Correcting the 2019 survey abundance of Bering-Chukchi-Beaufort Seas bowhead whales for disturbance from powered skiffs

Geof H. Givens[1], J. Craig George[2], Robert Suydam[2], Barbara Tudor[2], Andrew Von Duyke[2] and Brian Person[2]

Contact email: geof@geofgivens.com

ABSTRACT

The 2019 ice-based survey of Bering-Chukchi-Beaufort Seas bowhead whales was challenged by missed survey effort, unusual ice conditions and frequent use of motor-powered skiffs by hunters. All three of these likely led to downward bias in the abundance estimate of Givens *et al.* (2021). Data were collected about boat excursions during the survey period. Indices of short-term whale abundance at the survey perch and short-term boat noise disturbance were computed from the available data, where 'short term' refers to a few hours. A generalised additive model (GAM) was fitted to the results, predicting short-term whale abundance as a smooth function of the boat noise disturbance index, after controlling for long-term variation in the whale passage rate over the course of the season. The fitted GAM was then used to predict passage with and without the presence of boat noise. The ratio of the integrals of these two predicted passage curves provided a correction factor which can be applied post hoc to the abundance estimate of Givens *et al.* (2021). Variance of this correction factor was estimated using two approaches and found to be small. A wide array of sensitivity analyses was conducted to examine the robustness of the result to potential changes in methodology and the correction factor was found to be quite stable. The estimated correction factor would inflate the original abundance estimate by about 12%, yielding a corrected abundance of 14,025 (CV = 0.228). We recommend this replace the original abundance estimate.

KEYWORDS: ARCTIC; BOWHEAD WHALE; ABUNDANCE ESTIMATE; SURVEY – SHORE-BASED; WHALING – ABORIGINAL

INTRODUCTION

In Spring 2019, an ice-based survey of bowhead whale (*Balaena mysticetus*) abundance was conducted near Utqiaġvik, Alaska. A survey platform (hereafter, 'perch') was built on the edge of the shore-fast ice to observe the open lead system as whales migrated northeast past the site. Bowheads that passed within 4 km of the perch were observed, with detection probabilities modeled to depend on pod size, sighting distance and amount of ice in the lead. A correction factor for availability was applied to account for the proportion of the population that passes beyond 4 km visual range.

Givens *et al.* (2021) describe the statistical analysis of the resulting data, and provide an abundance estimate of 12,505 (CV = 0.228). The confidence interval for this 2019 estimate fully contains the confidence interval for the 2011 estimate produced by Givens *et al.* (2016) using nearly the same methods. However, the point estimate for 2019 is notably lower than the 2011 point estimate of 16,820. Givens *et al.* (2021) listed several reasons why

[1] Givens Statistical Solutions, 195 Moran Way, Santa Cruz CA 95062 USA.
[2] North Slope Borough Department of Wildlife Management, Utqiaġvik AK 99723 USA.

the 2019 estimate was likely biased downward, including missed survey effort at the start of the season and highly unusual ice conditions leading the whales to be unusually far offshore.

Givens *et al.* (2021) suggested that another major reason for the bias was the unprecedented frequent use of motor-powered skiffs by subsistence hunters during the survey period. They wrote:

> *In most previous survey years, hunters quietly paddle traditional umiaqs (i.e., seal skin covered boats) for much of spring whaling. However, in 2019, hunters consistently used power boats during the spring whaling season, in all regions of the coast near Utqiaġvik including a camp just 200 meters north from the perch. The disturbance to the survey was especially severe starting 7 May to the end of the season when the entire coast was opened to powered skiffs. The increased use of power boats was necessary for the community to land whales because of the unusual distribution of bowheads. The whales were inaccessible to skin boat hunters at the shore-fast lead edge, and much longer distances needed to be covered to hunt bowheads (and tow them back to the shore-fast ice). Bowheads avoid power boats by diverting further offshore and/or reducing surface times (Richardson et al., 1993). This behaviour persists for a period of time after the disturbance. Perch observers reported a downturn in the whale count in the hours after power boats launched.*

Indigenous hunters have known for centuries that noise disturbs migrating bowhead whales. Ethnographic studies at Utqiaġvik in 1881 described local practices to reduce noise disturbance during bowhead hunting: 'During this period, and while the whaling is going on, no pounding must be done in the village [several miles away], and it is not allowed even to rap with the knuckles on wood for fear of frightening away the whales' (Murdoch, 1892, p.274).

Numerous studies have shown strong reactions of bowheads to seismic surveys and large commercial vessels (National Research Council 2003; Richardson *et al*., 1993). Robertson *et al.* (2013) found surface-respiration-dive (SRD) behaviours changed when exposed to noise from seismic operations. Specifically, the number of blows per surfacing and surface time were significantly lower when seismic operations were ongoing. Robertson *et al.* (2015) further determined that changes in SRD affected visual detection probabilities of bowheads, including those exposed to seismic operations. Density and localised abundance estimates of disturbed whales increased from 3% to 63% when corrected for changes in SRD behavior data.

Reactions to seismic noise tend to be more overt and at a greater distance because of the extremely high sound source level. However, bowheads also react to commercial vessels and may show similar behaviours but at shorter distances (Richardson *et al*., 1993; McDonald *et al.*, 2012; Würsig *et al*., 2021). Whale hunters describe similar reactions of bowheads to perceived threats, where whales surface less and become skittish when powerboats are in the leads. While deploying satellite tags on bowhead whales, it has been noted that bowheads avoided skiffs when approached, particularly if they were not feeding. Carroll *et al*. (1980, p.83) described bowhead reactions to whale boats on the migration at Utqiaġvik: 'When bowheads are pursued by Eskimo whalers their usual reaction is docile escape. When they perceive something as threatening, bowheads quickly dive and do not resurface in the immediate area,' and outboard motors are therefore seldom used by hunters in spring because they 'cause whales to avoid the area'.

Currently, bowhead hunters at Utqiaġvik try to minimise noise along the lead edge during the migration by limiting snow machines and refraining from hunting belugas or waterfowl as well as restricting large vessel traffic (ice-breakers) in leads per the Conflict Avoidance Agreement (Lefevre, 2013). In the Barrow Whaling Captains Association (BWCA) guidelines, powerboats are typically not allowed in the leads until late in the season. Also, it was previously common to request that the BWCA crews reduce or suspend the use of powered skiffs for the entire season when an ice-based bowhead survey was conducted, particularly in the 1980s and 1990s.

In this paper, we attempt to correct the 2019 abundance estimate for bias caused by power boat disturbances. We seek a bias factor that can be estimated without modifying the analysis of Givens *et al.* (2021), so that it can be applied post hoc to their result. To the best of our knowledge, there is no standard way to make such a correction, so we spend considerable effort exploring the sensitivity and robustness of our approach (see Discussion).

DATA

The new data used in our analysis are records of powered skiff use by subsistence whalers during the survey season. No other power boats or commercial vessels were in the northeast Chukchi Sea near Utqiaġvik during the 2019 survey period. These data were obtained from the Barrow Volunteer Search and Rescue group who keep detailed records on hunting activities, including the locations of whaling crew camps, access trails and boating activities for each crew (Druckenmiller *et al*., 2013). Each entry in the dataset includes the date and time the power boat departed from the ice edge, the date and time it returned to the ice edge (at the same location as departure), and the distance of the launch/return site from the survey perch where observers were counting whales. There are no data available on where the boats travelled during their time at sea. In total, there were 199 boat excursions in the database, although several of these occurred after the survey had finished.

The other data used here are what Givens *et al.* (2021) call the Horvitz-Thompson contributions, denoted as w_i here. Each w_i corresponds to a single sighting of c_i whales at time t_i. The raw c_i counts are scaled by dividing by the estimated detection probability for that sighting, which was estimated from two-perch independent observer data (Givens *et al*., 2015). They are also multiplied by a factor of 0.5 if the sighting is 'Conditional' rather than 'New'. According to survey protocols, sightings are deemed Conditional if there is a 50% chance that the whale(s) was previously sighted. Aside from various technicalities, the total abundance estimated by Givens *et al*. (2021) is essentially the sum of the Horvitz-Thompson contributions divided by a factor accounting for availability (namely the proportion of whales that swim within visual sighting distance of the perch). Since this availability factor is applied globally to the sum of the w_i, we may view w_i as proportional to sighted abundance at time t_i. Thus, actual whale abundance passing the perch at a particular time may be high if a single large w_i was recorded then, or if a large number of smaller values of w_i were recorded all around the same time. By limiting our analysis to the boat data and Horvitz-Thompson contributions, we can estimate a boat disturbance correction factor without modifying the analysis of Givens *et al.* (2021).

Figure 1 shows the available data. The passage rate curve (right vertical axis, uncorrected for availability) is calculated by Givens *et al.* (2021) by binning the Horvitz-Thompson contributions in 12-hour bins, summing each

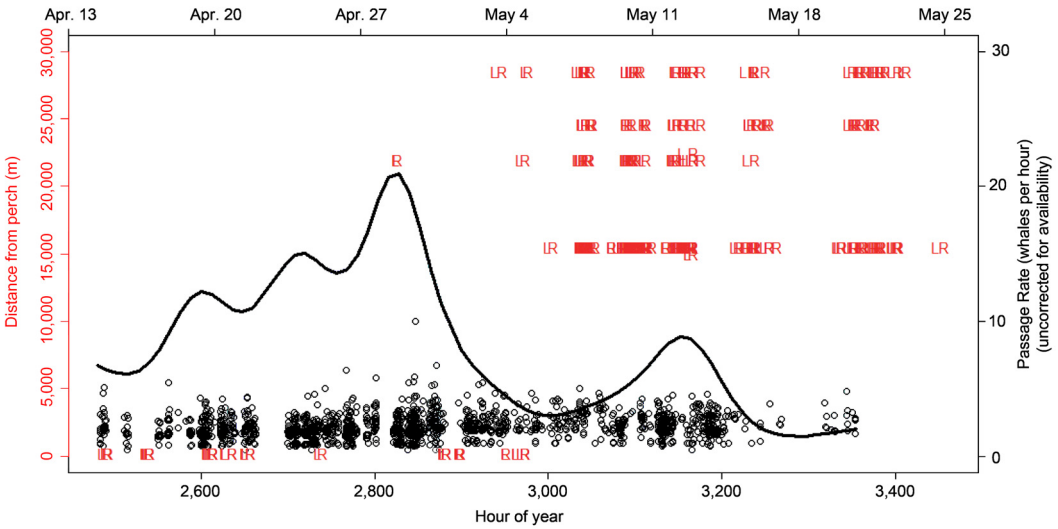

Fig. 1. Data used in analysis. Left axis and red data show the distance from the survey perch for each boat launch (L) and return (R) during the season. Right axis and black curve show the long-term passage rate curve, estimated by Givens *et al.* (2021). The circles represent individual sightings. The height of each circle, in reference to the right axis, is the Horvitz-Thompson contribution for that sighting, which is proportional to the number of whales represented by that sighting. The duration of the survey coincides with the extent of the passage rate curve. 'Uncorrected for availability' refers to the fact that the weighted Horvitz-Thompson contributions are later scaled by Givens *et al.* (2021) to account for the average proportion of whales that swam within 4 km of the survey perch.

bin, weighting each bin by the amount of qualifying survey effort during the period, and then smoothing the bin totals. The smooth was estimated from a shifted gamma generalised additive model (GAM) with log link, fit using the `mgcv` package in the **R** computing language (Wood, 2004, 2011, 2017; R Core Team, 2020), with $k = 20$ knots and the default generalised cross-validation approach to choose the smoothness penalty; see Givens *et al.* (2016, 2021) for more details. The boat data are also shown in Fig. 1. Each launch (L) and return (R) event is plotted, with the vertical axis corresponding to distance of the launch/return site from the survey perch.

STATISTICAL METHODS

Overview

Our analysis begins by recognising that the passage rate curve (Fig. 1) describes the long-term variation in the passage of bowheads during the season. The value of the passage rate curve at any point in time reflects a smoothing over many 12-hour bins, so the curve reflects a time scale of several days at any point. If there are short-term disturbances due to boat noise, these should be partly or mostly washed out from the passage rate curve due to its long-term focus.

Indeed, observers with decades of experience on the perch, and hunters with lifetimes of experience on the ocean, report that boat disturbances occur on a much shorter time scale. Perch observers believe that a boat noise event (or the striking of a whale by a hunter using an explosive weapon) results in a reduction of whales sighted from the perch for about 4–6 hours before returning to normal.

Thus, to estimate a boat effect, we first generate short-term indices of boat noise disturbance and whale passage over time. Our methods are described below. We then fit a generalised additive model (GAM) predicting the short-term whale passage. The predictors in this model are the long-term passage curve (to reflect long-term variation in sighted abundance, not impacted by boat noise) and the short-term boat noise disturbance index (to capture any short-term variation in sighted abundance attributable to boat noise). The fitted model tells us about the extent and nature of any boat noise effect, and from this we can estimate a correction factor using methods described later.

The season

Figure 1 shows that the use of power boats during the survey season was not homogeneous. During approximately the first half of the season (roughly 14 April to 4 May), essentially all boats operated (i.e., launched and returned) at a location only 200 m from the survey perch. Also, the total number of boat excursions during this period was relatively low compared to later. During the second half of the season (roughly 5–20 May), boat excursions were much more numerous and operated exclusively from sites many kilometers southwest of the perch. The actual long-term passage of whales followed the normal spring pattern, with most whales passing during the first half of the season, and fewer sighted later. This is consistent with the normal migratory pattern at Utqiaġvik and is not attributable to boats or other disturbance. Bowheads migrate in pulses of large numbers of subadult animals in the first half of the migration followed by lower numbers of reproductive animals passing later in the season (Koski *et al.*, 2006).

Because these two halves of the season were so different, we analysed them separately (splitting at hour 2986.67). The results are combined at the end of the analysis to compute the final correction factor. There were a couple of distant boat excursions during the first part of the season which we ignored. Any effect of these is likely dominated by the effect of the more numerous and disturbing near-perch boats, and eliminating these allows the GAM to accurately focus on the near-perch effect. Indeed, by splitting the season this way, we can have a single 'near-perch model' for the first half and a separate 'distant-perch model' for the second half. This makes GAM modeling simpler and more reliable.

Short-term whale abundance index

We use kernel methods to compute short-term indices of whale abundance and boat noise disturbance. In our usage, kernels are nonnegative functions that assign weight to a portion of the timeline. To generate the short-

term whale abundance index, we used a normal distribution with a standard deviation of 180 minutes as the kernel. This means we place this normal distribution shape centered over each t_i (and there are 2,008 Horvitz-Thompson contributions in total). Furthermore, each kernel is weighted by w_i, which is proportional to the number of whales it represents. The mathematical value of the short-term whale abundance index at any time t is:

$$W(t) = \sum_i w_i \phi(t; t_i, 180) / h_i$$

where $\phi(t;t_i,180)$ is the Gaussian density function with mean t_i (in minutes) and standard deviation 180, evaluated at t, and h_i is the proportion of the hour surrounding t_i during which qualifying survey effort occurred. (For example, if 1 new whale was seen with a detection probability of 1.0 but the perch was only operational for 40% of the hour, then that counts as 2.5 whales since $w_i = 1$ and $h_i = 0.4$.) The sum is taken over all Horvitz-Thompson contributions in the relevant half of the season, since *W(t)* is computed separately for each half.

The standard deviation of 180 was chosen for three reasons. First, experimentation with a wide variety of choices from 30 to 480 found that our choice yielded slightly stronger correlations with the boat noise disturbance index than narrower or wider bandwidths. Second, we wanted a sufficiently small value so that the computed whale index could reasonably be considered short-term; our choice ensures the kernel weight essentially vanishes six hours before/after the sighting. Third, our short-term boat noise index (below) is based on a six-hour washout period, and we believed it was sensible for both short-term indices to have equivalent bandwidths. The standard deviation of 180, the decision to correct for missed effort (via the h_i) and the time window for effort correction were all choices explored with sensitivity analyses described in the Discussion.

All the analyses in this paper (except for the very final correction factor) are computed on relative scales. For example, the passage rate curve in Fig. 1 does not scale for availability. Similarly, *W(t)* is only defined up to a proportionality constant, and thus serves as an index of relative abundance over time.

Short-term boat noise disturbance index

For brevity, we refer to this as the boat noise index, but it is important to stress that we are not computing an index of how much noise is in the water; instead, we are computing an index of how much the whales may be disturbed while boats are operating. The boat disturbance can be manifested by the whales changing path, delaying passage, speeding past and/or surfacing less, although survey observers believe the dominant disturbed behaviour is to change path to a more offshore route beyond visual sighting range, and also to make briefer and possibly fewer surfacing periods. Direct observations from boats, radio tagging and boat disturbance experiments confirm such responses (Richardson *et al*., 1993; National Research Council, 2003).

Figure 2 shows the kernels used to compute the boat noise index. The remainder of this section explains the approaches in detail.

For the first half of the season, all boat excursions started and ended at a site roughly 200 m from the survey perch, so we assume that the disturbance effect, if any, was immediate. The *i*th boat excursion has a launch time of L_i and return time of R_i (in hours). We again use a kernel approach to compute the index. The trapezoidal kernel shape is very simple: zero before L_i, a flat top from L_i to R_i, and then diminishing linearly to zero from R_i to $R_i + 6$. To express this mathematically, begin by defining the symmetric triangle density, evaluated at t, to equal $\tau(t;m,r) = (r - |t - m|)/r^2$ for $|t - m| \le r$ and zero elsewhere. This is a triangle shape, centered with a peak at m and diminishing linearly to zero at $m-r$ and $m + r$. Then the trapezoid kernel for the ith boat excursion has the following form:

$$b(t; L_i, R_i) = \begin{cases} 0 & t < L_i \\ \tau(L_i; L_i, 6) & L_i \le t \le R_i \\ \tau(t; R_i, 6) & R_i < t \le R_i + 6 \\ 0 & R_i + 6 < t \end{cases}$$

The short-term boat noise disturbance index for the first half of the season is the sum of these kernels for the 26 excursions during the period:

$$B(t) = \sum_{i=1}^{26} b(t; L_i, R_i)$$

This is a relative index, so scaling does not matter.

The choice of this kernel shape is driven by the experience of the survey observers. We have no data on where the boat goes during its excursion, and observers felt the most sensible model was constant disturbance during the excursion. As noted earlier, observers believe that conditions gradually return to normal about 4–6 hours after a disturbance, so the linear washout with $r = 6$ seemed to be the simplest model to reflect this experience. Sensitivity to the choice of the boat noise kernel is addressed in the Discussion.

The approach for the second half of the season is more complex. The sites where boats launched and returned were 15 to 28.5 km southwest of the perch. A whale present at the boat launch site at the time of launch may be disturbed and divert to a more offshore migration route, which it maintains thereafter, roughly parallel to the undisturbed route but too far offshore to be seen by the observers when it passes the perch. This behaviour was deemed most plausible by experienced survey observers. Our kernel accommodates this perception, but also reflects other possibilities. Since this hypothetical whale must first swim from the boat site to the perch before its unavailability to the visual observers occurs, we define the swim delay (D_i) to equal the distance (m) between the boat site and the perch, divided by 4,000. Zeh *et al.* (1993) report average bowhead swim speeds during the northbound migration, and 4km/h is a reasonable average value. Thus, the period of maximal disturbance is from $L_i + D_i$ to $R_i + D_i$. As previously, our kernel provides a six-hour linear washout period from $R_i + D_i$ to $R_i + D_i + 6$. However, there can also be disturbance from L_i to $L_i + D_i$. For example, boat noise travels well through water, so a whale at the perch at moment L_i may be disturbed then. Also, a whale somewhere between the boat launch site and the perch may also be disturbed at moment L_i, but then requires a swim delay of less than D_i to reach the perch (because the whale is closer). The disturbance in these types of cases may be less severe, or less frequent, because such whales are farther from the noise source. Therefore, we adopt the same linear tail leftward of $L_i + D_i$ that is used rightward of $R_i + D_i$. Of course, this tail is truncated at L_i because no disturbance of any sort occurs before the boat has launched, regardless of the whale's location. The result of this model is the trapezoid shape shown in the right panel of Fig. 2.

Mathematically, the trapezoidal kernel for the *i*th boat excursion during the second half of the season is:

$$b(t; L_i, R_i, D_i) = \left\{ \begin{array}{ll} 0 & t < L_i \\ \tau(t; L_i + D_i, 6) & L_i \le t < L_i + D_i \\ \tau(L_i + D_i; L_i + D_i, 6) & L_i + D_i \le t \le R_i + D_i \\ \tau(t; R_i + D_i, 6) & R_i + D_i < t \le R_i + D_i + 6 \\ 0 & R_i + D_i + 6 < t \end{array} \right\}$$

The short-term boat noise disturbance index for the second half of the season is the sum of these kernels for the 152 excursions during the period:

$$B(t) = \sum_{i=1}^{152} b(t; L_i, R_i)$$

Again, this is a relative index, so scaling does not matter. The *B(t)* indices are defined only during the survey period. Boats occurring before or after the survey are ignored.

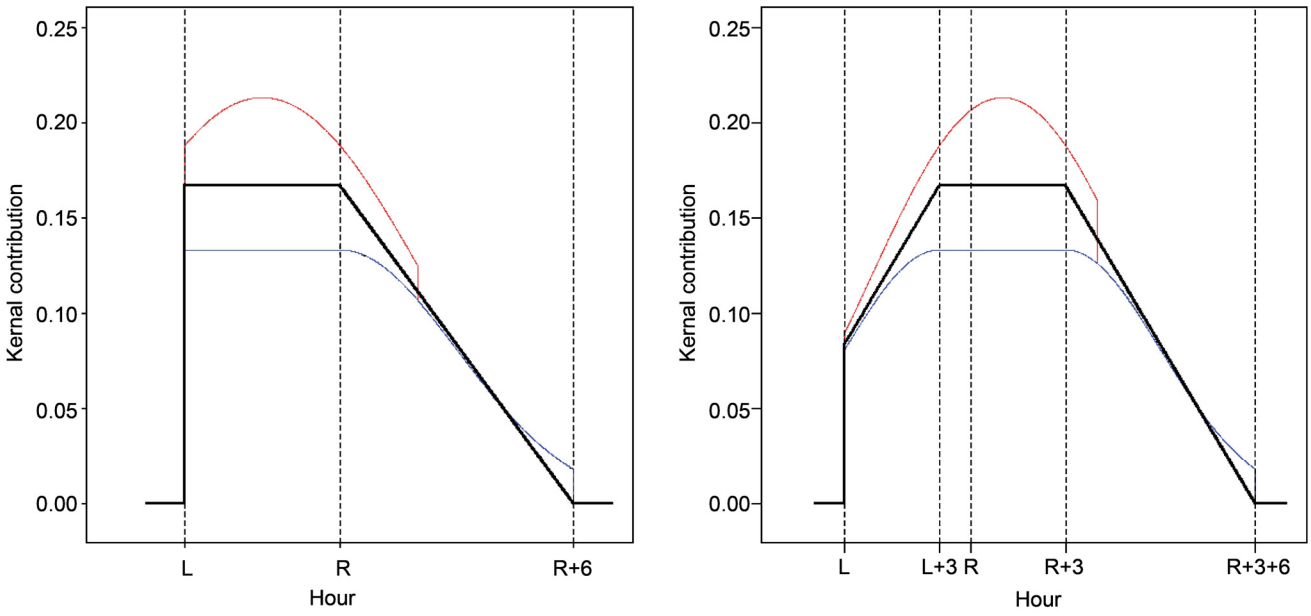

Fig. 2. Kernels used for the short-term boat noise disturbance index. The heavy black lines illustrate the kernels used for the main analysis for the first half of the season (left panel) and the second half of the season (right panel). The red and blue curves, which are partially superimposed in the right portion of each graph, correspond to sensitivity analyses mentioned in the Discussion. These kernels correspond to a hypothetical boat excursion lasting 4 hours from launch (L) to return (R). In the right panel, there is a 3-hour swim delay (i.e., 12 km from boat site to perch site).

Generalised additive model

We analysed the two halves of the season separately. Define *P(t)* to be the long-term passage rate curve shown in Fig. 1. We generated data for modeling by computing *P(t)*, *W(t)*, and *B(t)* at 10-minute intervals over the entire analysis period. Then the approach was to predict short-term whale passage using the long-term passage rate and boat noise index as explanatory variables. Specifically, we fit a quasipoisson GAM with log link:

$$\log \mu(t) = \beta_0 + \beta_1 P(t) + s(B(t))$$

where $\mu(t)$ denotes the theoretical mean of *W(t),* and *s( )* is a smooth function (spline). The model was fit using the `mgcv` package in the **R** computing language (Wood, 2004, 2011, 2017; R Core Team, 2020). We used the default thin plate regression spline with $k = 5$ knots for *s( )*, which proved suitably flexible without excessive variability. The default generalised cross-validation method was used to choose the smoothness penalty. Note that the model compensates for the unspecified scale parameters in the relative indices *W(t)* and *B(t)*. Furthermore, *s( )* is only determined up to a proportionality constant due to the presence of β_0 in the model. We used the default constraint on *s( )* to ensure identifiability: the function sums to zero over observed covariate values.

The quasipoisson model structure was chosen for consistency with the passage rate model used in the original Givens *et al.* (2021) analysis. Graphical analysis of the estimated variance-to-mean relationship, a diagnostic suggested by Ver Hoef *et al*. (2007), supports this choice for each half of the season.

Correction factor

In the GAM, *s( )* represents any change in short-term whale sighted abundance that can be attributed to noise, after controlling for long-term passage rate variation. The fitted GAM can be used to make predictions under different scenarios. First, we can simply compute predictions using the observed data, *P(t)* and *B(t)*. Denote these predictions as $\hat{W}(t)$. Alternatively, we can make predictions using *P(t)* but setting *B(t) = 0* always. This second set of predictions represents whale passage under the assumption of no boat noise disturbance, and by the design of the GAM above, it is log-linearly related to *P(t)*. Denote the no-boat-noise predictions as $\hat{W}_0(t)$.

Total abundance during the survey period is proportional to the integral of $\hat{W}(t)$. Total abundance that would have been observed without boat noise is proportional to the integral of $\hat{W}_0(t)$. Let

$$\alpha = \int \hat{W}_0(t)\,dt \,/ \int \hat{W}(t)\,dt$$

Then α is a correction factor that can be used to scale observed abundance to correct for the impact of boat noise disturbance. Note that there is no guarantee that $\alpha \leq 1$ since the value of α depends on the shape of *s( )*, which is entirely unconstrained in the GAM. Furthermore, since α is a ratio, the undetermined scale factor in the predicted relative indices of short-term sighted abundance cancels out.

Integrating *P(t)* over the two portions of the season reveals that 78.55% of whales passed during the first half of the season. Now we introduce superscripts on α to denote the first and second halves of the season since these were analysed separately. Then the corrected total abundance is:

$$0.7855\hat{N}\alpha^{(1)} + 0.2145\hat{N}\alpha^{(2)}$$

where $\hat{N}$ denotes the abundance estimate of Givens *et al.* (2021). Note that this can be expressed as $\lambda(p)\hat{N}$ where $\lambda(p) = p\alpha^{(1)} + (1-p)\alpha^{(2)}$ and, here, $p = 0.7855$. Thus, $\lambda(0.7855)$ is the total boat noise disturbance correction factor. We write the correction factor as a function of *p*, namely $\lambda(p)$, because our variance estimation approach accounts for uncertainty about *p*.

Variance estimation

We applied two variance estimation methods. Consideration of the merits of these and other options is given in the Discussion.

Two important issues must be emphasised. First, we sought methods that did not require changing or recomputing $\hat{N}$ from Givens *et al*. (2021). This analysis can therefore provide a standalone post hoc correction. Second, it was possible that $\lambda(p)$ and $\hat{N}$ are correlated. Estimating the variance of their product must consider this possibility.

Our first variance estimate applies an approach to GAM uncertainty recommended by Wood (2017) which assumes the correlation is zero. We applied the `mgcv` package to obtain the linear predictor matrix for the GAM, and then generated a posterior sample of predictions by sampling from the posterior multivariate normal distribution of the parameter vector, pre-multiplying by the linear predictor matrix, and applying the inverse link function. This posterior sample of predictions can be computed in both cases: with and without boat noise disturbance. Thus, a posterior sample of α can be obtained from these predictions, and hence a posterior for $\lambda(p)$. Finally, assuming independence:

$$\mathrm{var}\{\lambda(p)\hat{N}\} = \mathrm{var}\{\lambda(p)\}\,\mathrm{var}\{\hat{N}\} + \lambda(p)^2\,\mathrm{var}\{\hat{N}\} + \hat{N}^2\,\mathrm{var}\{\lambda(p)\}$$

where $\mathrm{var}\{\hat{N}\}$ is taken from Givens *et al.* (2021) and $\mathrm{var}\{\lambda(p)\}$ is the variance of the posterior sample described above.

The second approach is a conditional parametric bootstrap that accounts for potential correlation. We adopted the approach of simulating new response values (short-term sighted whale abundance) for given values of the predictor (short-term boat noise disturbance). We condition on whether a whale was seen at time *t*, in the sense that we generate new bootstrap nonzero weighted Horvitz-Thompson contributions at each t_i where a sighting was originally made, and do not generate any such contributions at other times. The generating distribution at time t_i is truncated Poisson(w_i), where the w_i are the weighted Horvitz-Thompson contributions, and the truncation prevents draws of 0. Thus, each bootstrap sampling iteration provides a new set of weighted Horvitz-Thompson contributions, and from this we can compute a bootstrap long-term passage rate curve $P^*(t)$ and a bootstrap short-term passage curve $W^*(t)$. Each bootstrap iteration then fits the GAM using the bootstrap data, producing a bootstrap correction factor $\lambda^*(p^*)$. Note that the proportion of the population passing during the first half of the season varies with each bootstrap iteration; the bootstrap proportion is p^*, and the bootstrap correction factor depends on this. We used 10,000 bootstrap iterations. This bootstrap required formidable

computing time, which we managed using the parallel computing capabilities of the `doParallel` and `foreach` packages in R (Wallig *et al.*, 2020a, 2020b).

To obtain a bootstrap estimate of the correlation, we note that the $\hat{N}$ of Givens *et al.* (2021) is approximately proportional to the sum of the weighted Horvitz-Thompson contributions. Therefore, without replicating the entire analysis of Givens *et al.* (2021) within the bootstrap, we can obtain bootstrap samples of a quantity that is proportional to abundance by summing the bootstrapped weighted Horvitz-Thompson contributions. The correlation between these bootstrap sums and the $\lambda^*(p^*)$ can then be estimated and serves as an estimate of $\mathrm{cor}\{\lambda(p),\hat{N}\}$. Finally, having bootstrap estimates for $\mathrm{var}\{\lambda(p)\}$ and $\mathrm{cor}\{\lambda(p),\hat{N}\}$, we can apply a standard, complex formula to compute the variance of the product $\lambda(p)\hat{N}$. We will report below that the bootstrap estimate of correlation was essentially zero, in which case the complex formula simplifies to the ordinary one used earlier.

RESULTS

Figures 3 and 4 show the short-term whale abundance and boat noise indices, *W(t)* and *B(t)*, plotted against time and against each other, for each half of the season. The right panels of these figures suggest relationships between these indices. This is desirable because it means we can estimate how boat noise disturbance impacts abundance. These panels suggest that the correlation between *W(t)* and *B(t)* is negative; indeed, the correlation is –0.193 for the first half of the season and –0.075 for the second. This suggests that the impact of boat noise disturbance is to reduce whale sightings.

Figure 5 shows the estimates of *s(B(t))* for each half of the season, using the `mgcViz` package (Fasiolo *et al.*, 2018) for visualisation of GAMs. For the first half of the season, low to moderate *B(t)* is associated with minor up-and-down fluctuations in *s( )*, which we interpret as sampling variability. For large *B(t)*, the effect of noise is strong downward impact on sightings. For the second half of the season, the results suggest an even more consistent downward effect as *B(t)* increases. Recall that the smooths *s( )* include an arbitrary identifiability constraint, so what matters is the curve shape, not whether values are positive or negative. These results clearly show that the effect of increasing boat noise disturbance is reduced short-term whale sighted abundance. The amount of uncertainty in these estimates, represented by the posterior samples shown in Fig. 5, is modest compared to the degree with which the effect itself varies.

Figure 6 shows the estimates $\hat{W}(t)$ and $\hat{W}_0(t)$ for the two halves of the season. Basically, the results show that boat noise had a small net impact on short-term abundance during the first half of the season when most of the

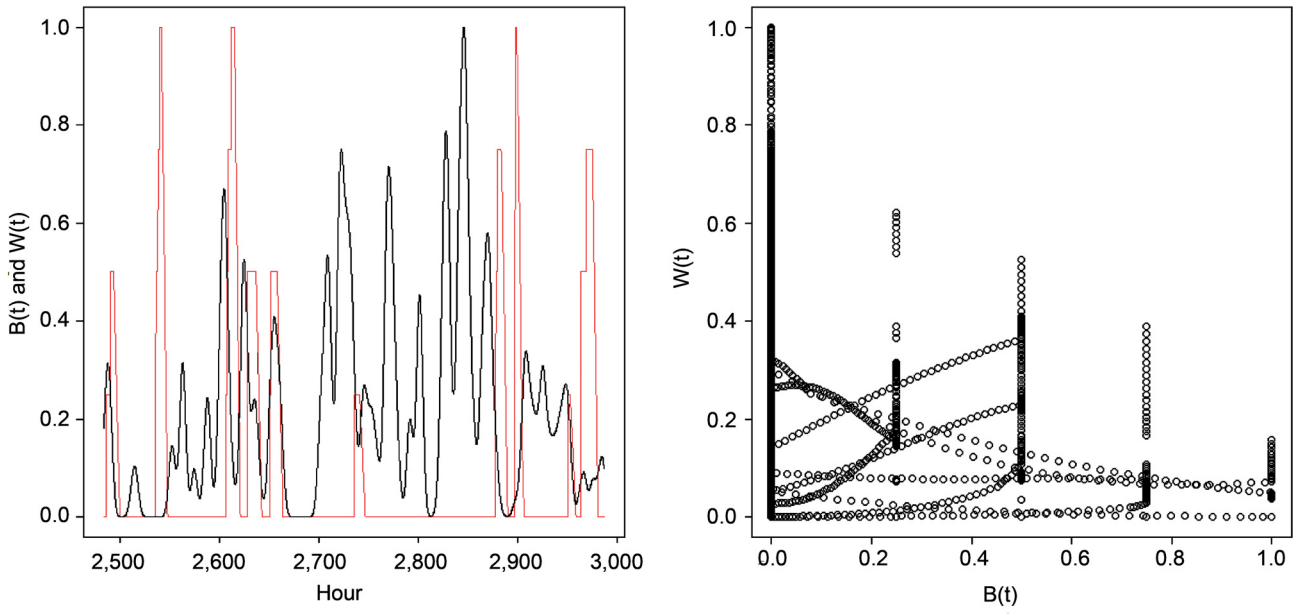

Fig. 3. Short-term boat noise disturbance index B(t) and short-term whale abundance index W(t), for the first half of the season. These indices are plotted against t in the left panel, with black for W(t) and red for B(t), and against each other in the right panel. Since these indices are relative, not absolute, we have normalised each for plotting by dividing by the maximums so the ranges are 0 to 1.

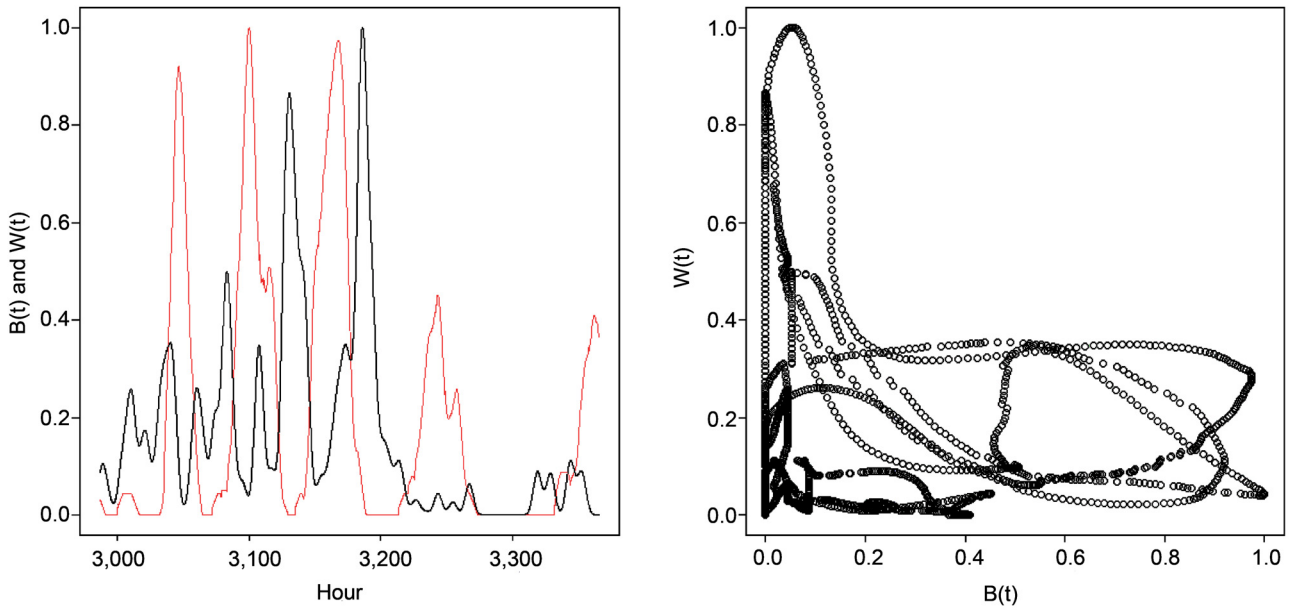

Fig. 4. Short-term boat noise disturbance index B(t) and short-term whale abundance index W(t), for the second half of the season. These indices are plotted against t in the left panel, with black for W(t) and red for B(t), and against each other in the right panel. Since these indices are relative, not absolute, we have normalised each for plotting by dividing by the maximums so the ranges are 0 to 1.

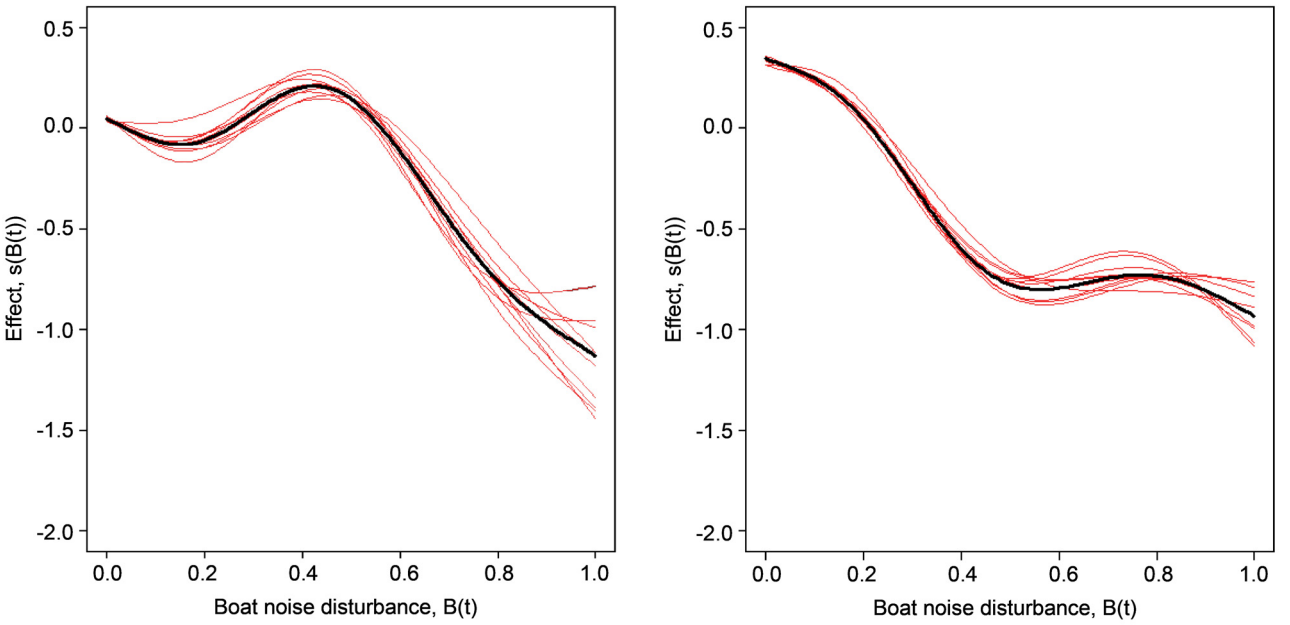

Fig. 5. Estimated smooth effects, *s(B(t))*, from the GAM, for the first (left panel) and second (right panel) half of the season. The black curve is the estimate. Also shown (red) are 10 draws from the posterior distributions of the estimates.

whales (78.55%) passed, and a large impact during the second half of the season when fewer whales passed. Indeed, numerically integrating the estimated curves (as described above) yields correction factors $\hat{\alpha}^{(1)} = 1.0230$ and $\hat{\alpha}^{(2)} = 1.4821$. Thus, the overall correction factor is $\lambda(0.7855) = 1.1215$. In other words, correcting for boat disturbances inflates the abundance estimate by about 12%. The corrected abundance is 14,025.

The variance of this estimated correction factor is quite small. Using the posterior variance method, the standard error of the correction factor is 0.0073. Applying the formula for the variance of a product of independent quantities, we obtain $\text{var}\{\lambda(0.7855)\hat{N}\} = 3202.731^2$. Table 1 compares the original and corrected abundance results. The bootstrap standard error of the correction factor is 0.0086, quite similar to the result of the first approach. The bootstrap estimate of correlation is $\text{cor}\{\lambda(p),\hat{N}\} = -0.0062$. This correlation is so tiny that we deemed it unnecessary to treat $\lambda(p)$ and $\hat{N}$ as dependent. Therefore, since the standard errors from the two approaches are similar and we prefer the first method (see Discussion), there was no need to further pursue the bootstrap.

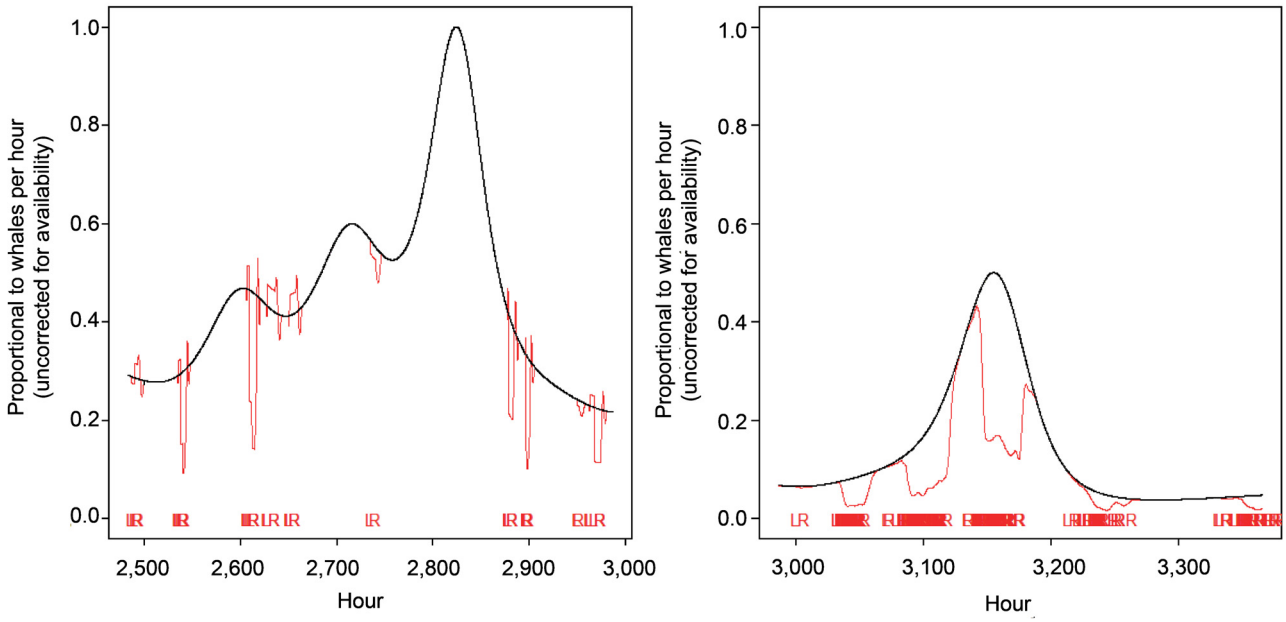

Fig. 6. GAM predictions of short-term whale abundance with boat noise ($\hat{W}(t)$, red) and without ($\hat{W}_0(t)$, black). Red and black are superimposed whenever $B(t) = 0$. The left panel corresponds to the first half of the season, and the right panel to the second half. Along the bottom of each plot, L and R denote boat launch and return events. The vertical scales are arbitrary; we have set the axes so the two panels together resemble Figure 1. 'Uncorrected for availability' refers to the fact that the weighted Horvitz-Thompson contributions are scaled by Givens *et al.* (2021) to account for the average proportion of whales that swam within 4 km of the survey perch, but that scaling was not necessary for this analysis and therefore not done here.

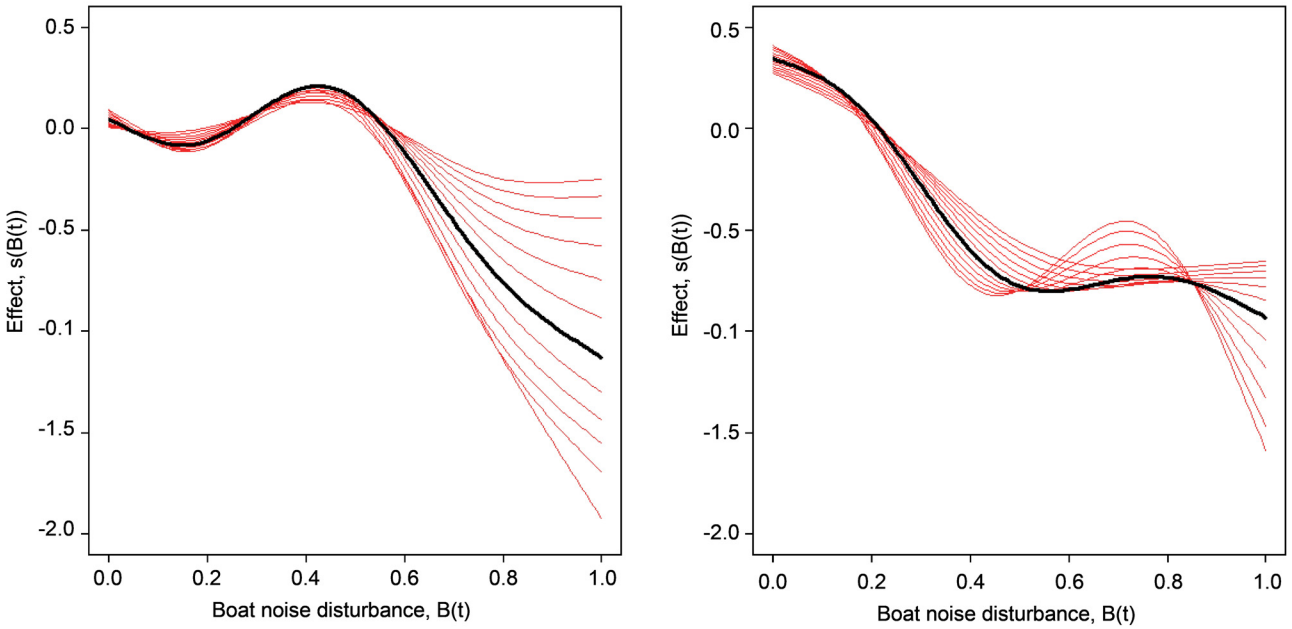

Fig. 7. Estimated smooth effects, *s(B(t))*, from the GAM, for the first (left panel) and second (right panel) half of the season. The red curves correspond to what would have been obtained if we had used kernel standard deviations for *W(t)* ranging from 30 to 360 minutes; black is our chosen estimate (180 minutes).

Table 1

Comparison of original and boat disturbance corrected abundance estimates. The CVs are virtually equal because the standard error of the correction factor is so small.

	Givens *et al.* (2021)	Corrected for boats
Abundance	12,505	14,025
Standard error	2,854	3,203
CV	0.228	0.228
95% confidence interval	7,994 to 19,560	8,964 to 21,942

DISCUSSION

Due to the nature of the problem and the limitations of the available data, our analysis is necessarily heuristic. Therefore, we devoted an unusually large amount of effort to studying the sensitivity of our results to various choices made during the analysis.

We mentioned above three reasons why we chose a kernel standard deviation of 180 minutes for the short-term whale abundance index: strength of signal, consistency with the boat noise disturbance index, and survey observers' sense of what constitutes 'short-term' variation in whale passage. As a sensitivity analysis, we also completed the analysis with kernel standard deviations ranging from 30 to 480 minutes in 30-minute increments. The high end of this range is surely too large, since a standard deviation of 480 minutes means that 'short-term' corresponds to a kernel spanning +/−16 hours. Figure 7 shows the estimated smooth noise effects, *s( )*, obtained after computing *W(t)* with various choices for the kernel standard deviation. Clearly, this choice is a source of greater uncertainty than the posterior distributions shown in Fig. 5. However, the curves are all broadly similar: downward effects on abundance for large values of *B(t)*. Moreover, the actual impact on the estimated correction factor is relatively small; see Table 2. Even for extremely (and in our view unreasonably) narrow or wide kernel spans, the final correction factor only ranges from 1.15 to 1.09; our estimate is 1.12.

Table 2

Correction factors obtained if the kernel standard deviation for computing $W(t)$ is changed. Our main analysis uses a kernel standard deviation of 180.

Kernel SD	30	60	90	120	150	180	210	240	270	300	330	360
$\hat{\alpha}^{(1)}$	1.05	1.04	1.04	1.03	1.03	1.02	1.02	1.01	1.01	1.01	1.00	1.00
$\hat{\alpha}^{(2)}$	1.55	1.54	1.53	1.52	1.50	1.48	1.47	1.45	1.44	1.42	1.41	1.40
$\lambda(0.7855)$	1.15	1.15	1.14	1.14	1.13	1.12	1.11	1.11	1.10	1.10	1.09	1.09

We did not repeat an analogous experiment for the width of the boat noise disturbance index kernel, because the washout period (six hours) was chosen on a firm scientific basis: the most experienced survey observers report that it takes about four to six hours for whale distribution and sightings to get back to normal after a noise disturbance from a boat or hunting weapon discharge. Therefore, allowing the left and right tails of this kernel to extend up to six hours is sensible. However, the shape of the kernel itself could be questioned. We experimented with two other approaches (Fig. 2). The first alternative kernel had a flat top from L to R, and Gaussian tails to the left and right, extending up to six hours, and truncated to zero beyond that. These are the blue examples in Fig. 2. The second alternative was based on the hypothesis that the actual events of launching and returning are the most disturbing, since they occur right at the edge of the open lead. The kernel was the sum of two sub-kernels: a Gaussian kernel centered at L, and another centered at R. The tails of each sub-kernel were truncated at six hours. The example resulting kernels are red in Fig. 2. Both alternatives assigned no kernel weight before L and applied the swim delay approach discussed previously.

We combined the sensitivity analysis about kernel shape for *B(t)* with another issue: the correction for missed survey effort when calculating *W(t)*. Recall that our analysis weights *W(t)* kernels by the inverse of h_i, the proportion of the hour surrounding t_i during which qualifying survey effort occurred. As a sensitivity analysis, we re-analysed using a two-hour interval centered at t_i, and again using no missed effort correction whatsoever. Table 3 shows the results of this combined sensitivity analysis. All these trials used 180 minutes as the bandwidth for the *W(t)* kernels. The results show that these variations have almost no effect on the estimated correction factors. It is well known that kernel shape usually has only minor impacts in density estimation and similar applications (Givens *et al*., 2013). The correction for missed effort has little effect because there is not much missed effort surrounding any sighting.

The bias correction factor, α, is a ratio of two integrals of predictions that have been back-transformed by inverting the log link in the quaisipoisson models. In principle, a back-transformation bias is introduced into the predictions from loglinear models when predictions on the log scale are back-transformed to the original scale of the response variable (e.g., Smith, 1993). Although one might attempt a complex bias correction separately for both components of the ratio before computing α, we believe this is overkill due to the heuristic/approximate

Table 3

Sensitivity analysis regarding boat noise kernel and effort correction for short-term whale abundance index. Three kernels were tried for computing $B(t)$: the sum of normal sub-kernels (red in Fig. 2), the flat top with normal tails (blue in Fig. 2), and the trapezoid shape used in our main analysis (black in Fig. 2). Our main analysis scales the kernels using the h_i to correct for missed effort when computing $W(t)$, using a 1-hour time window. Sensitivity analyses tabled here change that to a 2-hour window, or eliminate that aspect of the analysis. The asterisked row corresponds to our main analysis.

Kernel for $B(t)$	Correction for missed effort	$\hat{\alpha}^{(1)}$	$\hat{\alpha}^{(2)}$	$\lambda(0.7855)$
Sum of normals	None	1.0192	1.5062	1.1237
Sum of normals	1-hour window	1.0214	1.5245	1.1293
Sum of normals	2-hour window	1.0232	1.5533	1.1369
Flat top normal	None	1.0247	1.4535	1.1159
Flat top normal	1-hour window	1.0257	1.4649	1.1199
Flat top normal	2-hour window	1.0273	1.4896	1.1265
Trapezoid	None	1.0210	1.4707	1.1175
Trapezoid*	1-hour window	1.0230	1.4821	1.1215
Trapezoid	2-hour window	1.0247	1.5070	1.1282

nature of our analysis, and because Fig. 6 shows that the two integrated functions are extremely similar except for a few abrupt blips during periods with boat noise effects (especially in the first half of the season), so the bias correction factors would likely also be similar, thereby cancelling in the ratio.

We carried out two variance estimation approaches. The first, based on the posterior from the GAM, has several advantages. It is easy to compute, with well-tested software, and has a firm basis in statistical theory (Wood, 2004, 2011, 2017). This is the standard approach most applied statisticians would employ, except that it does not account for potential correlation between $\lambda(p)$ and $\hat{N}$. To address that, we conducted the conditional parametric bootstrap approach. The disadvantage of the particular bootstrap we employed is that it is rather ad hoc; our sampling strategy was the best that could be done in the circumstances, but less sophisticated than what we would do for a simpler problem. The variance results were very similar for the two approaches: the $\mathrm{CV}(\lambda(p))$ was 0.007 and 0.008 for the posterior and bootstrap methods, respectively. Furthermore, the correlation between $\lambda(p)$ and $\hat{N}$ was estimated to be negligible. Therefore, the absence of a correlation term in the posterior approach is unproblematic. Considering all the factors mentioned here, we believe the posterior approach is superior.

One might ask why a dependent data bootstrap approach like the block bootstrap was not applied (Givens *et al*., 2013). Among the complexities that prevent this approach is the nature of the boat excursion data. Those data, in the second half of the season, are extremely clustered. Blocks would either be empty or full of boat excursions. Moreover, it is not clear how to compute *B(t)* from bootstrap data if a boat excursion has *L* in one bootstrap block and *R* in another. There would be many such problems.

A potential concern about our approach is that the boat disturbance effect is also incorporated into the long-term passage rate, *P(t)*, rather than being isolated in *W(t)*. In principle, this is a valid concern. However, *P(t)* describes variation in whale sighted abundance at a time resolution of several days or more, due to the binning and smoothing used. In contrast, *W(t)* captures short-term variation on the order of hours. There is considerable independence between these, and the GAM estimate of *s(B(t))* controls for *P(t)*. Moreover, if part of the disturbance effect on sighted abundance is captured in *P(t)* rather than *W(t)*, the effect would likely be to underestimate the correction factors, leading to a smaller increase in corrected abundance.

Survey observers have reported for decades that the noise from boats and hunters disturbs the passage of bowheads past the perch, but this effect was not accounted for in past abundance estimates. In past surveys (1980–2011), analysts were aware that the use of powered skiffs caused reductions in whale sightings. However, the use of skiffs in those years was relatively infrequent, and while it imposed a negative bias on the abundance estimates, that bias was considered acceptable as it was limited and conservative in nature (i.e., resulted in a lower abundance estimate). Traditional indigenous knowledge also stresses the need for quiet when hunting. Despite understanding that noise disturbs the whales, we were frankly surprised that such a clear, robust signal could be obtained from our analysis of the limited available data. The finding is quite insensitive to the major methodological choices we made. We recommend that the IWC Scientific Committee adopt our corrected abundance estimate of 14,025 (CV = 0.228) to replace the original estimate of Givens *et al.* (2021). Also, if another

ice-based bowhead survey near Utqiaġvik is conducted in the future, we recommend that contemporaneous boat excursion data be collected so a similar correction can be made. It would be even better if that boat data included information on where the boats travelled during excursions, but this would present substantial practical, logistical and analytical challenges.

ACKNOWLEDGMENTS

We gratefully acknowledge the many researchers who contributed to the 2019 ice-based surveys under very difficult field conditions (see Acknowledgements of Givens *et al.*, 2021). We thank the Barrow Whaling Captains Association and the Alaska Eskimo Whaling Commission for allowing us to conduct these studies near their camps during the whaling season, which can interfere with their activities. In particular, we thank the crews of Little Whaler, Hopson 1 and Quuniq for allowing us to share their trail. The survey would have failed without their assistance. We thank Barrow Volunteer Search and Rescue for providing the boat excursion data. We thank an anonymous reviewer whose suggestions and comments substantially improved our original manuscript. Finally, we thank the North Slope Borough, the AEWC and NOAA for providing funding for the surveys. We also thank many people at the North Slope Borough for their support that allowed for the successful completion of the survey including Taqulik Hepa (Director of the NSB Department of Wildlife Management), Nicole Kanayurak (Deputy Director of the NSB Department of Wildife Management), Dolores Vinas, Christina Aiken, Martha Kaleak. Billy Adams, Bobby Sarren and others in Department of Wildlife Management. We were assisted with management of funding from the NOAA award NA16NMF4390120 to the AEWC by Sarah Ellis (NSB Grants), and Sarah Espelin (AEWC).

REFERENCES

Carroll, G. M., Smithhisler, J. R., 1980. Observations of bowhead whales during spring migration. *Mar. Fish Rev.* 42(9–10):80–85.

Druckenmiller, M. L., Eicken, H., George, J. C., Brower, L., 2013. Trails to the whale: reflections of change and choice on an Iñupiat icescape at Barrow, Alaska. *Polar Geog.* 36(1–2):5–29. [Available at: *https://doi.org/10.1080/1088937X.2012.724459*]

Fasiolo, M., Nedellec, R., Goude, Y., Wood, S. N., 2018. Scalable visualisation methods for modern generalized additive models. Arxiv preprint. [Available at: *https://doi.org/10.48550/arXiv.1707.00330*]

Givens, G. H., Edmondson, S. L., George, J. C., Tudor, B., DeLong, R. A., Suydam, R., 2015. Weighted likelihood recapture estimation of detection probabilities from an ice-based survey of bowhead whales. *Environmetrics* 26:1–16. [Available at: *https://doi.org/10.1002/env.2293*]

Givens, G. H., Edmondson, S. L., George, J. C., Suydam, R., Charif, R. A., Rahaman, A., Hawthorne, D., Tudor, B., DeLong, R. A., Clark, C. W., 2016. Horvitz-Thompson whale abundance estimation adjusting for uncertain recapture, temporal availability variation, and intermittent effort. *Environmetrics* 27:134–146. [Available at: *https://doi.org/10.1002/env.2379*]

Givens, G.H., George, J.C., Suydam, R., Tudor, B., 2021. Bering-Chukchi-Beaufort Seas bowhead whale (*Balaena mysticetus*) abundance estimate from the 2019 ice-based survey. *J. Cetacean Res. Manage.* 22 (1):61–73 [Available at: *https://doi.org/10.47536/jcrm.v22i1.230*]

Givens, G. H., Hoeting, J. A., 2013. *Computational Statistics (2nd Edition)*. John Wiley & Sons.

Koski, W. R., Rugh, D. J., Punt, A. E., Zeh, J., 2006. An approach to minimize bias in estimation of the length frequency distribution of bowhead whales (*Balaena mysticetus*) from aerial photogrammetric data. *J. Cetacean Res. Manage*. 8 (1):45–54. [Available at: *https://doi.org/10.47536/jcrm.v8i*]

Lefevre, J. S., 2013. A pioneering effort in the design of process and law supporting integrated Arctic Ocean management. *Envtl. L. Rep.* 43:10893–10908.

McDonald, T. L., Richardson, W. J., Greene Jr., C. R., Blackwell, S. B., Nations, C. S., Nielson, R. M., Streever, B., 2012. Detecting changes in the distribution of calling bowhead whales exposed to fluctuating anthropogenic sounds. *J. Cetacean Res. Manage.* 12(1):91–106. [Available at: *https://doi.org/10.47536/jcrm.v12i1.596*]

Murdoch, J., 1892. *Ethnological results of the Point Barrow Expedition*. Smithsonian Institution Press.

National Research Council 2003. *Cumulative environmental effects of oil and gas activities on Alaska's North Slope.* The National Academies Press.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Richardson, W. J., Malme, C. I., 1993. Man-made noise and behavioral responses. In: J. J. Burns, J. J. Montague, C. J., Cowles (Eds.), *The Bowhead Whale*. Allen Press.

Robertson, F. C., Koski, W. R., Thomas, T. A., Richardson, W. J., Würsig, B., Trites, A.W., 2013. Seismic operations have variable effects on dive-cycle behavior of bowhead whales in the Beaufort Sea. *Endanger. Species Res.* 21(2):143–160. [Available at: *https://doi.org/10.3354/esr00515*]

Robertson, F., Koski, W., Brandon, J., Thomas, T. A., Trites, A., 2015. Correction factors account for the availability of bowhead whales exposed to seismic operations in the Beaufort Sea. *J. Cetacean Res. Manage.* 15:33–44. [Available at: *https://doi.org/10.47536/jcrm.v15i1.513*]

Smith, R. J., 1993. Logarithmic transformation bias in allometry. *Am. J. Phys. Anthropol*. 90: 215–228. [Available at: *https://doi.org/10.1002/ajpa.1330900208*]

Ver Hoef, J. M., Boveng, P. L., 2007. Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* 88(11):2766–2772. [Available at: *https://doi.org/10.1890/07-0043.1*]

Wallig, M., Microsoft Corporation, Weston, S., Tenenbaum, D., 2020a. Package doParallel. https://cran.r-project.org/package=doParallel

Wallig, M., Microsoft Corporation, Weston, S., 2020b. Package foreach. https://cran.r-project.org/package=foreach

Wood, S. N., 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *JASA* 99:673–686. [Available at: *https://doi.org/10.1198/016214504000000980*]

Wood, S. N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Series B* 73:3–36. [Available at: *https://doi.org/10.1111/j.1467-9868.2010.00749.x*]

Wood, S. N., 2017. *Generalized Additive Models: An Introduction with R (2nd edition)*. Chapman & Hall/CRC. [Available at: *https://doi.org/10.1201/9781315370279*]

Würsig, B., Koski, W. R., 2021. Natural and potentially disturbed behavior of bowheads. In: J. C. George, J. G. M. Thewissen (Eds.), *The Bowhead Whale, Balaena mysticetus, Biology and Human Interactions* (339–363). Elsevier Press. [Available at: *https://doi.org/10.1016/B978-0-12-818969-6.00023-6*]

Zeh, J. E., Clark, C. W., George, J. C., Withrow, D., Carroll, G. M., Koski, W. R., 1993. Current population size and dynamics. In: J. J. Burns, J. J. Montague, C. J., Cowles (Eds.), *The Bowhead Whale*. Allen Press.