

**EFFECTS OF VESSELS ON BEHAVIOR OF SOUTHERN RESIDENT KILLER
WHALES (*ORCINUS* SPP.)**

David E. Bain,
Friday Harbor Labs,
University of Washington,
Friday Harbor, WA

Jodi C. Smith,
Friday Harbor, WA

Rob Williams,
Alert Bay, BC

and

David Lusseau,
University of Aberdeen
Cromarty, UK

NMFS Contract Report No. AB133F03SE0959 and AB133F04CN0040

March 4, 2006

EFFECTS OF VESSELS ON BEHAVIOR OF SOUTHERN RESIDENT KILLER WHALES (*ORCINUS* SPP.)

David E. Bain, Jodi Smith, Rob Williams and David Lusseau

ABSTRACT

Vessel traffic may have contributed to Southern Resident Killer Whales becoming endangered. To determine the importance of this threat, we measured behavior of Southern Residents in the presence and absence of vessels in 2003 and 2004 at two different sites along San Juan Island. Data collected include: theodolite tracks of focal individuals, along with observations of their behavior; and scan sampling of activity states of subgroups, along with counts of vessels at various distances from each subgroup. Theodolite tracks were summarized in terms of directness and deviation indices, and travel speed. Rates of respiration and display behaviors were also determined for each focal sample. Vessel number and distance were used as candidate explanatory variables for differences in track indices and other behavior, along with natural factors such as sex, age, pod membership, time of day, time of year, geographic location, etc. As with Northern Residents, directness index decreased significantly in the presence of vessels. This increase in distance traveled in the presence of vessels would result in increased energy expenditure relative to whales who can rest while waiting for affected whales to catch up. Rates of display behavior increased significantly as the distance to the nearest vessel decreased. Transitions between activity states were significantly affected by vessel traffic, indicating a reduction in time spent foraging as was observed in Northern Residents. If reduced foraging effort results in reduced prey capture, this would result in decreased energy acquisition. Each subgroup was within 400m of a vessel most of the time during daylight hours from May through September.

INTRODUCTION

The Eastern North Pacific Southern Resident Stock of killer whales declined to fewer than 80 individuals in 2001, resulting in their listing as “Depleted” under the Marine Mammal Protection Act and a proposed rule to declare them “Threatened” under the Endangered Species Act. It has been declared “Endangered” by the State of Washington and Canada. The causes of this decline are uncertain, but many scientists consider a combination of reduction in prey resources, toxic chemicals, disturbance from vessel traffic, and other factors to have contributed (Bain *et al.* 2002, Wiles 2004, Krahn *et al.* 2002 and 2004, Federal Register 2004, Killer Whale Recovery Team 2005).

Krahn *et al.* (2004) noted that the Southern Resident killer whale population increased at a normal rate in the late 1980’s. Growth began to slow in the early 1990’s and was followed by a decline of 20% from 1996 to 2001. J and K pods exhibited no growth during this period, so the decline in the population size can be considered to be due to the decline in L Pod. Factors in the inshore waters of Washington and British Columbia, such as declines in prey abundance, toxins and vessel traffic may be responsible for the lack of growth in all three pods. Differences in usage patterns of the inshore waters among the different pods (Bigg *et al.* 1990, Olesiuk *et al.* 1990) may account for some of the additional decline experienced by L Pod alone, but factors

external to these waters (perhaps entanglement, exposure to oil, etc.) are likely to be of similar importance to factors in inshore waters, as roughly half of the change in population growth rate resulted in the failure to increase, and the other half resulted in population decrease.

Vessel traffic may have contributed to the decline through a variety of mechanisms. Collisions between vessels and killer whales occur occasionally in Northern Residents and other killer whales and result in injury or death, and one case was observed in Southern Residents in 2005 (Visser 1999, Ford *et al.* 2000, G. M. Ellis pers. comm., K. C. Balcomb pers. comm.). Chemicals such as unburned fuel and exhaust may contribute to toxin load. The presence of vessels may contribute to stress (Romano *et al.* 2004). Noise from vessel traffic may mask echolocation signals (Bain and Dahlheim 1994) reducing foraging efficiency. Behavioral responses may result in increased energy expenditure, or disrupt feeding activity, which may reduce energy acquisition (Bain 2002, Bain *et al. unpublished ms*). Energetic mechanisms for impact are of particular concern, since Southern Resident Killer Whales may be food limited (Ford *et al.* 2005).

It stands to reason that repeated disturbance of wild animals could be implicated as a factor reducing the quality of life, foraging efficiency, fitness, or reproductive success of individual animals. Examples in the wildlife literature link anthropogenic disturbance to changes in foraging behavior (*e.g.*, Galicia and Baldassarre 1997), reproductive success (*e.g.*, Safina and Burger 1983), and mating system and social structure (*e.g.*, Lacy and Martins 2003). These in turn, either singly or synergistically, could influence population dynamics (Bain *et al. unpublished ms*).

Effects of vessel traffic have been studied in a range of cetacean species, including *Cephalorhynchus*: Bejder *et al.* (1999); *Delphinus*: Constantine (1997); *Eschrichtius*: Jones (1988), Duffus *et al.* (1998); *Globicephala*: Heimlich-Boran (1993), Heimlich-Boran *et al.* (1994); *Megaptera*: Corkeron (1995); *Orcinus*: Kruse (1991), Williams *et al.* (2002ab), Foote *et al.* (2004); *Physeter*: Fleming and Sarvas (1999); *Sousa*: Van Parijs and Corkeron (2001); *Stenella*: Angradi *et al.* (1993), Ritter (2003); *Tursiops*: Janik (1996), Allen and Read (2000), Nowacek *et al.* (2001), Constantine (2001); and *Ziphius*: Ritter (2003). Effects vary within and between species, and included changes in respiration patterns, surface active behaviors, swimming velocity, vocal behavior, activity state, inter-individual spacing, wake riding, approach and avoidance, and displacement from habitat. Collisions may result in injury or death (Wells and Scott 1997, Laist *et al.* 2001). More detailed reviews of vessel effects can be found in Lien (2001) and Ritter (2003).

Kruse (1991) and Williams *et al.* (2002ab) demonstrated short-term behavioral changes in Northern Resident killer whales associated with vessel traffic. Kruse (1991) found Northern Residents increased swimming speed as vessel number increased. Nowacek *et al.* (2001) found *Tursiops* also increased swimming speed in the presence of vessels. Williams *et al.* (2002ab) found Northern Residents swam in less predictable paths in the presence of vessels, and *Tursiops* exhibit similar behavior (Nowacek *et al.* 2001). Williams (unpublished data) found Northern Residents were less likely to forage in the presence vessels, and *Tursiops* exhibit the same change in parts of their range (Allen and Read 2000). Adimey (1995) found percussive behavior of Northern Residents was inhibited in the presence of vessels, though Williams *et al.* (2002ab) found no significant differences. However, for Southern Resident killer whales in the waters of Washington and British Columbia, even subtle behavioral responses to boats have not been

reported in the primary literature. This is a critical area of study because the San Juan and Gulf Islands are a region with high vessel traffic.

In this region, the commercial whale watching day runs from about 0900-2100 in summer, and until sunset in spring and early fall. In addition to commercial whale watching vessels, other vessels are also in contact with whales throughout the day. Early in the morning (sunrise), whales are approached by recreational vessels transiting the area, scientific research vessels, and sport fishing vessels. For part of the season, seiners and gill netters are also present. In the middle of the day, these boats are joined by the commercial whale watching fleet, and a few of these commercial whale watching vessels remain with whales until near sunset. Homeland security vessels are on the water much of the day, and sometimes approach whales or vessels near whales (pers.obs.). Further, commercial freight traffic is intermittently present 24 hours a day. Due to the variety of vessels observed in the presence of whales, the term whale watching as used in this paper refers to all whale-oriented vessel traffic, regardless of whether the vessels are commercial whale watching vessels or not. Because these whales are in the presence of vessels, including those not focused on whale watching during much of the day, it is important to investigate whether the behavior of killer whales is altered in the presence of vessels. This study addresses relationships between vessel activity and Southern Resident killer whale behavior.

METHODS

Study areas

From 28 July to 30 September 2003, and 1 May to 31 August 2004, a land-based team of observers monitored behavior of whales and activity of boats from two study sites (Figure 1). One site (hereafter referred to as the North Site) was located at 48° 30.561' N, 123° 8.494' W at an altitude of approximately 99m above mean lower low water. This site was chosen because its height offered an expansive and unobstructed view of the central and southwestern portions of Haro Strait, whales were known to pass it frequently while traveling close to shore, and it was located adjacent to the voluntary no-boat zone at Lime Kiln State Park.

The other (South) site was located at Mt. Finlayson, near the southeast corner of San Juan Island. The South site was located at Mt. Finlayson (48° 27.421' N, 122° 59.401' W) at a height of 72m and the view of the eastern portion of Juan de Fuca Strait was unobstructed. Further, whales have been reported to use this area heavily for foraging, whereas the North site appeared to be used primarily for travel and socializing (Felleman *et al.* 1991, Hoelzel 1993, Heimlich-Boran 1988). Together, these sites were chosen to maximize sample size and to allow the behavioral observations to include the entire repertoire of the population.

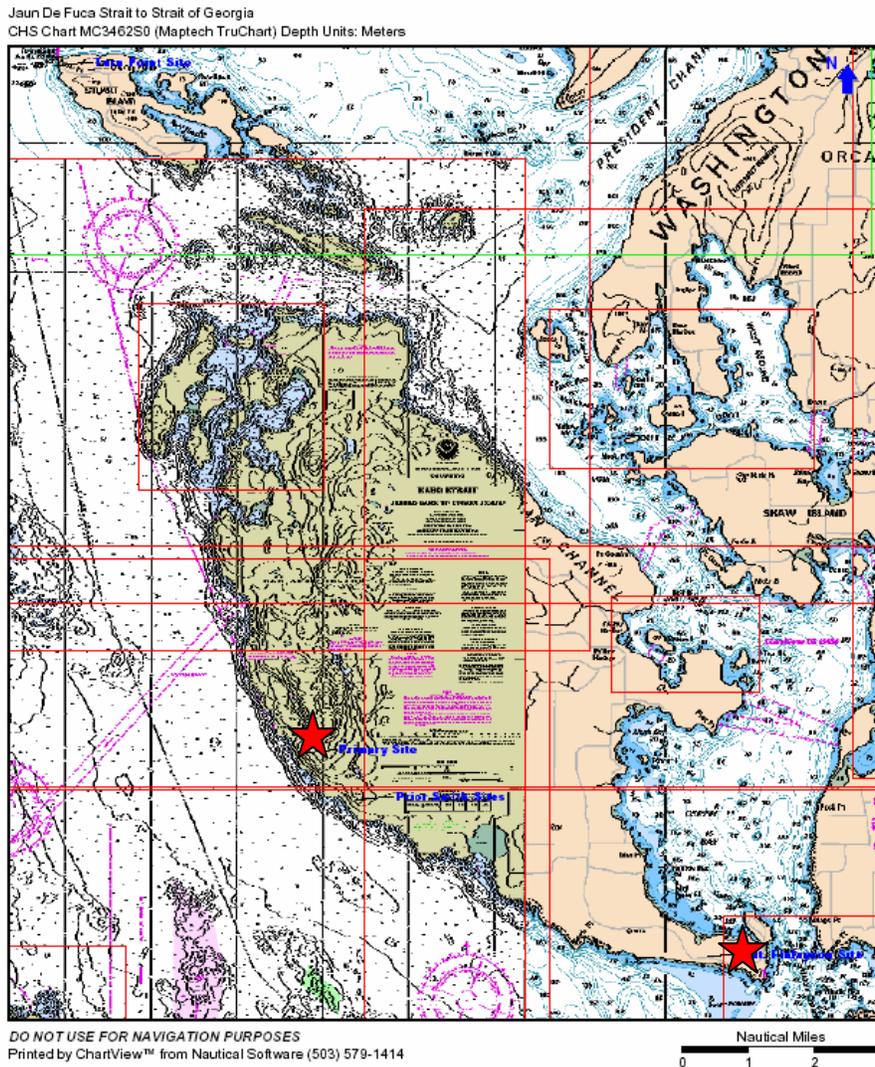


Figure 1. The study area, with the North and South theodolite sites marked with stars.

Research Teams

The team worked for 60 of 64 days in the summer of 2003. In total, 412 hours were spent searching for whales, or monitoring their behavior. Of these 60 days of research effort, whales were present on 38 days and absent on 22 days, or data were lost due to inclement weather (rain, fog, or Beaufort sea state 3 while whales were present). The team worked 6 days a week in May of 2004. From June through August 2004, the group divided into two teams to allow data collection every day. However, effort varied with 8 hours a day effort on three days of the week, 12 hours a day two days a week, and 14 hours a day the other two days a week. Data were obtained on 60 of 118 days in 2004.

The study design involved two simultaneous data collection protocols. One observer collected broad-scale samples of the activities of all whales in the study area at 15-minute intervals. The rest of the team collected fine-scale, continuous, observations of a focal animal. The two methods will be referred to subsequently as scan-sampling and theodolite tracking respectively, and are described in greater detail below. In 2003, the team worked from 6 a.m. until 10 a.m., and then worked on an on-call basis daily until approximately 6 p.m. The exact timing of the research schedule was modified on an *ad hoc* basis from one day to the next, based on a combination of reports from monitoring of VHF commercial traffic and the local sighting network and weather conditions, in order to maximize time spent observing whales in the absence of boats. In 2004, the research day was extended until 8 p.m., although the number of individuals working varied from three to six, and not all hours were covered every day.

Collection of scan-sampling data from focal groups

Scan sampling was also conducted at 15 minute intervals to characterize subgroup size (ranging from one to the size of the school in the study area), activity state, and the number of vessels within 100, 400 and 1000 meters. Vessels were counted separately depending on whether or not they were engaged in whale watching, although commercial and recreational whale watching boats were not distinguished in scan sample counts. Distances were estimated by eye, and checked against measurements with a theodolite when possible to improve observer reliability with experience. Sequential observation of focal groups allows modeling the probability of animals' switching from one coarse activity state to another as a function of vessel traffic. This aspect of the study complements the fine-scale focal animal studies by including all age-sex classes, and all activity states.

A focal group was defined as animals within 10 body lengths of one another at the time of a scan-sample observation, using a chain rule. That is, each individual was within 10 body lengths (approximately 80-100m) of another individual in the group, but large groups could extend over 100's of meters. Thus, our subsequent use of the terms group or school implies nothing about the relatedness of animals within a group and whether all group members were engaged in the same behavior. Similarly, focal groups could be of size one. Group membership was recorded for each identifiable individual. When individuals were too far away to be identified, their identity was assigned to categories based on size (e.g., calf, juvenile, medium sized whales [large juveniles or adult females], subadult male, adult male). When group composition remained unambiguous over time, but individual identity was unknown within the group, groups were given arbitrary labels (a, b, c...) in order to track their activity over time.

The activity of the focal group was recorded every 15min using the following definitions: The sub-categories (1-9) could be combined to either match the categories described by Ford *et al.* (2000) or Smith and Bain (2002) and Waite (1988).

Rest: characterized by prolonged surfacing in contrast to the rolling motion typically observed during travel

1. Deep rest, hanging, logging: whales do not progress through the water
2. Resting travel, slow travel: whales progress through the water, although they may not make forward progress over the ground.

Travel: characterized by a rolling motion at the surface, progress through the water, and membership in a subgroup of more than four individuals

3. Moderate travel, medium travel: travel in which whales do not porpoise
4. Fast travel: travel which includes porpoising

Forage: characterized by progress through the water by lone individuals or while a member of a subgroup of four or fewer individuals

5. Dispersed travel: foraging in a directional manner
6. Milling, feeding, pursuit of prey: foraging involving changes in direction

Socialize: interaction with other whales, or other species in a non predator-prey context

7. Tactile interactions: socializing that involves touching another whale, such as petting or nudging
8. Display: socializing that does not involve touching, but may include behaviors such as spy hops, tail slaps and breaches

Object play: tactile interaction with an object such as kelp, wood or fish (in a manner not related to feeding)

9. Kelping, object play: (when kelping also involved tactile interaction, it was counted as tactile interaction rather than object play.)

These definitions are shown in “dimensional” format (Marsh 2004) in Table 1. A subgroup size dimension was added, as it formed part of the operational distinction between states 3 and 5. These definitions are the product of a workshop attended by experienced killer whale observers and are intended to standardize definitions and allow comparison between studies. Workshop participants recognized that observers may not be able to record all aspects of behavior. Thus some dimensions of behavior are not listed in the table and data on those aspects of behavior were not recorded (e.g., orientation, acoustics), while other aspects of behavior were recorded, although they did not distinguish among behavior states (e.g., respiration). While the relationship between respiration rate and activity states were not analyzed for this report, the data could be applied to energetic studies addressing activity state, respiration rate, and swimming speed (e.g., Kriete 1995), and the table helps identify the suitability of the data for other purposes. Further, studies focusing on other events (e.g., prey capture, Hanson pers. comm.) could be used to assess the appropriateness of the definitions used here (e.g., for foraging).

Table 1. Activity state definitions using the dimensional system. All behavior states could consist of any orientation of individuals, degree of respiratory synchrony, acoustic behavior, and respiration rate, so these dimensions are not shown in the table. Distinctive characteristics of behavior states are highlighted with bold type.

State	Directionality	Inter-individual Distance	Speed	Events	Time	Subgroup Size
1	N/A	>0	motionless	Respiration only	>=1 surfacing	Any
2	Directional	>0	Slow	Respiration only	>=1 surfacing	Any
3	Directional	>0	Medium	Respiration only	>=1 surfacing	>4
4	Directional	>0	Fast, Porpoising	Respiration, porpoising	>=1 surfacing	Any
5	Directional	>0	Medium	Respiration only	>=1 surfacing	<=4
6	Non-Directional	>0	Medium, Fast	Any	>=2 surfacings	Any
7	Any	Contact	Any	Any	>=1 surfacing	>=2
8	Any	>0	Any	At least Percussive, fluke displays, or spy hops, No objects	>=1 surfacing	Any
9	Any	>0	Any	At least contact with objects	>=1 surfacing	Any

Analysis of scan-sampling data from focal groups

We sampled behavior every 15 minutes, allowing us both to consider current behavior and how behavior changed over 15 minute intervals. This additional information is rarely tapped into, yet it lends itself very well to impact studies because it allows one to directly assess what the likelihood is for animals to go from one state to another depending on the occurrence of a potential impact between two samples. Understanding the recurrence of activity states allows one therefore to understand the likelihood that a state will be disrupted by, in our case, boat presence. The data was divided into a series of scan samples of a focal group which were treated as samples of activity state sequences. A sequence stopped when sampling stopped on a given day or when a focal group ceased to exist due to changes in group membership (through fission or fusion with other individuals) or because they left the study area. Activity states were defined in Ford *et al.* (2000). For the purpose of this study we were only interested in understanding the change in the likelihood that when a group was in State A that they would be in State B 15 minutes later (*i.e.*, at the next scan). These are called first-order transitions in activity. This sequence of discrete time samples could be treated as a Markov chain (Lusseau 2003, 2004) because it was ergodic. A time series is ergodic when transitions between all states are possible; in this study a group could transit from any state to another (there was no biological constraint preventing whales from switching between some states and others). The other requirement for a time series to be ergodic is that there cannot be negative values for transition probabilities; since the sequence was bounded by time, sequences could only move in one way; that is forward in time, and therefore no negative values could be expected.

To understand the effect of boat interactions on the state transitions, the number of vessels in the field of view was counted, as these vessels may have contributed to ambient noise in the area (Bain, pers. obs.). The number of vessels within 100 m, 400m, and 1000m of subgroups were also counted. Distances were estimated visually as range rings around individuals or groups, but checked with a theodolite when possible. The numbers within specific distances were used as candidate explanatory covariates, to assess whether the probability of animals switching among activity states varied as a function of boat traffic.

We therefore constructed a transition matrix, representing the probabilities for whales to be observed in a State i at time t and subsequently in State j at the next sampling event ($t + 15$ minutes):

$$p_{ij} = \frac{e_{ij}}{\sum_k e_{ik}}$$
 where e_{ij} is the total number of times the transition was observed and $\sum_k e_{ik}$ is

the total number of time State i was observed as the starting state.

This transition matrix is based on an ergodic time series which means that eigenanalysis of this matrix reveals several properties of activity states. Applying the Perron-Frobenius theorem we show that the transition matrix long-term behavior, *i.e.*, the amount of time that the whales spent in each activity state can be approximated by the left eigenvector of the dominant eigenvalue of the matrix (Lusseau 2003). Ultimately, this approach can be used to calculate stable, unbiased time-activity budgets.

We were able to explore the effects of several parameters on the likelihood to go from one state to another (Lusseau 2003). We used log-linear analysis, LLA (SPSS algorithm), to test whether Site (North/South), Year (2003/2004), Pod (J, K and L), or Vessel Traffic (boat present/absent within 100, 400 and 1000m) affected transitions in activity states, which was the likelihood that focal groups went from a preceding behavior (state at time t) to a succeeding behavior (state at time $t+15\text{min.}$). Log-linear analyses can be thought of as generalized linear models for count data.

In a simple case in which we only have three independent variables (for example: **Boat presence**, **Preceding behavior**, and **Succeeding behavior**), we can assess the three-way effect by comparing the model containing all two-way effects (preceding behavior by succeeding behavior, preceding behavior by boat presence, succeeding behavior by boat presence) to the fully saturated model. This three-way interaction corresponds to the effect of boat presence on the state transition. In each case, the only difference between a candidate model and the fully saturated model is the effect we are trying to assess (the three-way interaction). An objective means of model selection is achieved by subtracting the maximum likelihood (approximated using G^2) of the two-way model from the one of the fully saturated model and testing the significance of this difference. This technique is described more in detail in Lusseau (2003) and (2004).

To understand the interaction between the different independent variables another LLA was carried out which included all independent variables for which there was enough samples to study their interactions.

We analyzed the scans containing distances between vessels and groups to determine the proportion of time groups spent within 100, 400, or 1000m of the nearest vessel (e.g., proportion of time within 100m = the number of scans with boats within 100 m / the number of scans in which vessel distances were recorded).

Theodolite tracking of focal individuals and boats

The theodolite tracking team consisted of three individuals who moved opportunistically between the two study sites to maximize sample size. The team recorded boat and whale positions and activity using a Pentax ETH-10D theodolite interfaced to a PC-compatible computer running Theoprog (Williams *et al.* 2002ab), a Bushnell 40x spotting scope, binoculars, and a mini-DV camera (see DeNardo *et al.* 2001).

As whales entered the field of view from a study site, a focal individual was selected. This individual was identified based on Ford *et al.* (2000) and more recent catalogs (van Ginneken *et al.* 2000 as updated annually by the Center for Whale Research) and tracked for at least 15 minutes. After a tracking session was completed, a new focal individual was selected, if possible. Individuals were selected haphazardly, but were drawn as evenly as practicable from all pods, age, and sex classes (that is if recent tracks had been of adult males, then subsequent selections were biased toward females and juveniles and vice versa, and whales from pods rarely

present were selected over whales from a pod consistently present). We attempted to choose individuals that would not be confused with other individuals nearby, and that were sufficiently close to shore to be accurately identified (typically within 3 km, although this varied with lighting, fog, and individual distinctiveness). Over 40% of the population was sampled at least once during these two seasons, and approximately equal numbers of tracks of males and females were obtained in 2004, indicating we were successful in getting a diverse and balanced sample. The theodolite was used to record position of the focal individual at as many surfacings as possible, and the spotting scope and computer operators, who had a wider field of view, watched for surfacings missed by the theodolite operator, to ensure an accurate record of respiration rate and surface active behavior. We typically collected data only when it was not raining and the sea state was less than three, as whitecaps made tracking significantly more difficult, and rain typically impaired visibility to the point that it was impossible to identify individuals.

While the focal whale appeared to be down on a long dive, the theodolite operator recorded vessel positions. Vessels were classified as commercial whale watching vessels, research and management vessels, commercial fishing vessels, recreational motor boats, sail boats, kayaks, or freight vessels. Estimated size and vessel type was also recorded (small = under 20', medium = 20-40', and large = over 40', inflatable or hard-bottomed).

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) and other notes (Williams *et al.* 2002ab). Boat and whale data were summarized for each track, such that each track was represented only once in the analyses. Independent variables included those related to: Time (Year, Month and Time of Day); Location (Site); Focal Animal (Age, Sex); and Vessel Traffic (Point of Closest Approach, Number of Different Boats observed near the focal during the track, Number of boats within 100, 400 and 1000m of the focal whale, and Number of boats observed within the observers' field of view during the track). Calculation of these candidate explanatory variables is described in greater detail in Williams *et al.* (2002ab).

The five dependent (*i.e.*, whale response) variables included:

1. Inter-breath interval [**RESP**]: A mean time between breaths was calculated (in seconds) for each track. The mean inter-breath interval was defined as the number of intervals (one less than the number of breaths) divided by the time from the onset of the first breath to the onset of the last breath. Only tracks lasting more than 15 minutes were included in the analysis to ensure the data reliably reflected the ongoing breathing pattern (Bain 1986, Kriete 1995).
2. Swimming Speed [**SPEED**]: The average swimming speed of the whale was obtained by dividing the total distance travelled 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. Speed was not corrected for the vertical component or underwater meandering, as underwater behavior was generally unknown, nor was it

corrected for current, which is highly variable spatially in the study areas, so tabulated current only serves as an approximation.

Two measures of path predictability were calculated: a *directness index* and a *deviation index*.

3. Directness Index [**DI**]: The directness index measures 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 and multiplying by 100. 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).
4. Deviation Index [**DEV**]: The deviation index measures 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. 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 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.

Analysis of theodolite data from focal individuals

Theodolite heights were measured using the Survey program in the Theoprog package (Williams *et al.* 2002ab). A tape measure was stretched along the shoreline at sea level, and theodolite readings were taken of the end points. A tide table was used to estimate tide height at the time of the measurement. The length of the tape measure, theodolite readings, and tide height were entered into Survey (a part of the Theoprog package), which calculated the theodolite height above mean lower low water. This process was repeated ten times and the resulting heights averaged. In a previous study, this method was compared against a measurement by a professional surveyor using GPS technology, and produced agreement within 5 cm (Smith and Bain 2002 and see also Bailey and Lusseau 2004).

These heights were entered into Theoprog to convert theodolite readings to X-Y coordinates. Theodolite height was corrected for tide using interpolations between tabulated values updated every ten minutes. The accuracy of the calculated heights and tidal corrections was verified by "tracking" the shoreline and other charted landmarks and plotting the resulting locations on a nautical chart.

For each track, the location of each surfacing by the focal individual was calculated. In addition, locations of vessels marked with the theodolite were calculated. The sequence of surfacing locations was used to calculate the distance and direction traveled between successive surfacings. The time between the first and last point in the theodolite track was the elapsed time. In turn, these values were used to calculate swimming speed (surface speed was the sum of the distances traveled between each pair of surfacings divided by elapsed time, while progressive speed was the distance between the first and last point divided by elapsed time), directness index, and deviation index. Breaths missed by the theodolite operator but observed by another member of the research team were added to breaths observed by the theodolite operator to determine the number of breaths during the track. One was subtracted from this number to determine the number of intervals, and divided by the elapsed time between the first and last point in the theodolite track to calculate the mean inter-breath-interval. Surface Active Behavioral bouts by the focal whale were counted and divided by the elapsed time to determine the mean rate (per hour) of this behavior.

The boat count for a track was the maximum of three types of values. First, the computer operator did boat counts when there were breaks in the tracking (e.g., at the start and end of a track, and occasionally during long dives if boats weren't being marked). Second, the scan sampler did boat counts every 15 minutes, so normally one of these took place during a track (sometimes more for longer tracks). These are both instantaneous counts. The third count was the number of different vessels actually tracked. This number was cumulative, so was potentially greater than the maximum present (from the researchers perspective, though not necessarily the whales perspective) at any given instant, but would be an undercount when not all vessels were tracked.

For number of vessels at specific distances (100, 400, 1000), only the scan sample count was used, so these were instantaneous counts that took place at a moment that was independent of the start and end times of the track and trends in vessel number.

Many vessels are present intermittently. For the instantaneous counts, if a vessel happened to be present when the count was made, it got counted. Otherwise, it did not. For the third count, whether the vessel got counted depended on whether the theodolite tracker marked it. That depended on how close to the focal it got, and how many other vessels were closer. We used a single value, the maximum, to represent the whole track--we did not try to analyze tracks based on whether vessel numbers were consistent or variable.

A spreadsheet was then prepared containing candidate explanatory variables and the five response variables (plus progressive speed, although this is redundant once surface speed and directness have been calculated) for each track.

Only tracks lasting more than 15 minutes were included in the analysis. If a whale was lost briefly (e.g., behind a boat or in glare, or was missed when first surfacing after a long dive), the track was used. Respiration rate was corrected for surfacings observed by members of the team other than the theodolite operator. No corrections were made to deviation and directness indices. As a result, tracks with missed breaths would have artificially low deviation and artificially high directness indices, but the error was small as long as the proportion of breaths

missed was small (on the order of 33%). We tested for bias by comparing results with percentage marked to determine whether tracks with a higher percentage of missed breaths were suitable for use. If too many surfacings were missed, bad portions were eliminated from the record, and whether the track was used at all depended on whether there was a 15 minute segment within the track that met the criteria for use.

These data did not lend themselves to straightforward analysis. We approached the analysis in phases. The first was a naive, preliminary, binary analysis. Values for each track were assigned to a vessel present or vessel absent condition. Tracks were considered to have vessels present if either of the following conditions were met: 1) the interpolated position of at least one vessel was within 1000m of the focal whale at any time during the track, or 2) the scan sampler recorded at least one vessel within 1000m of the focal individual. The binary analysis ignored the potential for factors other than vessel traffic to have influenced the values in the vessel present and vessel absent datasets, but since the sampling protocol was designed to be as representative as possible of real world conditions, these values provide a best estimate of average behavior in the presence and absence of vessels. That is, this analysis provides good descriptive statistics, but for reasons discussed below, the statistical significance of the binary analysis should be treated with extreme caution.

We tested the data for normality, but since they were not normally distributed, we ruled out the use of simple statistics like a t-test. Due to the limited power of data sets with small sample sizes, we elected not to use non-parametric statistics, either. Therefore, we performed a Monte Carlo simulation (1000 iterations) to determine the probability, given the distribution in the vessel absent data, that a sample the size of the vessel present data would have means at least as divergent as those observed, if they had been drawn from the same distribution as the vessel absent data. This level of analysis simply determines whether the no-boat and boat data are drawn from the same population. A result indicating they are from the same population could be misleading, because effects could cancel out to give the appearance of no effects. Similarly, since Williams et al. (2002b) found a variety of variables other than vessel traffic influenced behavior, if all other things are not equal, a factor other than vessel traffic could be responsible for differences between the two datasets.

Thus we performed a more detailed analysis to test whether potentially confounding variables provided a better explanation for differences between the two datasets. Each track was considered an independent sample of animal behavior. It is unlikely that repeated observations of the same individual under different traffic conditions are statistically independent in the strictest sense, however in a small, endangered population, sample size will always be limiting. To that extent, we chose an analysis framework that accounted for as much of the individual variability as possible, holding these natural covariates constant while modeling effects of the variables of interest. We knew, *a priori*, that our modeling approach would have to be a flexible one. Candidate explanatory variables included: binary variables (Year, Site, Sex); factors with varying numbers of levels (Month, Day, Hour, Pod and Age); continuous variables (Point of Closest Approach, Tide height, Current speed, and two measures of data quality--the Percentage of surfacings successfully located with the theodolite and the Duration of the track) and count data from the variable of interest (boat counts at four spatio-temporal scales: the three range rings and the number of unique boats identified by the theodolite operator as having interacted

with the whale). Similarly, the five response variables were all bounded by zero. They included those that might be expected to have derived from: a Gamma or log-normal distribution (perhaps swimming speed and inter-breath intervals); a quasi-Poisson distribution (expected number of surface-active events per hour); and two artificially constructed variables whose theoretical underlying distribution is not intuitive (deviation and directness indices), but are known to be bounded (between 0 and 180°, and between 0 and 100, respectively). Many of these variables can be expected to have violated assumptions underlying traditional linear modeling, such as homoscedasticity and normality. Sample size will not be equal, given the unpredictability of the movements of both people and free-ranging cetaceans. Finally, there is no reason to assume that any relationships between human activity and whale behavior ought to be linear, but neither can one derive from first principles the predicted shape that these relationships ought to follow.

We attempted to address as many of these problems as possible by describing heterogeneity in whale behavior using generalized additive models, GAMs (Venables and Ripley 2002). Generalized additive models (GAMs) were fitted in package **mgcv** (multiple generalized cross-validation) for program R (Wood 2001). Unlike the GAM implementation in S-Plus, the **mgcv** approach uses thin-plate regression splines (Wood 2003) for the smooth terms of each explanatory variable, but each spline carries a penalty for excessive flexibility (Wood 2000). Flexibility is determined by the number of ‘knots’ (approximately one higher than the estimated degrees of freedom, edf) for each model term, between which the functional, or smoothed, relationship was modeled. Smoothing splines were fitted using multiple generalized cross-validation (GCV). In other words, the amount of flexibility given to any model term was determined in a maximum likelihood framework by minimizing the GCV score of the whole model (i.e., given the other terms in the model), rather than each component score. In other words, models were penalized for being over-parameterized, and the degree of smoothing was automated for each model term simultaneously. This avoided the problem common to many step-wise procedures, whereby the order in which terms are presented to the model influences the apparent significance of subsequent terms.

The default smoothing value used for splines was the default value set by package **mgcv**, 10 knots in each spline, corresponding to 9 degrees of freedom (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. Histograms of the response variables were used to determine the appropriate family distribution and link function. Variables that approximated a normal distribution were modelled using the quasi family. Rates of surface-active behavior were expected to approximate a Poisson distribution, given that they derived from count data. A quasi family with a log link was chosen for this analysis, which allowed the dispersion parameter to be modelled from the data. All others were fitted using the quasi-likelihood family with an identity link, which allows the underlying distribution to be modelled in a maximum-likelihood framework.

While determination of the optimal amount of smoothing is automated by **mgcv**, the decision whether to include or drop a model term is not, so the decision whether to do so was guided by a set of criteria described below. Potential explanatory variables considered for inclusion in the model were *Year, Month, Julian Day, Time, Current, Site, Pod, Age, Sex, Point*

of Closest Approach (*PCA*), number of boats within 100m (*SUM100*), number of boats within 400m (*SUM400*), number of boats within 1000m (*SUM1000*), number of unique boats identified by the theodolite operator during the tracking session (*BOATS*). Factor variables were entered as linear or grouping terms. Continuous variables were entered as candidates for smoothing ($s(x)$) by mgcv.

However, the above suite of candidates pushes the limit of the analysis given our sample size, so we analyzed the remaining three parameters separately. We examined the relationship between percentage of surfacings marked and the five behavioral parameters to determine which tracks had acceptable accuracy, and excluded tracks with fewer than 2/3 of surfacings marked. We did not consider the percentage marked in subsequent analyses.

We performed a similar analysis based on tide height, and found no obvious trends, so eliminated it from further consideration.

We performed a similar analysis based on track duration, and found no obvious trends, except in the case of SAB, so we excluded duration from further consideration except in the analysis of surface active behavior. For SAB, we examined both rates per unit time, which was negatively correlated with track duration, and probability of occurrence (one-zero sampling) which was positively correlated with track duration. This is discussed further below.

We have identified another limitation to the use of GAMs with small data sets such as ours, namely their inability to incorporate mixed-effects (*i.e.*, to account for repeated measures of individuals). We addressed this by including candidate covariates, such as *Age* and *Sex* that were likely to have made pseudoreplication an issue. However, the overwhelming advantage of the mgcv approach in R is that it assesses the contribution of each term to the model given the effects of the other terms simultaneously. We believe that avoiding the problem common to many step-wise procedures (*i.e.*, conflating importance of each term with the order in which it enters the model) was important enough to justify using this technique.

The following summarizes our 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: $\{y \sim Year + Month + JDay + Time + Site + Pod + s(Age) + Sex + s(PCA) + s(BOATS) + s(SUM100) + s(SUM400) + s(SUM1000) + Current$ 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 R-squared score) 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.

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 whether they included zero across the range of observations. If so, the term was dropped temporarily, to see whether 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 values for R-squared and deviance explained). 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.

RESULTS

SCAN-SAMPLING OF FOCAL GROUPS

Over the two field seasons we observed 352 behavioral transitions (135 in 2003, 217 in 2004, out of 373 and 1058 scans, respectively). The difference between number of transitions and the number of scans is due to two factors. One, it takes two scans to obtain a transition. A transition consists of two observations of the same group 15 minutes apart, and may or may not include changes in behavior state. Second, groups may cease to exist due to fission and fusion, or leaving the study area (either being so far away that they are no longer recognizable as the same group or being out of sight altogether).

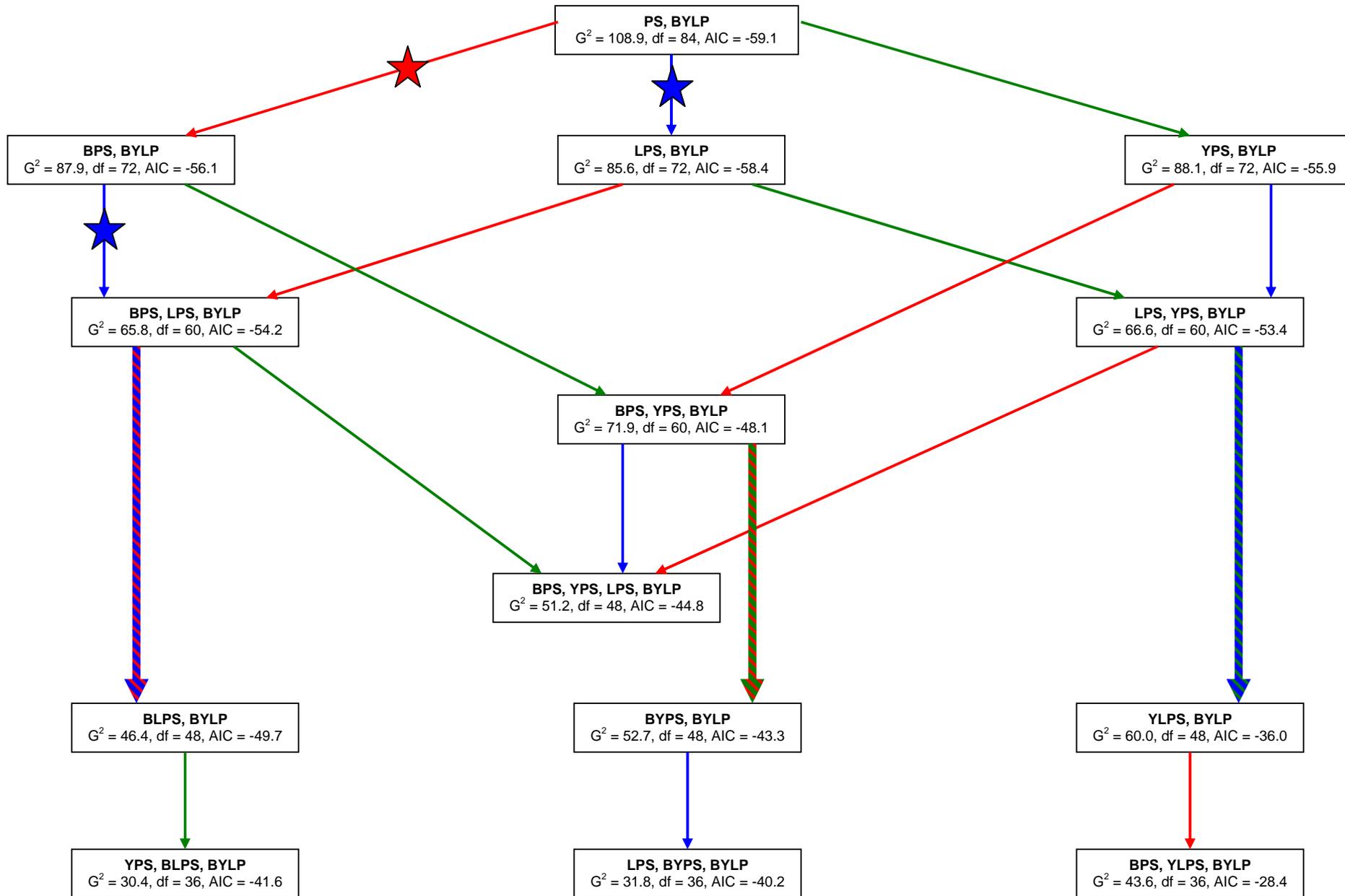
We assessed the effects of Year (2003/2004), Site (North/South), and Vessel Traffic (no boat within 100m, boat present within 100m) on behavioral transitions using a five-way log-linear analysis (LLA) (see Table 2 for sample size). Due to small sample size the full interaction of the three independent variables could not be quantified. Figure 2 is designed to present models going from the most simple one at the top (the null model), to the more complex ones at the bottom by increasing the number of parameters involved in the models as one moves away from the null model. Each model builds on a previous, simpler one by adding new effects to it. The effects added have been color coded: blue for a site effect, red for a boat effect, and green for a year effect. Interactions terms could also be added and those are represented by striped arrows (colors are the 2 effects interacting). This analysis reveals that three models provided more information on the data's variance (Figure 2). The null model (*i.e.* no effects from independent variables: PS, BYLP), the model considering a site effect (LPS, BYLP), and the model considering a boat effect (BPS, BYLP) all had lower Akaike Information Criteria (AIC) than the other models (Table 3). In addition, adding a boat and site effect to the model provided significantly more explanation of the data variance (significant effects represented by stars on Figure 2, and see Table 3), the site effect being still significant after the boat effect was taken into consideration. The significance of the terms being derived from the maximum likelihood estimates derived as described in the methods. From this analysis, we can conclude both that boat presence within 100m from the focal whales affected their behavioral transitions and that the whales behaved differently on both sites.

To elucidate these differences, we calculated the activity budgets of the whales at both sites, as well as in the presence and absence of boats. Whales tended to spend significantly more time traveling when boats were present than when they were absent within 100m (Figure 3, all data pooled with site effect ignored). So, while the boat presence significantly affected the activity state transition matrix as the log-linear analysis revealed, it did not affect the whole activity budgets significantly. However, the power to reject a difference in the proportion of time spent foraging with and without boats present was low (0.185) highlighting the need for more samples. The power for less common behaviors was even lower (Resting: 0.057 and Socializing: 0.050; power to test the difference of two given proportions with the available sample size [sing Minitab]). The difference in activity budgets between both sites also involved foraging and traveling, but was more pronounced than the differences with vessel presence (Figure 4, all data pooled with boat effect ignored). Whales spent significantly more time foraging at the south site and significantly less time traveling.

Table 2. The number of activity state transitions observed in the presence/absence of boats within 100m.

Site	2003		2004	
	No boat	Boat	No boat	Boat
North site	49	30	121	30
South site	45	11	46	20

Figure 2. (next page). Tests of boat presence within 100m (*B*), site (*L* for location to avoid confusion in abbreviations), and year of sampling (*Y*) effects on behavior transitions (*PS*) using log-linear analyses. Models and their respective goodness-of-fit G^2 statistics, degrees of freedom, and AIC values are shown in the boxes (adapted from Caswell 2001). Terms added are color-coded. Blue arrows represent the addition of a site effect (*LS*, *LPS* terms added to the previous model), red arrows represent the addition of a boat effect (*BS*, *BPS*), and green arrows represent the addition of a year effect (*YS*, *YPS*). To those terms correspond an increment in G^2 and degrees of freedom, which are used to test for the significance of the term addition. Arrows are marked with a star when the term addition is significant ($p < 0.05$).



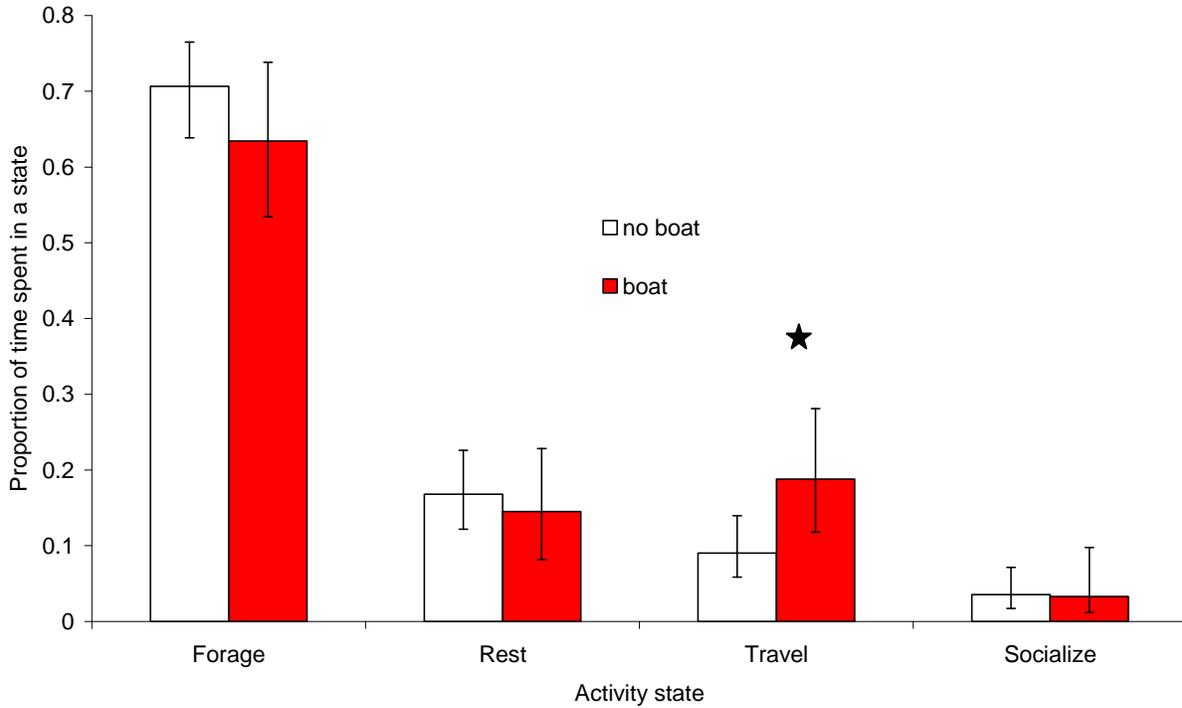


Figure 3. The proportion of time focal killer whales spent in each activity state (their activity budget) depending on the presence of boat within 100m of them. Error bars are 95% confidence intervals. Black stars indicate differences that are significant at the 0.05 level.

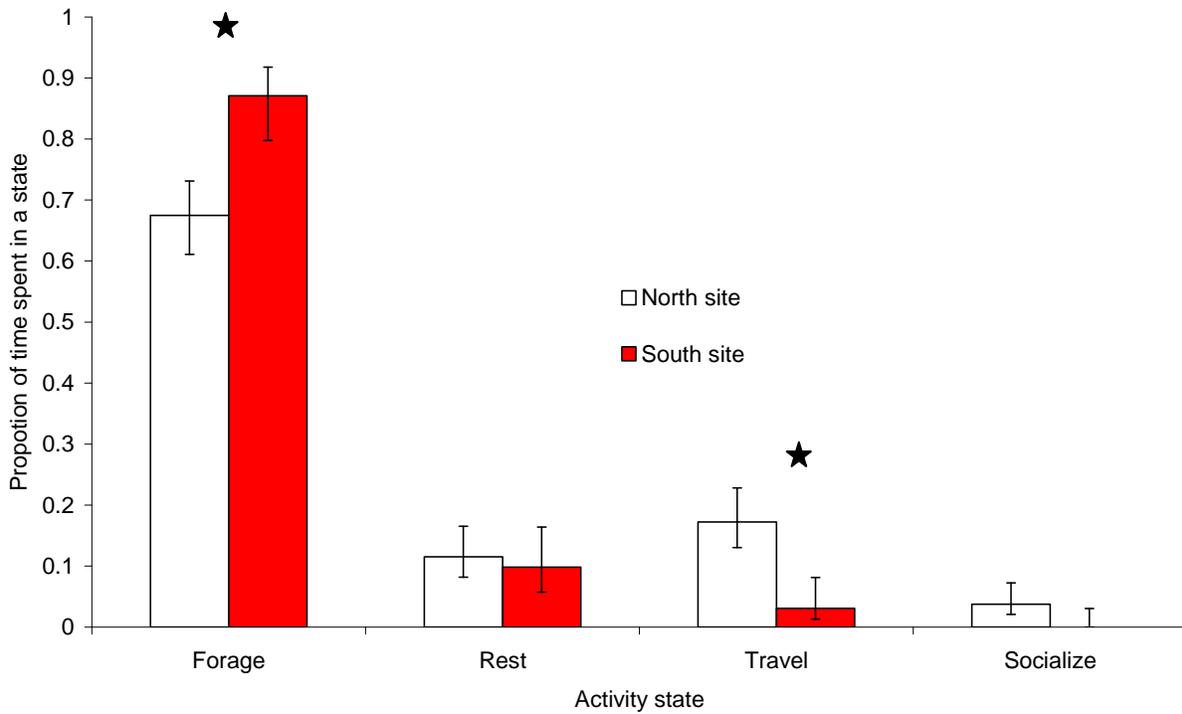


Figure 4. The proportion of time focal killer whales spent in each activity state (their activity budget) depending on the site sampled. Error bars are 95% confidence intervals. Black stars indicate differences that are significant at the 0.05 level.

Table 3. Information theoretic approach used to select models, from figure 2, providing the most parsimonious explanation for the variation in the scan-sample dataset. The models are described in figure 2. The difference between the best fitting model and the other models, ΔAIC , helps defining models that are less plausible (usually $\Delta AIC > 4$ to 8). The likelihood of the model given the data can be approximated using an exponential transformation of ΔAIC : $\ell(\text{model}_i | \text{data}) = e^{(-0.5\Delta AIC_i)}$. The weight of evidence provided by each model can be obtained by normalizing these likelihoods so that they sum to 1.

Model	AIC	ΔAIC	weight
Null model	-59.1	0	0.437
Site	-58.4	0.7	0.308
Boat	-56.1	3	0.098
Year	-55.9	3.2	0.088
Boat + site	-54.2	4.9	0.038
Site + year	-53.4	5.7	0.025
Boat + year	-48.1	11	0.0018
Boat + year + site	-44.8	14.3	0.0003
Boat x site	-49.7	9.4	0.0040
Boat x year	-43.3	15.8	0.0002
Year x site	-36	23.1	<0.0001
Year + (boat x site)	-41.6	17.5	<0.0001
Site + (boat x year)	-40.2	18.9	<0.0001
Boat + (year x site)	-28.4	30.7	<0.0001

The difference in weight between the site effect model and the boat effect model cannot be distinguished from model uncertainty: $\frac{\text{weight}_{\text{location}}}{\text{weight}_{\text{boat}}} = 3.2$ which is not significant (likelihood ratio test with 1 degree of freedom, $\alpha = 0.07$). In addition, the introduction of a boat effect explains significantly more variation in the dataset than the null model alone (as shown by the red star in figure 2). We can therefore conclude that while site is explaining much of the variation in the dataset, a boat effect cannot be excluded (because of the AIC weights) and should be included as well (because this parameter is providing significantly more information than the null model alone).

The influence of pod identity

The southern resident community of killer whales is composed of three pods (J, K, and L), which do not spend the same amount of time within Puget Sound. This may result in differences in the cumulative exposure of each pod to whale watching and therefore may lead to variation in the way these pods respond to boat presence. We therefore assessed whether the identity of the focal whales affected the behavioral response observed in relation with boat presence. To do so, we conducted a log-linear analysis including pod identity and boat presence

within 100m as independent variables. Sample size restricted the analysis. For this reason focal schools in which more than one pod was represented were counted for each of the pods present (see Table 5 for sample size). This analysis must therefore be treated with caution as observed effects might be a result of interactions between pods present in the same group. A larger sample size would allow treating single-pod schools and mixed schools separately, hence reducing also potential pseudoreplication issues. In addition, it would allow considering site effect concurrently with those two other effects. Table 4 shows that the pods use the sites differently.

Table 4. Number of samples (taken every 15 minutes) in which the following school structure (pod composition) was observed at each site for 2003 and 2004 combined. The table also summarizes the proportion of schools observed at each site with the different pod composition. Proportions in bold represents school composition more likely to occur at one site than the other (proportion test using Z statistic, $\alpha = 0.05$).

pod present in the school	Number of samples		Proportion of samples	
	North site	South site	North site	South site
J	233	83	0.263	0.177
K	20	7	0.023	0.015
L	194	212	0.219	0.451
J and K	62	18	0.070	0.038
J and L	90	42	0.102	0.089
K and L	0	47	0.000	0.100
J, K, and L	286	61	0.323	0.130

Table 5. Number of activity state transitions observed with and without boat present in the vicinity of each pod for both sites in 2003 and 2004 combined.

Pod	No boat present	Boat present
J	178	47
K	76	40
L	146	60

The log-linear analysis showed that there does not appear to be any variation in the way that whales responded to boat presence depending on their pod (Figure 5, Table 6). As in Figure 2, the results are presented as adding terms starting from the simplest model (the null model). Each time a term is added, the significance of this addition is tested (tests on the side of the arrows in Figure 5) and the resulting model is tested for its significance and its parsimony using AIC (Table 6). The best model included only boat presence as an effect (AIC=-11.6, Table 6), which was statistically significant (Figure 5). That is, adding a boat effect, whether to the null model or to a model already including a pod effect, provided a significant contribution to the explanation of the data variance; which the addition of a pod effect did not.

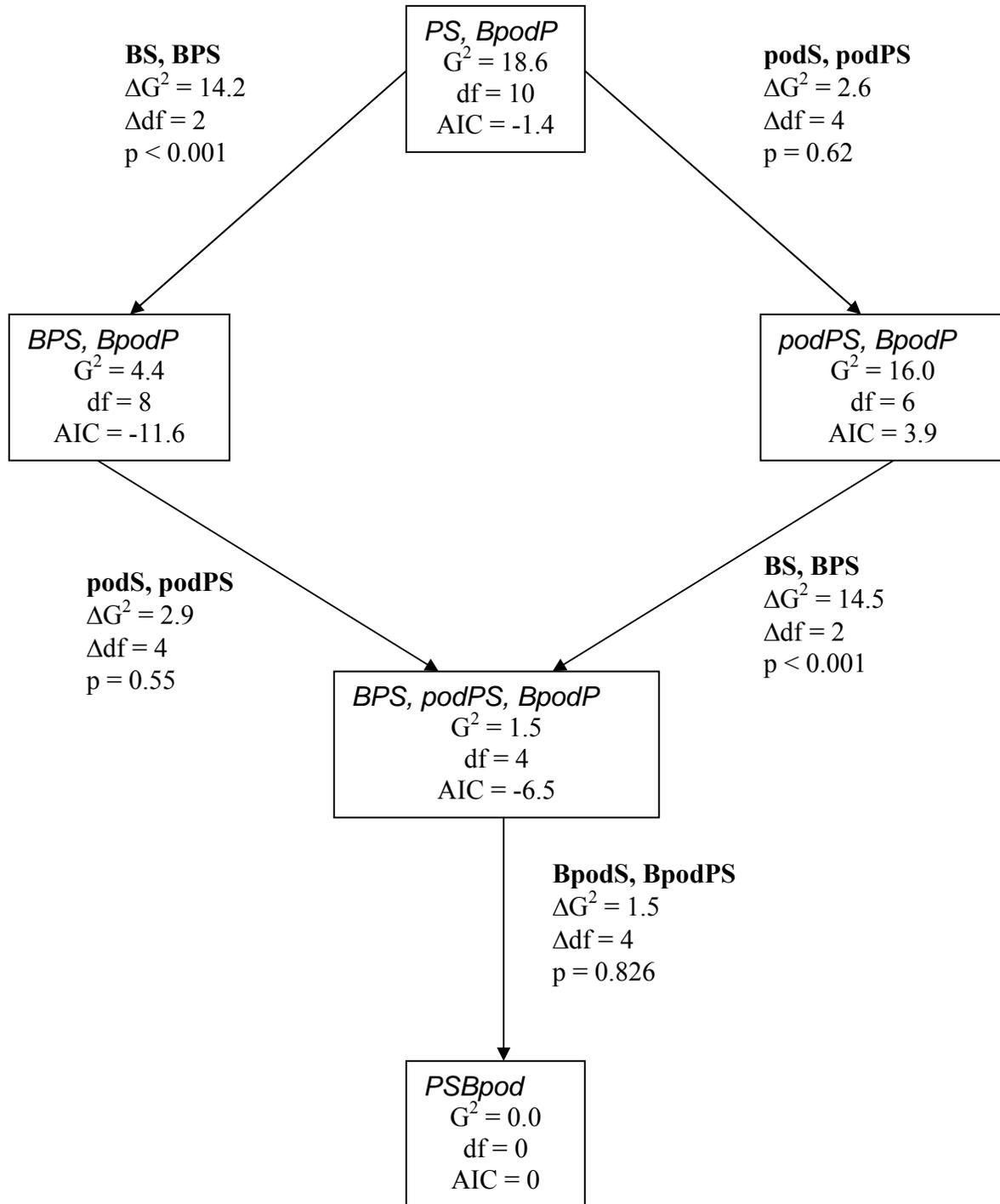


Figure 5. Tests of boat presence (*B*) and pod identity (*pod*) effects on activity state transitions (*PS*). Models and their respective goodness-of-fit G^2 statistics, degrees of freedom, and AIC values are shown in the boxes. Terms added, with the corresponding increment in G^2 and degrees of freedom, are shown along the arrows (result presentation adapted from Caswell 2001).

Table 6. Information theoretic approach used to select models, from Figure 5, providing the most parsimonious explanation for the variation in the scan-sample dataset. The models are described in Figure 5. The difference between the best fitting model and the other models, ΔAIC , helps defining models that are less plausible (usually $\Delta AIC > 4$ to 8). The likelihood of the model given the data can be approximated using an exponential transformation of ΔAIC : $\ell(\text{model}_i | \text{data}) = e^{(-0.5\Delta AIC_i)}$. The weight of evidence provided by each model can be obtained by normalising these likelihoods so that they sum to 1.

Model	AIC	ΔAIC	weight
Boat (BPS,BpodP)	-11.6	0	0.919
Pod + Boat (BPS,podPS,BpodP)	-6.5	5.1	0.072
Null model (PS,BpodP)	-1.4	10.2	0.006
Pod x Boat (PSBpod)	0	11.6	0.003
Pod (podPS,BpodP)	3.9	15.5	0.000

The influence of the distance between the focal school and boats

While an effect of boat presence was apparent when boats were within 100m of the focal schools, we wanted to assess whether more distant boats also influenced the activity states of the schools. Control samples are more difficult to obtain when considering boat presence at distances of 400m and 1000m, because boat traffic around the two sampling sites is consistently high. For this reason the current samples available did not allow us to assess the site effect in relation to boat presence within 400m of the whales. Similarly, we could not assess the effects of boats within 1000m of the whales because of the lack of a minimum amount of control samples. Therefore, we present only the results of the log-linear analysis assessing the effects of boat presence within 400m of the whales without considering the other potential effects (*i.e.*, merging samples obtained during both years and at both sites). Given that the site effect was found to affect the activity budget in previous analyses, the following results need to be interpreted with caution.

As in previous models, we compared the model, obtained from log-linear analyses, containing all two-way interactions (**Boat presence** by **Preceding behavior**, **Boat presence** by **Succeeding behavior**, **Preceding behavior** by **Succeeding behavior**) to the fully saturated model (**Boat presence** x **Preceding behavior** x **Succeeding behavior**) to assess the effect of boat presence on behavioral transitions (Lusseau 2003). This comparison, based on the difference in maximum likelihood estimates of both models using G^2 statistics, did not reveal an effect of boat presence within 400m on behavioral transitions ($\Delta G^2 = 10.88$, $\Delta df = 9$, $p = 0.28$). However, trends in behavioral budgets depending on boat presence within 400m were in the same direction as those when boats were within 100m (Figure 3 and Figure 6), with a whales spending significantly more time traveling and tending to spend less time foraging, although this difference was not significant.

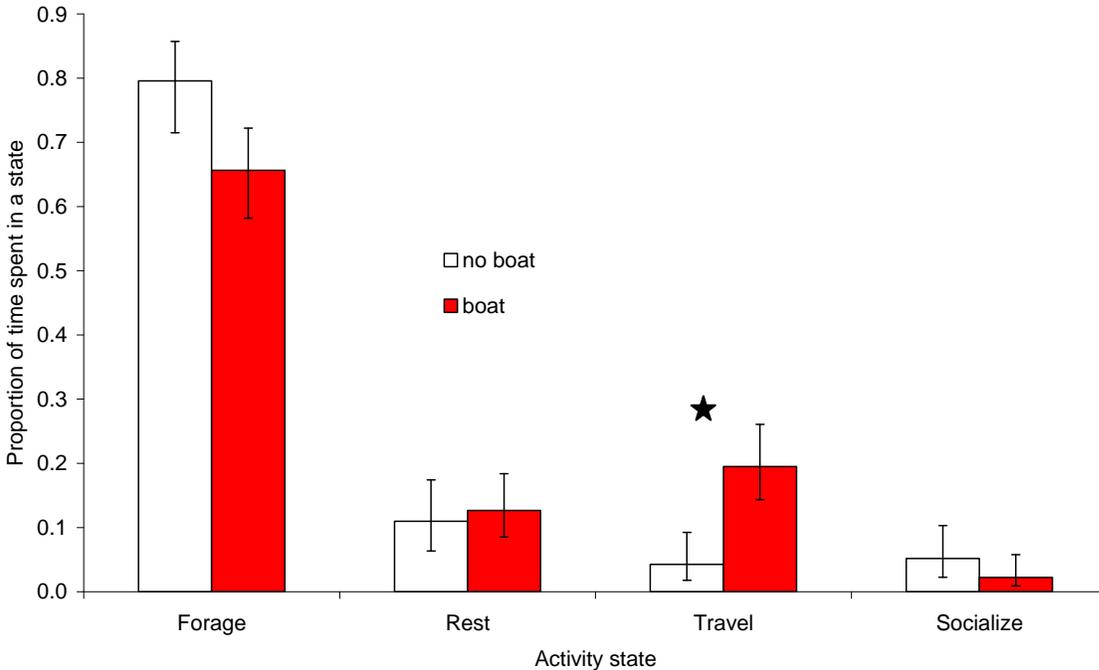


Figure 6. The proportion of time focal killer whales spent in each activity state (activity budget) depending on the presence of boat within 400m of them, pooling across years and sites. Error bars are 95% confidence intervals. Black star indicates differences that are significant at the 0.05 level.

It is worth noting that, while we had a relatively similar number of samples for transitions with a boat present coming from both sites, most of the control behavioral transitions (no boat present) were obtained at the North site. Whales spend less time foraging and more time traveling at the North site than at the South site (Figure 4). This sampling bias could have affected the results since it counters the apparent boat effect. We therefore repeated the same analysis taking into consideration only North site samples. Socializing became rare in the censored dataset and therefore was discarded from the analysis. The resulting comparison between 2-way and fully saturated models resulted in a significant effect of boat presence ($\Delta G^2 = 9.62$, $\Delta df = 4$, $p = 0.047$). The same results emerged in the variation in behavioral budget -- the whales spent significantly less time foraging and more time traveling when boats were present than in their absence (Figure 7). While this result is not based on the whole behavioral repertoire, and the sample size was relatively smaller, it does suggest that the previous analysis most likely suffered from sampling bias and that boat presence within 400m of the whales is highly likely to be affecting the whales' activity budgets as well.

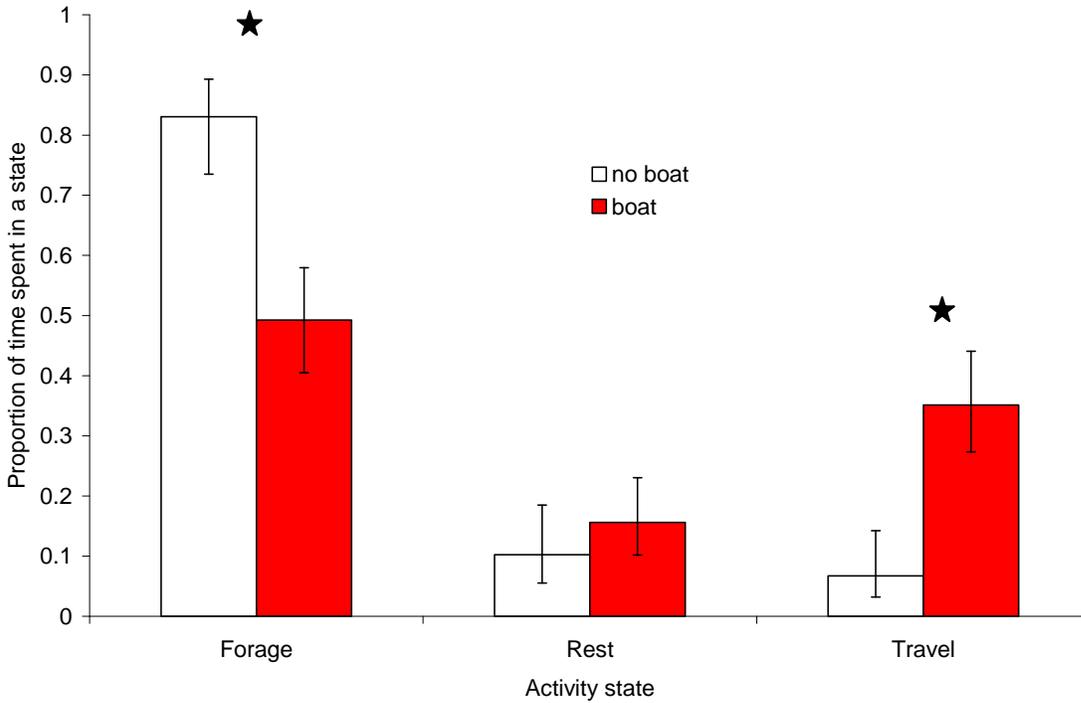


Figure 7. The proportion of time focal killer whales spent in each activity state (activity budget) depending on the presence of boat within 400m of them at the North site. Error bars are 95% confidence intervals. Black stars indicate differences that are significant at the 0.05 level.

To avoid the site sampling bias we censored the dataset to only include North site data and assessed the effect of boat presence within 100, 400 and 1000m on the probability to stay foraging when foraging. The effect size of boat presence, *i.e.* the difference in the likelihood to stay foraging when foraging between control and impact situation (vessels present within the specified distance), decreased with the distance to boat present increasing (Figure 8). The effect of boat presence appeared to be only significant when boats were within 100m (Figure 8, note the star and the confidence intervals), yet sample size might be preventing the detection of smaller effect size for the other treatments (Figure 8, 400 and 1000m). In addition, the likelihood to stay foraging when foraging increased as the distance between the focal group and boats present in the study area increased (Figure 9). Whales were significantly more likely to stay foraging when foraging if boats were 1000m away from them than if boats were 100m from them. That is, the probability to stay foraging when boats were at 1000m was significantly greater than when boats were within 100m.

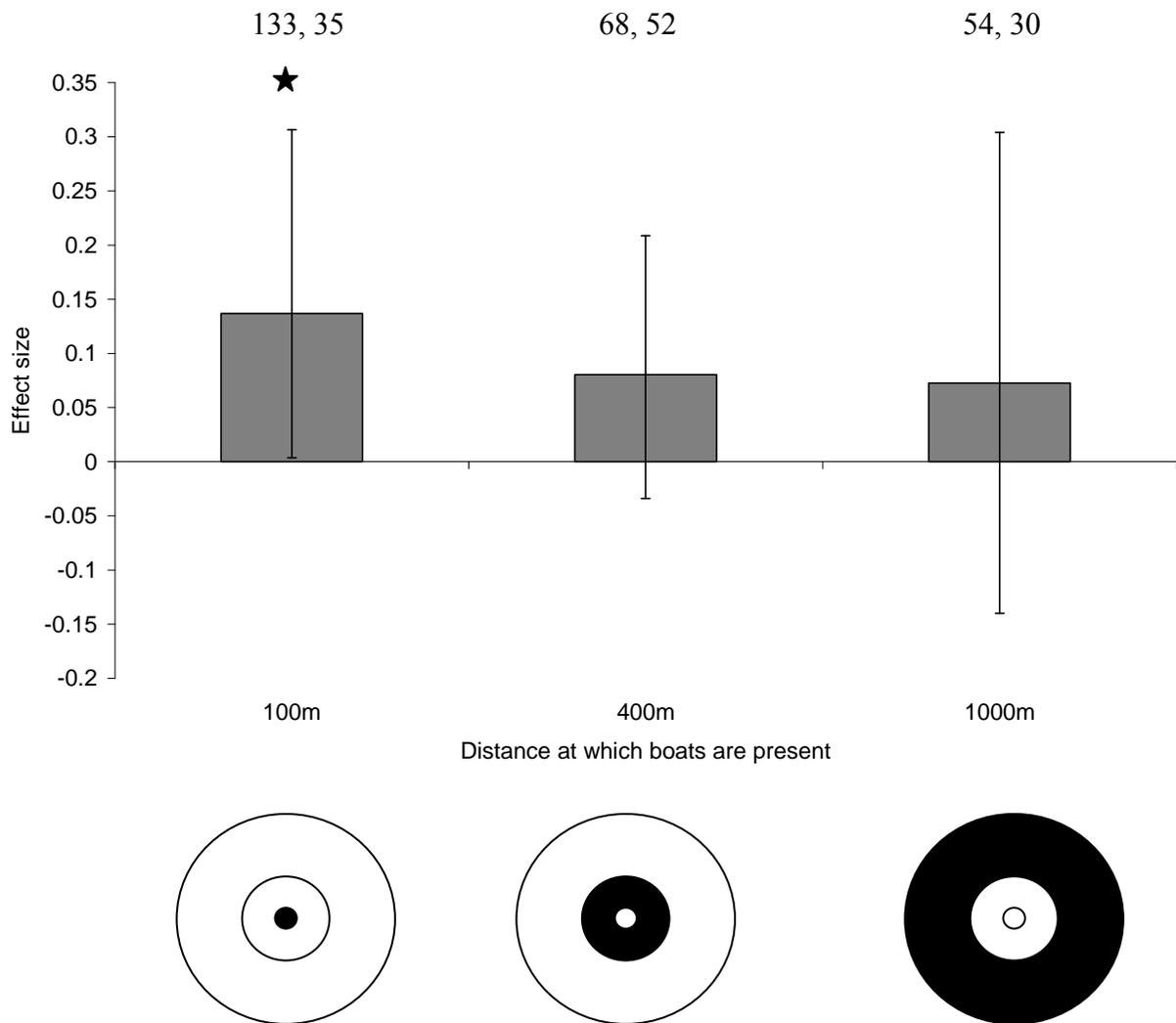


Figure 8. Difference in the likelihood to stay foraging when foraging ($p_{F \rightarrow F}$) between control and impact situations ($p_{\text{control}} - p_{\text{impact}}$), i.e. effect size, depending whether some boats were present within 100m, 400m, or 1000m of the focal whales at the North Site. The diagram describes these three treatments in which the focal group is at the centre of the concentric doughnuts (100m, 400m, and 1000m radii) and the boat exposure is represented in black. In order to test the effect of distance to boats, this analysis only takes into consideration samples when there were boats present only at the given distance. Error bars are 95% confidence intervals for the difference. If the interval includes 0, the difference is not significant at the conventional ($p < 0.05$) level (indicated by a star). The number of transitions observed is given above each bar (n_{control} , n_{impact}).

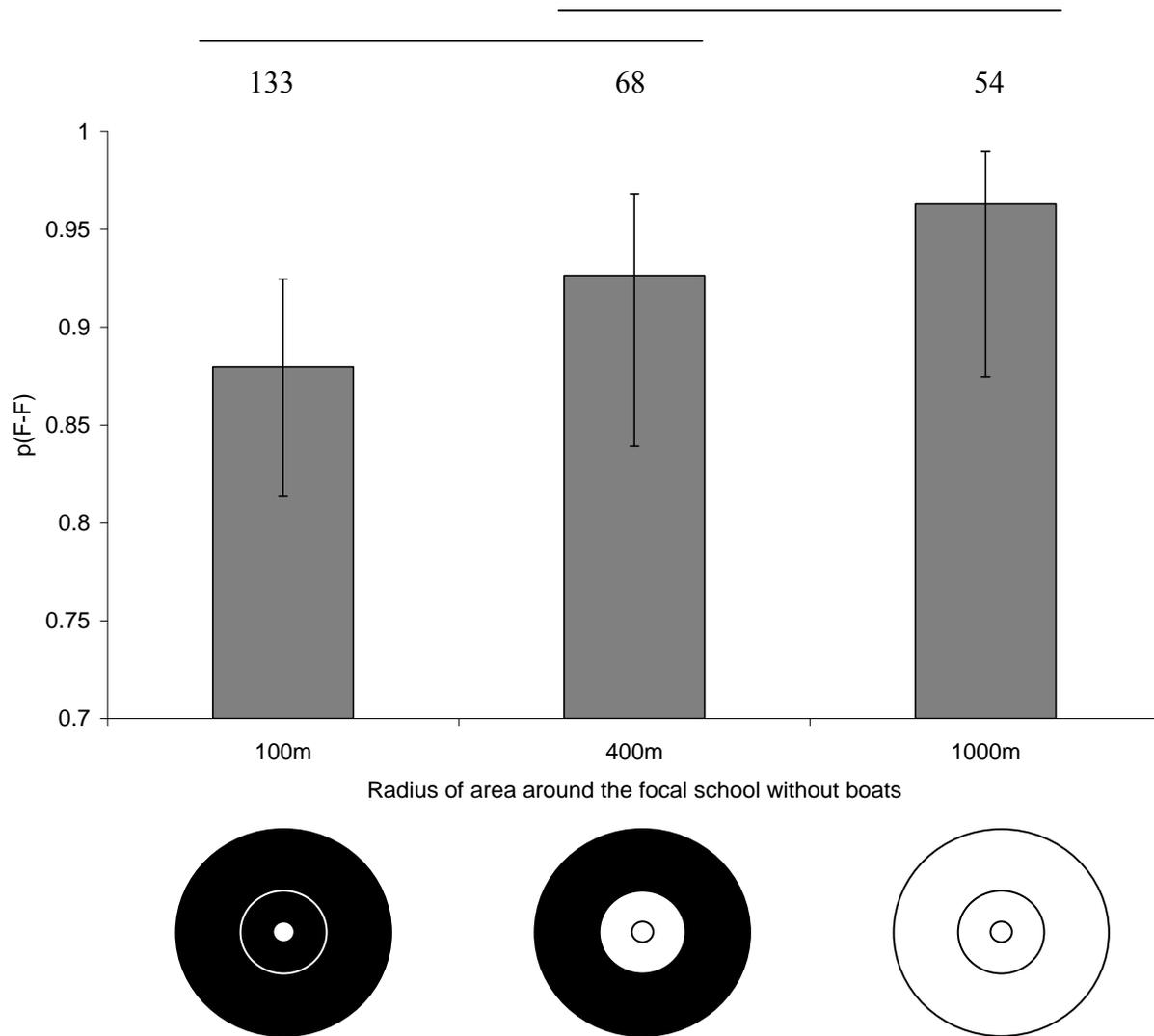


Figure 9. Likelihood that whales stay foraging when foraging, $p(F-F)$, when there is no boat present within 100m, 400m, or 1000m of the focal school at the North site. The diagram describes these three treatments in which the focal group is at the centre of the concentric doughnuts (100m, 400m, and 1000m radii) and the boat exposure is represented in black. Error bars are 95% confidence intervals. Whales are significantly more likely to stay foraging when foraging when there are no boats within 1000m as opposed to when there are no boats within 100m of them (as shown by the horizontal lines above the figure which link situations that could not be significantly distinguished from one another). The number of transitions observed is given above each bar.

THEODOLITE TRACKING OF FOCAL INDIVIDUALS

We collected 44 tracks in 2003 and 77 tracks in 2004 that were of sufficient quality to use in the analysis. Whales were tracked for an average of 24.9 minutes over 2.6 km. Behavioral indices derived from theodolite tracks are summarized in Table 7. More details on track characteristics (e.g., breakdowns by individual, pod, location, vessel presence, etc.) are shown in Appendices 1-3. Note that only two tracks in 2003 met the criteria for good boats-absent tracks.

The relationship between percentage of surfacings marked and directness index shows no bias, supporting inclusion of all these tracks except the one in which fewer than 60% of surfacings were marked with the theodolite (an outlier both in terms of proportion marked and biased in the direction expected when surfacings are missed so it was excluded from the analysis, see Appendix 4).

There was no evidence in scatterplots to suggest that whale behavior varied as a function of absolute tide height. Consequently, we dropped TIDE from subsequent GAM analyses (which were in danger of failing to converge given the large number of candidate covariates and relatively small number of observations). Whale behavior did vary with track duration in the case of surface active behavior, and our method for addressing that is described in the section on that response variable. DURATION was dropped from the list of candidate covariates in the case of the other four response variables, in order to assist model convergence, but retained for analysis of SAB.

Example tracks are shown in Appendix 5.

Results of binary (baseline/exposure) analyses

The mean values of the behavioral indices employed in this study and the probability of the with-boat and without-boat data being drawn from the same distribution based on the Monte Carlo simulation are shown in Table 7.

Table 7. Behavioral indices for Southern Residents in the absence (distance to nearest vessel > 1000m) and presence of vessels.

Behavior	Value		Significance
	Without Boats N = 21	With Boats N = 100	
Directness Index	83.28	74.82	P < .001
Deviation Index	25.63	28.58	P < .01
Respiratory Interval (sec)	42.58	49.15	P < .001
Surface Active (/hr)	0.62	2.60	P < .001
Surface Speed (m/hr)	6638	6404	P > .1

Directness Index. Whales tended to travel in more direct paths when vessels were absent than when they were present. The magnitude of this effect was 11%.

Deviation Index. Whales made slightly larger course changes when vessels were present than when vessels were absent. The magnitude of this effect was 12%.

Breathing Patterns. The average time between breaths was significantly longer in the presence of boats than in their absence. The magnitude of this effect was 15%

Surface Active Behavior. Surface active behavior was significantly more frequent in the presence of vessels than in its absence. This was due to surface active behavior being roughly twice as likely to occur in the presence of vessels, and roughly twice as many events occurring given that at least one occurred when vessels were present.

Swimming Speed. There was no significant difference in swimming speed due to vessel traffic found in the binary analysis.

Vessel Proximity. The proportions of subgroups scanned in proximity of vessels are shown in Table 8. The primary component of the reduction in scans with boats within 400 and 1000m of whales in 2004 is probably due to extending the field season prior to the opening of salmon fishing season.

Table 8. Percentage of scan samples with at least one vessel within the distances shown of the subgroup. WOT = Whale Oriented Traffic, NOT = Not whale Oriented Traffic, and ANY refers to the presence of at least one of either type of vessel.

Year	Within 100 m			Within 400 m			Within 1000 m			Field of View
	ANY	WOT	NOT	ANY	WOT	NOT	ANY	WOT	NOT	ANY
2003	23.3	22.5	0.8	81.0	75.6	9.7	90.1	84.2	19.0	99.5
2004	28.8	26.7	4.9	56.7	51.6	11.2	75.6	69.4	21.5	98.5

Results of GAM-based analyses of focal animal behavior

Respiration analysis. The model that fitted the respiration data best included all five candidate vessel traffic variables, whale related variables (age, sex and pod), and temporal variables (time of day and month of year, Table 9). The model described the variation in mean respiratory interval well, in that it was able to account for 72.6% of the deviance, and had an R-squared value of 0.565. Term-wise parameter estimates indicate that male whales tended to have longer inter-breath intervals than females, after all other effects were included in the model. As number of boats (the maximum of the total number of vessels identified by the theodolite operator,

instantaneous counts by the theodolite and computer operators, and counts by the scan sampler within 1000m) increased, the inter-breath interval tended to decrease.

The smooth term describing covariation of SUM400 and inter-breath interval indicates that the relationship is not as straight-forward as suggested by the linear terms above. Dive times tended to be shorter when no boats were present within 400m of the focal whale, and increased as number of boats increased to approximately 6 or 7 boats. When many more boats were present within this range, this relationship became non-significant (*i.e.*, the confidence intervals comfortably spanned zero).

It may be biologically significant to note that the average respiratory interval reached a maximum in late July through mid-August.

Table 9. Summary of selected model describing heterogeneity in mean respiration rate as linear (top) and smooth (*i.e.*, non-linear, bottom) functions of covariates selected by **mgcv**.

Family: quasi
Link function: identity

Formula:
RESP ~ MONTH + POD + SEX + s(AGE) + SUM100 + s(SUM400) + SUM1000 +
s(PCA) + s(BOATS) + s(JDAY) + TIME

Parametric coefficients:

	Estimate	std. err.	t ratio	Pr(> t)
(Intercept)	92.469	35.59	2.599	0.012104
MONTH	-4.8686	5.247	-0.9279	0.35766
PODK	-13.042	8.039	-1.622	0.11068
PODL	-12.308	6.657	-1.849	0.070069
SEXM	-6.5008	5.754	-1.13	0.26365
SUM100	-7.4932	4.191	-1.788	0.079545
SUM1000	1.3885	0.6838	2.03	0.04736
TIME	-0.4301	0.8298	-0.5183	0.60642

Approximate significance of smooth terms:

	edf	chi.sq	p-value
s(AGE)	7.864	27.469	0.0027565
s(SUM400)	7.37	29.453	0.0011990
s(PCA)	2.33	4.3267	0.16024
s(BOATS)	3.869	22.118	0.00075994
s(JDAY)	2.675	13.449	0.005201

R-sq.(adj) = 0.565 Deviance explained = 72.6%
GCV score = 343.56 Scale est. = 213.78 n=85

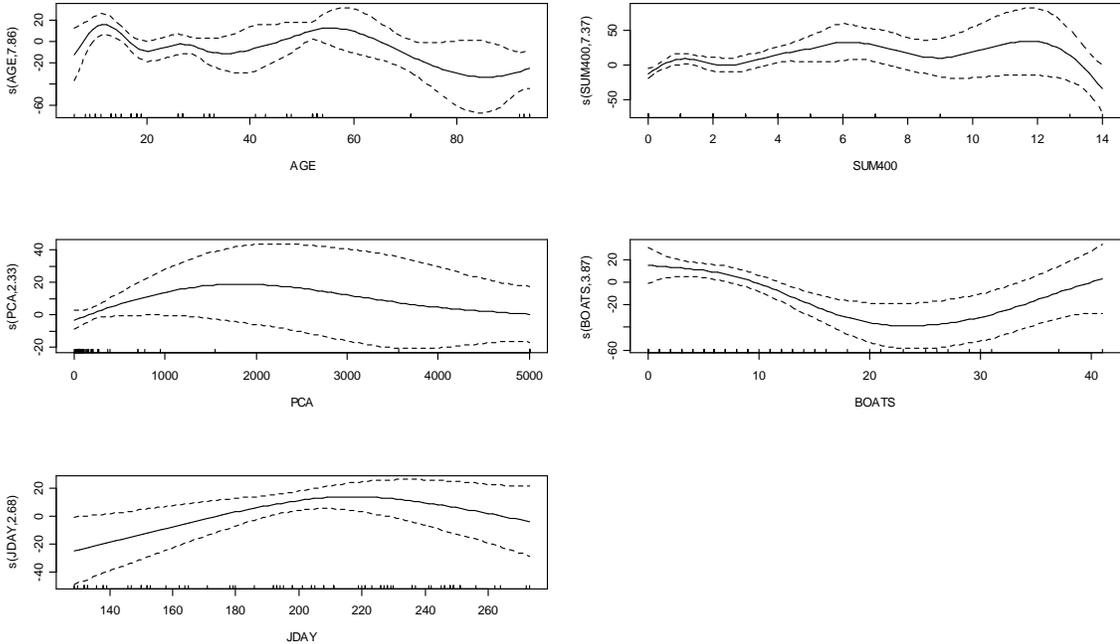
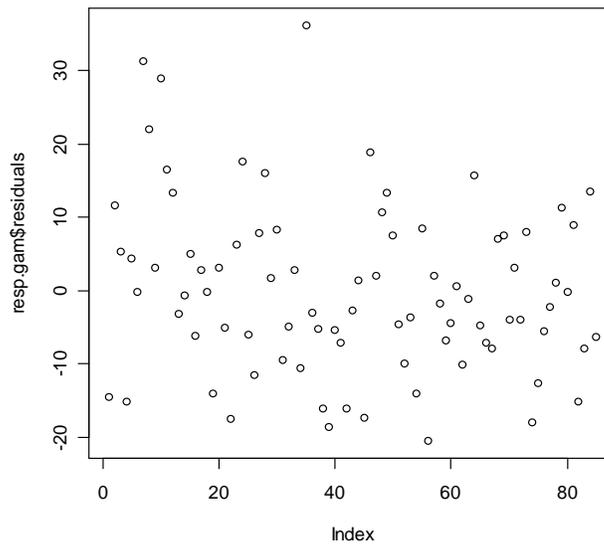


Figure 10. Relationships between smoothed components (solid lines) of the explanatory variables (x-axis) used in the fitted GAM (after accounting for the linear effects of MONTH, POD, SEX, SUM100, SUM1000 and TIME) and the response variable, mean time between breaths. The explanatory variable, SUM400, represents the maximum number of boats ever observed within 400m of the whale. SUM400 was allowed up to 9df in model selection, and the degree of smoothing was automated by mgcv. The x-axis contains a rugplot, in which small ticks mark locations of observations. Zero on the y-axis corresponds to no effect of the covariate on the estimated response (here, inter-breath interval). Values above zero on the y-axis indicate positive correlation, *i.e.*, long inter-breath intervals. The y-axis is labeled $s(\text{covariate name}, \text{estimated degrees of freedom})$ indicating the curve is smoothed. The dashed lines represent ± 2 standard errors, or roughly 95% confidence intervals.

The bottom plot shows the residuals of the fitted model. The slight positive skew (asymmetry about zero on the y axis) suggests that some over-dispersion in respiration rate remained unmodeled by the selected GAM, but that the model provided a reasonably good fit to the data overall.



Swimming speed. The selected model included the maximum number of boats scanned within 100m and 400m of the focal animal, as well as the theodolite count of number of boats, although the latter term contributed weakly and with a low probability of being statistically valid. In addition, model fit was improved significantly by including information on the age of the focal animal and the month of the observation (Table 10). The model described variation in swimming speed moderately well, in that it was able to account for 25.7% of the deviance, and had an R-squared value of 0.17. Term-wise parameter estimates indicated that older whales tended to swim more slowly than younger whales, and that all animals tended to swim more slowly as the season progressed. The relationship between boat count and swimming speed was negative for close boats (within 100m) and was positive (*i.e.*, whale swimming speed tended to increase) as number of boats within 400m of the whale also increased).

Table 10. Summary of selected model describing heterogeneity in mean swimming speed as linear (top) functions of covariates selected by **mgcv**.

Family: quasi

Link function: identity

Formula:

SPEED ~ MONTH + AGE + SUM100 + s(SUM400) + BOATS

Parametric coefficients:

	Estimate	std. err.	t ratio	Pr(> t)
(Intercept)	10.48	1.49	7.033	8.0121e-10
MONTH	-0.49739	0.182	-2.733	0.0078305
AGE	-0.012396	0.01104	-1.123	0.26503
SUM100	-0.90962	0.5018	-1.813	0.073858
BOATS	0.027528	0.03486	0.7896	0.43226

Approximate significance of smooth terms:

	edf	chi.sq	p-value
s(SUM400)	4.884	10.525	0.06949

R-sq.(adj) = 0.169 Deviance explained = 25.7%
 GCV score = 5.4775 Scale est. = 4.8406 n = 85

Figure 11 shows the smooth spline relating swimming speed to the maximum number of boats scanned within 400m of the whale and the residuals of the fitted GAM.

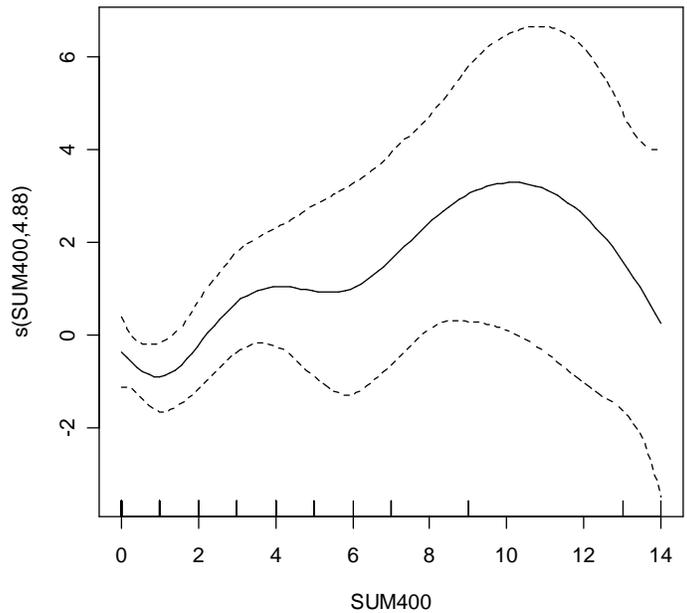
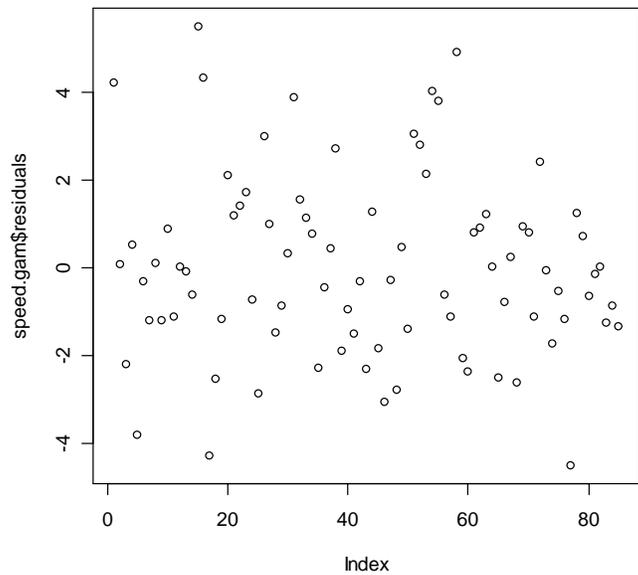


Figure 11. TOP – the smooth spline relating swimming speed to the maximum number of boats scanned within 400m of the whale suggests that whales tended to swim slowly when one boat was scanned within 400m, and then swimming speed tended to increase as boat number increased. BOTTOM – the residuals of the fitted GAM after accounting for the linear effects of Month, Age and a smooth spline of the maximum number of boats within 400m of the whale. No pattern is evident in the residuals, indicating that the model fitted the data well, and that the maximum likelihood approach was able to account for any overdispersion in the response data.



Deviation index. The model that fitted the path deviation index data best included three boat count variables (SUM100 and SUM1000, and boats), as well as Site (Table 10). The model demonstrated poor power to describe variation in deviation index, accounting for only 8.74% of the deviance explained, with an adjusted R-squared value of 0.01.

Term-wise parameter estimates indicate that whales adopted smoother paths (*i.e.*, lower deviation index) at the South site than at the North site. The maximum number of boats within 100m (SUM100) entered the model as a linear term, indicating that whales exhibited, on average, smoother paths when few boats were observed close to the whale (Table 10).

Table 11. Summary of selected model describing heterogeneity in path deviation index as linear (top) and smooth (*i.e.*, non-linear, bottom) functions of covariates selected by **mgcv**.

Family: quasi
Link function: identity

Formula:

$$\text{DEV} \sim \text{SITE} + \text{SUM100} + \text{s}(\text{SUM1000}) + \text{PCA} + \text{BOATS} + \text{JDAY}$$

Parametric coefficients:

	Estimate	std. err.	t ratio	Pr(> t)
(Intercept)	45.826	9.551	4.798	7.1377e-06
SITESouth	-3.4762	3.371	-1.031	0.30548
SUM100	-2.2121	2.687	-0.8234	0.41269
PCA	-0.0021413	0.001069	-2.002	0.048563
BOATS	-0.30088	0.2408	-1.25	0.21504
JDAY	-0.053465	0.03958	-1.351	0.18048

Approximate significance of smooth terms:

	edf	chi.sq	p-value
s(SUM1000)	1.479	1.182	0.41908

R-sq.(adj) = 0.0149 Deviance explained = 8.74%
GCV score = 202.59 Scale est. = 185.56

n = 89

The smooth term in the selected model is shown in Figure 12. The spline shows weak evidence that swimming paths were relatively smooth when few boats were observed within 1000m of the whale, becoming more erratic when moderate numbers (5-17 or so) boats were present, and becoming smooth again when many (>17) boats were present. However, this U-shaped curve should be interpreted with caution for two reasons. First, the confidence intervals span zero across the range of X. Secondly, the rugplot reveals that the observations were not uniformly spread across X. In other words, few observations were made when large numbers of boats were present within 1000m of the whale. However, the model fit was improved significantly by including this term, despite the penalty for overparameterization.

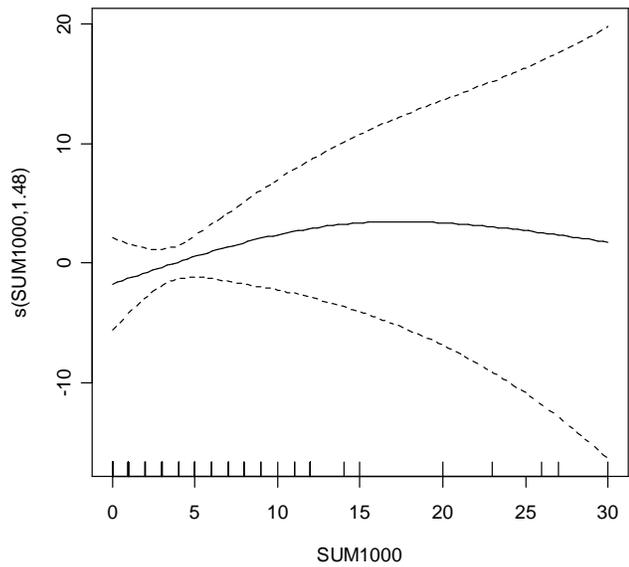
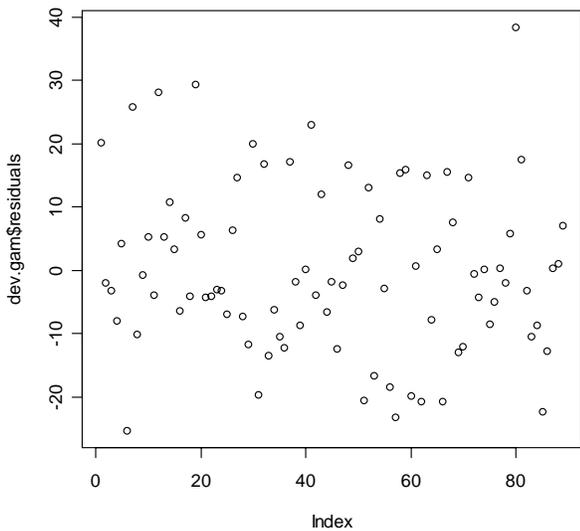


Figure 12. TOP-- Relationship between smoothed component (solid line) of the one non-linear explanatory variable (SUM1000) selected in the fitted GAM (after accounting for the linear effects of SITE, JDAY, PCA, SUM100 and BOATS), and the response variable, path deviation index. The explanatory variable, maximum number of boats observed within 1000m, was allowed up to 9df and the degree of smoothing (≈ 1.48 df) was automated by **mgcv**. The x-axis contains a rugplot, in which small ticks mark locations of observations. Zero on the y-axis corresponds to no effect of the covariate on the estimated response (deviation index). Values above zero on the y-axis indicate positive correlation, *i.e.*, an erratic, or above-average, path deviation index. The y-axis is labeled $s(\text{covariate name}, \text{estimated degrees of freedom})$. The dashed lines represent ± 2 standard errors, or roughly 95% confidence intervals. BOTTOM – The plot below shows the residuals of the fitted model. No pattern is evident, indicating that the model fitted the data well, and that the maximum likelihood approach was able to account for any overdispersion in the response data.



Directness index.

The model that fitted the path directness index data best included two vessel traffic variables (SUM1000 and BOATS), as well as SITE, POD, and AGE (Table 12). The model demonstrated some power to describe variation in directness index, accounting for 93% of the deviance, but an R-squared value of 0.0326.

Most variables entered the model as linear terms.

Table 12: Summary of selected model describing heterogeneity in path directness index as linear functions of covariates selected by **mgcv**.

Family: quasi

Link function: identity

Formula:

DI ~ SITE + POD + s(AGE) + s(SUM1000) + BOATS - 1

Parametric coefficients:

	Estimate	std. err.	t ratio	Pr(> t)
SITENorth	0.76008	0.0488	15.58	< 2.22e-16
SITESouth	0.81064	0.06072	13.35	< 2.22e-16
PODK	-0.11062	0.08244	-1.342	0.18358
PODL	-0.097665	0.05897	-1.656	0.10174
BOATS	0.0063302	0.003879	1.632	0.10678

Approximate significance of smooth terms:

	edf	chi.sq	p-value
s(AGE)	1.779	2.1643	0.29833
s(SUM1000)	1.228	1.9209	0.21773

R-sq.(adj) = 0.0326

Deviance explained = 93%

GCV score = 0.055435

Scale est. = 0.050213

n = 85

Figure 13 shows the non-linear smooth terms that entered the model and the residuals of the fitted GAM. While there is no dramatic evidence of a pattern in the residuals, there is some suggestion of asymmetry about zero. Recall that directness index was bounded between 0 and 1, but that direct paths (*i.e.*, those near 1) were much more common than those near 0. We tried using a log link with constant variance term to account for this deviation from a normal distribution, but it did not do as well as the quasi-likelihood approach with an identity link (possibly because the log link brought the negative values lower, but could not improve the symmetry about zero). Model fit and convergence may have been constrained by the asymmetry of this response variable.

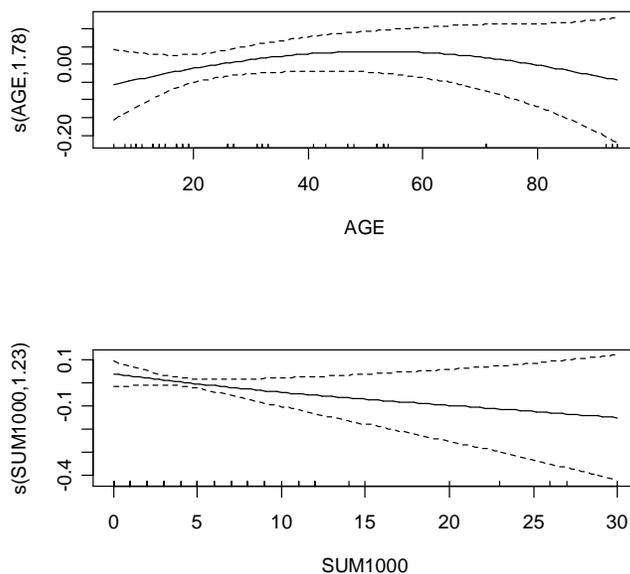
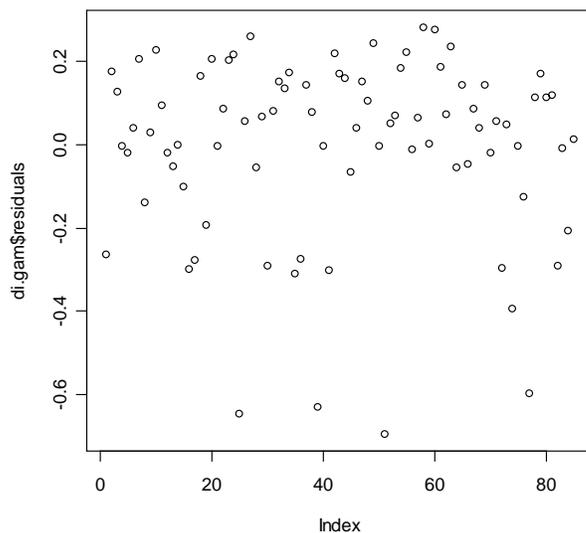


Figure 13. TOP-- Relationship between smoothed component (solid line) of the explanatory variables (AGE, SUM1000) selected in the fitted GAM, and the response variable, path directness index. The explanatory variables, age and maximum number of boats observed within 1000m, was allowed up to 9df and the degree of smoothing was automated by **mgev**. The x-axis contains a rugplot, in which small ticks mark locations of observations. Zero on the y-axis corresponds to no effect of the covariate on the estimated response (deviation index). Values above zero on the y-axis indicate positive correlation, *i.e.*, a more direct path. The y-axis is labeled $s(\text{covariate name, estimated degrees of freedom})$. The dashed lines represent ± 2 standard errors, or roughly 95% confidence intervals. BOTTOM-- Residuals of the fitted model. The model fitted the data reasonably well, and the maximum likelihood approach (with a constant variance term) was able to account for overdispersion in the response data. Some evidence of asymmetry about zero in the y-axis reinforces the conclusion of model parameters that model fit was relatively poor. [Note that while approximately half of the values are positive, and half negative, the largest negative values are larger than the largest positive values – this reflects the boundaries of the original scale, which was bounded by 0 and 1, but values of DI tended to be nearer 1 than 0.]



Surface active behavior Surface active behavior tended to occur in bouts widely separated in time. As a result, many tracks had no surface active behavior. Those that did have any at all, tended to have at least a few events and could have many. We normalized the rate of SAB to number of events per hour, but found the model was unable to fit high rates of SAB, perhaps due to disproportionately large corrections in short tracks (i.e., if the interval between bouts is large compared to the sampling period, the correction for sample period would bias the data). Then we tried to treat SAB as either present (1) or absent (0) during a track (SAB.1.0). We found this value was positively correlated with track duration, as expected, but the GAM analysis could not correct for this when considering other parameters. As a result, we analyzed both SAB and SAB.1.0, in hopes that asking two variations on a common question would elucidate underlying trends.

The results for SAB are shown in Table 13 and Figure 14. The results for SAB 1.0 are shown in Table 14 and Figure 15.

The analysis of SAB suggests young animals are highly active, but this rate slows as animals reach sexual and physical maturity, and by senescence, SAB is rare, although the trend is insignificant at most ages. The effect of boats (Figure 14, upper right) was linear, but model fit was improved by retaining it as a smoothed relationship. SAB was most common when the number of boats was small (<10), and those boats that were present were within 400 m.

The analysis of SAB.1.0 suggests SAB is most likely to occur when boats are close (within 200m) and the number of boats is small. Surface behavior was also most likely to occur in the late afternoon and evening. The fit and explanatory power was much better for SAB1.0 than for SAB.

The consistent trend for both SAB and SAB.1.0 to be maximized when the number of boats is small and they are in close proximity suggests that bias from sample duration was overcome in this aspect of the analysis.

Age was not a factor in SAB.1.0, perhaps suggesting that the probability of engaging in SAB is equal for all age classes, but that younger animals tend to do more once they get started. Similarly, pods may be equal in their probability of initiating SAB, but differ in the number of events once it is initiated.

Other factors that may affect the probability of initiating SAB include current, site, and month of the year, with males more likely than females to initiate SAB.

Table 13. Summary of selected model describing heterogeneity in rates of surface active behavior as linear (top) and smooth (bottom, *i.e.*, non-linear) functions of covariates selected by **mgcv**.

Family: quasi
 Link function: identity

Formula:
 SAB ~ TIME + POD + s(AGE) + SUM400 + s(BOATS) + -1

Parametric coefficients:

	Estimate	std. err.	t ratio	Pr(> t)
TIME	0.33341	0.2981	1.118	0.26694
PODJ	-2.5341	3.381	-0.7495	0.45586
PODK	3.8221	4.207	0.9084	0.36652
PODL	-1.0006	3.701	-0.2704	0.78762
SUM400	0.10228	0.3319	0.3081	0.75883

Approximate significance of smooth terms:

	edf	chi.sq	p-value
s(AGE)	3.005	6.2031	0.11188
s(BOATS)	1	0.59526	0.44279

R-sq.(adj) = 0.0676 Deviance explained = 26.9%
 GCV score = 53.113 Scale est. = 47.487 n = 85

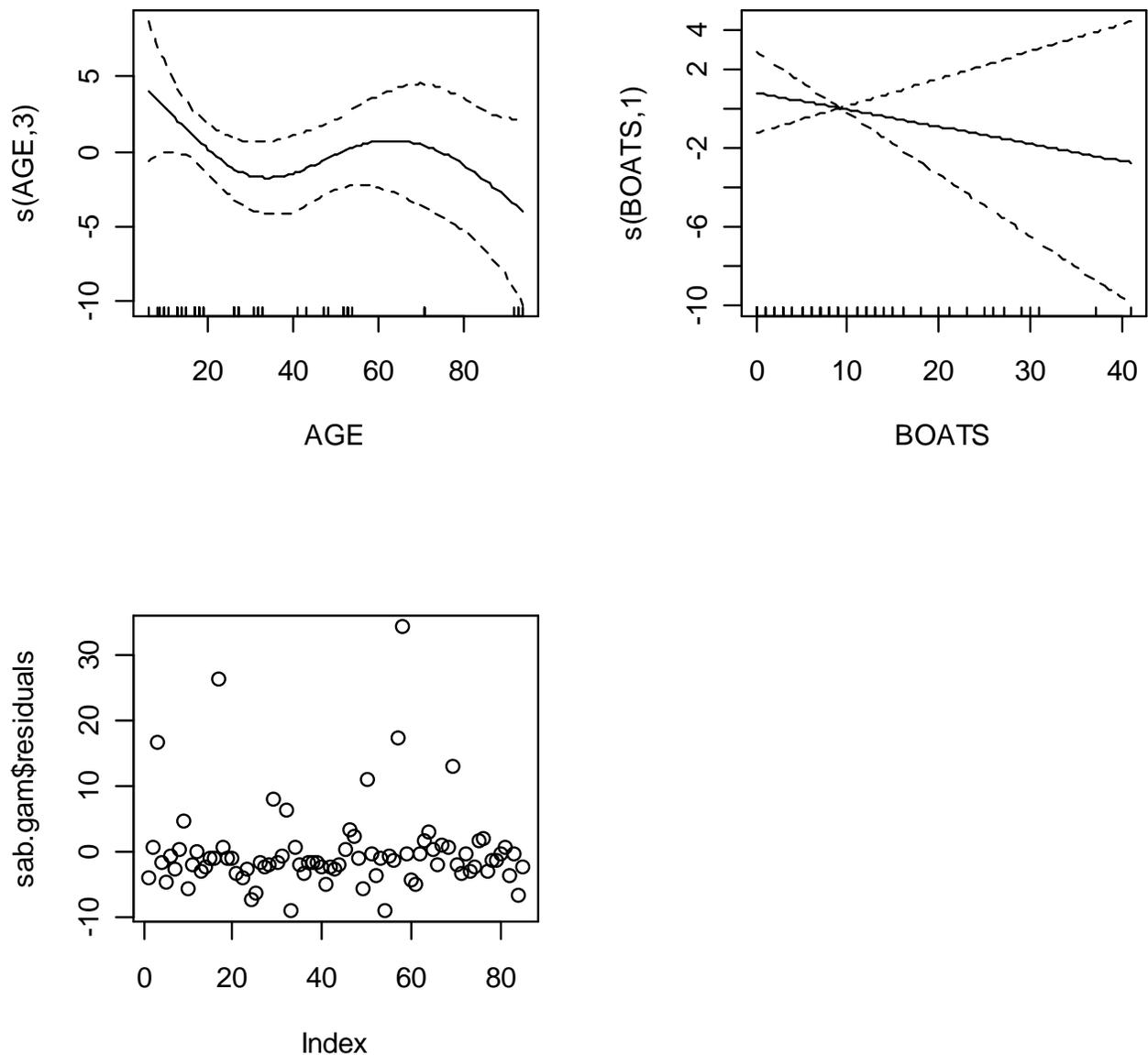


Figure 14. TOP – Relationship between smoothed component (solid line) of the explanatory variables (age, boats) selected in the fitted GAM, and the response variable, rates of surface active behavior. The explanatory variables, AGE and number of boats observed by the theodolite team (BOATS), was allowed up to 9df and the degree of smoothing was automated by **mgcv**. The x-axis contains a rugplot, in which small ticks mark locations of observations. Zero on the y-axis corresponds to no effect of the covariate on the estimated response (rate of surface active behavior). Values above zero on the y-axis indicate positive correlation, *i.e.*, higher rates of surface active behavior. The y-axis is labeled s(covariate name, estimated degrees of freedom). The dashed lines represent ± 2 standard errors, or roughly 95% confidence intervals. BOTTOM – Residuals of the fitted model. The plot of residuals indicates poor ability to explain high rates of SAB.

Table 14. Summary of selected model describing heterogeneity in likelihood of surface active behavior as linear (top) and smooth (bottom, *i.e.*, non-linear) functions of covariates selected by **mgcv**.

Family: binomial
Link function: logit

Formula:
SAB.1.0 ~ s(PCA) + CURRENT + s(TIME) + SITE + MONTH + SEX + SUM100 +
s(SUM400) + s(DURATION) - 1

Parametric coefficients:

	Estimate	std. err.	t ratio	Pr(> t)
CURRENT	-0.76706	0.345	-2.223	0.026212
SITENorth	-3.187	2.161	-1.475	0.14029
SITESouth	-7.1128	2.577	-2.76	0.0057867
MONTH	0.29718	0.2691	1.104	0.2695
SEXM	1.167	0.7488	1.559	0.11909
SUM100	-0.75986	0.7529	-1.009	0.31285

Approximate significance of smooth terms:

	edf	chi.sq	p-value
s(PCA)	1.527	3.5892	0.10950
s(TIME)	2.994	12.013	0.0073003
s(SUM400)	1.444	2.9071	0.14730
s(DURATION)	2.107	7.4764	0.026584

R-sq.(adj) = 0.318

Deviance explained = 46.5%

UBRE score = -0.063291

Scale est. = 1

n = 83

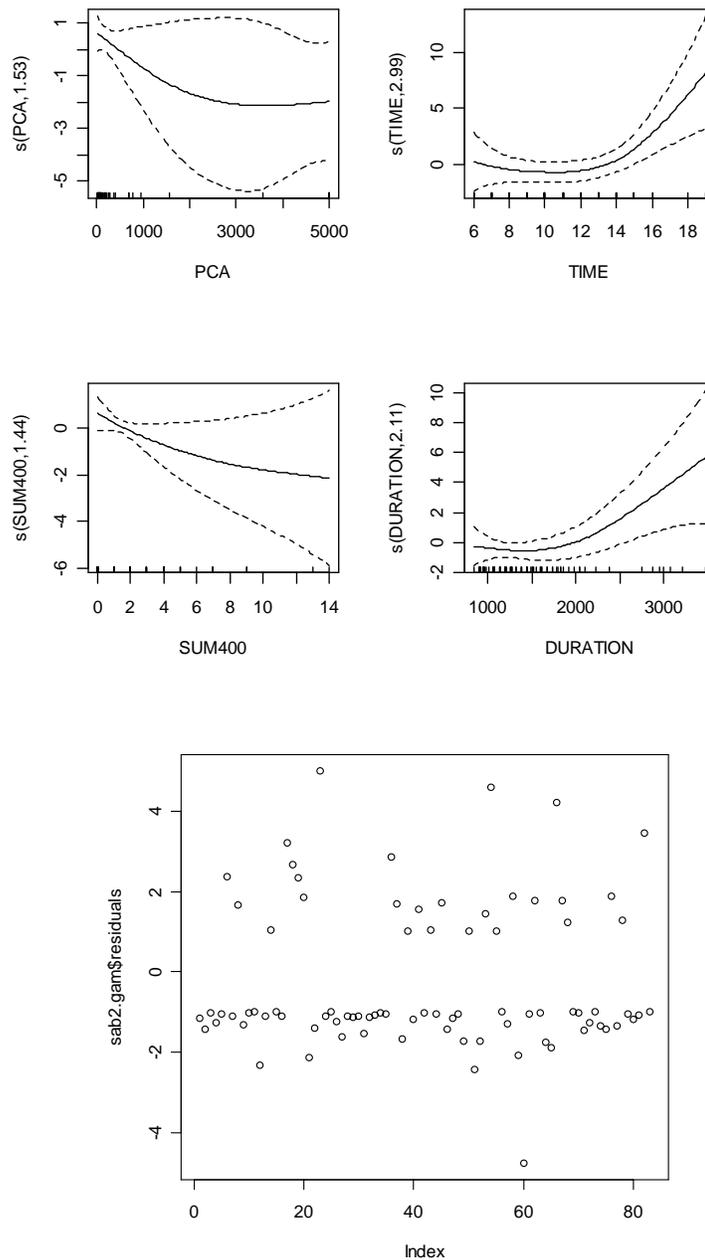


Figure 15. TOP – Relationship between smoothed component (solid line) of the explanatory variables (PCA, time of day, number of boats within 400m, and track duration) selected in the fitted GAM, and the response variable, likelihood of surface active behavior. The explanatory variables were allowed up to 9df and the degree of smoothing was automated by **mgcv**. The x-axis contains a rugplot, in which small ticks mark locations of observations. Zero on the y-axis corresponds to no effect of the covariate on the estimated response (likelihood of surface active behavior). The y-axis is labeled $s(\text{covariate name}, \text{estimated degrees of freedom})$. The dashed lines represent ± 2 standard errors, or roughly 95% confidence intervals. BOTTOM – Residuals of the fitted model. The plot of residuals indicates a good ability to classify samples into those with a high versus those with a low probability of having a surface active event occur.

DISCUSSION

COMMENTS ON GENERAL ANALYSIS APPROACH

Due to sample size constraints, it was not possible to use an ideal approach to data analysis. With 10 classes of independent variables (Year, day of season, time of day, site, age, sex, pod, individual, distance to boats, and number of boats), there would be over 1000 possible conditions, even if variables with large ranges like distances to boats, number of boats, day of season, and individual were placed into binary categories, as would only be appropriate for variables like year, site, or sex. Ideally there would be a sample of at least 10 observations in each cell (over 10,000 tracks). That is, even ignoring potentially important variables (e.g., behavior states during tracks, individual variation) and over-simplifying others (e.g., number of boats and distance), far more data would be needed than we were able to collect. Indeed, for some analyses, far more data would be required than there are whales to be studied in the population of interest.

We chose two different approaches to deal with these limitations. One was to reduce data to baseline and exposure conditions. While this ignored a great deal of potentially important information, the assumption made would be that the data were representative of overall conditions. That is, data were collected during five months of the year, 14 hours of the day, at two different sites, individuals ranging in age from new calves to the oldest individuals in the population, from both males and females, 38 individuals from all three pods (over 40% of the whales in the population), and with boat densities ranging from 0 to the highest in the area.

The other was to use Log-Linear Analysis and General Additive Models. These approaches look at the relative importance of various factors. “Importance” is a combination of magnitude of effect size and statistical significance. These approaches are implemented sequentially. That is, analysis begins anew after each important variable is identified. As a result, only one of a group of variables correlated with each other might be identified, even though another might be important in its own right. While such compromises limit the interpretations that can be drawn from the analysis, these techniques are suitable for a sparse data set such as ours.

Finally, we combined these two approaches to a limited degree. For example, site was found to be a significant factor. Therefore, we looked at data separately by site. Such results cannot be extrapolated to a large portion of the range, and such reanalysis needs to be pursued cautiously. As the number of analyses increases, the probability of obtaining spurious results increases.

SCAN-SAMPLING OF FOCAL GROUPS

General conclusions. Behavioral sequences varied significantly between locations, as expected. They also differed significantly with the presence of vessels. However, differences among pods were insignificant.

Southern resident killer whales behaved differently at the two study sites. They spent significantly more time foraging at the South site and significantly more time traveling at the North site. This is likely related to prey density. The waters off the south site have been identified as possible critical habitat for salmon (Protected Resources Division 2004). In addition, the three pods used these two sites with different intensities; J pod was more likely to be observed at the North site while L pod was more likely to be at the South site. Similarly, schools composed of members of all three pods were more likely to be observed at the North site.

Boats interacting within close vicinity of the whales (within 100m) affected their activity budget in a similar fashion at both sites. Based on data from the north site, whales were less likely to be foraging and significantly more likely to be traveling when interacting with vessels. This finding is in agreement with previous studies undertaken with the northern resident population (Williams, unpublished data). This effect raises concerns about the implications of this short-term displacement for the ability of individuals to acquire prey and the potential for long-term repercussions at the level of the population; especially in the light of the level of whale-watching activities carried out with Southern Residents. Vessel activity is also believed to reduce foraging success in other species (*Tursiops*, Allen and Read 2000).

After controlling for effects of site and boats, there was no significant difference in the data between years. Additional years of study will be needed to determine whether the two years happened to be similar in factors that vary on an annual time scale (e.g., prey abundance), or if our results will be robust across a range of conditions.

This study shows whales are displaced short distances by the presence of vessels. Thus whales may be displaced from optimal foraging routes. Further, Bain and Dahlheim (1994 and see Bain *et al.* unpublished ms.) suggested noise would mask echolocation signals and reduce foraging efficiency. These data are also consistent with observations of Northern Residents (Williams 2003; Williams *et al.* in review.). Thus we would encourage further study to determine how noise and proximity interact to reduce foraging effort.

The influence of pod identity. Boat interactions appear to have the same effect on all three pods, yet more sampling is required to fully understand the interaction between the composition of focal groups and the influence of boats on their activity state. Since pods appear to use certain areas preferentially (Hauser *et al.* 2005), and whales use both study sites differently, it would be necessary to assess the interaction between these two factors. Small sample size, especially in terms of control sequences, precludes this in the present analyses.

The influence of the distance between the focal school and boats. Boats within 100m had a clear effect on behavior states. Boats out to between 100m and 400m have effects as well, although we cannot state whether this is true throughout the range, or only in specific portions of the range. Similar but smaller effects were observed out to 1000m, but the sample with no boats within 1000m is too small to address whether this is simply due to chance. That is, there appears to be a negative trend between the boat effect size and the distance at which boats approached the focal groups. These results suggest the zone of influence of vessels in this area exceeds the 100m radius in current guidelines, and that more extensive guidelines such as those developed by the Whale Watch Operators Association NorthWest (2003), or those proposed by Orca Relief Citizens Alliance (2005) will be necessary to completely prevent harassment.

One potential explanation for these results is that noise impairs the ability to forage using echolocation. Previous studies have shown that the active acoustic foraging range of killer whales can be reduced or masked by boat noise (Erbe 2002; Bain and Dahlheim 1994, and see Bain *et al.* unpublished ms.), and since received noise levels typically decline with distance, the closer the boats are the more echolocation range is reduced (Williams *et al.* 2002b), potentially leading to foraging disruption, as suspected here. To test this hypothesis, acoustic monitoring would be required, as noise produced varies with engine type, and the speed at which boats operate. These data were beyond the scope of this study. The changes in deviation and directness indices observed here and reported by Williams *et al.* (2002ab) reflect increases in non-directional movement that would make behavior tend to more closely resemble foraging, indicating movements to avoid vessels cannot account for a *decrease* in foraging. Since this was an observational study, we cannot rule out that boats were attracted to whales unlikely to remain foraging, although it is unclear how boaters would be able to predict whale behavior.

THEODOLITE TRACKING OF FOCAL INDIVIDUALS

General conclusions. Despite a model specification approach that penalized over-parameterization, all five models fitted the data better with boat variables included than when they were excluded. The models lend support for concluding that boats exerted a small but significant effect on behavior of southern resident killer whales in 2003 and 2004, but that the relationships were complex and often non-linear.

Directness Index. The decrease in directness of travel with vessel traffic has appeared consistently in studies such as this one (Williams *et al.* 2002ab, Smith and Bain 2002). This pattern is consistent with whales making concerted efforts to evade boats. The sample size achieved in the 2004 field season resulted in the difference between the vessel present and vessel absent condition being significant. Intriguingly, the effect in this season was of similar magnitude to that reported for one boat paralleling the path of Northern Resident killer whales (Williams *et al.* 2002a): in the presence of boats, Northern Residents adopted a swimming path that was 13% less direct than during preceding, no-boat conditions, compared with 11% in Southern Residents.

Deviation Index. The increase in deviation with vessel traffic reported here based on the binary analysis is consistent with results of experiments conducted on female Northern Residents in

1995-6 (Williams *et al.* 2002a) and male Northern Residents in 1998 (Williams *et al.* 2002b). However, the GAM analysis found limited influence of vessel traffic. This is not surprising in light of the small overall effect found here, and that the deviation index in Northern Resident males only showed a change in response to intense whale watching (leapfrogging).

Breathing Patterns. Breathing changes have been inconsistent from one study to another. The GAM analysis suggests that interbreath-interval increases with increasing vessel number when the number of vessels is small, but decreases when the number of vessels is large. This “U-shaped” response pattern may account for the inconsistent results. There may be alternative tactics employed that vary depending on vessel number and proximity. Additional data will be needed to confirm whether the result reported here is robust in a wider range of conditions.

Surface Active Behavior. Changes in surface active behavior have been significant in many studies, although the direction of the change varies from one study to another. Our results suggest the inconsistency may be due to differences in methodology. For example, our work, and that of Williams (1999) suggest that SAB is maximized when one or a small number of boats approach closely, but SAB may be inhibited by other configurations of vessels. Data collected when boats are primarily in an inhibitory configuration may find vessels reduce rates of SAB. Alternatively, studies that pool all configurations may find no effect.

The analysis is further complicated by the relationship between track duration and measured values. Analysis of rates may need to be limited to longer tracks than some of those used here. Longer tracks would also be helpful for one-zero sampling, as that would allow subdividing tracks into multiple short segments. There is potential for binary analyses to be confounded by track duration, and other variables to be obscured in the GAM analysis.

It is also apparent that environmental factors, such as time of day and time of year, influence rates of surface active behavior. These need to be specifically addressed, as was done here in the GAM analysis, or samples need to be matched for these values.

One could speculate that threat displays consisting of surface active behaviors such as breaches, slaps, and fluke lifts (Tavolga 1966, Norris *et al.* 1994, Bain, pers. obs.) increased when vessels were close but not close enough to trigger an escape response. At greater distances, surface active behavior could be reduced to avoid attracting the attention of vessel operators. Baseline rates would reflect the use of surface active behavior for purposes independent of vessels such as communication among whales, foraging, and non-communicative purposes such as self-grooming.

Since Surface Active Behavior includes agonistic displays, it may reflect stress. Therefore, it is important to gain a better understanding of this complicated index of behavior.

Swimming Speed. The trend in swimming speed with respect to vessel traffic has been inconsistent from one study to another (e.g., contrast Kruse 1991 with Williams *et al.* 2002a). Month of year was found to be a significant factor in swimming speed in the GAM analysis. This is an important point to consider as vessel traffic also varies with time of year, making it an important confound for binary analyses such as the one performed here. Given the potential for

changes in swimming speed to carry energetic costs to whales, as well as reflecting their physical condition, the factors influencing swimming speed deserve more careful assessment.

Vessel Proximity. Killer whales had vessels in close proximity for a large proportion of daylight hours. Even with stewardship programs in place, whales spent over 20% of their time with at least one vessel closer than the 100 meters allowed under current guidelines, and over 75% of the time within a quarter mile of vessels.

RECOMMENDATIONS FOR FUTURE RESEARCH

This study found preliminary evidence consistent with small changes in behavior in the presence of whale watching vessels.

The effects support the development and enforcement of regulations for whale watchers, both recreational and commercial. Future research could address whether different approaches to whale watching have different degrees of impact (as paralleling and leap-frogging were compared by Williams *et al.* 2002ab). However, since it has proven difficult to demonstrate significant differences in behavioral responses to currently accepted practices and no disturbance, it could be expected to take carefully controlled experiments or many years of observation to compare the implications of proposed guidelines to current guidelines. Future research could also attempt further elucidation of age, sex, pod, and individual differences in responses to vessels.

It is possible that some may object to regulations and point to inconsistency of results among studies to question whether science adequately supports regulation. In that case, additional seasons of intensive effort such as those reported here could be justified. However, another approach to establishing which results are reproducible is to take advantage of existing data, such as those collected by Smith and Kriete (Smith and Bain 2002, Kriete 1995 and 2002). If new data are to be collected, we would point out that roughly 90% of our observations of whales without boats were obtained in May and June, despite extensive effort in July and August, and suggest that effort be focused early in the season. Given that variables like year and time of year are likely to be important, it would be valuable to mount an intensive effort so that changes through time do not confound the results. It may be valuable to add a third site (e.g., near the sites used by Kriete (2002) and Smith and Bain (2002), where the field of view is longer than at our North Site but whales tend to travel closer to shore than at the South Site. This site would be especially valuable for scan sampling work.

It is likely to be more productive to conduct research on impacts of vessel traffic through other mechanisms. Bain *et al.* (unpublished ms.) have suggested that the magnitude of impact through reduced foraging efficiency due to vessel noise is likely to be much larger than impact due to the increase in energy expenditure. Stress is a potential mechanism that has not been addressed. Similarly, Williams *et al.* (in review) suggested that for Northern Residents, reduction of energy acquisition for whales as vessels disrupt feeding activity has the potential to be 4-6 times greater than the increased energetic cost of avoiding boats. We urge additional analyses of this and other datasets to place subtle, short-term behavioral changes in an energetic

context. Any attempt to link vessel traffic to the decline of Southern Residents will require the synthesis of behavioral, acoustic, physiological and population dynamics studies in a quantitative modeling framework.

Studies on acoustic mechanisms could take several approaches. Noise could be measured with an acoustic tag placed on a whale. A less intrusive approach would be to measure noise with a hydrophone towed near whales. A third approach would be to incorporate noise monitoring into whale watch stewardship programs such as Soundwatch and M3. Shore-based hydrophones would allow monitoring changes in noise and behavior simultaneously. These data would complement source level and fixed-point ambient noise measurements already obtained. Data on how whales locate prey would also be needed, which might be accomplished through use of a Crittercam or imaging sonar.

Habituation to whale watching has been identified as a potential research area. However, the similarity of these results to those for Northern Residents, and the presence of significant effects in both Northern and Southern Residents after decades of intensive whale watching suggest habituation is far from complete and further habituation is unlikely to be substantial, although no data exist to address whether habituation occurred prior to Kruse's (1991) work in 1983 with Northern Residents or Kriete's (1995) work in 1987 with Southern Residents. Having said that, data from a NMFS-funded experimental study on Northern Residents in summer 2004 were collected to maximize sampling of individuals tracked in 1984, 1985, 1995-8, and 2002. This represents a time-series spanning 20 years, and would present an ideal opportunity to assess whether individuals change their behavior around boats over time in a way that is consistent with habituation, versus alternative explanations for changing behavior, such as aging or changes in whalewatching practices.

Strong behavioral responses of animals to disturbance do not always indicate population-level effects. Indeed, inter-specific variability in site fidelity and availability of alternative suitable habitat make it difficult to infer population-level consequences from inter-specific variability in sensitivity to disturbance (Gill *et al.* 2001). Thus it will be important to develop the link between short-term behavioral effects and population dynamics (see Bain *et al.* unpublished ms.).

SUMMARY AND CONCLUSIONS

As with Northern Residents, the presence of vessels inhibited foraging behavior. This may lead to a reduction in energy acquisition, and a priority research area would be to address directly through field studies whether prey capture actually is affected by vessel presence. In addition, modeling exercises should be carried out to identify potential mechanisms and the biological significance of any effects found.

Horizontal avoidance (evidenced by changes in Directness and Deviation indices) appears consistently across studies, although the statistical significance may depend on the sample size of the study and the pattern of vessel traffic experienced by the exposure group.

This may lead to an increase in energy expenditure, and again, research designed to test this hypothesis should be pursued.

Surface active behavior often shows significant differences depending on vessel activity, although such results are inconsistent in their magnitude and direction. Surface active behavior is largely composed of threat displays, so a relationship to vessel traffic is not unexpected. Non-linear responses to changes in vessel proximity and number probably account for the inconsistent results. Since many surface active behaviors are threat displays, they may be indicative of stress, and we urge additional research on captive and free-ranging killer whales to assess potential linkages among anthropogenic activity, stress and rates of surface active behavior. Some surface active behaviors like breaching require increased energy expenditure, so should be considered when calculating cumulative effects.

Average inter-breath interval and swimming speed do not show consistent changes across studies, and probably are not useful measures of the impact of vessel traffic. This is not to say that other measures of breathing patterns might not vary with vessel traffic. Alternatives to mean IBI, such as applying chaos theory to sequences of respiratory intervals (Bain 1995), have not been examined, to the best of our knowledge, but should be.

The behavior of Southern Residents in the presence of vessels is consistent with that observed in Northern Residents. This increases the confidence that can be placed in cross-population extrapolations, and in using individual Northern Residents as proxies for Southern Residents when conditions preclude experimentation on Southern Residents. Indeed, it is time for a meta-analysis of existing data from both populations, given the potential to increase statistical strength through the larger sample size to answer questions that small sample size precludes addressing through single studies alone.

Future research should focus on prey acquisition, and potential impact through other mechanisms such as noise and stress.

ACKNOWLEDGMENTS

We would like to thank Bill Gleason of San Juan Island National Historic Park for permission to use the South Site (permit SAJH-2004-SCI-0003), and the Reiss family for permission to use the North Site. We thank Erin Ashe, Amanda Coleman, Billy Cuddihy, Andy Foote, Megan Foss, Jeff Hogan and Derin Ross for assistance with data collection. The Whale Museum, Marine Mammal Monitoring Program (M3) and The Whale Watch Operators Association Northwest assisted with increasing the quantity of no-boat data at our North Site. We give special thanks to Dawn Noren for her help with the design of this project and review of the manuscript. We also thank the anonymous reviewers who made many helpful suggestions on earlier drafts of this manuscript.

REFERENCES

- Adimey, N. M. (1995). A descriptive study of the percussive behavior of orcas, *Orcinus orca*, in Johnstone Strait, British Columbia. MS. thesis, Nova Southeastern University, Fort Lauderdale, Florida.
- Allen, M. C. and A. J. Read. 2000. Habitat selection of foraging bottlenose dolphins in relation to boat density near Clearwater, Florida. *Mar. Mamm. Sci.* 16:815-824.
- Angradi, A.M., Consiglio, C. & Marini, L. 1993. Behaviour of striped dolphins (*Stenella coeruleoalba*) in the Central Tyrrhenian Sea in relation with commercial ships. *Proc. 7th Ann. Conf. ECS, Inverness, Scotland.* 77-79.
- Bailey, H. and D. Lusseau. 2004. Increasing the precision of theodolite tracking: Modified technique to calculate the altitude of land-based observation sites. *Mar. Mamm. Sci.* 20:880-885.
- Bain, D. E. 1986. Acoustic behavior of *Orcinus*: sequences, periodicity, behavioral correlates and an automated technique for call classification. In (B. C. Kirkevold and J. S. Lockard, eds.) *Behavioral Biology of Killer Whales*. Alan R. Liss. New York. 335-371.
- Bain, D. E. 1995. Breathing patterns and health: a place for chaos theory? Poster presented to the International Marine Animal Trainers Association Conference. Las Vegas, NV.
- Bain, D. E. 2002. A model linking energetic effects of whale watching to killer whale (*Orcinus orca*) population dynamics. Friday Harbor Laboratories, University of Washington, Friday Harbor, Washington.
- Bain, D. E., W. Anderson, F. Felleman, M. Harris and P. Higgins. 2002. Orca Recovery Conference Report 2002. Earth Island Institute. 30 pp.
- Bain, D. E. and M. E. Dahlheim. 1994. Effects of masking noise on detection thresholds of killer whales. Pages 243-256 in T. R. Loughlin, editor. *Marine mammals and the Exxon Valdez*. Academic Press, San Diego, California.
- Bain, D. E., R. Williams and A. W. Trites. Unpublished ms. Potential effects of whale watching on killer whale (*Orcinus orca*) population dynamics: insights from three models.
- Bejder, L., S. M. Dawson and J. A. Harraway. 1999. Responses by Hector's dolphins to Boats and swimmers in Porpoise Bay, New Zealand. *Marine Mammal Science.* 15(3):738-750.
- Bigg, M. A., P. F. Olesiuk, G. M. Ellis, J. K. B. Ford and K. C. Balcomb III. 1990. Social organization and genealogy of resident killer whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington State. *Rep. IWC Special Issue* 12:383-405.
- Caswell H. 2001. *Matrix Population Models*. Sinauer Associates, Boston, Massachusetts.

- Constantine, R. 1997. Monitoring the commercial swim-with-dolphin operations in the Bay of Islands. Science for Conservation. Dept. of Conservation, New Zealand (56): 59pp.
- Constantine, R. 2001. Increased avoidance of swimmers by wild bottlenose dolphins (*Tursiops truncatus*) due to long-term exposure to swim-with-dolphin tourism. Mar. Mam. Sc. 17 (4): 689-702.
- Corkeron, P. 1995. Humpback whales (*Megaptera novaeangliae*) in Hervey Bay, Queensland. Behaviour and responses to whale watching vessels. Working paper for the Workshop on the Scientific Aspects of Managing Whale Watching, Montecastello Di Vibio, Italy, 30 March – 4 April, 1995. Canadian Journal of Zoology, 73:1290-1299
- DeNardo, C., Dougherty, M., Hastie, G.D., Leaper, R., Wilson, B., and Thompson, P.M. 2001. A new technique for investigating variability in spatial relationships within groups of free ranging cetaceans. J. Appl. Ecol. 38: 888-895.
- Duffus, D. A., J. Bass, J. S. Dunham and C. D. Malcolm. 1998. Ecology and recreational use of gray whales in Clayoquot Sound, Vancouver Island, Canada 1991-1997. Abstract submitted to the Whale watching Research Workshop: Biology, politics, esthetics and human attitudes. Monaco, January 18, 1998. Available at: <http://office.geog.uvic.ca/dept/abstract.html>
- 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 18:394-418.
- Federal Register. 2004. Endangered and Threatened Wildlife and Plants: Proposed Threatened Status for Southern Resident Killer Whales. 69:76673-76682.
- Felleman, F. L., J. R. Heimlich-Boran and R. W. Osborne. 1991. Feeding Ecology of the killer whale (*Orcinus orca*). In (K. W. Pryor and K. S. Norris, eds.) Dolphin Societies: Discoveries and Puzzles. University of California Press. Berkeley. 113-147.
- Fleming, V. and T. H. Sarvas. 1999. Effects of whale watching on sperm whale (*Physeter macrocephalus*) behavior off Andoya, Norway. Proc. 13th Ann. Conf. ECS, Valencia, Spain. 103-107.
- Foote, A.D., R.W. Osborne, and A.R. Hoesel. 2004. Whale-call response to masking boat noise. *Nature* 428(April 29): 910.
- Ford, J.K.B., Ellis, G.M. and Balcomb, K.C. 2000. Killer Whales: The Natural History and Genealogy of *Orcinus orca* in British Columbia and Washington. 2nd Edn. UBC Press, Vancouver. 102pp.

- Ford, J. K. B., G. M. Ellis and P. F. Olesiuk. 2005. Linking prey and population dynamics: did food limitation cause recent declines of 'resident' killer whales (*Orcinus orca*) in British Columbia? Canadian Science Advisory Secretariat Research Document 2005/042. 31 pp.
- Galicia, E. and G. A. Baldassare. 1997. Effects of motorized tourboats on the behavior of nonbreeding American flamingos in Yucatan, Mexico. *Conservation Biology*. 11:1159-1165.
- Gill, J. A., Norris, K. and Sutherland, W. J. 2001. Why behavioural responses may not reflect the population consequences of human disturbance. *Biological Conservation*. 97: 265-268.
- Hauser, D. D., M. G. Logsdon, G. R. VanBlaricom, and E. E. Holmes. 2005. Southern resident killer whale (*Orcinus orca*) summer distribution patterns: effects of changing temporal scales. Paper presented at the 16th Biennial Conf. Biol. Mar. Mammals. San Diego, CA. p. 124.
- Heimlich-Boran, J. R. 1988. Behavioral ecology of killer whales (*Orcinus orca*) in the Pacific Northwest. *Can. J. Zool.* 66:565-578.
- Heimlich-Boran, J.R. 1993. Social organisation of the short-finned pilot whale, *Globicephala macrorhynchus*, with special reference to the comparative social ecology of delphinids. Ph.D.Thesis, Univ. of Cambridge. 132pp.
- Heimlich-Boran, J.R., S. L. Heimlich-Boran, R. Montero, and V. Martin. 1994. An overview of whale-watching in the Canary Islands. Proc. 8th Ann. Conf. ECS, Montpellier, France. 37-39.
- Hoelzel, A.R. 1993. Foraging behaviour and social group dynamics in Puget-Sound killer whales. *Anim. Behav.* 45(3):581-91.
- Janik, V. M. 1996. Changes in surfacing patterns of bottlenose dolphins in response to boat traffic. *Mar. Mam. Sci.* 12: 597-602.
- Jones, M. L. 1988. Response of gray whales to whale watching in San Ignacio Lagoon, Mexico. Page 16 in *Proceedings of the Workshop to Review and Evaluate Whale Watching Programs and Management Needs*. 14-16 Nov. 1988, Monterey, CA., Center for Marine Conservation, Washington, DC.
- Killer Whale Recovery Team. 2005. Draft National Recovery Strategy for Northern and Southern Resident Killer Whales (*Orcinus orca*). Prepared for Public Consultations, Spring 2005, for Fisheries and Oceans Canada, on behalf of the Resident Killer Whale Recovery Team. 70 pp.

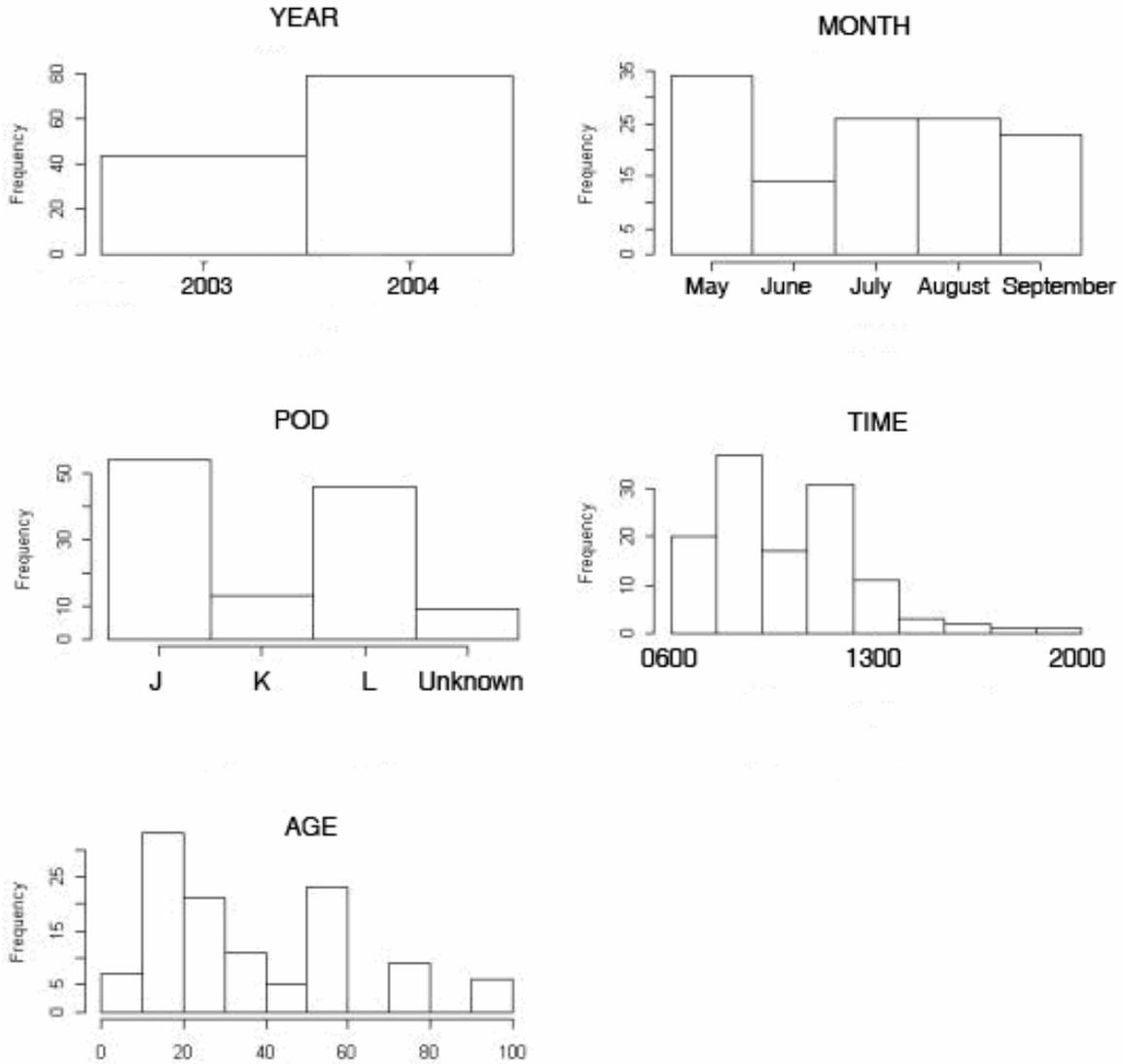
- Krahn, M.M., M.J. Ford, W.F. Perrin, P.R. Wade, R.P. Angliss, M.B. Hanson, B.L. Taylor, G.M. Ylitalo, M.E. Dahlheim, J.E. Stein, and R.S. Waples. 2004. 2004 Status review of southern resident killer whales (*Orcinus orca*) under the Endangered Species Act. U.S. Dept. Commer., NOAA Tech. Memo. NMFSNWFSC-62, 73 p.
- Krahn, M. M., P. R. Wade, S. T. Kalinowski, M. E. Dahlheim, B. L. Taylor, M. B. Hanson, G. M. Ylitalo, R. P. Angliss, J. E. Stein, and R. S. Waples. 2002. Status Review of Southern Resident killer whales (*Orcinus orca*) under the Endangered Species Act. U.S. Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC-54.
- Kriete, B. 1995. Bioenergetics in the killer whale, *Orcinus orca*. Ph.D. Thesis, University of British Columbia, Vancouver, BC. 138pp.
- Kriete, B. 2002. Bioenergetic changes from 1986 to 2001 in the southern resident killer whale population, *Orcinus orca*. Orca Relief Citizens' Alliance. Friday Harbor, WA.
- Kruse, S. 1991. The interactions between killer whales and boats in Johnstone Strait, B.C. pp. 149-59. In: K. Pryor and K.N. Norris (eds.) *Dolphin Societies, Discoveries and Puzzles*. University of California Press, Berkeley, California, USA. 397pp.
- Lacy, K.E., and E.P. Martins. 2003. The effect of anthropogenic habitat usage on the social behavior of a vulnerable species, *Cyclura nubila*. *Animal Conservation* 6: 3-9.
- Laist, D.W., A. R. Knowlton, J. G. Mead, A. S. Collet, and M. Podesta. 2001. Collisions between ships and whales. *Mar. Mam. Sci.* 17: 35-75.
- Lien, J. 2001. The Conservation Basis for the Regulation of Whale Watching in Canada by the Department of Fisheries and Oceans: A Precautionary Approach. *Can. Tech. Rep. Fish. Aquat. Sci.* 2363: vi + 38 p.
- Lusseau, D. 2003. The effects of tour boats on the behavior of bottlenose dolphins: Using Markov chains to model anthropogenic impacts. *Conservation Biology*. 17: 1785-1793.
- Lusseau D. 2004. The hidden cost of tourism: Effects of interactions with tour boats on the behavioural budget of two populations of bottlenose dolphins in Fiordland, New Zealand. *Ecology and Society* 9(1): 2 [online]
URL: <http://www.ecologyandsociety.org/vol9/iss1/art2>
- Marsh, J. A. 2004. Southern resident killer whale behavior workshop final report. NMFS NWFSC. 5 pp.
- Norris, K. S., B. Wursig and R. S. Wells. 1994. Social Behavior. In (K. S. Norris, B. Wursig, R. S. Wells and M. Wursig, eds.). *The Hawaiian Spinner Dolphin*. University of California Press. Berkeley, CA. 103-121.

- Nowacek, S.M., Wells, R.S. & Solow, A.R. 2001. Short-term effects of boat traffic on bottlenose dolphins, *Tursiops truncatus*, in Sarasota Bay, Florida. *Mar. Mam. Sci.* 17: 673-688.
- Olesiuk, P. F., M. A. Bigg and G. M. Ellis. 1990. Life history and population dynamics of resident killer whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington State. *Rep. IWC Special Issue* 12:209-243.
- Orca Relief Citizens Alliance. 2005. The best view of our whales is from shore. Orca Relief Citizens Alliance. Friday Harbor, WA. 2 pp.
- Protected Resources Division. 2004. Initial Assessment of NOAA Fisheries' Critical Habitat Analytical Review Teams For 13 Evolutionarily Significant Units of Pacific Salmon and *O. mykiss*. NOAA Fisheries Protected Resources Division. Portland, Oregon. 569 pp.
- Ritter, F. 2003. Interactions of Cetaceans with Whale Watching Boats – Implications for the Management of Whale Watching Tourism. M.E.E.R. e.V., Berlin, Germany, 91 pp.
- Romano, T. A., M. J. Keogh, C. Kelly, P. Feng, L. Berk, C. E. Schlundt, D. A. Carder and J. J. Finneran. 2004. Anthropogenic sound and marine mammal health: measures of the nervous and immune systems before and after intense sound exposure. *Can. J. Fish. Aquat. Sci.* 61:1124-1134.
- Safina, C., and J. Burger. 1983. Effects of human disturbance on reproductive success in the black skimmer. *Condor* 85:164-171.
- Smith, J. C. and D. E. Bain. 2002. Theodolite study of the effects of vessel traffic on killer whales (*Orcinus orca*) in the near-shore waters of Washington State: 2001 field season summary. Unpublished report.
- SPSS, Inc. 2003. Systat. www.systat.com.
- Tavolga, M. C. 1966. Behavior of the bottlenose dolphin (*Tursiops truncatus*): social interactions in a captive colony. In (K. S. Norris, ed.) *Whales, Dolphins and Porpoises*. University of California Press, Berkeley, CA. 718-730.
- Van Ginneken, A., D. Ellifrit, and K. C. Balcomb, III. 2000. Official orca survey field guide. Center for Whale Research, Friday Harbor, Washington.
- Van Parijs, S.M. and P. J. Corkeron. 2001. Boat traffic affects the acoustic behaviour of Pacific humpback dolphins, *Sousa chinensis*. *J. Mar. Biol. Assoc. U.K.* 81: 533-538.
- Venables, W.N. and Ripley, B.D. 2002. *Modern Applied Statistics with S*. New York: Springer.
- Visser, I. N. 1999. Propeller scars on and known home range of two orca (*Orcinus orca*) in New Zealand waters. *N. Z. J. Mar. Freshwat. Res.* 33:635-642.

- Waite, J. M. 1988. Alloparental care in killer whales (*Orcinus orca*). Master's thesis. University of California at Santa Cruz.
- Wells, R. S. and M. D. Scott. 1997. Seasonal incidence of boat strikes on bottlenose dolphins near Sarasota, FL, *Marine Mammal Science* 13(3):475-480.
- Whale Watch Operators Association Northwest. 2003. Best practices guidelines - 2003. <<http://www.nwwhalewatchers.org/guidelines.html>>
- Wiles, G.J. 2004. Washington State status report for the killer whale. Washington Department Fish and Wildlife, Olympia. 106 pp.
- Williams, R. M. 1999. Behavioural responses of killer whales to whale-watching: opportunistic observations and experimental approaches. Master's thesis. University of British Columbia. 61 pp.
- Williams, R., D. E. Bain, J. K. B. Ford and A. W. Trites. 2002. Behavioural responses of killer whales to a "leapfrogging" vessel. *J. Cet. Res. Manage.* 4:305-310.
- Williams, R., D. Lusseau and P. Hammond. In Review. "A framework for estimating energetic cost of disturbance to killer whales (*Orcinus orca*)"
- Williams, R., A. Trites and D. E. Bain. 2002. Behavioural responses of killer whales (*Orcinus orca*) to whale-watching boats: opportunistic observations and experimental approaches. *J. Zool. (Lond.)*. 256:255-270.
- Williams, R. 2003. Cetacean studies using platforms of opportunity. Ph.D. thesis. St. Andrews University. St. Andrews, Scotland.
- Wood, S. N. (2000). "Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties." *Journal of the Royal Statistical Society, Series B.* 62: 413-428.
- Wood, S. N. (2001). "mgcv: GAMs and Generalized Ridge Regression for R." *R News* 1(2): 20-25.
- Wood, S. N. (2003). "Thin-plate regression splines." *J. R. Statist. Soc. B* 65(1): 95-114.

APPENDIX 1. Sample size information.

Relative frequencies of samples are broken down by year, month, time of day, and age.

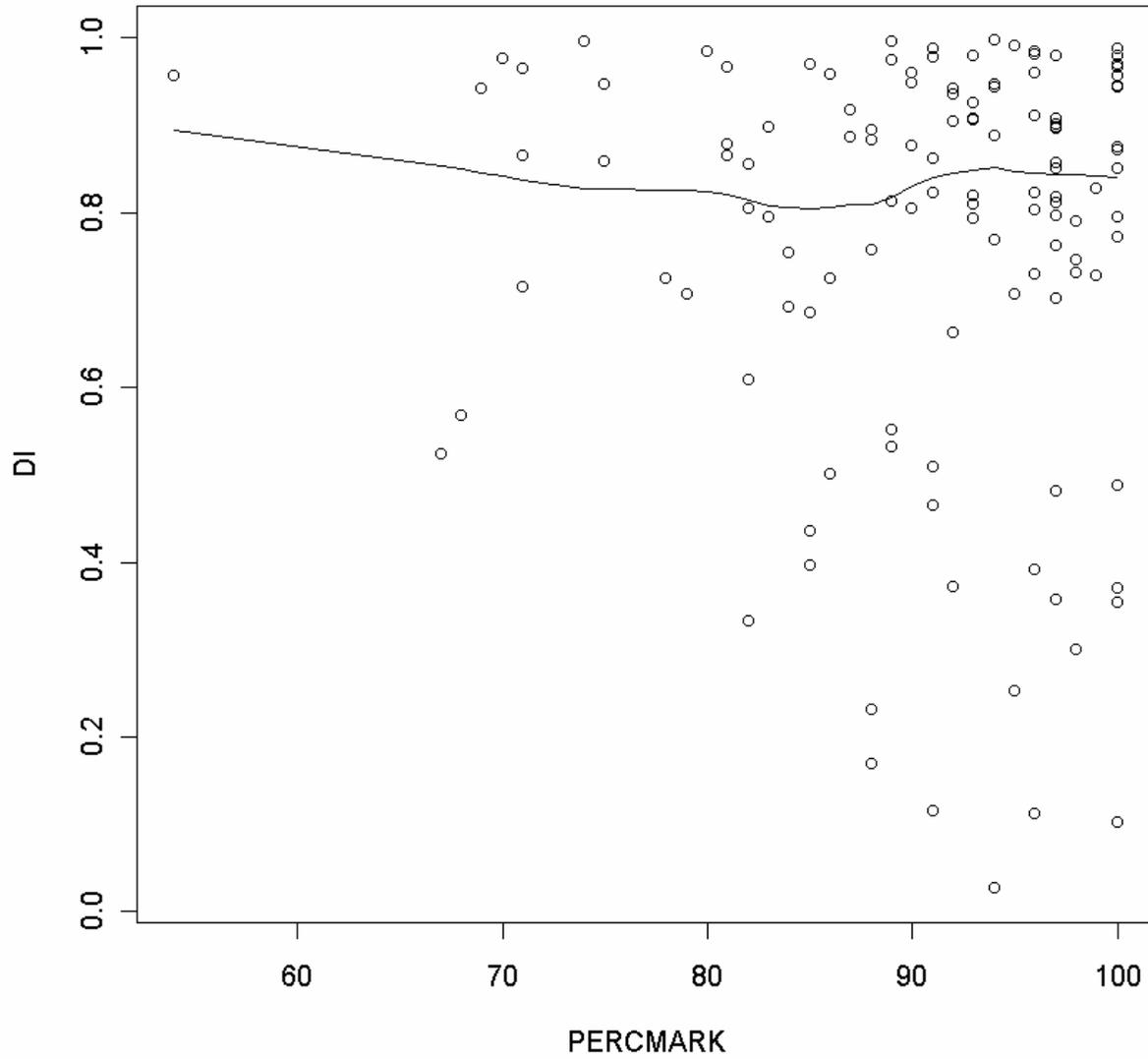


APPENDIX 2. Sample sizes broken down by various parameters.

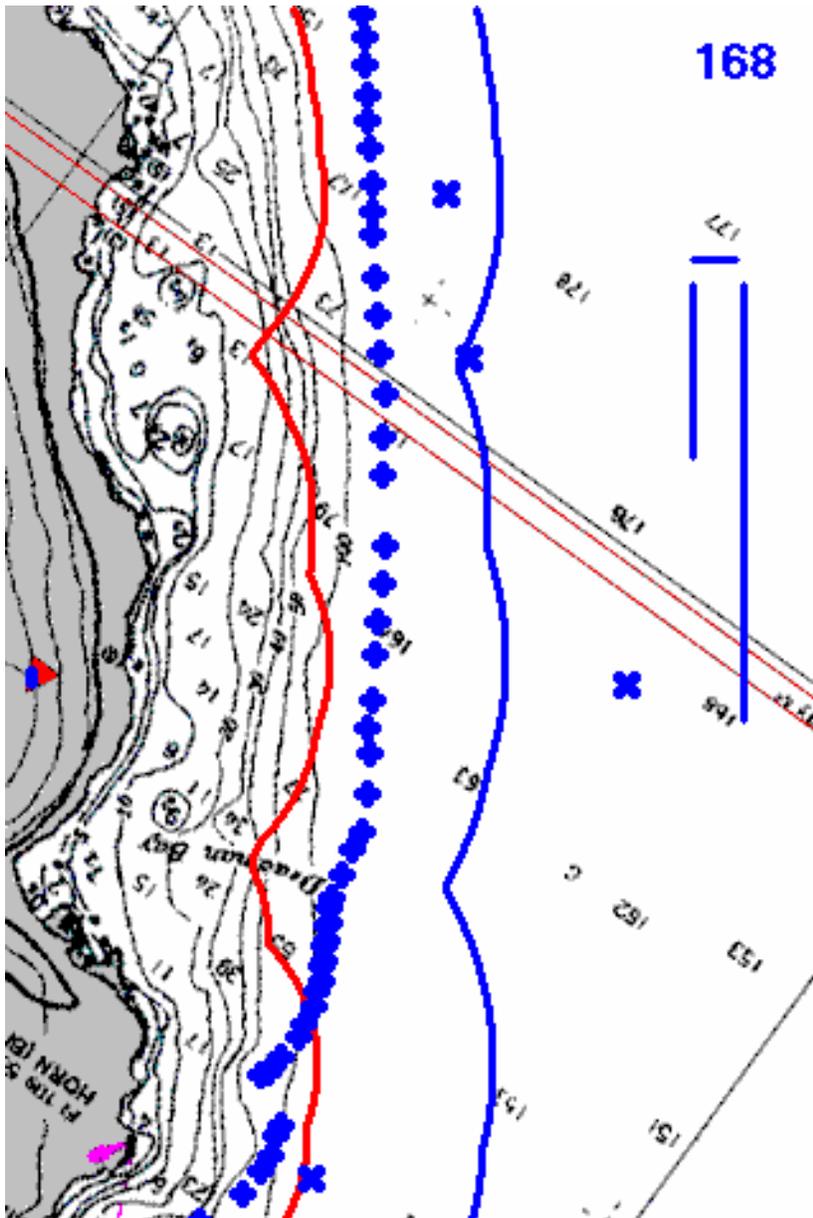
Sample size (number of tracks)

	2003	2004	Total
Study site			
North	30	52	82
South	14	25	39
Month			
May	0	33	33
June	0	13	13
July	7	19	26
August	14	12	26
September	23	0	23
Track duration			
15-20 min	13	27	40
>20 min	31	50	81
Sex of focal animal			
Female	8	39	47
Male	28	38	66
Unknown	8	0	8
Pod of focal animal			
J	9	44	53
K	4	9	13
L	22	24	46
Unknown	9	0	9
Traffic (number of unique boats in theodolite track)			
0	2	19	21
1	2	2	4
2	0	4	4
3	4	6	10
4	1	6	7
5	6	2	8
6-10	17	15	32
11-15	7	11	18
16-20	1	3	4
21-25	2	2	4
26-30	2	3	5
31-35	0	1	1
36-40	0	1	1
41-45	0	2	2
Total	44	77	121
Minimum number of focal individuals sampled (ignoring unknowns)			
	14	34	38

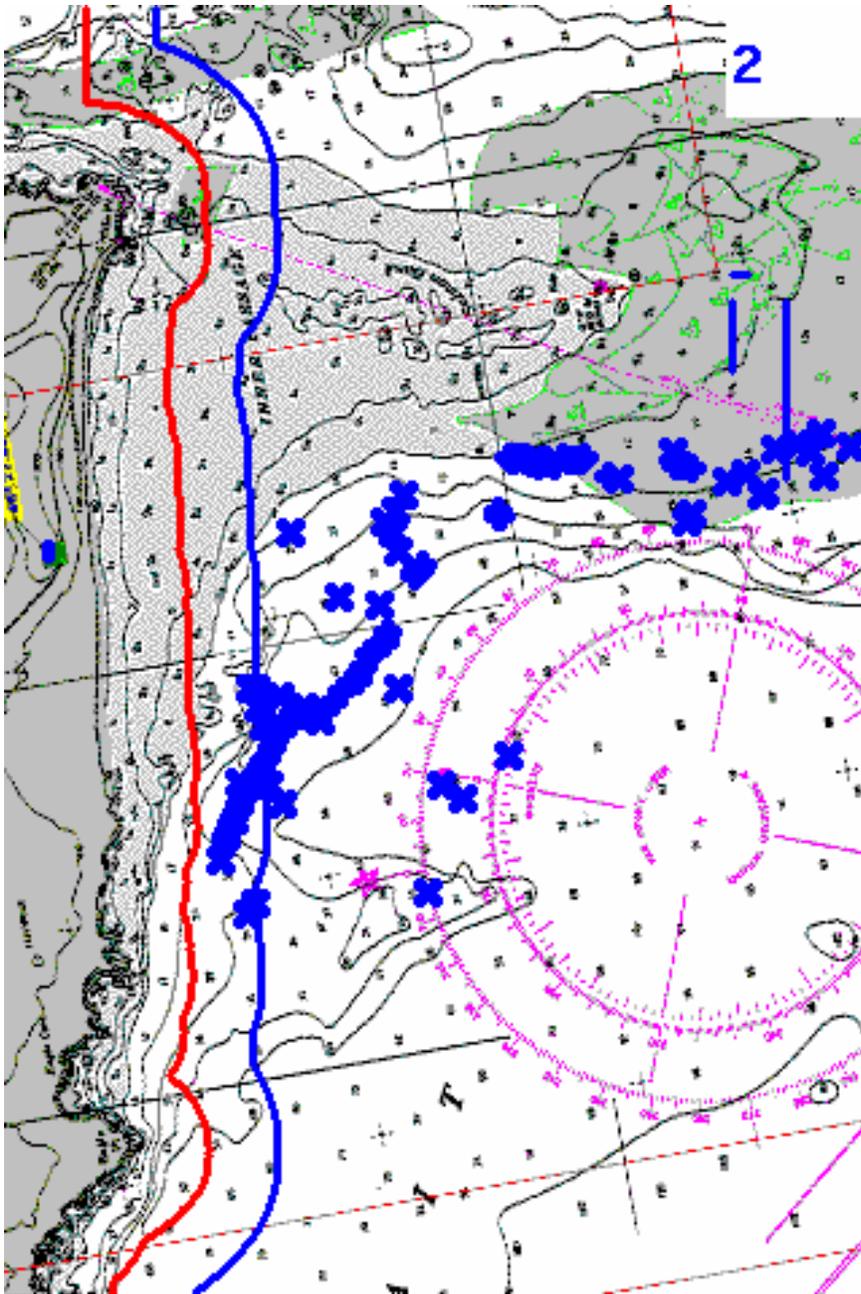
Appendix 4. Relationship of Directness Index to Percentage of Surfacing Marked. The leftmost point was identified as an outlier to be removed from the dataset.



Appendix 5 Sample Tracks Showing the Study Areas



Sample Track at the North Site. Lime Kiln Lighthouse is near the bottom of the chart. The red line is one-quarter mile from shore, and the blue line is one half mile from shore. +’s indicate whale positions and x’s represent boat positions. Scale bars are 100m, 400m, and 1000m. Number in the upper right corner is a real-time report of the distance between the last two marks (useful for checking the distance between a vessel and a whale, or how far a whale traveled underwater, and to help train observers in distance estimation). Blue spot on the red triangle is the theodolite position. As can be seen, tracks sometimes extended off the chart, but this approximates the study area well, as visibility is restricted beyond Lime Kiln Point and Edwards Point (top of image).



Sample Track at the South Site. Cattle Point Lighthouse is near the top of the chart. The red line is one-quarter mile from shore, and the blue line is one half mile from shore. +’s indicate whale positions and x’s represent boat positions. Scale bars are 100m, 400m, and 1000m. Number in the upper right corner is a real-time report of the distance between the last two marks (useful for checking the distance between a vessel and a whale, or how far a whale traveled underwater, and to help train observers in distance estimation). Blue spot on the green mark is the theodolite position. As can be seen, tracks sometimes extended off the chart, but this approximates the study area well, as visibility is restricted beyond Eagle Point and distance offshore becomes limiting.