The following supplements accompany the article

Whale watching disrupts feeding activities of minke whales on a feeding ground

Fredrik Christiansen[1,*], Marianne Rasmussen[2], David Lusseau[1,3]

[1]Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 2TZ, UK
[2]Húsavik Research Center, University of Iceland, Hafnarstett 3, 640 Húsavik, Iceland
[3]Marine Alliance for Science and Technology for Scotland, University of Aberdeen, Aberdeen AB24 2TZ, UK

*Email: f.christiansen@abdn.ac.uk

Marine Ecology Progress Series: 478: 239–251 (2013)

Supplement 1. The sensitivity of model outcomes to sampling errors

INTRODUCTION

The horizontal movement metric variables, directness index (DI) and deviation index (DEV), used in the generalised least squares (GLS), linear mixed effects (LME) models and generalised estimation equations (GEE) are estimated from measurements collected by various instruments (GPS, theodolite, range finder and photogrammetry), which all come with their own intrinsic measurement errors (instrument precision). When several measurements are used to calculate variables, the individual errors propagate in the model. If sufficiently large, these errors can ultimately affect the values and standard errors (SE) of the model parameters, and thus the results and conclusion of the entire analysis.

Apart from measurement errors, data can also be subject to observer errors. The detection probability of minke whales during sighting surveys in the Antarctic was found to be reverse-correlated to observer experience, with sighting rates by beginners being 42% lower than that of more experienced observers (Mori et al. 2003). In the context of the present study, such an observer bias would lead to an increased probability of missing surfacings, which would lead to an overestimation of inter-breath intervals (IBI) by inexperienced observers. Because the experience of observers increases through the field season, it could further be expected that observer errors due to inexperience would be highest at the beginning of the field season and then decrease towards the end.

We estimated the effect of measurement and observer errors on model parameter values and SE, using resampling methods.

MATERIALS AND METHODS

Propagating measurement errors

Individual measurement errors for each measurement tool (instrument) were estimated. For the land (control) data, theodolite measurement errors were assessed by taking repeated measurements of an object of known position (i.e. research boat) at various distances from the land-based research platform. The relationship between the measured (theodolite) distance and the true (GPS) distance was investigated using linear models (LM) in R (R version 2.12 2011). The main article describes how correction factors were added to the distance and bearing estimation formulas for both the land (control) and boat (impact) data.

To estimate the effect of distance on the distance and bearing estimate errors, a loop function was used to estimate the cumulative SE for different subsets of the data (the SE of the absolute values of the difference in distance and bearing), each using different lower threshold values for distance to be included. By starting with far distance values only and stepwise-reducing the lower distance threshold to include closer values, the cumulative SE throughout the entire range of distances could be calculated. The resulting error structure was extrapolated to all distances measured from land and sea respectively, so that an individual SE was provided for each original distance and bearing estimate to the whales. During the extrapolation process, only SE values estimated using more than 3 measurements were used, to avoid the effect of small sample size on SE estimates.

The final model of the model selection process (Model 12 in Table 3 in the main article) was then bootstrapped 1000 times, with distance and bearing values to the whales for each itinerary being taken randomly from a distribution of values, in which the means corresponded to the originally measured values and the standard deviations were the ones obtained from the error structure estimates. For the boat data, a SE for the position of the research platform itself (Lat_P and $Long_P$) was also estimated, since the GPS unit used also comes with an intrinsic error. The bootstrapping process generated a density distribution around the estimate of each model parameter, as well as their associated SE.

Apart from intrinsic measurement errors, due to the measuring instruments themselves, data can also suffer from extrinsic errors. For cetaceans, a common factor affecting the accuracy of behavioural data is sea conditions, and in particular sea state (Beaufort number). Detection probabilities of many cetaceans decline with sea state (DeMaster et al. 2001, Hammond et al. 2002), meaning that the probability of missing a surfacing, and thus overestimating IBI, will increase with sea state. The effect of sea state on IBI was investigated by adding sea state as a variable to the final model, and comparing model fit using Akaike's information criterion (AIC) and Bayesian information criterion (BIC).

Observer errors

To estimate the observer errors in the data, a jackknifing resampling method was used, where observers were removed from the data set one at a time and the model parameter values and SE were recomputed for the remaining subset of the data.

To test the effect of observer experience on errors through the field seasons, a loop function was used, where the cumulative SE was calculated for different subsets of the IBI data, starting with only data collected at the end of the season, and stepwise reducing the lower threshold for day, until the entire length of the field season was included. An increase in SE at lower threshold values would indicate an increase in observer error towards the beginning of the field season. Analyses were run separately for land- and boat-based data, as well as years.

RESULTS

Measurement errors

For the land-based (control) data, 184 theodolite measurements were taken of a known object (research boat), across a range of distances between 2500 and 6000 m. The loop functions showed an increase in measurement errors (SE) with distance both for distance (Fig. S1a) and bearing (Fig. S1b). For both measurements, the rate of increase in SE was relatively small up to about 4000 m, after which the SE increased drastically (Fig. S1).

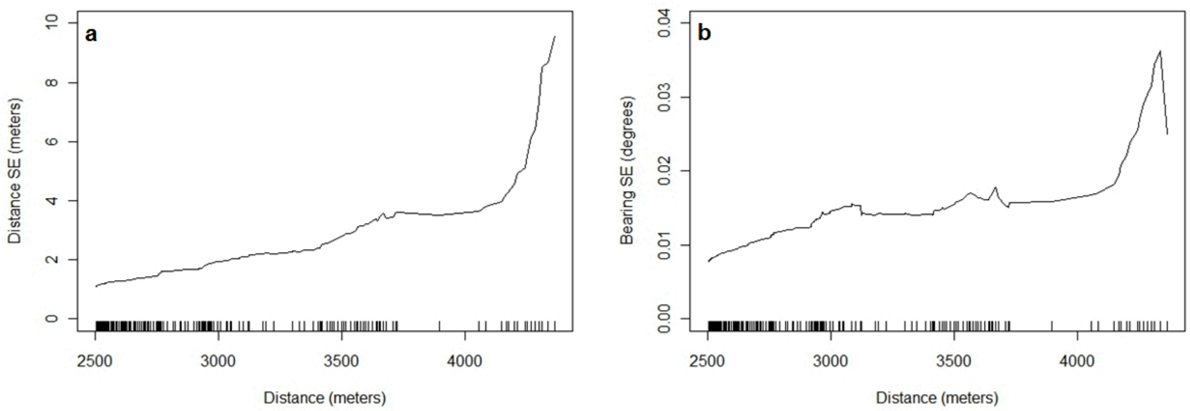

Fig. S1. Control (land) data measurement errors, SE of absolute difference between measured (theodolite) and true (GPS) (a) distance (n = 184) and (b) bearing (n = 184) at different distances (GPS). A rug plot showing the distribution of the data points for each figure is also shown

For the boat-based (impact) data, distance was measured from 61 photographs of a known object (whale-watching boat) at a distance between 80 and 900 m. For bearing, 48 measurements of true bearing were recorded between 130 and 1800 m with the range finder. For distance, the measurement error (SE) showed a slow increase with distance up to 700 m, after which the SE increased sharply (Fig. S2a). For bearing, the measurement error (SE) had a constant rate of increase with distance up to 1000 m (Fig. S2b), after which it levelled off. For the impact data, the accuracy of the Garmin eTrex H GPS unit was 4 m, giving an SE of 2 m for the observer position in any direction.

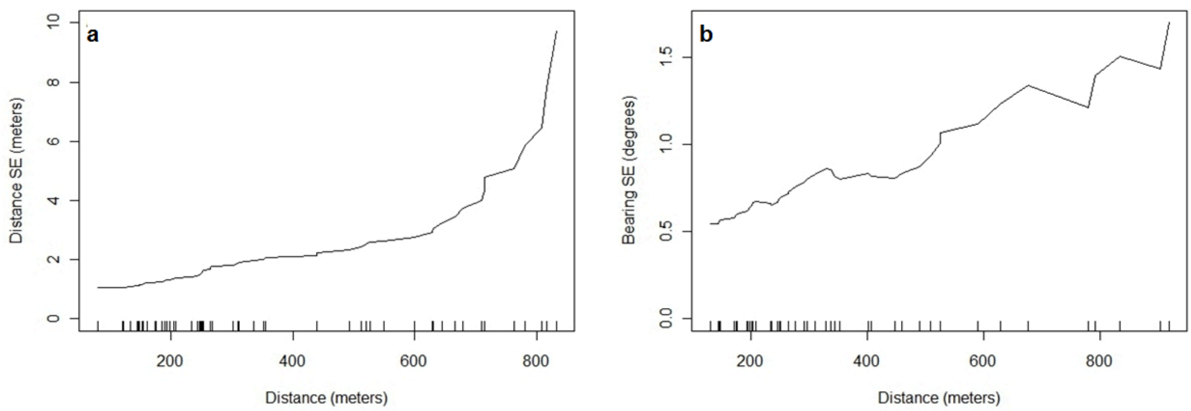

Fig. S2. Impact (boat) data measurement errors, SE of absolute difference between measured (photogrammetry/range finder) and true (GPS) (a) distance (n = 61) and (b) bearing (n = 48) at different distances (GPS). A rug plot showing the distribution of the data points for each figure is also shown

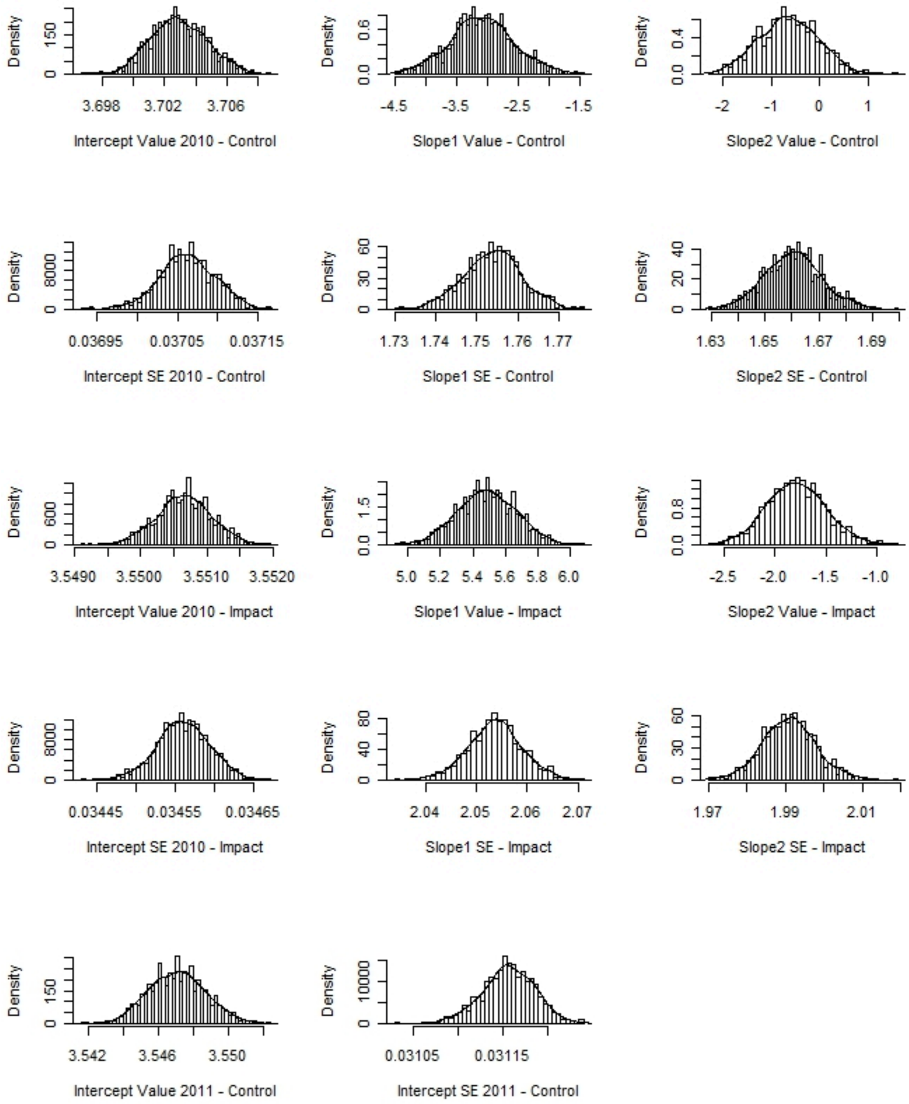

Fig. S3. Density distributions of the parameter values (Value) and their associated standard errors (SE) for the best-fitting model based on 1000 bootstrapping iterations

Parameter values estimated from the bootstrapping procedure were relatively narrow in their distribution, indicating that measurement errors had little effect on parameter estimates (Fig. S3). The SE for the parameter estimates were all normally distributed and did not cross zero, again indicating minor effects from measurement errors (Fig. S3). There was no overlap between intercept values and Slope 1 values, in the presence and absence of boats, as well as between years (Fig. 6 in the main article), demonstrating that results were robust to measurement errors. Measurement errors also had a negligible effect on the variance estimate of the random effect (follow) as well as the residual variance estimate of the best-fitting model (Fig. S4).

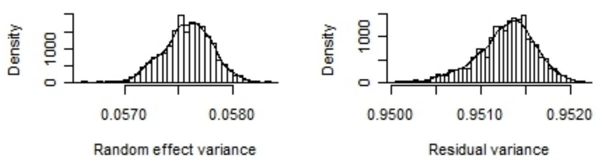

Fig. S4. Density distributions of the random effect and residual variance for the best-fitting model based on 1000 bootstrapping iterations

Sea state had no significant effect on IBI ($F_{1,4775} = 1.62$, p = 0.203) when added as a variable to the final model, and both AIC ($AIC_{Sea\ state} = 14730$, $AIC_{Original} = 14723$) and BIC ($BIC_{Sea\ state} = 14809$, $BIC_{Original} = 14796$) gave strong support for the original model, indicating that there was no bias in the data caused by sea state.

Observer errors

The re-estimated parameter and SE values generated from the jackknifing resampling method showed a relatively narrow distribution, similar to those of the full model (Table S1). The only exceptions were for the parameter values of directness index (Slope 1 and Slope 2), which had greater SE around the estimates, indicating that this variable was most sensitive to observer errors. Observer errors also had a negligible effect on the variance estimate of the random effect (follow) as well as the residual variance estimate of the best-fitting model (Table S1).

Table S1. Parameter values and standard errors (SE) and random effect and residual variance for the final model of the model selection process (Model 12 in Table 3 in the main article), as well as resampling mean and SE of the jackknifing resampling method. During the resampling process, observers were removed one at a time and the model parameter values, SE and variance were re-estimated for the remaining data subset

Parameter value	**Full model**	**Resampling mean**	**Resampling SE**
Intercept 2010, control	3.728	3.727	0.006
Slope 1, control	−7.013	−6.687	0.300
Slope 2, control	−5.638	−5.355	0.170
Intercept 2010, impact	3.551	3.549	0.004
Slope 1, impact	5.373	5.078	0.218
Slope 2, impact	−1.930	−1.898	0.048
Intercept 2011, control	3.572	3.572	0.006
Parameter SE			
Intercept 2010, control	0.037	0.040	0.002
Slope 1, control	1.733	1.731	0.046
Slope 2, control	1.685	1.681	0.042
Intercept 2010, impact	0.035	0.038	0.002
Slope 1, impact	2.042	2.060	0.017
Slope 2, impact	2.020	2.034	0.015
Intercept 2011, control	0.031	0.034	0.002
Variance			
Random effect	0.056	0.058	0.002
Residual	0.945	0.944	0.007

Both the control and impact data showed no visible temporal trends in SE within field seasons, indicating that there was no temporal bias in observer errors due to experience (Fig. S5). There was an increase in SE when the number of excluded days from the sample reached above 60–80 days; however, this was a consequence of small sample size associated with higher exclusion thresholds rather than an observer effect (Fig. S5).

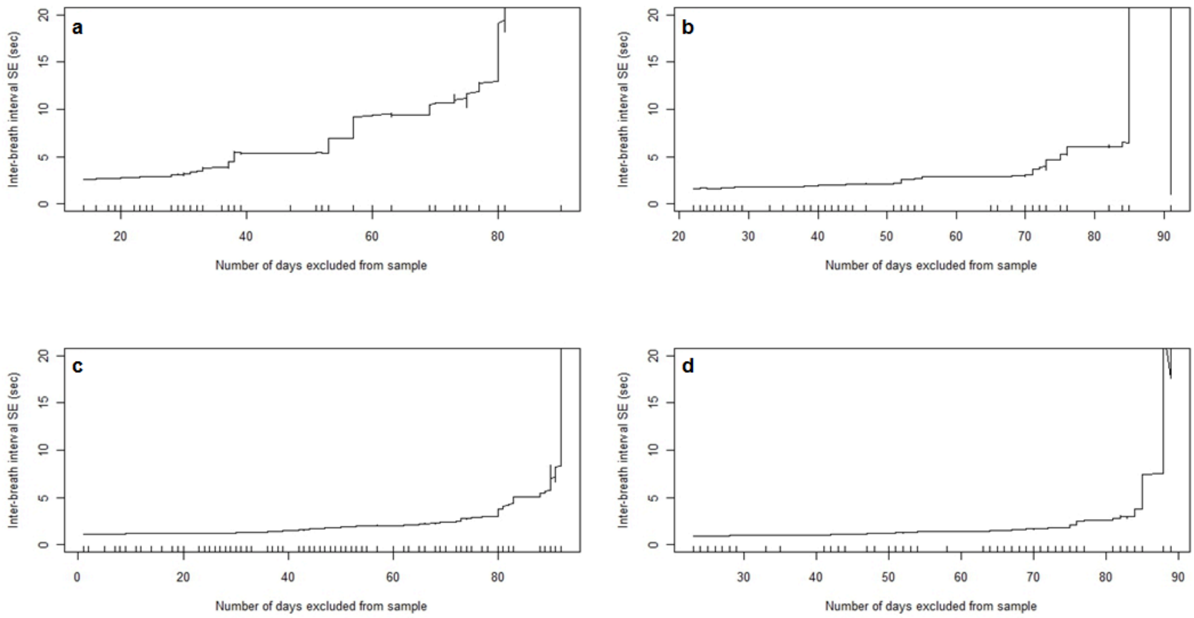

Fig. S5. Cumulative SE of inter-breath interval for different data subsets for (a) control 2010 and (b) 2011, and (c) impact 2010 and (d) 2011. The values on the *x*-axis indicate the lower threshold value from where data were included in the SE estimate. A rug plot showing the distribution of the data points for each figure is also shown

DISCUSSION

Data quality assessment is an important part of any study looking at the movement of animals. This study shows how measurement and observer errors can be accounted for and incorporated into the data analysis, to investigate the effects on model parameters. This approach makes it possible to evaluate the accuracy, or reliability, of the data, thus minimising the probability of making a type II error. In this study, the error distributions around the parameter values were relatively narrow, so the entire data set could be kept in the final analysis (Fig. S3). However, error propagation can sometimes be so large that it alters the conclusions of the entire study. In such a situation, this approach makes it possible to identify where errors are located and how they are expressed, so that appropriate actions can be taken to minimise them (e.g. by increasing research effort or limiting data analysis to a smaller, more reliable, subset of the data). Measurement errors increased drastically at distances above 4000 and 700 m for the control and impact data, respectively (Figs. S1 & S2). Measurement errors could have been reduced further, by setting an upper limit for distance, above which data would be excluded from analysis. A similar approach could also have been used to minimise the effect of observer errors due to inexperienced observers, by excluding data collected during the beginning of the field season. However, since the effect of measurement errors on parameter value estimates were rather small (Fig. S3), and there was no apparent temporal bias in SE of IBI (Fig. S5) due to observer errors, all the data could be kept in the final analysis.

The bootstrapping and jackknifing resampling methods highlighted which model parameters were most sensitive to measurement and observer errors, respectively (Fig. S3, Table S1). Measurement errors were largest around the slope parameter values, whilst the parameters influencing the intercept were narrower in their distributions (Fig. S3). The same holds for observer errors, which showed a higher SE around the slope parameter values than intercept (Table S1). Slope parameters are needed whenever continuous explanatory variables are used in a model. In the final model of the model selection procedure, only DI was a continuous variable. Thus, it becomes apparent that the major cause of measurement and observer errors in the model was due to errors around the DI estimate, rather than IBI. Since DI was estimated from measured positions (theodolite angles, photograph pixel count and range finder compass), while IBI was recorded from visual cues, efforts to improve data accuracy should be focused on reducing measurement errors (theodolite, photogrammetry and range finder), rather than observer errors (number of missed surfacings, measuring the wrong animal, etc.). Looking at error propagation in the data is thus very useful to highlight where research effort should be focused, to improve the accuracy of future studies.

LITERATURE CITED

DeMaster DP, Lowry LF, Frost KJ, Bengtson RA (2001) The effect of sea state on estimates of abundance for beluga whales (*Delphinapterus leucas*) in Norton Sound, Alaska. Fish Bull 99:197–201

Hammond PS, Berggren P, Benke H, Borchers DL and others (2002) Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. J Appl Ecol 39:361–376

Mori M, Butterworth DS, Brandão A, Rademeyer RA, Okamura H, Matsuda H (2003) Observer experience and Antarctic minke whale sighting ability in IWC/IDCR-SOWER surveys. J Cetacean Res Manag 5:1–11

Supplement 2. Effects of whale watching using deviation index as explanatory variable

INTRODUCTION

Directness index (DI) and deviation index (DEV) both provide useful horizontal movement metrics to describe the movement of animals (Williams et al. 2002). DI gives a measure of the linearity of movement while DEV gives a measure of the predictability of movement. The activity of minke whales can be described in terms of movement metrics using either of these 2 measurements. We showed in the main article that the effects of whale-watching boats on minke whale foraging activity could be inferred from changes in the relationship between inter-breath intervals (IBI) and DI. In this supplementary analysis, we investigated if the same held for DEV, i.e. if whale-watching boat interactions affected the relationship between IBI and DEV.

MATERIALS AND METHODS

As in the main article (see Table 1 in the main article), IBI, DEV, as well as the observation of surface feeding events (SFE) were used to define the 3 activity states used in the study: surface feeding (SF), foraging and non-feeding (NF). The same model selection procedure was used as in the main article; however, in this analysis, IBI was modelled as a function of DEV instead of DI to infer the effects of whale-watching boats on minke whale activity states.

RESULTS

Model selection and validation

Table S2 shows the results of the generalised least squares (GLS) and linear mixed effects (LME) models in R (R version 2.12 2011). Just as for DI in the main article, the polynomial model tests revealed a quadratic polynomial linear relationship between the log(IBI) and DEV (Model 10 in Table S2). DEV ($F_{2,4822}$ = 4.16, p = 0.0157) and vessel presence ($F_{1,795}$ = 14.79, p < 0.0001) both had a significant effect on minke whale IBI. In addition, the interaction between the 2 variables was significant ($F_{2,4822}$ = 13.27, p < 0.0001), suggesting that the effect of whale-watching boats on minke whale IBI is depending on the DEV, and hence the activity of the animal. As in the main article, IBI also varied between years ($F_{1,795}$ = 27.21, p < 0.0001). Again, SF activity could not be distinguished from the other activity states based on movement metrics alone, and adding SFE as a variable did not improve the fit of the model (Model 5 in Table S2). This justified the use of a separate generalised estimation equations (GEE) model to measure the effect of whale-watching boats on SF activity (see 'Measuring whale-watching effects on movement metrics' in the main article). As in the main article, the best-fitting model included an auto-correlation structure of IBI within follows, variance heterogeneity between treatment levels (vessel presence) ($\varepsilon_{ij} \sim N(0,\sigma_j^2)$) and follow as a random effect (Model 10 in Table S2). Model validation tests showed no violation of model assumptions.

The effects of whale watching on activity states

The fitted control data showed a quadratic relationship between IBI and DEV (Fig. S6a). During linear movement (low DEV values), IBI was relatively low, but started to increase when movement became more erratic and DEV values increased, indicating the presence of foraging dives in the absence of whale-watching boats (Fig. S6a). In contrast, the impact data showed a decrease in IBI at higher DEV values (Fig. S6b). This indicated a reduction in minke whale foraging activity during interactions with whale-watching boats (Fig. S6b).

Table S2. Selection of models explaining the observed variance in log-transformed inter-breath interval (IBI), using deviation index (DEV) as explanatory variable. GLS = generalised least squares, LME = linear mixed effects, Boat = vessel presence, SFE = surface feeding event, Follow = follow number, AR = auto-regression, AIC = Akaike's information criterion, BIC = Bayesian information criterion

Model no.	Model type	Fixed effects	Correlation structure	Variance structure	Random effects	df (among)	df (within)	AIC	BIC	ΔAIC	ΔBIC
1	GLS	Boat				2	5566	14981	15001	269	216
2	GLS	DEV				2	5566	15005	15024	292	239
3	GLS	DEV + Boat				3	5565	14998	15024	285	239
4	GLS	DEV × Boat				4	5564	14986	15019	273	234
5	GLS	DEV × Boat + SFE				5	5563	14991	15030	278	245
6	GLS	DEV × Boat + Year				5	5563	14949	14989	237	204
7	GLS	DEV × Boat + Year	~Follow (AR = 1)			5	5563	14919	14965	206	180
8	GLS	DEV × Boat + Year	~Follow (AR = 1)	σ^2_{Boat}		5	5563	14877	14930	164	144
9	LME	DEV × Boat + Year	~Follow (AR = 1)	σ^2_{Boat}	~Follow	5	5563	14763	14823	51	38
10	LME	(DEV + DEV^2) × Boat + Year	~Follow (AR = 1)	σ^2_{Boat}	~Follow	7	5561	14712	14785	0	0

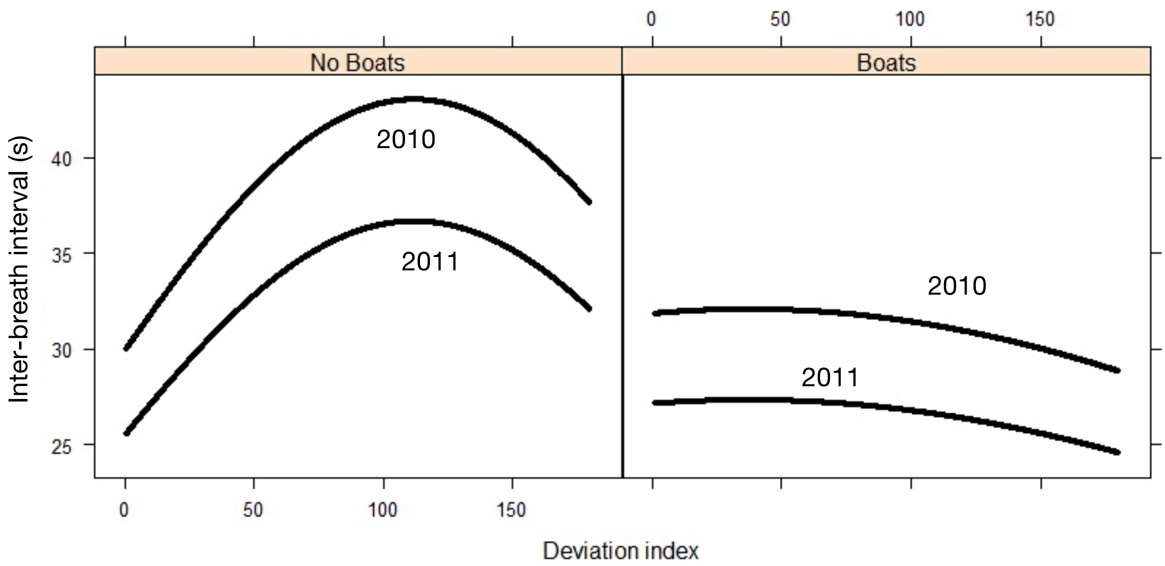

Fig. S6. *Balaenoptera acutorostrata*. Back-transformed fitted values of inter-breath interval (IBI) from the best-fitting model (Model 10 in Table S2) as a function of deviation index (DEV) in the absence (control; n = 1645) and presence (impact; n = 3979) of whale-watching boats in 2010 and 2011

DISCUSSION

The relationship between IBI and DEV was markedly different between control and impact situations (Fig. S6). The long and relatively erratic dives observed during foraging behaviour decreased in relative proportion during interactions with whale-watching boats. This indicated that whale-watching boats disrupted the foraging activity of minke whales. Further, the impact data actually showed a small decrease in IBI at higher values of DEV, even below the values normally observed during NF activity (Fig. S6). The lower values of IBI at higher DEV (higher turning angles) were most likely defining an avoidance behaviour of minke whales towards whale-watching boats, resulting from an increase in metabolic rate and an increase in erratic movement.

This analysis shows that DEV can be used in the same way as DI to model the effect of whale-watching boat interactions on minke whale movement metrics, from which effects on activity states can be inferred. The results give further support to the findings of the main article, that whale-watching boats disrupt the foraging activity of minke whales in Faxaflói Bay, Iceland. Effects on SF activity, however, could not be captured by using movement metrics (IBI and DEV), which indicated that a separate analysis (as described in the main article) was needed to measure the effect of whale watching on this activity state.

LITERATURE CITED

Williams R, Trites AW, Bain DE (2002) Behavioural responses of killer whales (*Orcinus orca*) to whale-watching boats: opportunistic observations and experimental approaches. J Zool (Lond) 256:255–270