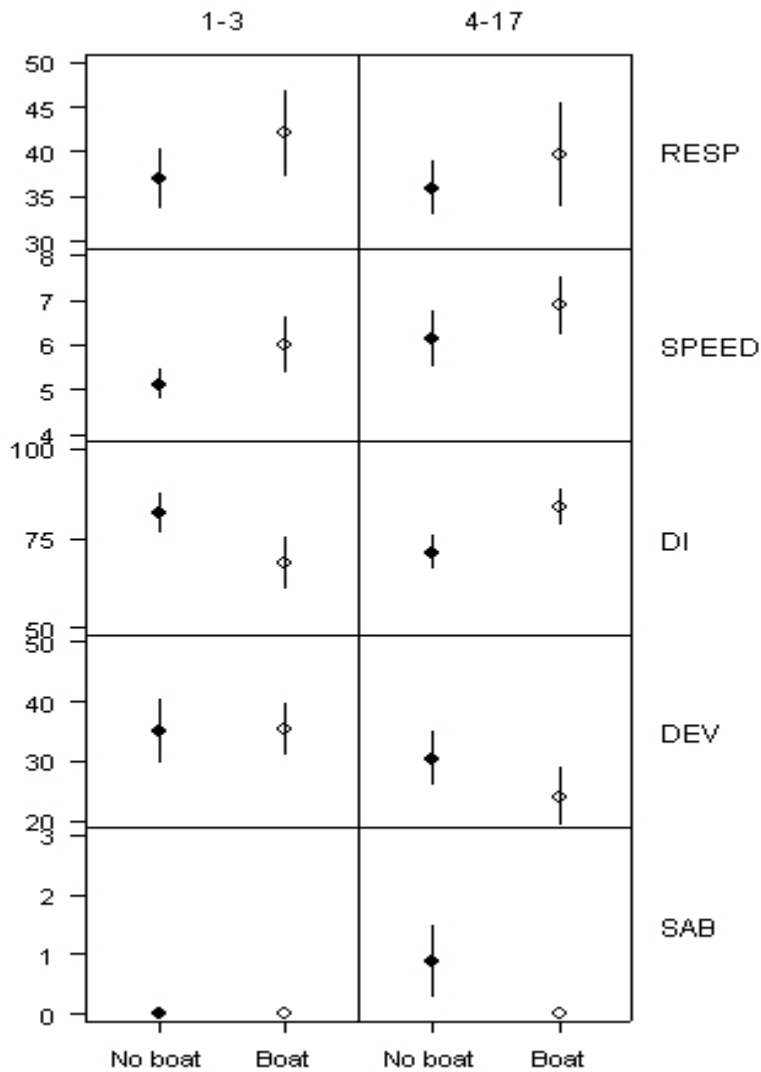


Figure 3. Behavioral responses (mean \pm se of the difference in whale behavior during control and treatment phases) of whales to experimental approach under two treatment conditions: approach to within 1000m by 1-3 boats, or by 4-17 boats. (a) RESP: Average time between respirations, or inter-breath interval(s); (b) SPEED: swimming speed (km/h); (c) DI: directness index (unitless); (d) DEV: deviation index (degrees); (e) SAB: surface active behavior (bouts/h). Note that all sample sizes are the same: 8 no-boat observations followed immediately by 8 treatment observations, for each of the two treatment levels, for a total of 32 observations.

Responses of whales to experimental approach by any number of boats (i.e., comparing whale behavior during absence and presence of boats)

Rather than looking at how cetacean behavior varies across a number of boats, vessel impact studies often contrast behavior in a presence-versus-absence framework. The analyses below demonstrate that when the response is dependent on number of vessels present, reducing that experimental treatment to a presence-absence framework can mask those effects. The following summarizes the results of a comparison of the behavior of whales that were approached by any number of boats (from 1-17 boats approaching the whale within 1000m) and the behavior of whales in the absence of boats.

1. *Inter-breath interval [RESP]*

In the absence of boats, the mean inter-breath interval of focal whales was 36.5s. When approached by boats, the mean inter-breath interval was 40.9s. This difference was not statistically significant ($p=0.252$). The sum of all positive ranks (*i.e.*, cases where experimental approach resulted in longer mean dive times) was 45, while the sum of all negative ranks was -91.

2. *Speed [SPEED]*

In the absence of boats, the mean swimming speed of focal whales was 5.6km/h. When approached by boats, the mean swimming speed was 6.4km/h. This difference was not statistically significant ($p=0.117$).

3. *Directness index [DI]*

In the absence of boats, the mean path directness index of focal whales was 76.4. When approached by boats, the mean directness index was 75.8. This difference was not statistically significant ($p>0.9999$).

4. *Deviation index [DEV]*

In the absence of boats, the mean deviation index (*i.e.*, average angle between successive dives) of focal whales was 32.8. When approached by boats, the mean deviation index was 29.7. This difference was not statistically significant ($p=0.404$).

5. *Surface Active Behavior [SAB]*

Bouts of surface activity were observed twice out of 16 trials during the control phase, and never during the treatment phase. The proportion of observations in which surface activity occurred was unrelated to the presence or absence of boats ($p=0.484$).

Opportunistic observations

The scatterplot matrix shown in Figure 4 displays the relationships among the eight predictor variables and the five response variables, without implying any causal link between the two. Lowess (LOcally WEighted Sums of Squares) curves, *i.e.*, moving averages, were fitted through each scatterplot to illustrate smoothed pairwise correlations, but these should be interpreted with caution. The subsequent section reports the results from the GAM-fitting for each of the five response variables, in which: not all candidate explanatory covariates shown in Figure 4 actually entered the selected models; those terms that did enter the models may have done so as either linear or smooth terms; and model smoothing and selection was done conditional on all selected terms being in the model. In other words, the scatterplot matrices were produced to illustrate the distributions of and relationships among the original variables, but pairwise comparisons are unreliable and should be interpreted with caution. Finally, sex may have entered the selected model. In such cases, the smoothed relationships shown in Figure 4, which is an average of observations of adult male, adult female and juvenile focal animals, may look different than those in the selected model.

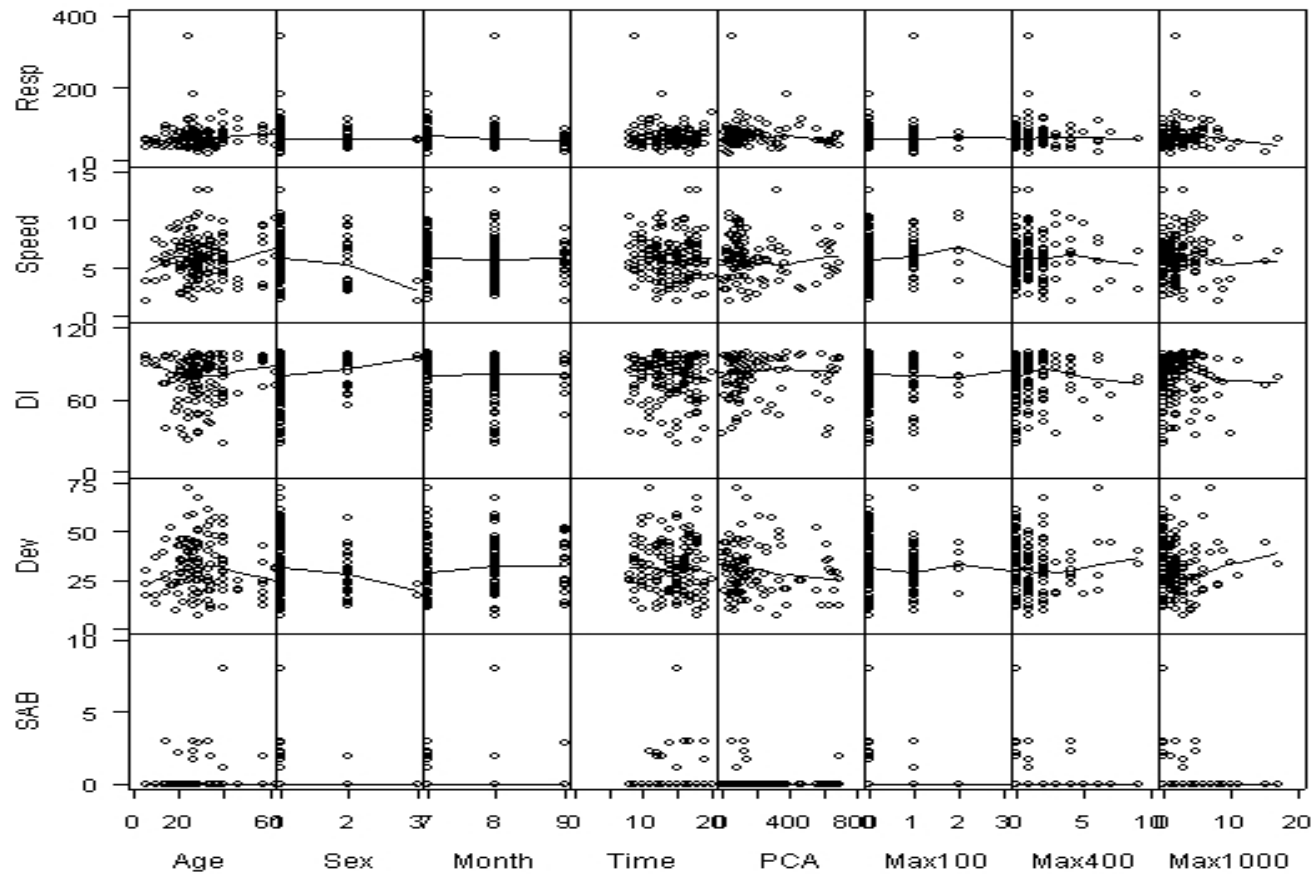


Figure 4. A scatterplot matrix illustrating the relationships among the eight predictors and five responses. Lines represent standard lowess (smooth) curves: *i.e.*, no attempt was made to regress the y-variables on the x-variables, but rather simply to show patterns apparent in moving averages across the ranges of x. For detailed descriptions of the variables and their units of measurement, please refer to the text, noting that the 13 variables are labeled using the abbreviations given in the methods section. *Sex* is coded using 1 for males, 2 for females and 3 for juvenile whales of unknown sex. *Time* refers to the hour in which the observation began.

Results of generalized additive model (GAM) fitting

The following summarizes the selected models for each of the five response variables. The P-values listed are approximate, and should be interpreted with caution (Wood 2001). For linear terms, they represent the P-values for the null hypothesis that the corresponding parameter is zero. Significance of linear terms is calculated with reference to the t-distribution with the estimated residual degrees of freedom for the model fit. For smooth terms, significance is estimated as an approximate P-value for the null hypothesis that each smooth term is zero. They are obtained by comparing the chi.sq statistic given above to the chi-squared distribution with degrees of freedom given by the estimated degrees of freedom for the term. The P-values are known to be somewhat unreliable, because they are conditional on all other terms being in the model, the smoothing parameters, and any scale parameter estimate. As a result, some model terms were retained even though their P-values were >0.05, for the reasons outlined in the description of the GAM-fitting process in the methods section.

The explanatory power of the model is summarized in two ways. The adjusted R-squared of the model is defined as the proportion of variance explained, where original variance and residual variance are both estimated using unbiased estimators. It is a relative measure, so it can have a negative value if the model explains less of the variance than a constant term. For models with non-normal errors, the “deviance explained” is a better descriptor of model fit than the R-squared score.

1. Inter-breath interval [RESP]

Average inter-breath interval, or mean dive time, was found to vary significantly in conjunction with one traffic-related (*MAX1000*) and a variety of traffic-unrelated variables (*MONTH*, *TIME*, *SEX* and *AGE*). The model specified a quasi family with a log link, which is designed to address skewed distribution, and allowed package *mgcv* to estimate the appropriate dispersion parameter.

The coefficients of the selected model’s linear terms and the smoothing parameters of non-linear terms are given in Table 2. The explanatory power of the selected model was moderate: the model explained 18.5% of the deviance (*i.e.*, residual sums of squares).

In general, males tended to show longer mean dive times than females. The two animals of unknown sex had even stronger coefficients suggesting long dives, however it must be borne in mind that these represent (by definition) juvenile whales, so long dive times may instead reflect times when breaths were missed. Mean dive times tended to be shorter as the season progressed.

Figure 5 shows how mean dive time related to the smoothed components of *TIME*, *AGE* and *MAX1000* in the selected model.

Table 2. Summary of the selected GAM describing variation in mean dive time.

Family: quasi
Link function: log

Formula:

$RESP \sim MONTH + s(TIME) + SEX + s(AGE) + s(MAX1000)$

Parametric coefficients:

	Estimate	Std. Err	t ratio	Pr(> t)
(Intercept)	4.8346	0.05435	8.896	4.7537e-15
MONTH	-0.10214	0.06731	-1.517	0.13161
SEXM	0.15643	0.1656	0.9447	0.34661
SEXU	0.50973	0.4994	1.021	0.30936

Approximate significance of smooth terms:

	Edf	chi.sq	p-value
s(TIME)	1.81	3.6145	0.14431
S(AGE)	4.329	6.7682	0.18366
S(MAX1000)	20.69	3.1578	0.22077

R-sq.(adj) = 0.0609 Deviance explained = 18.5%
 GCV score = 14.998 Scale est. = 13.69 n = 140

The terms not selected for the model included PCA, MAX100 and MAX400.

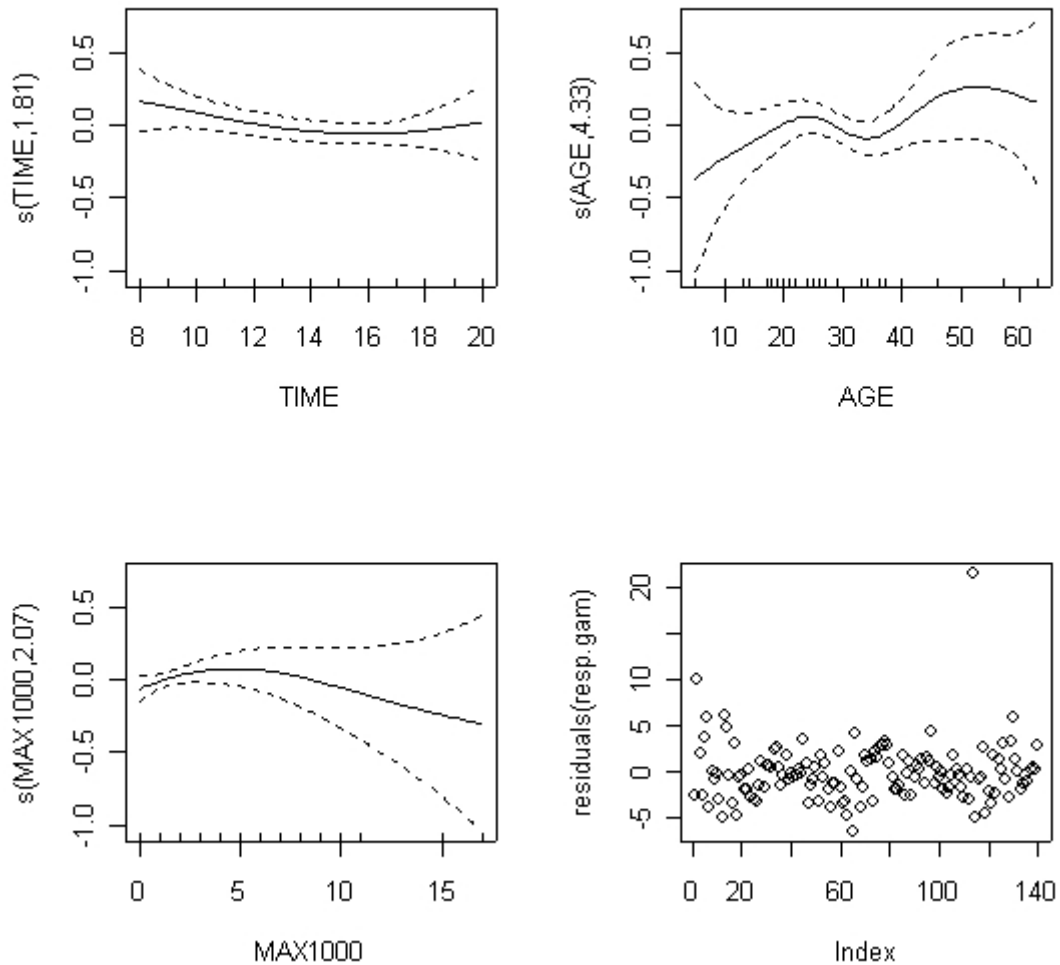


Figure 5. A plot showing the three component smooth functions that made up the selected GAM describing variation in the response variable, mean dive time, with the x axes shown on the scale of the predictors. The x-axis of each plot is labeled with the covariate name, while the y-axis is labeled $s(\text{covariate name}, \text{edf})$, where s refers to a smoothing spline and edf to the estimated degrees of freedom selected by mgcv 's automated smoothing process. The plot in the lower right shows the residuals (data minus fitted values). The symmetry about $y(0)$ and constant variance along the range of x indicate that the model fitted the data well, except for one potential outlier.

2. Speed [SPEED]

Average swimming speed was found to vary significantly in conjunction with a variety of traffic-related variables (*PCA*, *MAX400* and *MAX1000*) and with the sex of the focal whale. The model specified a quasi family with a log link, which is designed to address skewed distribution, and allowed package *mgcv* to estimate the appropriate dispersion parameter.

The coefficients of the selected model's linear terms and the smoothing parameters of the non-linear term are given in Table 3. The explanatory power of the selected model was reasonably good: the model explained 30.3% of the deviance (*i.e.*, residual sums of squares). Note that sample size is smaller than the number of tracks collected, because *PCA* (point of closest approach) entered the model as a selected term. Given that true no-boat tracks have missing values for the *PCA*, all associated covariates had to be treated as missing values for these observations as well.

In general, males tended to swim faster than females. Animals of unknown sex tended to swim more slowly than animals of known sex, which is as one would expect from juvenile animals. Boat counts entered the model as linear terms. Generally, whales had a weak tendency to swim more slowly as number of boats increased within 400m, but more quickly as number of boats within 1000m increased. This apparent inconsistency may reflect sample size. In the scatterplots of original data (Figure 4), it would seem that these are likely to reflect U-shaped curves that may appear linear on the scale of 0-3 or so. Large numbers of boats within 400m were very rarely seen. (While the maximum number of boats ever observed within 400m was 9, the third quartile of the histogram of the maximum number of boats within 400m was 2.) Number of boats within 1000m showed much more variability. (The maximum number of boats ever observed within 1000m was 17, and the third quartile of the histogram of the maximum number of boats within 1000m was 3.)

Figure 6 shows how mean swimming speed related to the smoothed component in the selected model of the point of closest approach (*PCA*) by a boat. While *MAX100*, *MAX400* and *MAX1000* represent the maxima of boat counts ever observed within three ranges of the whale, *PCA* represents the absolute minimum distance ever observed between the whale and the nearest boat. This plot shows that whales tended to swim slowly when the nearest boat approached closely (*i.e.*, <50m) and quickly when boats stayed approximately 100m or farther from the whale. Beyond this point, the confidence intervals span zero, and the relationship becomes unconvincing.

Table 3. Summary of the selected GAM describing variation in swimming speed.

Family: quasi

Link function: log

Formula:

SPEED ~ SEX + s(*PCA*) + *MAX400* + *MAX1000*

Parametric coefficients:

	Estimate	Std. err.	t ratio	Pr(> t)
(Intercept)	1.681	0.09846	17.07	< 2.22e-16
SEXM	0.11505	0.0978	1.176	0.24268
SEXU	-0.70791	0.3615	-1.958	0.053437
<i>MAX400</i>	-0.057745	0.03613	-1.598	0.11367
<i>MAX1000</i>	0.040625	0.01974	2.058	0.042639

Approximate significance of smooth terms:

	edf	chi.sq	p-value
S(<i>PCA</i>)	7.475	21.51	0.0081236

R-sq.(adj) = 0.186
 GCV score = 0.73666

Deviance explained = 30.3%
 Scale est. = 0.64384

n = 99

The terms not selected for the model included AGE, MONTH, TIME and MAX100.

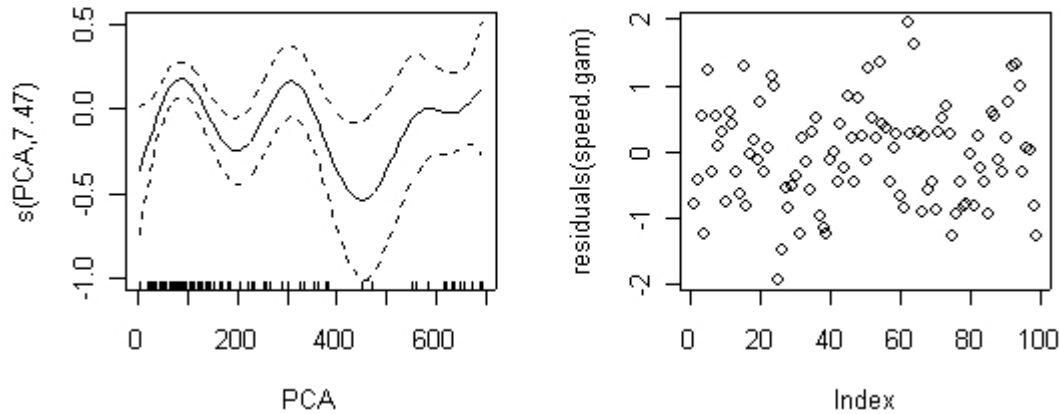


Figure 6. Plot showing the component smooth function that made up the selected GAM describing variation in swimming speed, on the scale of the linear predictor, point of closest approach. The y-axis is labeled $s(\text{covariate name}, \text{edf})$, where edf is the estimated degrees of freedom selected by mgcv 's automated smoothing process. The plot on the right hand side shows the residuals (data minus fitted values). The symmetry about $y(0)$ and constant variance along the range of x indicate that the model fitted the data well.

3. Directness index

Poor model fit was found (as evidenced by patterns in residual plots) when specifying a quasi or Gamma distribution in path directness data. This was addressed by standardizing the data – expressing them as a proportion by dividing by 100, so that 0 represented a circular path and 1 a straight line. The response was then modeled in a quasibinomial framework with a logit (1/0) link.

Path directness index was found to vary significantly in conjunction with a variety of traffic-related (MAX400 and MAX1000) and traffic unrelated variables (MONTH and AGE). The GCV score was improved by dropping the intercept term (which is represented as -1 in the model definition, below). The coefficients of the selected model's linear terms and the smoothing parameters of the non-linear term are given in Table 4. The explanatory power of the selected model was good: the model explained 67.7% of the deviance (*i.e.*, residual sums of squares).

In general, paths became more direct as the season progressed. Paths also became more direct as number of boats within 400m increased. Age and number of boats within 1000m entered the selected model as smooth terms, and Figure 7 shows how directness index related to the smoothed components of these two variables.

Table 4. Summary of the selected GAM describing variation in directness index.

Family: quasibinomial
 Link function: logit

Formula:
 $\text{DI} \sim \text{MONTH} + s(\text{AGE}) + \text{MAX400} + s(\text{MAX1000}) - 1$

Parametric coefficients:

	Estimate	std. err.	t ratio	Pr(> t)
MONTH	0.15381	0.01857	8.281	1.14E-13
MAX400	0.084724	0.09028	0.9384	0.34972

Approximate significance of smooth terms:

	edf	chi.sq	p-value
s(AGE)	2.04	3.0592	0.22666
s(MAX1000)	2.784	5.1208	0.14729

R-sq.(adj) = 0.0377 Deviance explained = 67.7%
 GCV score = 0.19017 Scale est. = 0.18091 n = 140

The terms not selected for the model included SEX, TIME, PCA and MAX100. (The intercept term was also dropped.)

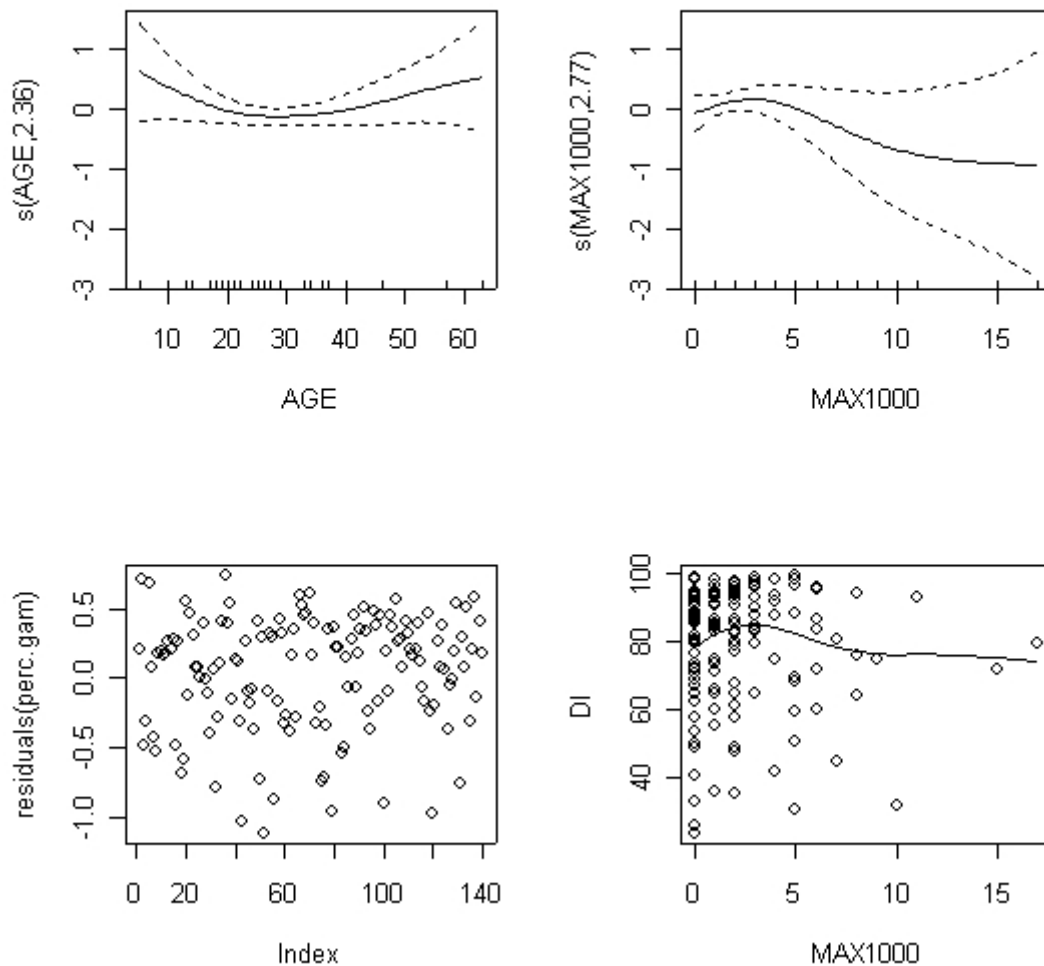


Figure 7. Plots (top) showing the component smooth functions (age, and maximum number of boats approaching the whale within 1000m) that made up the selected GAM describing variation in directness index, on the scale of the predictors. The y-axis is labeled $s(\text{covariate name}, \text{edf})$, where edf is the estimated degrees of freedom selected by mgcv's automated smoothing process. The plot on the lower left-hand side shows the residuals (data minus fitted values). The symmetry about

$y(0)$ and constant variance along the range of x indicate that the model fitted the data well. Given its importance to the relationship of interest, the lower right-hand side shows an expanded (LOWESS-smoothed) scatterplot of the relationship between directness index and the number of boats within 1000m of the focal whale on the original scales of measurement. (This is simply an expanded look at the same data presented in Figure 4 relating boat number and path directness.) Note the non-linearity of the relationship, with an apparent peak at around three boats.

4. Deviation index

Deviation index (*i.e.*, average angle between successive dives) was found to vary significantly in conjunction with all four candidate traffic-related variables (*PCA*, *MAX100*, *MAX400* and *MAX1000*) and three traffic-unrelated variables (*SEX*, *MONTH* and *TIME*). The model specified a quasi family with a log link, which is designed to address skewed distribution, and allowed package *mgcv* to estimate the appropriate dispersion parameter.

The coefficients of the selected model's linear terms and the smoothing parameters of the non-linear term are given in Table 5. The explanatory power of the selected model was reasonably good: the model explained 29.5% of the deviance (*i.e.*, residual sums of squares). Note that sample size is smaller than the number of tracks collected, because *PCA* entered the model as a selected term. Given that true no-boat tracks have missing values for the point of closest approach, all associated covariates had to be treated as missing values for these observations as well.

In general, paths showed a weak tendency toward erratic paths as the season progressed, and males tended to show slightly more jagged paths overall than females. Recall that directness and deviation indices represent path parameters on two different spatial scales: beginning-to-end point sinuosity; and dive-to-dive smoothness. Paths can be smooth, but circuitous, or jagged but relatively direct. Figure 8 shows how path deviation index related to the smoothed components in the selected model of: time of day; point of closest approach by a boat; and maximum number of boats within 1000m. The 95% confidence intervals of the smoothed term describing time of day comfortably included zero across the range of observed values. As a result, no consistent trend can be described with respect to diel patterns in whale swimming paths, however model fit was improved by including the term – that is, holding this weak effect constant while examining other relationships of interest.

Including the point of closest approach improved model explanatory power, as well as model fit as indicated by residual plots, even though its contribution was decidedly linear (Figure 8). Consequently, there was a linear relationship between point of closest approach and deviation index. In general, whales tended to show jagged, zig-zag paths as boats approached closely (within approximately 200m). When the point of closest approach was greater than that, whale paths tended to be smooth.

Relationship between boat number and path directness showed similar trends. When number of boats within 1000m of the whale was two or fewer, paths tended to be jagged. When approximately three to five of boats approached, paths tended to be smooth. Beyond five boats, the relationship between boat number and path deviation index comfortably included zero. The residual plot (Figure 8) suggested that the selected model provided a good fit to the data.

Table 5. Summary of the selected GAM describing variation in deviation index.

Family: quasi

Link function: log

Formula:

$DEV \sim MONTH + s(TIME) + SEX + s(PCA) + MAX100 + MAX400 + s(MAX1000)$

Parametric coefficients:

	Estimate	std. err.	t ratio	Pr(> t)
(Intercept)	2.721	0.5991	4.542	1.91E-05
MONTH	0.070133	0.07363	0.9525	0.34365
SEXM	0.10224	0.113	0.9048	0.36823
SEXU	-0.47815	0.3694	-1.295	0.19914
MAX100	-0.11988	0.09062	-1.323	0.18957
MAX400	0.056605	0.04275	1.324	0.18922

Approximate significance of smooth terms:

	edf	chi.sq	p-value
s(TIME)	4.851	7.9228	0.16241
s(PCA)	1	1.2387	0.269
s(MAX1000)	5.579	13.652	0.035104

R-sq.(adj) = 0.166 Deviance explained = 29.5%
 GCV score = 5.5138 Scale est. = 4.543 n = 99

The only term not selected for the model was AGE.

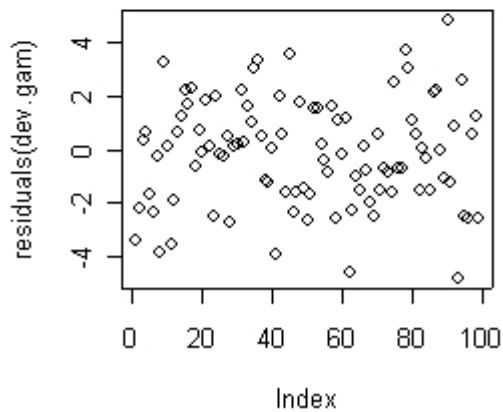
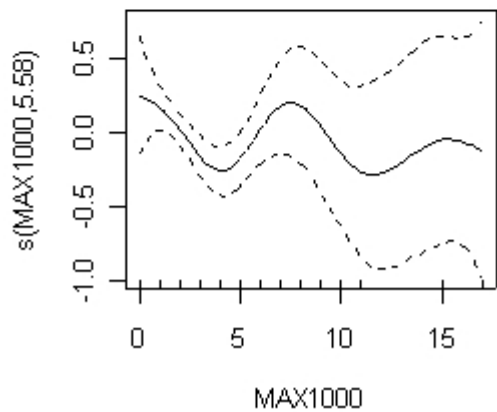
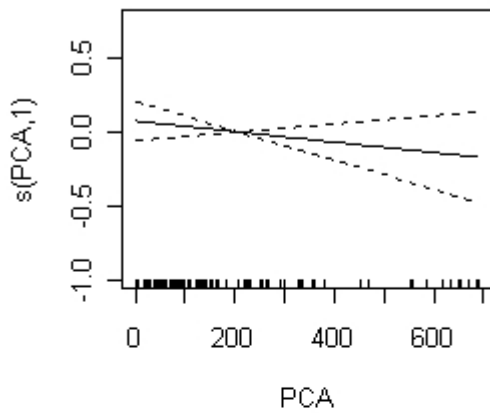
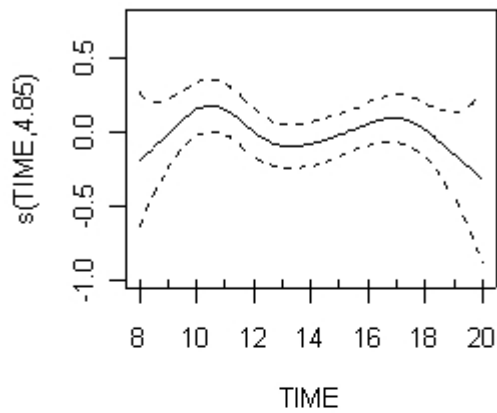


Figure 8. A plot showing the component smooth functions that made up the selected GAM describing variation in deviation index, on the scale of the predictors. The x-axis of each plot is labeled with the covariate name, while the y-axis is labeled $s(\text{covariate name}, \text{edf})$, where edf is the estimated degrees of freedom selected by mgcv 's automated smoothing process. While mgcv assigned only 1 degree of freedom to the vector describing the point of closest approach by any boat, model fit was noticeably worsened by dropping the term or replacing it with a linear term. In the plot at lower left, note that deviation index tended to be high (*i.e.*, erratic paths) when boat number was low, then showed no consistent trend as boat number increased beyond approximately 4-5 boats. The plot in the lower right shows the residuals (data minus fitted values). The symmetry about $y(0)$ and constant variance along the range of x indicate that the model fitted the data well.

5. Surface Active Behavior

Analysis of surface active behavior was problematic, due primarily to its rarity of occurrence. For all opportunistic and experimental trials combined, surface activity was observed in only 11 out of 140 samples. Describing factors that influence the frequency of rare events is conceptually and logistically challenging. We chose instead to reduce this problem to a logistic one. We asked, "What are the factors that influenced the probability that a bout of surface-active behavior occurred?" To that end, all measurements of the rate of surface active behavior (*i.e.*, the expected number of bouts per hour) to a binary response variable – one if any bout occurred during the observation and zero if it did not.

The probability of a bout of surface active behavior occurring was found to show very weak and inconsistent non-linear trends with time of day and two boat traffic variables (Table 6; Figure 9). The explanatory power of the selected model was reasonably good: the model explained 33.6% of the deviance (*i.e.*, residual sums of squares). Note that sample size is smaller than the number of tracks collected, because PCA entered the model as a selected term. Given that true no-boat tracks have missing values for the point of closest approach, all associated covariates had to be treated as missing values for these observations as well. However, all of the relationships in the selected model comfortably include zero in the response variable, which suggests that many more observations of surface activity would be required to gain sufficient statistical power to assess whether any of these apparent trends were real. The strongest effect was simply the intercept term, which tells us what we already know: on average, the probability of a bout of surface active behavior occurring in a sample was exceedingly low. Ultimately, the data were not sufficient to assess how vessel traffic affects surface active behavior.

Table 6. Summary of the selected GAM describing the probability of a bout of surface active behavior occurring during an observation.

Family: binomial
Link function: logit

Formula:
SAB ~ s(PCA) + s(TIME) + s(MAX1000)

Parametric coefficients:

	Estimate	std. err.	t ratio	Pr(> t)
(Intercept)	-3.9315	0.8989	-4.374	1.22E-05

Approximate significance of smooth terms:

	edf	chi.sq	p-value
s(PCA)	3.389	7.1508	0.0883
s(TIME)	1	2.7554	0.09692
s(MAX1000)	1.765	2.4264	0.25253

R-sq.(adj) = 0.199
UBRE score = -0.45075

Deviance explained = 33.6%
Scale est. = 1

n = 99

The terms not selected for the model included SEX, AGE, MONTH, MAX100 and MAX400.

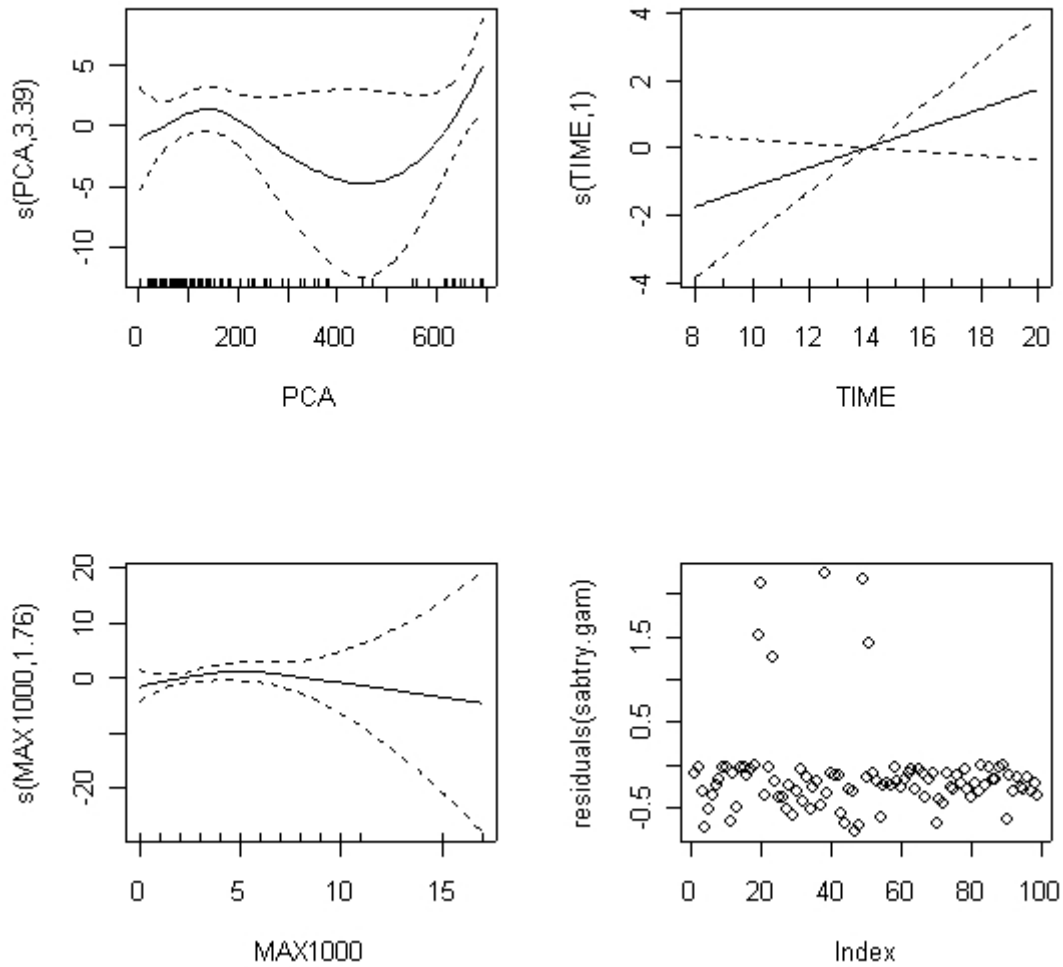


Figure 9. A plot showing the component smooth functions that made up the selected GAM describing variation in the probability of a bout of surface activity occurring, on the scale of the predictors. The x-axis of each plot is labeled with the covariate name, while the y-axis is labeled $s(\text{covariate name}, \text{edf})$, where edf is the estimated degrees of freedom selected by mgcv 's automated smoothing process. While mgcv assigned only 1 degree of freedom to the vector describing time of day, model fit was noticeably worsened by dropping the term or replacing it with a linear term. The 95% confidence intervals of all three covariate smooths comfortably include zero, which suggests that these apparent trends are weak, although there is strong statistical support for including them in the model (in terms of penalized UBRE score). The plot in the lower right shows the residuals (data minus fitted values). The clusters about $y(0)$ and $y(1)$ reflect the logistic nature of the response variable, and the constant variance along the range of x indicates that the model fitted the data well.

Using bsplines to determine the node between ‘few’ and ‘many’ boats

The analysis of experimental data relied on correctly identifying the node that separated the treatment levels: few and many boats. This node was identified at 2-3 by Williams *et al.* (2002a) based on patterns apparent in the relationship between boat number and path directness index. However, those data were collected in 1995 and 1996. We looked at the opportunistic data collected in 2004 to test: (1) that the 2004 data still retained an inflection point (Figure 7, lower right); and (2) to identify whether the inflection point ought to be placed at 2 or 3 boats, that is, to identify whether our *a priori* experimental treatment level specification was appropriate. Note that this was a confirmatory exercise that followed the statistical hypothesis-testing analysis of the experimental data. It was not used to identify the treatment levels, because that would have biased the experimental analysis in favor of finding significant results, by identifying where a difference seemed to lie and then testing whether that difference was real. Instead, we defined the experimental treatment levels as best we could using previously published results, and then evaluated whether that apparent inflection point was evident in the much larger sample of data collected opportunistically.

One way to test objectively where a curve changes is to use bsplines. Bsplines are an advanced curve form that allows for the order to be set by the user instead of being defined by the number of control points. Control points, called knots in GAM fitting, refer to the places along the curve between which the GAM interpolates smooth values. The GAM-fitting procedure in *mgcv* identifies the number of knots objectively, using the observed data. We conditioned on this selected model describing variation in path directness index, but manually placed knots at every observed value of MAX1000 (that is, following the hypothesis made in Williams *et al.* 2002a). A plot of AIC of the model describing path directness index against the knot position is shown (Figure 10). AIC was lowest when the knot was placed at MAX1000=3. This provides an objective, if simplistic, identification of the optimal placement of the node separating few from many boats. It provides additional support for our decision to use 3 as the cutoff in our treatment categories for the experimental data.

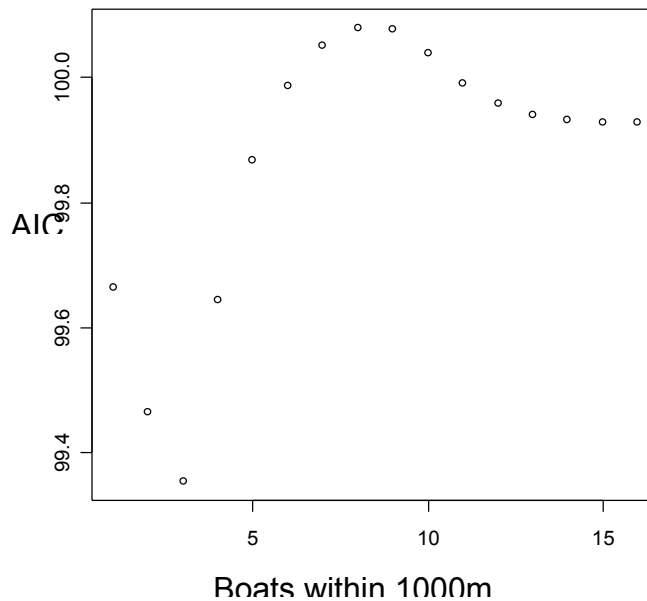


Figure 10. Bsplines allow for user-defined cutoffs between which values of y are interpolated for values of x , instead of being defined by the number of control points. The x -axis represents the point at which a knot was placed manually, when fitting a bspline through the opportunistic data, conditional on the selected model (Table 4 and Figure 7). The y -axis represents the AIC value of that model. AIC was lowest when a knot was placed manually at MAX1000=3. This may be interpreted as an objective identification of a node separating “few” and “many” boats at a

maximum of 3 boats ever approaching the whale to within 1000m. This supports the use of 3 boats as the cutoff in treatment levels in the experimental data.

Discussion

Our results add to a growing body of experimental and opportunistic studies that suggests that northern resident killer whales show a suite of stereotyped responses to boat traffic. The most consistent of these trends is that whales approached by experimental boats tended adopt paths that were less predictable (*i.e.*, less direct on the scale of a 20-min observation period) than those observed during the preceding control phase during which no boats were observed within 1000m. This pattern has been demonstrated experimentally now for whales responding to: one boat following whalewatching guidelines (*i.e.*, paralleling at 100m; Williams *et al.*, 2002a); one boat violating whalewatching guidelines (*i.e.*, leapfrogging within 100m; Williams *et al.*, 2002b); and to a few (1-3) boats approaching simultaneously within 1000m (Figure 3, Figure 10).

Path Directness Index

Male northern resident killer whales responded with 13% increase in distance covered when experimentally treated with a 100m parallel vessel approach (Williams *et al.*, 2002a). An increase of 17% in the distance a male whale has to swim to cover the 100m distance was measured when exposed to an experimental vessel that leapfrogged the whale's path within 100m (Williams *et al.*, 2002b). This study measured a 21% increase in distance required to travel 100m when focal animals were approached by 1-3 experimental boats. However, the opposite of this effect was seen when several (4-17) boats approached within 1000 m of the whale. Of course, these rough calculations are two-dimensional simplifications of a three-dimensional process. Whales could compensate for longer 2-D distances traveled by making shallower dives. This could be tested experimentally by tagging whales, if it were deemed a high conservation priority, but we do not have the data to test this at present.

Whales approached by many boats adopted a path that was straighter than that observed during preceding, control conditions. This increase in path directness when 4-17 experimental boats approached resulted in a 16% *reduction* in the distance a whale would have to swim in order to travel 100m, when compared with the preceding control phase.

Were the observed alteration of path directness consistent with avoidance tactics?

Models described by Howland (1974) and Weihs & Webb (1984) attempt to explain optimal strategies for predator evasion. Both models link successful escape from predation to the simultaneous fluctuation in velocity and turning radius. Speed adopted by prey is varied as well as the degree to which they diverge from the path of the predator in this horizontal avoidance tactic. Prey may increase in speed to adjust for a larger turning radius or may increase maneuverability to compensate for slower movement. Fast predators can be avoided by slower prey if the prey can turn more sharply (Howland 1974).

Our observed responses of killer whales to few boats, then, may be considered loosely analogous to a predator-prey interaction. In fact, some tracks of killer whales and experimental boats (Williams 1999, Williams *et al.* 2002a) are reminiscent of long-exposure photographs of moths evading bats (Roeder 1967). This resemblance provides a framework for interpreting behavior of whales around a boat in the context of the tactics that some prey use to escape predation.

This predator-prey analogy also offers a plausible framework for context-specific avoidance tactics. Whales would be expected to display a variety of responses to a variety of traffic scenarios, depending on the speed and maneuverability of the whale and vessel(s) involved. Certainly, the avoidance responses generated by experimental approach by few boats were reaffirmed by the retention of the boat number term in the model describing path directness. This adds increasing evidence to our earlier suspicion (Williams *et al.* 2002a) that an irregular path may be a useful avoidance tactic with a single boat but ineffective with more than one. In a multiple-vessel scenario, a dive that takes a whale farther from one boat may bring it closer to

another. Taken as a whole, it would seem that the tendency to adopt a more circuitous path is an effective evasive tactic with 1-3 boats, but ineffective with more boats. Perhaps as boat number increases, the animals switch to some other avoidance tactic, or tolerate a disturbance that they can mitigate at lower levels. A simple turning gambit is an effective means of evading fast predators, but it requires the *simultaneous* increase in speed and tightening of turning radius [*i.e.*, deviation index].

These statements suggest that the best descriptor of whale behavior is a multivariate one. But the amount of variance explained by our models was typically only 20-30% (Tables 2-6). Thus, other variables not measured by this study may play a larger role in influencing whale behavior. Indeed, distribution of salmon may be the best determinant of whale activity in Johnstone Strait (Nichol & Shackleton 1996), however no such information is available at the temporal and spatial resolution that we would need for this study.

Swimming speed and deviation index

While the average swimming speed of a whale increased with increasing numbers of boats within 1000m, it decreased with number of boats within 400m. We suspect that the former relationship is more plausible, given the better span of data sampled within a 400m-1000m radius of the whale than that within the 400m radius. Alternatively, this could mean that animals were slowing down to avoid close boats, but that animals did not need to divert their paths when boats stayed relatively far (*e.g.*, >400m) away. Nonetheless, these relationships were weak. Whales also adopted paths with higher deviation indices when approached by few boats than by many (Figure 8).

The smoothed, highly non-linear relationship between swimming speed and point of closest approach is an interesting one. The maximum swimming speeds observed were those measured when boats approached closer than 100m of the whale (Figure 6). In other words, swimming speeds tended to be highest when boats approached to within approximately 100m of the whale. When boats were much farther than 100m, or when they stayed approximately 200m from the whale, whale swimming speeds tended to be relatively slow. Future work might examine the boat activity at these distances (for example, leapfrogging or paralleling) that accounted for differences in whale behavior. Unfortunately, the emphasis on boat number in this study precluded consistent recording of boat activity with the level of staffing available. (Our decision to put higher priority on measuring boat number more precisely than boat activity reflects our *a priori* intention to model average whale behavior as a function of point values rather than average values for boat data (point of closest approach, and maximum number of boats within three radii of the whale). Previous analyses (Williams et al. 2002a) suggested that these maximum and minimum values explained more of the variation in whale behavior than average boat traffic values, but this assertion should continue to be tested. For the purposes of the present study, however, it was important for our traffic variables to be as consistent with previous studies as possible. It may be important for studies in other areas to assess the percentage or absolute duration of time that boats spend within certain distances of whales, but this was beyond the scope of the current study.

Whales adopted more erratic paths (that is, higher deviation indices) on average when boats approached within 100m than when they stayed farther away (Figure 8) – a trend that became non-significant as the number of boats within 1000m (MAX1000) increased beyond approximately 5 boats. However, the variability in the data made it difficult to ascertain the inflection point in the trend with any degree of accuracy. Similarly, the trends between boat number and whale behavior can be discussed with greater confidence for MAX1000 than for boat number within 100 and 400m, because of the models' need for observations across a reasonable range of values. Increasing our sample size, particularly by increasing range of number of boats observed within 100 and 400m during the experimental tracks, would allow us to assess whether whale behavior varied with varying numbers of boats within a radius that is of greater management concern than 1000m. Upon first glance it may appear that occurrence of multiple vessels within 100 and 400m is rare due to the small sample size of multiple vessels recorded in these radii during the

experimental section of a track. However, opportunistic tracks reveal that northern resident killer whales are exposed to multiple vessels within 100 and 400m with some degree of regularity. Therefore, increasing the sample size of boats observed within 100 and 400m of the focal would not represent rare traffic conditions for northern resident killer whales. In the meantime, a reasonable summary of the results of our analyses of whale swimming speed and path smoothness (deviation index) would indicate that whales appeared to swim faster and more erratically when a few boats approached closely than when many boats stayed far away.

Inter-breath interval

Our model was particularly poor at explaining variability in mean dive time (RESP: mean time between respirations). Experimental approaches did not alter mean dive time, either in terms of presence/absence of boats overall, or in terms of approach by few/many boats. The selected GAM fit the data best when including linear terms for month and sex (Table 2), but neither term was statistically significant. The smooth terms that entered the model (Age, Time of Day and Maximum Number of Boats within 1000m) comfortably spanned zero across the ranges of x (Figure 5). Mean dive time, then, seemed to be largely unrelated to boat traffic, and only weakly related to temporal variables in this study.

This is the third study on northern resident killer whales to suggest that they are not using vertical avoidance tactics to evade boats (Williams *et al.* 2002ab, Table 2). A boat-based study on northern resident killer whales in the 1980s also failed to find boat-related variation in whale respiration rate despite analyses of thousands of dives (Dr. Dave Duffus, pers. comm.). However, it does raise an important point: if mean inter-breath interval was the only response variable that we had measured, we might falsely have concluded that vessels do not alter whale behavior, simply because the animals do not alter that one behavioral parameter that we chose to measure. This finding points to the need to recognize this issue as inherently multivariate.

Similarly, mean or median inter-breath interval may be the wrong measure to use for assessing evidence for vertical avoidance, possibly because there is so much variability in dive time. *Maximum* (rather than *average*) inter-breath intervals may well have changed with boat traffic, which could have been masked by presenting only the average dive rate. Alternatively, it may be the *complexity* of diving patterns that changes with respect to boat traffic, in which case fractal analyses may be required to determine whether some point exists at which whales switch from horizontal to vertical avoidance tactics. However, this does not gain much support from the distance data: point of closest approach did not enter the model describing mean dive time.

Finally, it may be that vessel traffic affects killer whale dive time on a temporal scale that was well beyond the scope of this single-season study. Baird *et al.* (2005) examined diving behavior of southern resident killer whales between 1993 and 2002. Like the current study, Baird and colleagues also found that number of dives per hour (dive rate) did vary significantly with age, sex and time of day (Table 2, Figure 5), although our sampling design was unable to monitor diving rate at night. Of particular interest is their suggestion that rates of deep diving in southern resident killer whales have declined since 1993 (Baird *et al.*, 2005). While this trend is consistent with whales' catching up on resting activity at night, when boat traffic is minimal, the authors were unable to discriminate among the many competing explanations for such an effect.

In summary, no consistent pattern between boat traffic and whale diving patterns was observed in our study. However, we recognize that several other analytical tools could have been used. The analyses used here might have resulted in our failure to detect a real effect.

Experimental design

In the process of testing the hypothesis that a node exists at approximately 2-3 boats experimentally, we first had to define *post-hoc* from our trials whether to use 2 or 3 as that node. The bspline-fitting exercise suggested that the node was best placed at 3 boats. This lends support for our decision to use 3 boats as the cutoff between few and many boats in the

experimental data. More importantly, it allows for an objective way of testing whether similar cutoffs exist in other behavioral impact assessment studies.

Intriguingly, the distribution of boaters that happened to approach focal animals during the trials also showed an apparent discontinuity at 2-3 boats (Table 1). It may be that the existence of a node that was hinted at in the earlier study, and demonstrated more convincingly in the current one, may reflect a tendency for whales to respond one way when exposed to what they perceive as typical or average boat traffic. Above that level, they may switch their avoidance tactics. This may well reflect the peak observed in the relationship between swimming speed and the point of closest approach near 100m. Again, perhaps whales behave “normally” when they experience well-behaved whalewatching, but react differently when boats get very close.

Pooling treatment levels

While previous work has shown that focal animals responded to the presence of one experimental boat paralleling (Williams et al., 2002a) or one experimental boat leapfrogging (Williams et al., 2002b) a whale's path, no response was found in this study when the experimental treatment was approached by 1-17 boats. We infer from this that northern resident killer whales perceive boats as a continuum from benign to distracting, and that by pooling a wide range of these stimuli, evidence of whale responses can be masked. In other words, the average of at least two behavioral responses can look very convincingly (Wilcoxon's test: $P > 0.999$) like no response at all.

Ignoring the multivariate nature of responses might lead a researcher to falsely conclude, with a great deal of statistical confidence, that boats do not alter the behavior of killer whales, when in fact, boats altered the behavior of whales in at least two opposing ways. A classic flaw in ecological studies is an attempt to describe a non-linear relationship using a linear model. The model will fit the data poorly, and one will falsely conclude, with a high degree of statistical certainty that there is no effect. In fact there is one, but the effect is just not a linear effect. We believe that something similar may happen in vessel impact studies, where presence/absence analyses may mask effects, simply because the average of a multivariate or non-linear response may look like no response.

General conclusions/recommendations

Experimentation is a powerful way to detect effects of human activity on whales, but only when the interpretation is restricted to effects that the experiment was designed specifically to test. However, in true experimental situations one would control for all other variables and only manipulate one variable at a time. This obviously was not feasible for this study. Ultimately, a study design that includes both experimental and opportunistic data may be best able to allow analysis that can account for confounding effects that remain in our inability to control all other variables in the rare opportunities one has to conduct experiments on free-ranging cetaceans. We believe that we have demonstrated convincingly that yes, northern resident killer whales did react differently to approach by many boats than by few boats, and urge that similar experimental studies begin on other populations with equally modest, measurable research questions and objectives.

This study adds to a growing body of evidence that whales use evasive tactics to avoid boats, but the results have to be interpreted carefully. Consider this: whales responded to the experimental approach of a few boats by adopting a less predictable path; but adopted a *more* predictable path when many boats approached. One might interpret this result as building a management case for preferring many boats to few. Conversely, it makes intuitive sense that all other things being equal, many boats introduce more noise, distraction and pollution than few boats do. It is problematic when one result can be interpreted in such a way that it would allow one to make two opposing recommendations.

We believe it is important to identify whether these stereotyped responses to boats reflect avoidance of a perceived predator. These animals often swim parallel to shorelines – their ability

to follow the optimal predator-avoidance strategy of simultaneously changing velocity and turning radius is restricted in such an arrangement. Alternatively, if the whales are responding to some intervening cause-and-effect linkage such as salmon reaction to boats, then such vessel activity may actually increase whale foraging efficiency. Our ability to make concrete conclusions with confidence is strongly hampered by our lack of information on what animals are doing below the surface.

The type of vessels tracked in this study was highly variable. Those boats that participated in the experimental trials were engaged in whalewatching, but many of the boats in the opportunistic data were fishing boats. Previous work suggested that models fitted the data better when pooling information on whale-oriented and non-whale-oriented traffic (Williams *et al.* 2002a), but perhaps the more sophisticated analytical tools available today (such as the GAMs used in this study) could tease the two variables apart. Future research should try to ascertain whether information on vessel type explains enough of the variability in whale behavior to warrant further consideration. For these and other reasons, direct transfer of the results of this study to presumed impacts on southern resident killer whales, or other areas where vessel traffic is a concern for cetacean conservation, should be done with caution. It may be interesting to note that the focal whales in the experimental phase of this study were responding after just 20 minutes of traffic conditions immediately following 20 minutes in the absence of boats. With southern resident killer whales, a single pod may be followed for 10 hours/day. Our inability to monitor post-treatment recovery of whale behavior strongly hinders our confidence in extrapolating from northern to southern resident killer whale vessel impacts.

Another limitation of the study was its need to focus on male killer whales for the experimental approaches. Ideally, impact assessments ought to target the most vulnerable members of a population (which may be mothers with calves). But this creates an interesting suggestion of selection pressure of whalewatchers on male killer whales. Earlier, we speculated that Kruse's (1991) finding of increased swimming speed as number of boats increased was a statistical artifact of sampling bias: whalewatchers tend to approach male killer whales preferentially, due to their larger dorsal fins, and male killer whales can swim faster than females, although they may be less maneuverable (Williams *et al.* 2002a). So the strong sexual dimorphism in this species may confound our ability to measure impact. Males are more conspicuous than females, are targeted preferentially by whalewatchers, and the whales have sex-based differences in avoidance strategies (Williams *et al.* 2002a). If tall dorsal fins were selected for by females, but also made them more susceptible to being watched by whalewatching boats (and hence run greater risk of having their foraging efficiency reduced by masking effects of boat noise), then we may be seeing competing forces at play. It would be interesting to examine all longitudinally collected theodolite data on northern residents to see if whales' responsiveness to boats changes in a way that is consistent with increased sensitivity or habituation.

Meanwhile, we conclude that northern resident killer whales display a variety of stereotyped cues to demonstrate impact of boat presence, proximity and number on their behavior. Pooling observations of whale behavior under all traffic conditions can make it appear as though there is no impact, when in fact, many different impacts can be occurring simultaneously. We believe that when whale populations are thriving despite extensive habitat degradation, simple descriptive statistics of whale behavior during "average" or "typical" traffic conditions can give a rough proxy for what the animals can tolerate in terms of boat number, saturation of daytime hours with boat presence, or vessel proximity.

Acknowledgements

We would like to thank NOAA Fisheries (Northwest Fisheries Science Center) for funding to allow us to conduct this study, which was aided logistically by partnership with BC Parks and the Robson Bight (Michael Bigg) Ecological Reserve warden program and their staff. Data collection was assisted by Rachael Green and Elaine Leung, with occasional assistance from Jeff Hogan. Cheryl Ciccone in particular provided training and assistance with identification of focal animals. Paul Spong and Helena Symonds of OrcaLab loaned us a theodolite, and David Bain wrote

custom software to allow us to collect and process these data with that equipment. Jodi Smith assisted with processing of many of the raw tracks. Arliss Winship helped produce two of the figures. Constructive discussions with Lars Bejder, David Lusseau, Ben Wilson and Dawn Noren and other scientists from the Northwest Fisheries Science Center improved the manuscript. We thank them for their advice and their critical review of earlier drafts. We also thank Robin Baird and two anonymous reviewers for their thorough and helpful comments. And finally, the whalewatching community of Johnstone Strait provided invaluable assistance by providing both control and treatment observations for the experimental component of our study.

References

- Bain, D. E. and Dahlheim, M. E. (1994). "Effects of masking noise on detection thresholds of killer whales." *in* T. Loughlin (ed.) Marine mammals and the Exxon Valdez. San Diego, Academic Press: 243-256.
- Baird, R. W. (2001). "Status of killer whales in Canada." Canadian Field Naturalist 115(4): 676-701.
- Baird, R.W., Hanson, M.B., and Dill, L.M. (2005). "Factors influencing the diving behaviour of fish-eating killer whales: sex differences and diel and interannual variation in diving rates." Can. J. Zool. **83**: 257-267.
- Bejder, L. (2005). "Linking short and long-term effects of nature-based tourism on cetaceans." Unpublished PhD Thesis, Dalhousie University, Halifax, Nova Scotia, Canada.
- Bigg, M A., Olesiuk, P. F., Ellis, G. M., Ford, J.K.B. and Balcomb, K.C. (1990). "Social organization and genealogy of resident killer whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington State." *in* Hammond, P.S., Mizroch, S.A. and Donovan, G.P. (Eds.) Individual Recognition of Cetaceans: use of photo-identification and other techniques to estimate population parameters. Rep. Int. Whaling. Comm. Spec. Issue 12: 383-405.
- Boyd, I.L., McCafferty, D.J. and Walker, T.R. (1997). "Variation in foraging effort by lactating Antarctic fur seals: response to simulated increased foraging costs." Behavioral Ecology and Sociobiology 40: 135-144.
- Davis, R. E., Foote, F. S., Anderson, J. & Mikhail, E. (1981). "Surveying theory and practice." New York: McGraw Hill.
- DFO. (2002). "Protecting Canada's Marine Mammals: Proposed Regulatory Amendments." Marine Mammal Bulletin December 2002. Available from: http://www-comm.pac.dfo-mpo.gc.ca/pages/consultations/marinemammals/documents/bulletin-dec02_e.htm.
- Erbe, C. (2002). "Underwater noise of whale-watching boats and potential effects on killer whales (*Orcinus orca*), based on an acoustic impact model." Marine Mammal Science **18**(2): 394-418.
- Ford, J. K. B., Ellis, G. M. and Balcomb, K. C. (2000). Killer whales. 2nd Ed. Vancouver, UBC Press.
- Ford, J. K. B., Ellis, G. M., Barrett-Lennard, L. G., Morton, A. B., Palm, R. S. and Balcomb, K. C. (1998). "Dietary specialisation in two sympatric populations of killer whales (*Orcinus orca*) in coastal British Columbia and adjacent waters." Canadian Journal of Zoology 76(8): 1456-1471.
- Hammond, P. S., Mizroch, S. A. and Donovan, G. P. (1990). Individual recognition of cetaceans: Use of photo-identification and other techniques to estimate population parameters (Special Issue 12). Cambridge: International Whaling Commission.
- Harwood, J. (2000). "Risk assessment and decision analysis in conservation." Biological Conservation 95: 219-226.
- Howland, H.C. (1974). Optimal strategies for predator avoidance: the relative importance of speed and maneuverability. J. Theor. Biol. **47**: 333-350.
- Hoyt, E. (1997). The potential of whale watching in Europe. Bath, Whale and Dolphin Conservation Society.
- International Whaling Commission (IWC). (2005). Journal of Cetacean Research and Management **7** (Supplement).

- Johnstone Strait Killer Whale Committee (cited as JSKWC). (1996). Johnstone Strait Whale Watching Guide. BC Min. of Environment, Land and Parks and Dept. of Fisheries and Oceans.
- Johnstone Strait Killer Whale Committee (cited as JSKWC). (1991). Background report. BC Min. of Environment, Land and Parks and Dept. of Fisheries and Oceans.
- Kriete, B. (1995). "Bioenergetics in the killer whale, *Orcinus orca*." Unpublished PhD thesis. Vancouver, University of British Columbia, 138 pages.
- Kruse, S. (1991). "The interactions between killer whales and boats in Johnstone Strait, B.C." in K. Pryor and K. S. Norris (eds.) Dolphin Societies: discoveries and puzzles. Berkeley, University of California Press: 149-159.
- Lusseau, D. (In press). "The effects of tour boats on the behaviour of bottlenose dolphins: Using Markov chains to model anthropogenic impacts." Conservation Biology **17(6)**: 000-000.
- Miller, M. L. (1993). "The rise of coastal and marine tourism." Ocean and Coastal Management **20(3)**: 181-199.
- Morton, A. B. and Symonds, H. K. (2002). "Displacement of *Orcinus orca* (L.) by high amplitude sound in British Columbia, Canada." ICES Journal of Marine Science **59**: 71-80.
- Møller, A. P. and Jennions, M. D. (2002). "How much variance can be explained by ecologists and evolutionary biologists?" Oecologia **10.1007/s00442-002-0952.2**.
- Nelson, J. G. (1994). "The spread of ecotourism - some planning implications." Environmental Conservation **21(3)**: 248-255.
- Nichol, L. M. and Shackleton, D. M. (1996). "Seasonal movements and foraging behavior of northern resident killer whales *Orcinus orca* in relation to the inshore distribution of salmon *Onchorhynchus* spp. in British Columbia." Canadian Journal of Zoology **74(6)**: 983-991.
- Orams, M.B. (2000). "Tourists getting close to whales, is it what whale-watching is all about?" Tourism Management **21**: 561-569.
- Roeder, K. D. (1967). "Nerve cells and insect behavior." Cambridge, MA: Harvard University Press.
- Scheidat, M., C. Castro, J. Gonzales & R. Williams. (2004). Behavioural responses of humpback whales (*Megaptera novaeangliae*) to whalewatching boats near Isla de la Plata, Machalilla National Park, Ecuador. Journal of Cetacean Research and Management **6(1)**: 63-68.
- Weihs, D. & Webb, P. W. (1984). "Optimal avoidance and evasion tactics in predator-prey interactions." J. theor. Biol. **106**: 189-206.
- White, D., Kendall, K. C. and Picton, H. D. (1999). "Potential energetic effects of mountain climbers on foraging grizzly bears." Wildlife Society Bulletin **27(1)**: 146-151.
- Williams, R. (2003). "Cetacean studies using platforms of opportunity." Unpublished PhD Thesis. University of St Andrews, St Andrews, Scotland, UK.
- Williams, R. (1999). Behavioural responses of killer whales to whale-watching: opportunistic observations and experimental approaches. Unpublished MSc thesis, University of British Columbia, Vancouver, BC, Canada. Available from: <http://www.marinemammal.org/pub.htm>.
- Williams, R., Trites, A. W. and Bain, D. E. (2002a). "Behavioral responses of killer whales to whale-watching traffic: opportunistic observations and experimental approaches." Journal of Zoology **256**: 255-270.
- Williams, R., Bain, D.E., Ford, J.K.B. and Trites, A.W. "Behavioral responses of male killer whales to a 'leapfrogging' vessel. (2002b). Journal of Cetacean Research and Management **4(3)**: 305-310.
- Wood, S. N. (2003). "Thin-plate regression splines." J. R. Statist. Soc. B **65(1)**: 95-114.
- Wood, S. N. (2001). "mgcv: GAMs and Generalized Ridge Regression for R." R News **1(2)**: 20-25.
- Wood, S. N. (2000). "Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties." Journal of the Royal Statistical Society, Series B. **62**: 413-428.
- Würsig, B., Cipriano, F. & Würsig, M. (1991). Dolphin movement patterns:

information from radio and theodolite tracking studies. In Dolphin societies: discoveries and puzzles: 79-111. Pryor, K. & Norris, K. S. (Eds). Berkeley: University of California Press.

Zar, J. H. (1996). Biostatistical Analysis. New Jersey, Prentice Hall.

Appendix 1: Additional tables of sample size by age, sex and individual identity.

	Males	Females	Unknown Sex	Total
Experimental No-boat	16	1	0	17
Experimental boat	16	1	0	17
Opportunistic No-boat	22	3	0	25
Opportunistic Boat	63	16	2	81
Total	117	21	2	140

Table 1 (Appendix): Sample size by sex and traffic.

ID	Year Born	Age	Sex	Total # Tracks Per Individual	# of Experimental Tracks	# of No Boat Tracks	# of Opportunistic Tracks
A32	1964	40	M	16	2	4	10
A37	1977	27	M	5	1	2	2
A46	1982	22	M	2	0	0	2
A12	1941	63	F	2	0	0	2
A33	1971	33	M	15	3	4	8
A34	1975	29	F	1	0	1	0
A55	1990	14	M	3	0	0	3
A30	1947	57	F	8	0	3	5
A38	1970	34	M	11	2	0	9
A39	1975	29	M	12	4	1	7
A72	1999	5	UNK	2	0	1	1
A11	1958	46	F	3	0	0	3
A13	1978	26	M	15	1	5	9
A73	2000	4	F	1	0	0	1
A24	1967	37	F	1	0	0	1
A42	1980	24	F	2	0	0	2
A43	1981	23	F	1	0	1	0
A61	1994	10	M	1	0	1	0
B7	1947	57	F	1	0	0	1
B10	1979	25	M	2	1	0	1
B12	1984	20	M	1	0	0	1
B13	1987	17	M	1	1	0	0
B14	1991	13	F	1	0	1	0
C14	1985	19	M	1	1	0	0
I16	1968	36	F	4	1	1	2
I41	1980	24	M	11	1	1	9
I43	1983	21	M	11	0	2	9
G38	1986	19	M	1	0	0	1

Table 2 (Appendix): Sample size by individual.

Appendix 2: Description of cliff height measurement procedure.

The theodolite was located approximately 50m above mean sea level. The theodolite-computer apparatus measured the length of a 30m rope to be 28.93m ($n = 20$, $SE = 0.18$) at a distance of 3.79km. This translates to measurement error of approximately 3.5% in terms of accuracy and <1% in terms of precision. Percent errors in measuring cliff height, distance traveled and speed tend to be approximately equal (Würsig *et al.* 1991).

Appendix 3: Description of experimental tracks.

For the experimental treatment, at least one primary experimental vessel was “watching” the focal whale. All boats additional to the primary experimental vessel were those vessels within 1000m of the focal whale.

During the no boat portion of the experimental tracks, no vessels were within 1000m of the focal whale for at least 20 minutes. Exceptions to this rule included vessels with engines shut down (eg the Reserve Warden boat with which we communicated and verified engine status) and “rafted” or stationary kayaks.

<u>Experimental Track</u>	<u># Boats</u>	<u>Boat Code</u>
1	3	clghb, cf, rsminf
2	2	clghb, cf
3	6	cf, 5 k
4	4	csminfl, 2 plghb, tug
5	8	cmedhb, clghb, 2 cf, rsminf, 2 k
6	3	clghb, psmhb, plghb
7	4	clghb, psmhb, cmedhb, k
8	9	3 clghb, 2 cf, 2 pmhb, psmhb, rsminf
9	2	rsminf, cf
10	16	13 k, 2 clghb, cmedhb
11	2	tc, rsminf
12	12	3clghb, csmhb, 2psmhb, cf, 4 k, rsminf
13	2	clghb, rsminf
14	1	Csminfl
15	3	cf, clghb, 2 k
16	11	2 clghb, 3 cmedhb, csminf, csmhb, 3 k, rsminf

Table 3 (Appendix): Vessel type in each experimental track

Boat Code	Description
Csmhb	commercial small hard bottom
Cmedhb	commercial medium hard bottom
Clghb	commercial large hard bottom
Csminf	commercial small inflatable
Rsminf	research small inflatable
Psmhb	private small hard bottom
Pmedhb	private medium hard bottom
Plghb	private large hard bottom
K	Kayak
Tc	tanker/cargo
Tug	Tug
Cf	commercial fishing
Sb	Sailboat
Col	commercial ocean liner

Table 4 (Appendix): Vessel code descriptions

Appendix 4: Model Specification and Selection Outline

Smooth and linear predictor terms were fitted using multiple generalized cross-validation (mgcv) in package mgcv for R (Wood 2000). The GAM-fitting approaches in R and S-Plus differ in two key ways. First, S-Plus allows users to automate the decision to add or drop model terms in a step-wise procedure that minimizes AIC score. Conversely, R-users must make that decision based on an understanding of the system that they are describing, with guidance from summary statistics describing model fit. Secondly, S-Plus allows users to specify how flexible the model can be, while program mgcv in R automates this process using the patterns that emerge from the data themselves.

If a model were allowed to fit smooth curves without penalty, it would fit a line through every data point: this is biologically implausible, and of no value for prediction. Instead, mgcv models are fit in a maximum likelihood framework in which the model likelihood is modified by the addition of a penalty for each smooth function (Wood 2000). This creates a mathematical disincentive for unnecessary flexibility. On the other hand, one does not want a model that ignores true complexity, (e.g., assuming a linear relationship when the 'true' underlying relationship is a non-linear one) or complex biological relationships due to unnecessarily harsh penalties for additional model terms or degrees of freedom to a given smooth term.

Package mgcv uses generalized cross-validation (GCV) to create a balance between penalizing the flexibility of smooths on the one hand and penalizing badness of fit on the other. In the simplest terms, this involves using part of the data to build a descriptive model, and testing its ability to predict response values of the remaining data. This process is repeated in an iterative function until the GCV score (analogous to AIC) is minimized, and the deviance explained (analogous to an R^2 value) is maximized. The estimation of the shape and flexibility of these smoothing parameters, and the appropriateness of mathematical penalties for bad fit and overparameterization, represent the main differences between GLMs and GAMs.

This minimization of a GCV score (or an UBRE (Un-Biased Risk Estimator criterion (UBRE) score, which also in practice is an approximation to Akaike's Information Criterion, or AIC), selects the best model among *user-specified* candidates. Models were penalized for being overparameterized, and the degree of smoothing was automated for each model term simultaneously using the default validation methods in mgcv. GAMs in mgcv are fitted using data from all variables at once, which avoids a problem with step-wise fitting (as seen in GAM-fitting in S-Plus, for example), whereby the order in which terms enter the model determines how much unexplained variance is left for which subsequent terms could potentially account, and consequently influences the apparent significance of subsequent terms. Therefore, the relationship between any one predictor and a response is assessed and plotted, *conditional on all the other terms being in the model*. It is an objective means of assessing the contribution of any one term to the model, and penalizes the model for unnecessary terms.

The default smoothing value used for regression splines (Wood 2003) was the default value set by package mgcv, 10 knots in each spline, corresponding to 9df (Wood, 2001). In practice, few biological relationships are expected to display this degree of complexity, but setting lower values can cause problems with model convergence. Response variables were modeled as combinations of linear and smooth functions of a suite of candidate predictors using the quasi family, which allowed the dispersion parameter to be modeled from the data in a maximum-likelihood framework.