



Gray whale habitat use and reproductive success during seismic surveys near their feeding grounds: comparing state-dependent life history models and field data

Lisa Schwarz · Elizabeth McHuron ·
Marc Mangel · Glenn Gailey · Olga Sychenko

Received: 11 June 2021 / Accepted: 5 February 2022
© The Author(s) 2022

Abstract We used a stochastic dynamic programming (SDP) model to quantify the consequences of disturbance on pregnant western gray whales during one foraging season. The SDP model has a firm basis in bioenergetics, but detailed knowledge of minimum reproductive length of females (L_{\min}) and the relationship between length and reproductive success (R_{fit}) was lacking. We varied model assumptions to determine their effects on predictions of habitat use, proportion of animals disturbed, reproductive success, and the effects of disturbance. Smaller L_{\min} values led to higher predicted nearshore habitat use. Changes in L_{\min} and R_{fit} had little effect on predictions of the

effect of disturbance. Reproductive success increased with increased L_{\min} and with higher probability of reproductive success by length. Multiple seismic surveys were conducted in 2015 off the northeast coast of Sakhalin Island, with concomitant benthic prey surveys, photo-identification studies, and whale distribution sampling, thus providing a unique opportunity to compare output from SDP models with empirical observations. SDP model predictions of reproductive success and habitat use were similar with and without acoustic disturbance, and SDP predictions of reproductive success and large-scale habitat use were generally similar to values and trends in the data. However, empirical estimates of the proportion of pregnant females nearshore were much higher than SDP model predictions (a large effect, measured by Cohen's d) during the first week, and the SDP model overestimated whale density in the south and underestimated density around the mouth of Piltun Bay. Such

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1007/s10661-022-10024-9>.

This article is part of the Topical Collection on *Western Gray Whales and Industry Seismic Operations*

L. Schwarz (✉) · E. McHuron · M. Mangel
Institute of Marine Sciences, University of California,
Santa Cruz, CA 95064, USA
e-mail: lisa.schwarz@ucsc.edu

E. McHuron
Department of Ecology and Evolutionary Biology,
University of California, Santa Cruz, CA 95064, USA

E. McHuron
Ocean, and Ecosystem Studies, Cooperative Institute
for Climate, University of Washington, Seattle, WA 98195,
USA

M. Mangel
Theoretical Ecology Group, Department of Biology,
University of Bergen, 9020 Bergen, Norway

M. Mangel
Puget Sound Institute, University of Washington, Tacoma,
WA 98402, USA

G. Gailey · O. Sychenko
Cetacean EcoSystem Research, Lacey, WA 98516, USA

differences in nearshore habitat use would not affect SDP predictions of reproductive success or survival under the current seismic air gun disturbance scenario.

Keywords Bayesian analysis · *Eschrichtius robustus* · Russia · Population consequences of disturbance · PCOD · Stochastic dynamic programming

Introduction

Each summer and fall, western gray whales (*Eschrichtius robustus*) forage off the NE Sakhalin coast in Russia in two defined areas, known as the offshore and nearshore feeding areas, both of which are close to long-term oil and gas developments and their associated activities. In 2015, two oil and gas companies conducted multiple seismic surveys near or overlapping those gray whale feeding areas, which could change gray whale behavior and distribution (Gailey et al., 2007, 2016; Muir et al., 2015, 2016; Yazvenko et al., 2007). Specifically, there was concern that the combined 2015 seismic surveys could lead to decreased gray whale foraging activity and reduce population growth (Aerts et al., 2022).

The population-level effects of previously documented changes in behavior and distribution remain unclear. Since they are capital breeders, gray whales would generally be less affected by lost foraging opportunities than income breeders (McHuron et al., 2017). However, several characteristics of the population make it susceptible to decline. Their two known feeding areas are small compared to that of eastern gray whales and many other baleen whales (Mate et al., 2011; Moore et al., 2003), and they exhibit a high degree of site fidelity to these areas (Fig. 1) (Bröker et al., 2020). In addition, the estimated population size is small for a K-selected species, with 175 identified non-calf individuals, of which 33 are known reproducing females (Cooke et al., 2017). Furthermore, environmental factors such as ice conditions can limit the duration of their foraging season, reduce energy intake, and lower calf survival (Bradford et al., 2012; Gailey et al., 2020; Perryman et al., 2020).

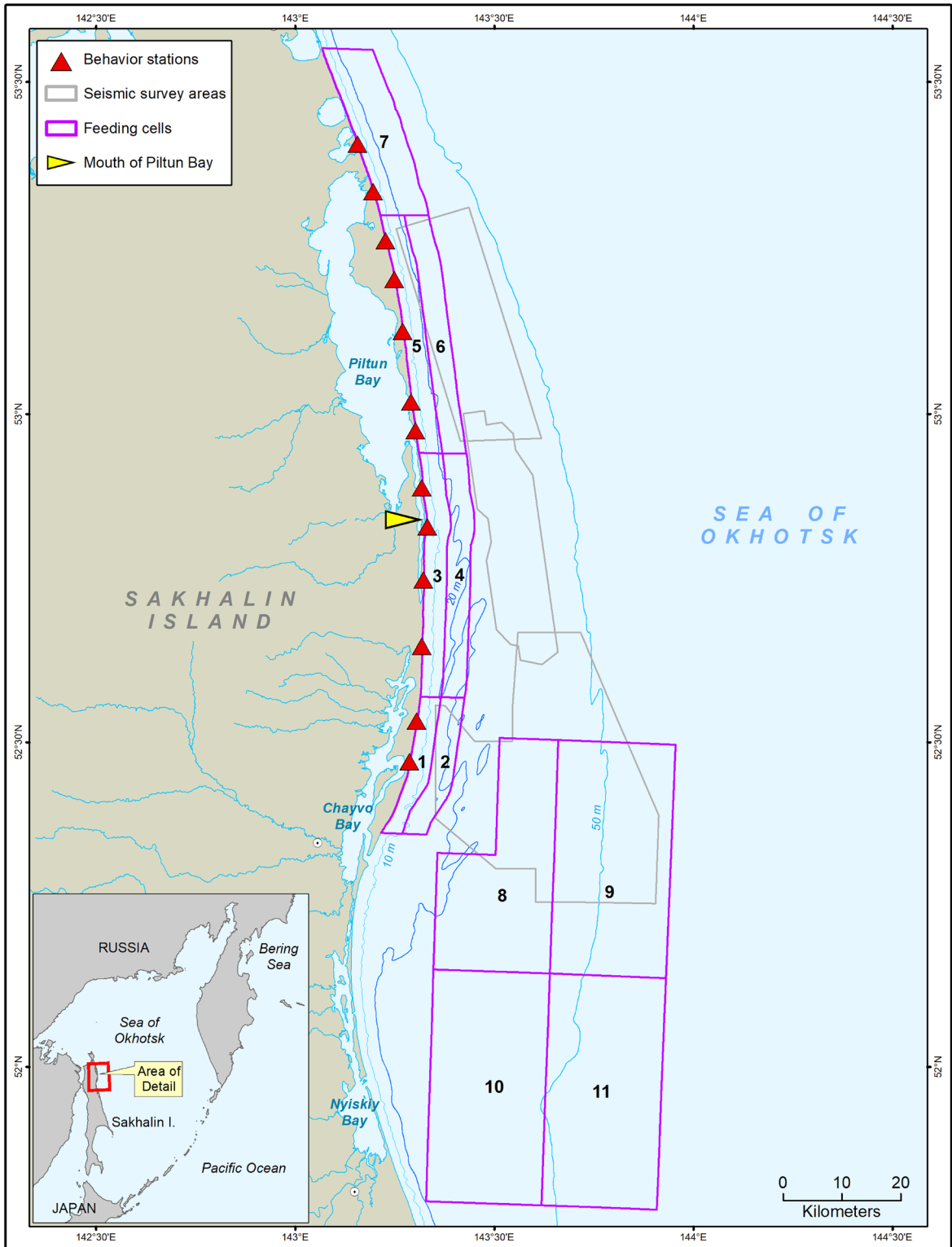
The population consequences of disturbance (PCOD) framework offers a conceptual model linking environmental impacts, changes in behavior, physiological health, and population growth through a series of

Fig. 1 Map showing the defined feeding cells (numbered pink polygons) and the 2015 seismic survey areas (gray polygons). Distribution data were collected at the 13 shore-based behavior stations (red triangles), and nearshore gray whale density was highest around the mouth of Piltun Bay (tip of yellow triangle)

transfer functions (New et al., 2014). Stochastic dynamic programming (SDP) modeling provides a way to simulate natural situations of disturbance in the presence of environmental variation while allowing individuals to vary in their behavioral responses to maximize reproductive fitness, ultimately linking behavior to individual health and health to demographic rates (Clark & Mangel, 2000; Houston et al., 2006; Mangel & Clark, 1988; Mangel & Ludwig, 1992). Although the application of SDP models to determine the effects of reduced marine mammal foraging is relatively new, the modeling approach has already provided novel insights on whale populations (Costa et al., 2016; McHuron et al., 2017; Pirota et al., 2019).

McHuron et al. (2021), using hypothetical acoustic disturbance data, developed the first SDP model for pregnant western gray whales. They showed that disturbances that excluded or reduced foraging in the nearshore area, where prey energy density is relatively low, did not affect predictions of female survival or reproductive success (carrying a fetus to term and weaning the calf back on the foraging grounds). Regardless of disturbance scenarios, access to the high-energy, amphipod-dense offshore feeding area was particularly important to successful reproduction (McHuron et al., 2021). However, the effects of important underlying model assumptions about minimum maternal length and the reproductive fitness functions were not investigated.

In 2015, benthic prey surveys, photo-identification studies, and distribution investigations occurred in conjunction with seismic surveys and intensive acoustic monitoring in and near the gray whale feeding areas, providing a unique opportunity to further develop, apply, and test the SDP model of McHuron et al. (2021). Specifically, the objectives of the work described here were to (1) perform sensitivity analyses on the western gray whale SDP model of McHuron et al. (2021) to determine how changes in minimum reproductive female length and the reproductive fitness function affect predictions of habitat use, probability of disturbance, and reproductive success; (2) compare disturbance-free SDP model predictions of habitat use and reproductive success with



predictions from models that include measured acoustic exposure during the 2015 seismic surveys; and (3) compare SDP model predictions that account for actual acoustic disturbance and the observed gray whale distribution and photo-identification data collected in 2015.

Methods

Study area

Following McHuron et al. (2021), we divided the feeding area into 11 cells (seven nearshore and four offshore) based on the 2015 seismic survey locations, gray whale density, and the locations of long-term benthic sampling stations (Fig. 1). Overall, prey energy density is much higher and a higher proportion of energy is provided by amphipods in the offshore feeding area (Maresh et al., 2022). Young of the year only utilize the nearshore area, and as they age, gray whales tend to increase their use of the offshore area (Schwarz et al., 2022).

SDP model

The SDP modeling development involves two parts, backward iteration and forward simulation, as described in detail in McHuron et al. (2021). Backward iteration determines the behavioral choices that maximize reproductive fitness using prey availability and energy requirements for survival and reproduction. Forward simulation predicts how a population will respond based on individuals' behavioral choices, prey availability, and disturbance.

Because they are capital breeders, pregnant gray whales on the foraging grounds have to procure sufficient energy to meet metabolic and reproductive needs for the foraging period, subsequent migrations, and time at the breeding grounds. The bulk of reproductive costs (most of gestation and almost all of lactation) are incurred during migrations and in the breeding area, when females are presumed to fast or have minimal foraging success (Nerini, 1984; Sanchez-Pacheco et al., 2001). McHuron et al. (2021) relied on quantification of metabolic rates of gray whales, cost of pregnancy, and cost of lactation from two previous bioenergetics modeling efforts (Villegas-Amtmann

et al., 2015, 2017). During foraging, energy allocation was prioritized in the following order: current metabolic needs, current fetal growth, future metabolic needs, and future gestation and lactation. If a foraging female could not meet her current energetic needs, fat reserves were used to maintain metabolism as well as allow for fetal growth at a reduced rate. Otherwise, surplus energy was stored as fat for future needs. The survival probabilities of the female and fetus during the non-foraging period were a function of the size of the fetus and the female's total stored fat energy when the female left the foraging ground. Smaller females could store less fat for use during the fasting period.

Given the fetus length and the female's fat energy stores at the end of the foraging season, fitness values of all potential behaviors at previous time steps were calculated via backwards iteration (Clark & Mangel, 2000; Mangel & Clark, 1988), thereby determining the behavior that optimized reproduction. At each daily time step, females could (1) leave the foraging area completely, (2) continue to feed at the current location, (3) travel within the current cell, or (4) travel to a different cell. In the modeling work described here, we used backward iteration to determine the probability of a whale selecting one of the four behavioral options given her fat mass, the length of the fetus, location, and food availability at each time step.

We calculated available benthic prey energy for each of the 11 cells (Fig. 1) for three periods (early, mid, and late season) using biomass values and energy density of six benthic prey groups collected in 2015 (Maresh et al., 2022). For use in the SDP model, prey energy distribution probabilities were modeled as lognormal distributions for each cell and time period and converted to equally spaced, discrete categories (McHuron et al., 2021).

In the forward Monte Carlo simulation, individual females behaved in one of the four ways at each 6-h time step. The behavioral state was determined using the probability of each behavior state conditioned on the female's fat mass, the length of her fetus, location, and food availability. Behavioral state probabilities were determined from the backward iterations. We drew both population and individual parameters at the start of the feeding period from appropriate distributions (described below). The method allows us to capture uncertainty and natural variability in

parameter estimates, including population size, body length of each female, starting fat mass, and starting fetal length. We also varied travel (speed, time, and linearity) and foraging parameters (length of dive cycle and percentage of time spent diving) for each individual and 6-h time step by drawing from appropriate distributions.

We assumed females died if their fat mass fell below 5% of length-specific total body mass based on studies from many mammalian species and cited in McHuron et al. (2021). Survival of a female to the following foraging season was a function of her length and fat reserves when she left the foraging grounds. Alternatively, a female could die at any time from random chance based on the probability of mortality for non-calves estimated in mark-recapture studies (Cooke et al., 2017). Because the SDP model of McHuron et al. (2021) considered only a single breeding event, we did not include the option of abortion during a foraging season, which would occur in a multi-year SDP model (e.g., McHuron et al., 2018). We follow the terminology of Cooke (2010) and define reproduction (or reproductive success) as successfully carrying a fetus to term and weaning the calf upon return to the foraging grounds.

We slightly modified the forward simulation by changing the arrival date of individuals to reflect a later arrival at the foraging grounds distributed over several days (normal distribution with a mean date of June 15 ± 5 days s.d.). The modification does not change overall survival or reproductive rates because females in the model compensate for lost foraging at the beginning of the season (McHuron et al., 2021), but the more dispersed, later arrival date better matches field observations. The model structure allowed females to arrive in the feeding area as early as mid-May and forage for up to 25 weeks.

For each population replicate, we calculated proportions of pregnant females in the nearshore feeding area (Cells 1–7), the offshore feeding area (Cells 8–11), or outside both areas (not arrived yet or left the area) over the foraging season for each 6-h bin. To determine the mean and variance of those proportions over time, 6-h bins of the calculated proportions were pooled by the three areas, population replicate, and week of year. Weekly time bins were used to reduce the effects of individual movements that created a high level of stochasticity at smaller time scales.

Objective 1: Application of SDP model with different reproductive female lengths and fitness functions

The probability of successful reproduction (reproductive fitness function; R_{fit}) is dependent on time of departure from the foraging grounds, maternal mass at departure, maternal length, and fetal length at departure. Because R_{fit} is based on functions with a high level of uncertainty, we tested three different fitness functions to determine how the shape of the function affected overall reproductive success and female behavioral choices, as indicated by differences in foraging location over time. We used (1) R_{fit} of McHuron et al. (2021), (2) R_{fit} with a higher probability of calf survival by maternal mass and a small change in the shape of the function defining calf survival vs. fetus length, and (3) R_{fit} with a higher probability of calf survival by maternal mass and a large change in the shape of the function defining calf survival vs. fetus length (Supp. Mat.: Variations in R_{fit} ; Table S1; Figs. S1 and S2). We refer to the fitness functions as low, medium, and high probability of reproductive success (Fig. 2).

McHuron et al. (2021) sampled female body length from a normal distribution (12.7 ± 0.6 m) with a range limit of 11–14 m. That distribution was based on lengths of southbound pregnant females from whaling data (Rice & Wolman, 1971; Villegas-Amtmann et al., 2015). However, their SDP model results indicated smaller females are unlikely to successfully reproduce. In addition, the length distribution of reproductive females in the field is most likely shifted higher compared to the distribution tested by McHuron et al. (2021) because photo-identification studies can only identify a reproductive female once she has successfully returned to the foraging grounds with a calf, perhaps not when she was first pregnant. Thus, for comparison with field data, three additional distributions of reproductive female length were tested in this paper. These length distributions had the same normal mean and standard deviation as in McHuron et al. (2021) but limited the minimum reproductive lengths (L_{min}) to 12.1, 12.7, or 13.0 m instead of 11.0 m. The lower limits were determined by estimating the length at which 50% of the females had successfully reproduced, based on recent growth rate information of gray whale females (Agbayani et al., 2020) and the three R_{fit} (Supp. Mat.: Determination of minimum limit on reproductive female length; Fig. S3).

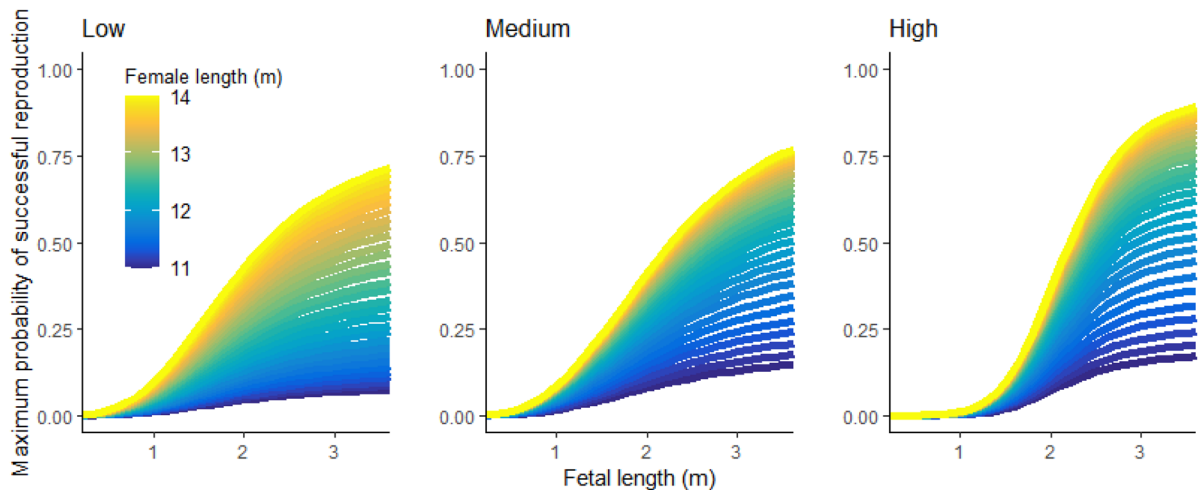


Fig. 2 Maximum probability of successful reproduction as a function of female length and fetal length for the three tested fitness functions. Values indicate the fitness function for maximum possible maternal fat mass on the last possible day of foraging

To provide a distribution of resulting reproductive rates from forward simulations, 100 populations were simulated for each of the twelve model permutations, that is, for three fitness (R_{fit}) functions with four minimum reproductive lengths (L_{min}) without disturbance.

The proportion of females nearshore and offshore was very similar for many of the twelve models. Therefore, to compare SDP model predictions of habitat use at a finer scale of SDP cell, we focused on a representative subset of three models: two models illustrating the highest and lowest nearshore proportions and one model characterizing a roughly average nearshore proportion. On a weekly scale, we used Cohen's d (Cohen, 1977; White et al., 2014) to compare results between SDP models. Cohen's d is calculated as the difference in mean predicted values divided by the pooled standard deviation. The magnitude (effect size) of Cohen's d values is defined as little-to-no effect ($|d| < 0.2$), small effect ($0.2 \leq |d| < 0.5$), intermediate effect ($0.5 \leq |d| < 0.8$), and large effect ($|d| \geq 0.8$).

Objective 2: Comparison of modeled habitat use and reproductive success without disturbance and with measured acoustic disturbance in 2015

Sound levels within and outside both foraging areas during the 2015 seismic surveys were recorded by 40 Autonomous Underwater Acoustic Recorders (AUARs) at 48 locations during the season (Rutenko

et al., 2022). Not all locations had recorders for the entire season, with some recorders moved to different locations relative to the seismic activity at the time. The measured air gun sound levels from the AUARs were used to calibrate acoustic propagation models (Rutenko et al., 2022). For the nearshore area, the maximum sound pressure level was estimated in 1-km² blocks over 5 min periods. A block was considered disturbed if the maximum SPL exceeded 163 dB re 1 μPa^2 mean square sound pressure level (SPL). For the offshore area, the estimation of acoustic metrics was performed on a broader spatial scale of 10 \times 10 km (or 100 km²) blocks because model calibration was based on records from only one AUAR. Although vessel sounds are known to cause gray whale behavioral and distribution changes (Gailey et al., 2022a, b), for simplicity they were not included in model iterations. Onshore pile driving also occurred in 2015, but it produced localized, low sound levels; like vessel sounds, pile driving sounds were not included in model iterations.

Forward simulations that include disturbance require quantification of the probability of exposure to a disturbance, the probability animals will respond to the disturbance, and the behavioral change that will occur if they do respond. For all twelve SDP models, air gun sound levels recorded during the 2015 seismic surveys were used to quantify disturbance. We computed the probability of exposure to air gun sounds in each SDP cell as the proportion of acoustic blocks

above the disturbance threshold in a given SDP cell during each 6-h bin.

We set the threshold for behavioral change at 163 dB re $1\mu\text{Pa}^2$ SPL, which has been used for mitigation purposes when minimizing disturbance to gray whales from air gun sounds in the nearshore feeding area (Aerts et al., 2022). The 163 dB re $1\mu\text{Pa}^2$ SPL threshold was based on a study in the Bering Sea, which estimated that 10% of gray whales stopped feeding and moved away from the area when exposed to received air gun sound levels of 163 dB re $1\mu\text{Pa}^2$ SPL (Malme et al., 1988). We conservatively assumed any whale exposed to ≥ 163 dB re $1\mu\text{Pa}^2$ SPL during a 6-h period responded by moving to the next closest undisturbed cell and lost the opportunity to forage for 6 h.

To compare disturbed versus undisturbed SDP model estimates of reproduction, the proportion of pregnant females that successfully reproduced (gave birth and returned to the foraging area the following spring with a calf) was pooled by population replicate for each of the twelve SDP model combinations of the three fitness functions and the four minimum reproductive lengths.

We used Cohen's d as described above to compare disturbed versus undisturbed SDP habitat use predictions (proportion of pregnant females in a cell) for the three representative subsets of SDP models over the field season. We also calculated the predicted proportion of animals disturbed in each 6-h bin using all three SDP models.

Objective 3: Comparison of model results with observed gray whale distribution and photo-identification data from 2015

During the 2015 foraging season, the most coastal nearshore cells (Cells 1, 3, 5, and 7) were monitored for gray whales via photo-identification and distribution surveys (shore-based scan sampling) on temporal and spatial scales that can be compared with SDP model results. Photo-identification data provided information about individual use of the entire nearshore area over the season (seen versus not seen), including specific determination of nearshore use by known reproductive females. Distribution data, collected using shore-based scan sampling techniques, can determine the spatial occurrence of animals within the nearshore area.

Schwarz et al. (2022) provide details on the photo-identification data collection process off Sakhalin Island. In brief, unique permanent body pigmentation patterns along with additional scarring and barnacle patches were used to identify individual gray whales. Data were collected using a combination of up to five stationary and two mobile shore-based teams as well as up to two inflatable boats. Although effort was high, environmental conditions prevented consistency in photo-identification data collection both temporally and spatially.

Distribution data were collected at hourly intervals to quantify gray whale occurrence patterns within the nearshore area throughout the season, which were compared with SDP model density predictions. Distribution data were collected at 13 shore-based stations by 13 observation teams using scan sampling techniques. Shore-based stations were strategically placed roughly 10 km apart that covered around 122 km of coastal habitat. A synchronized scan was conducted hourly (scan survey), weather permitting, at each station by two observers, equipped with 7×50 Fujinon FMTRC-SX reticle binoculars, that continuously scanned from the northern part of their observation area to the southern extent at a constant rate of $9.33^\circ/\text{min}$. For each gray whale(s) sighted during the scan, the magnetic bearing, reticle, and number of individuals were recorded. Environmental conditions (visibility, Beaufort sea state, swell height, etc.) were recorded prior to conducting a scan. All sighting data were recorded using the Pythagoras software system that calculates the geographic position of the animal using Lerczak and Hobbs' distance approximation equation combined with Leaper and Gordon's refraction correction (Gailey & Ortega-Ortiz, 2002; Leaper & Gordon, 2001; Lerczak & Hobbs, 1998). Further methodological details can be found in Gailey et al. (2022b).

We compared SDP model predictions with field estimates of (1) the proportion of pregnant females identified nearshore across the foraging season, (2) pregnant female whale density in the seven nearshore cells over the foraging season, and (3) reproductive rates. Sightings of all possibly pregnant females in 2015 were used in our analysis. The group consisted of known pregnant females that were seen with calves in 2016 and potentially pregnant females that were seen with a calf prior to 2015. Since females

known to have successfully reproduced (seen with a calf in 2016) may utilize the habitat differently than other reproductive females, the “known pregnant” or “successfully pregnant” group was also analyzed separately. “All potentially pregnant” females included all successfully pregnant and possibly pregnant females. Note that we analyzed habitat use for a subset of females that were seen in 2015 and 2016, since not all mothers were identified. Also, nearshore counts of pregnant females may be underestimated as photo-identification efforts were largely from shore, and within the nearshore area, older whales tend to forage farther from the coast where identification is not possible (Sychenko, 2011).

To determine the proportion of reproductive females seen in the nearshore area over time (p_{near}), the number of identified known pregnant or all potentially pregnant females identified nearshore was counted on a daily scale and divided by the total number in their respective group ($N=8$ and 18 , respectively). Zero-inflated beta regression was used to estimate p_{near} over time to allow for a daily proportion of zero (no whales of interest) nearshore for a given day. The regression estimates the probability of $p_{near} > 0$ (α_B) using the total number of days when photo-identification surveys occurred (S) and the number of days when at least one pregnant whale was seen ($S_{y \neq 0}$). Proportions greater than zero were then estimated from a beta distribution, leading to an overall likelihood of

$$bi_0(p_{near} | \alpha_B, \mu_B, \varphi) = \begin{cases} 1 - \alpha_B & \text{if } p_{near} = 0 \\ \alpha_B f(p_{near} | \mu_B, \varphi) & \text{if } 0 < p_{near} < 1 \end{cases}, \text{ where}$$

$$f(p_{near} | \mu_B, \varphi) \sim \text{Beta}(\mu_B, \varphi) \text{ and } S_{y \neq 0} \sim \text{Bin}(S, \alpha_B)$$

Both α_B and the mean of the beta distribution (μ_B) were independently estimated as a function of visibility, Beaufort sea state, week of year (week), and week of year squared (week²). The beta distribution precision parameter (φ) was estimated as a constant. Daily visibility and Beaufort sea states were calculated as the mean of daily station mean values from the 13 shore-based stations, where environmental data were collected during hourly scan samples (Gailey et al., 2022b). Regression results were compared with

SDP model predictions of p_{near} pooled weekly for Cells 1, 3, 5, and 7 from the disturbance scenario.

To estimate whale density within each of the seven nearshore SDP cells, we converted spatially detailed density estimates from scan surveys (D). Methodology for determining detailed density estimates can be found in Gailey et al. (2022b) and Muir et al. (2015). In brief, the nearshore area was divided into a grid of 1-km² blocks. For a given survey, gray whale sightings were assigned to a respective block and converted to density, accounting for the proportion of the block area covered and whale availability (not all whales are above water during a scan). Previous analyses suggest a flat detection probability curve equal to 1.0 out to the maximum distance analyzed (which is unique for each scan station height at 0.1 binocular reticles), so detection probability was set at 1.0 (Muir et al., 2016). Sightings from only one station were used to calculate density when scan stations had overlapping visible area.

To calculate density within the seven large nearshore SDP cells, each 1-km² block was assigned a cell based on its location. Blocks that overlapped cell boundaries were split into two blocks with smaller areas. Survey (i), block (j), and cell (k) specific density estimates ($D_{i,j,k}$) were reconverted to count estimates $N_{i,j,k}$ by multiplying by a block's area (A_j). Counts within a cell were summed over all blocks in that cell and divided by the total block area in that cell ($A_{i,k}$) covered during the survey to determine gray whale density in a cell for a given survey ($D_{i,j,k}$)

$$D_{i,k} = \frac{\sum_{j=1}^{J_{i,k}} N_{i,j,k}}{A_{i,k}}, \text{ where}$$

$$N_{i,j,k} = D_{i,j,k} \cdot A_{i,j,k}$$

$$A_{i,k} = \sum_{j=1}^{J_{i,k}} A_{i,j,k} \text{ and}$$

$$J_{i,k} = \text{total number of blocks in survey } i \text{ for cell } k.$$

We used a hurdle gamma regression model to analyze density as a function of time to allow for a density of zero (no whales). The hurdle gamma regression estimates the probability of the density being greater than zero (α_G) using the total number of surveys (R) and the number of surveys when at least one whale was seen ($R_{\hat{D} \neq 0}$).

Densities greater than zero were then estimated from a gamma distribution, leading to an overall likelihood of

$$Gamma_0(D|\alpha_G, \mu_G, \omega) = \begin{cases} 1 - \alpha_G & \text{if } D = 0 \\ \alpha_G f(D|\mu_G, \omega) & \text{if } D > 0 \end{cases}, \text{ where}$$

$$f(D|\mu_G, \omega) \sim Gamma(\mu_G, \omega) \text{ and } R_{D \neq 0} \sim Bin(R, \alpha_G)$$

Both α_G and the mean of the gamma distribution (μ_G) were independently estimated as a function of week, week², and week³ separately by cell. Not all blocks were observed during every survey, so the proportion of each cell covered in a survey was also included as a predictor variable in the models. The gamma distribution shape parameter (ω) was assumed constant.

The above density estimates include all nearshore gray whales, while the SDP predicts density for pregnant whales only. To convert regression estimates to pregnant female density, we used photo-identification data to calculate the proportion of pregnant females (D_{preg}) compared to the total number of identified whales in the coastal nearshore cells (p_{preg}) on a daily scale for both known pregnant females and all potentially pregnant females. Logistic regression was used to estimate p_{preg} as a function of week, week², daily Beaufort sea state, and visibility. Weekly estimates of D were multiplied by estimates of p_{preg} and compared with SDP predictions of D_{preg} in the nearshore (Cells 1–7) from the three SDP models with disturbance. SDP predictions of D_{preg} were calculated as the predicted proportion of females in a cell, multiplied by the number of known pregnant females (or all potentially pregnant females), and divided by total cell area.

The Bayesian R package “brms” was used to estimate posterior distributions of regression parameters and to determine the best-fitting models (Bürkner, 2017, 2018; Plummer, 2019). Priors on parameters were uniform with boundaries wide enough that posterior distributions of the parameters were not truncated at any prior boundaries. Standard practices (multiple independent chains with low lag-1 autocorrelation) ensured mixing, convergence, and stationarity in posterior samples using R packages “MCMCvis,” “coda,” and “bayesplot” (Gabry & Mahr, 2021; Gabry et al., 2019; Plummer et al., 2006; Youngflesh, 2018). Bayesian model stacking using the leave-one-out criteria was used to average over multiple models

when no one individual model was weighted > 0.9 (Yao et al., 2017). When Beaufort sea state or visibility was included as a predictor variable, the dependent variable was estimated using posterior samples of parameters and ideal environmental conditions (Beaufort = 0 and visibility = 1). In posterior estimates of density, the proportion of each cell covered in a survey was held constant at the seasonal mean value.

The reproductive success predictions from the twelve SDP model scenarios (with disturbance) were compared with the number of calves identified in 2016 taking into account the number of potentially pregnant females in 2015, defined as known reproductive females that were not observed with a calf in 2015. The reproductive success estimate assumes the inter-birth interval is at least 2 years (Rice & Wolman, 1971; Weller et al., 2009), and all known reproductive females not identified with a calf in 2015 were pregnant. Not all mothers were identified in 2015 and 2016. Nine out of the 11 calves (82%) in 2015 were identified with their mothers, while 8 out of 14 calves (57%) were identified with their mothers in 2016.

As described above, we used Cohen’s d (Cohen, 1977; White et al., 2014) to compare SDP model results with and without disturbance. We also used Cohen’s d to compare SDP model predictions with estimates derived from field data. The proportion of pregnant females nearshore over the season was compared between SDP predictions and results from regression analysis of photo-identification data on pregnant and potentially pregnant females. Within the nearshore habitat, whale density by cell was compared between SDP predictions and results from regression analysis of empirical whale densities, scaled to pregnant female whale density.

Results

Objective 1: Application of SDP model with different reproductive female lengths and fitness functions

The SDP model results indicated a relatively equal proportion of the pregnant female population occupying the nearshore and offshore areas (p_{near} and p_{off} , respectively) at the beginning of the foraging season (Fig. 3). The models predicted that p_{near} would steadily decline throughout the season, and p_{off} would show an opposite trend. The low reproductive fitness function produced

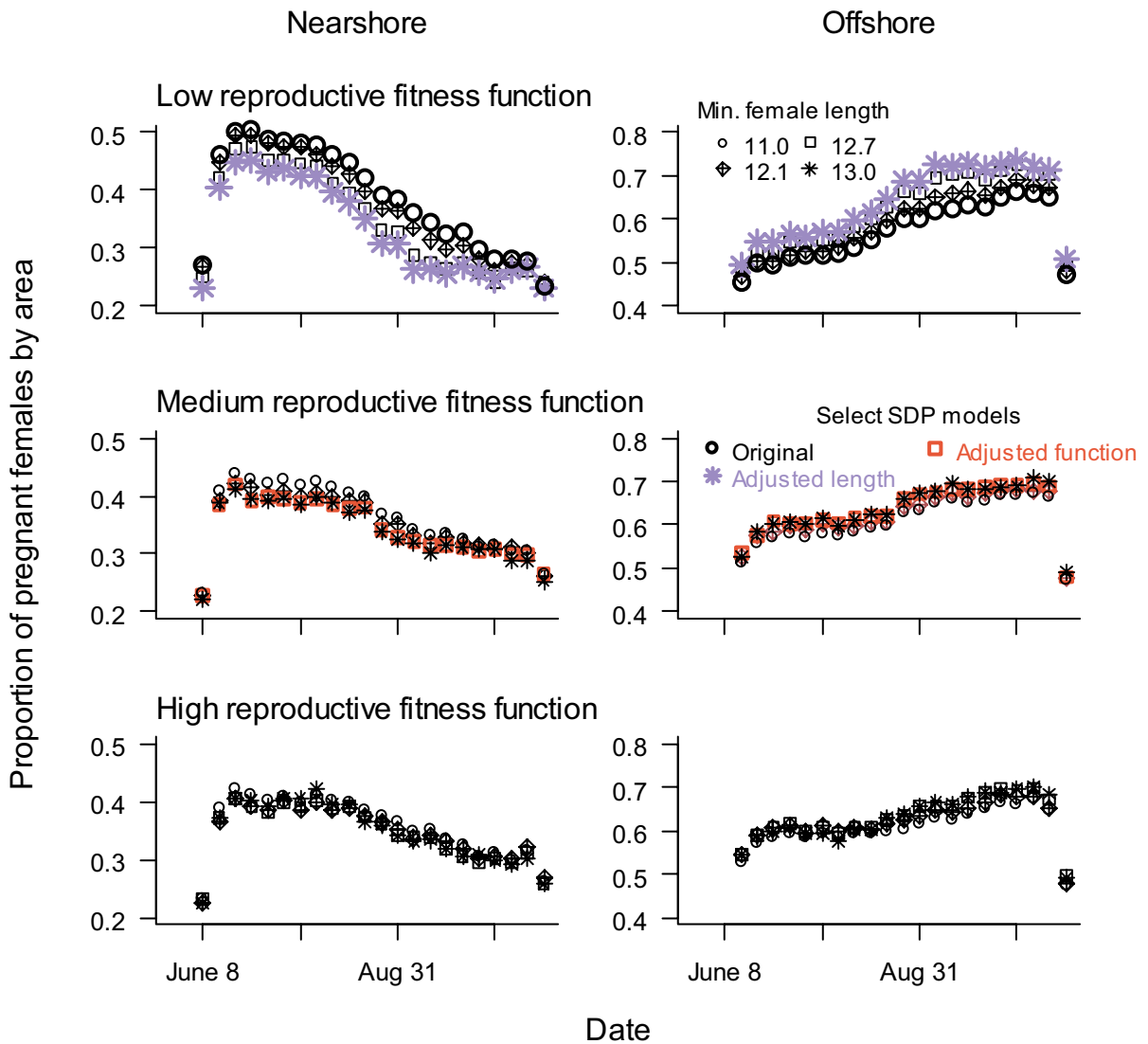


Fig. 3 SDP model results of mean weekly proportion of pregnant females nearshore (left column) and offshore (right column) without disturbance, using three different reproductive fitness functions and four minimum maternal lengths. Three combinations of reproductive fitness function and minimum maternal length were used in analyses of the consequences

of disturbance: Original=low fitness function, minimum length=11.0 (bold black circles), adjusted length=low fitness function, minimum length=13.0 (bold purple stars), and adjusted fitness=medium fitness function, minimum length=12.7 (bold orange squares)

higher p_{near} predictions earlier in the season compared to the medium and high reproductive fitness functions. Predictions of p_{off} were higher with increased minimum maternal length (L_{min}), although differences in habitat use for different L_{min} values became less pronounced with higher reproductive fitness functions. In fact, L_{min} did not affect habitat use with the high reproductive fitness function. Using the low reproductive

fitness function with the shortest L_{min} (11.0 m) led to the highest p_{near} early in the season (original SDP model). The low reproductive fitness function with the longest L_{min} (13.0 m) led to the highest p_{off} late in the season (adjusted length SDP model). Weekly habitat use patterns were approximately the same for the medium reproductive fitness function with $L_{min} > 12.1$ and for all high reproductive fitness function SDP

models. Therefore, we chose one of those models (adjusted function SDP model: medium fitness function, $L_{\min}=12.7$) as a representative model. The two low reproductive fitness function models represented the most extreme p_{near} and p_{off} values.

Comparing cell-use predictions between the three SDP models without disturbance, the original SDP model predicted a higher proportion of nearshore females, primarily in Cells 3 and 5 ($|d|$ between 0.2 and 0.5). For all three models, use in offshore Cells 8, 9, and 10 was generally constant once whales arrived but declined in the last several weeks (Fig. 4). In contrast, the proportion of the population in Cell 11 increased throughout the season. The adjusted function model had a higher proportion of females in Cells 8 and 9 at the beginning of the season and Cell 10 in the middle of the season with $|d|$ between 0.2 and 0.5 (Fig. 4). There was a moderate-to-high difference in the proportion of animals in Cell 11, with highest proportions for the adjusted length SDP model and lowest for the original SDP model (Fig. 4).

Objective 2: Comparison of modeled habitat use and reproductive success without disturbance and with acoustic disturbance data from 2015

The probability of disturbance in a 6-h period, i.e., the probability an individual could be exposed to air gun sound levels of 163 dB re $1\mu\text{Pa}^2$ or more and move away to the closest undisturbed cell if it was in that cell, was spatially and temporally variable and reflects the seismic survey activities (Fig. 5). Disturbance in the nearshore cells occurred earlier in the season, and coastal nearshore area (Cells 1, 3, and 5) overall had lower disturbance probabilities than their equivalent non-coastal area (Cells 2, 4, and 6, respectively). Offshore Cell 8 had some level of air gun sound exposure for almost the entire seismic survey period. The probability of disturbance was highest in offshore Cells 8 and 9. SDP models predicted the proportion of disturbed pregnant females increased through the beginning of the season and peaked around 0.035 throughout August then sharply declined (Fig. S4).

Using the 2015 measured acoustic values, forward simulations from the three representative models were used to predict the proportion of animals disturbed in 2015 and to compare cell use with and without disturbance. Although a higher proportion of pregnant

females were disturbed in the adjusted function model, Cohen's d values indicate no-to-little effect ($|d|<0.1$) of SDP model choice on the predicted proportion of pregnant females disturbed (Fig. S4). Cohen's d comparisons between the disturbed and undisturbed scenarios indicated there was little effect of disturbance on nearshore cell usage for all three SDP models (Cohen's d between -0.2 and 0.2). Regardless of SDP model type, disturbance had a no-to-small effect on use patterns in offshore Cells 10 and 11 and a small-to-medium effect on cell use in Cells 8 and 9, showing an increased proportion of females in cells with higher probability of disturbance (Fig. 6).

Objective 3: Comparison of model results with observed gray whale distribution and photo-identification data from 2015

Photo-identification surveys were conducted on 99 days from June 1 to October 1, 2015 (Supp Data) (Schwarz et al., 2022). Zero-inflated beta regression analysis was used to estimate the proportion of known pregnant and possibly pregnant females utilizing the coastal nearshore cells (p_{preg}) over the foraging season suggested a slight decline in the proportion with time (Tables S2 and S3, Fig. 7). Proportions increased with better environmental conditions (lower Beaufort sea state and better visibility).

The number of scan surveys conducted in each cell ranged from 962 to 1281 (Table S4). The number of surveys with whale density >0 ranged from 15 to 1019 depending on cell. Whale density was extremely low in Cell 2; of the 962 scan surveys that covered that cell, only 15 had a whale density >0 . Therefore, sample size limited the gamma regression analysis to one explanatory variable per model in Cell 2. Hurdle-gamma regression results of whale density by cell consistently indicated the probability of whale density >0 increased with the proportion of the cell covered during the scan survey (Tables S5 and S6). However, given that whales are seen, the density estimate declines with higher cell coverage (Tables S5 and S6). The proportion of the nearshore group that was known to be pregnant declined slightly over the season (Tables S7 and S8; Fig S4). The proportion of all potentially pregnant females in the nearshore group exhibited a u-shaped function over the season, although uncertainty was high (Tables S7 and S8; Fig S4).

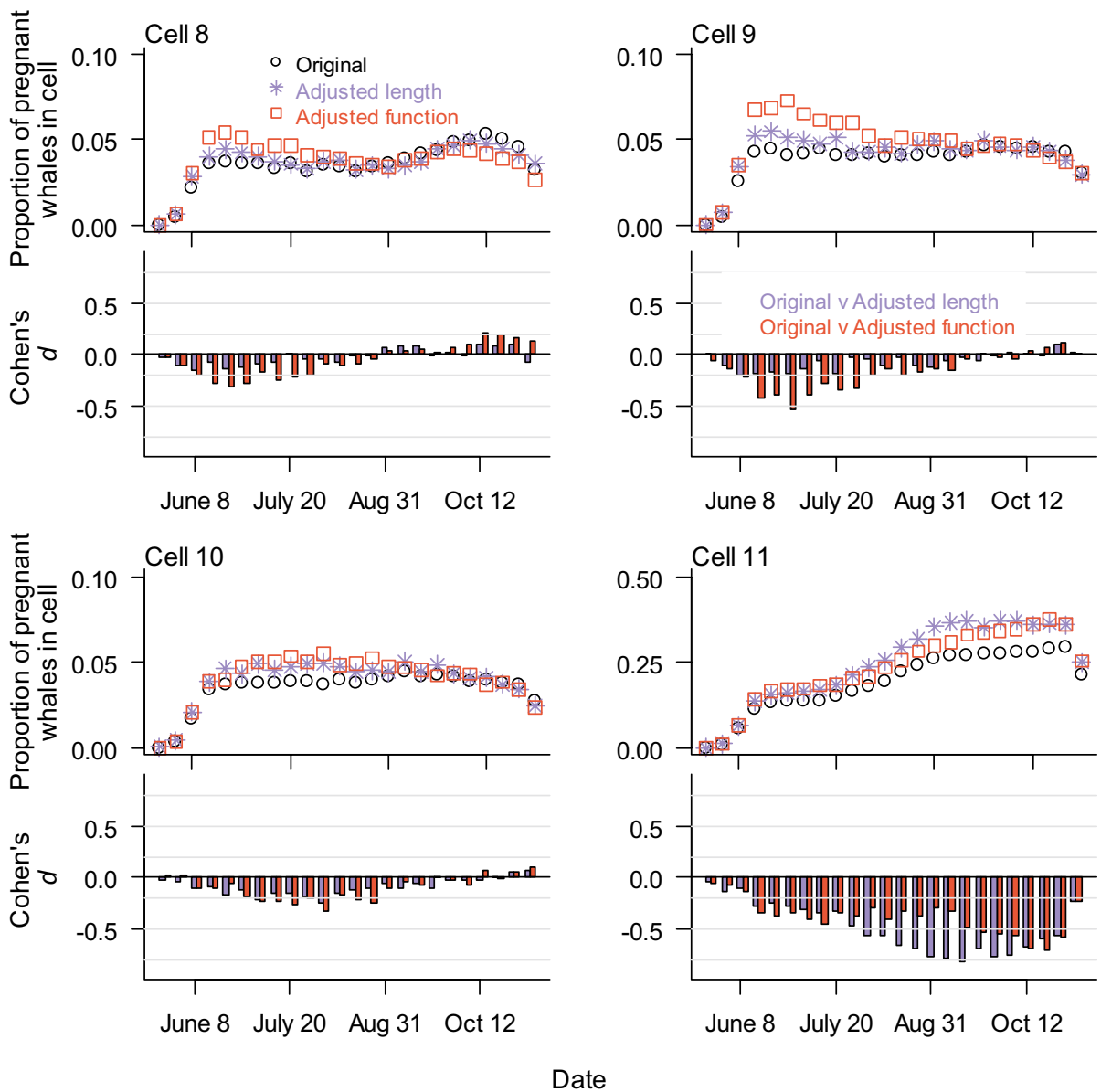


Fig. 4 SDP model results of mean weekly proportion of pregnant females in the four offshore cells without disturbance. Three combinations of reproductive fitness function and minimum maternal length were used in analyses of the consequences of dis-

turbance: Original=low fitness function, minimum length=11.0 (black circles), adjusted length=low fitness function, minimum length=13.0 (purple stars), and adjusted fitness=medium fitness function, minimum length=12.7 (orange squares)

Regression estimates indicated the highest density of successfully pregnant females was in the coastal nearshore Cells 3 and 5 (Fig. 8). Density estimates were highest in the coastal nearshore area (Cells 1, 3, and 5) and lowest in Cells 4, 6, and 7.

Fourteen new calves were identified in the Sakhalin feeding areas in 2016 (Tyurueva et al., 2018), and eleven

calves were identified in 2015 (Yakovlev et al., 2016). With a minimum inter-birth interval of 2 years, eleven reproductive females would not be able to return with a calf in 2016. Our dataset included a total of 27 reproductive females based on their previous reproductive history. An additional six reproductive females have been identified by another group (Cooke et al., 2017), meaning we

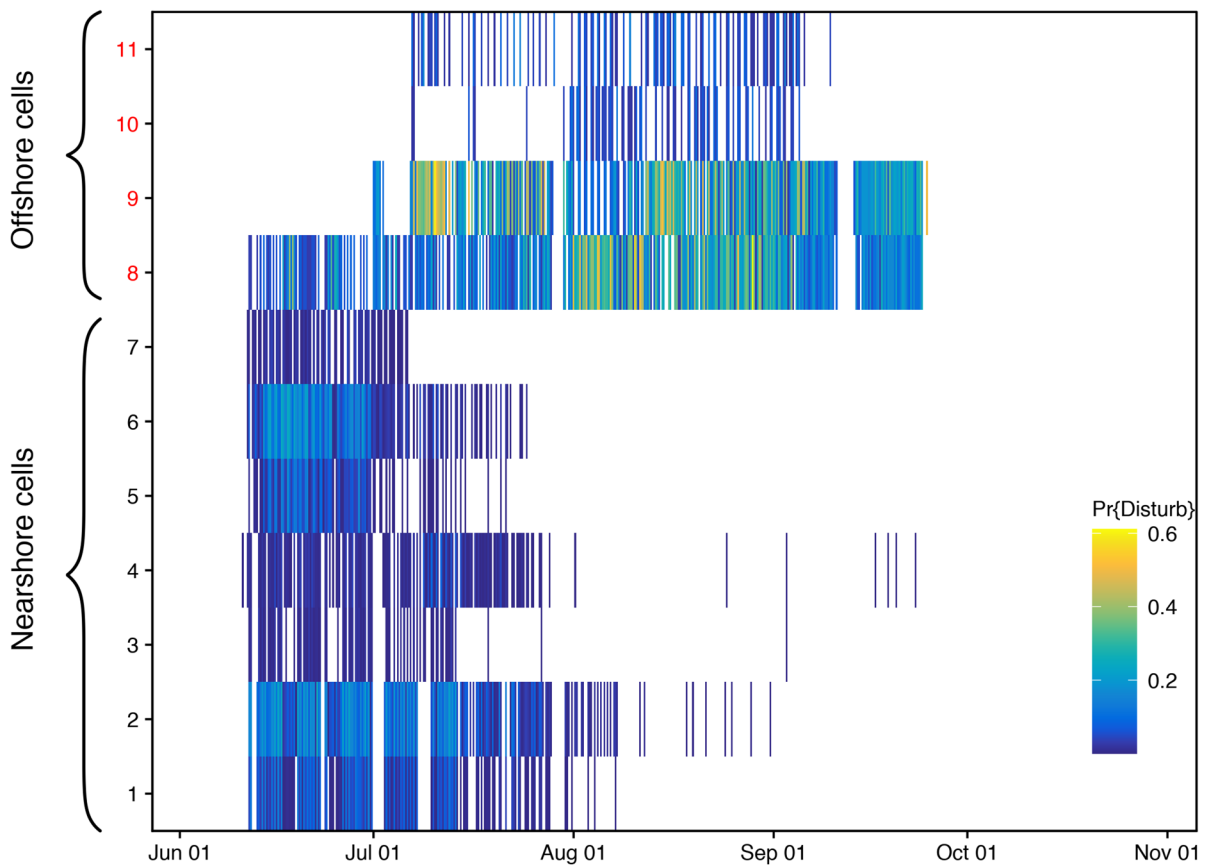


Fig. 5 Disturbance probability by cell. Probability of disturbance is defined as the proportion of acoustic data points within a cell ≥ 163 dB re $1\mu\text{Pa}^2$ SPL measured every 300 s, pooled in 6-h bins

never had confirmed sightings of those individuals with a calf. Assuming 27 reproductive females in the population, only 16 females could be pregnant in 2015 since the other 11 produced a calf in 2015, leading to a measured successful reproductive rate of 0.88. If we include the additional six females in the total, the measured successful reproductive rate is 14/22 or 0.64 (Fig. 9).

Given the differences in predicted cell use, the three representative SDP models were used to compare SDP predictions with empirical estimates of p_{near} , nearshore density, and reproduction. SDP predictions of p_{preg} in the coastal nearshore area (Cells 1, 3, 5, and 7) increased dramatically from June 1 through the first 4 weeks of the season followed by a gradual decline throughout the rest of the season (Fig. 7). Compared to other SDP models and empirically derived estimates, the original SDP model predicted the highest p_{preg} (Fig. 7). Empirical estimates under ideal conditions (Beaufort=0 and visibility=1) were higher than SDP model predictions at the

beginning and end of the season and lower during the middle of the season (Fig. 7). Cohen’s d values indicate empirical estimates were much higher than SDP model predictions (large effect) during the first week.

All three SDP models predicted a higher whale density in Cells 1, 2, and 4 than the regression estimates, with large Cohen’s d effects. The SDP models predicted lower whale densities in Cell 3 compared to regression results, with intermediate-to-high Cohen’s d values. Once the females arrived, all three SDP models predicted a decline in the nearshore density of pregnant females as the foraging season progressed, a similar trend generally mirrored in the regression results. Regression estimates indicated the peak whale density and the rate of the subsequent decline were higher in Cells 5 and 7 than what was predicted from the SDP model. In particular, Cell 5 Cohen’s d values indicated an intermediate effect in the first part of the season (Fig. 8). Trends in density estimates of all possible pregnant females were

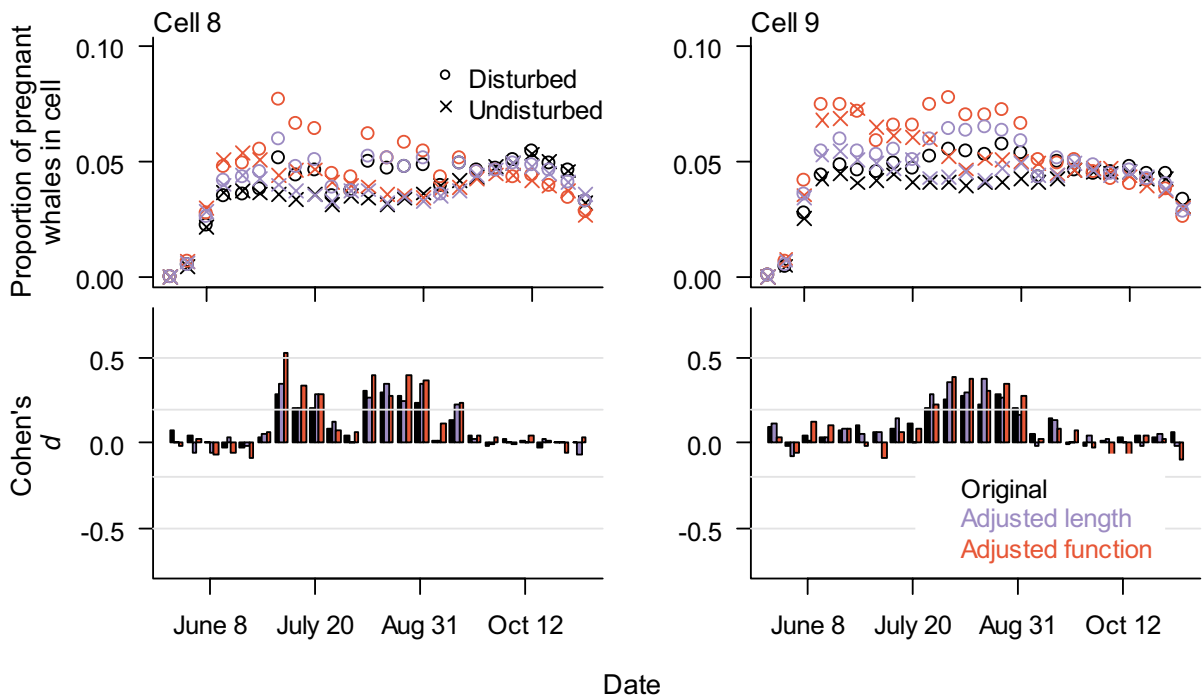


Fig. 6 Model results of proportion of pregnant females in offshore Cells 8 and 9, binning 6-h periods into weeks, with and without disturbance (open circles and crosses, respectively). Cohen's d values compare disturbed and undisturbed scenarios for three SDP models. Cohen's $d > 0$ indicates higher

proportions in a cell under disturbed conditions. Gray lines in Cohen's d graphs indicate thresholds between no-to-little effect ($|d| < 0.2$), small ($0.2 \leq |d| < 0.5$), intermediate ($0.5 \leq |d| < 0.8$), and large effect ($|d| \geq 0.8$). Disturbance had no-to-little effect in proportions of pregnant females in all other cells

similar to estimates for the smaller group of known pregnant females with the exception of Cell 3 where density of females increased at the end of the season (Fig. S5). The proportion of all possibly pregnant females within the photo-identified coastal nearshore area group of animals was a function of week² (Fig. S6), which when combined with scan regression analyses created the U-shaped density estimates in Cell 3.

SDP predictions of the probability of successful reproduction increased with increased minimum maternal length (L_{\min}) and with changes in fitness function (Fig. 9). Mean SDP predictions of reproductive rate ranged from 0.24 to 0.74, with two means of the models falling within the range of empirical values (high reproductive fitness function with $L_{\min} > 12.1$ m). Of the 12 SDP model simulations, the field reproductive rate range was included in the \pm two standard deviation span around SDP predicted means for all but three models (low reproductive fitness function with $L_{\min} < 13.0$ m). Cohen's d values indicated no-to-little effect of disturbance on reproductive rate predictions ($|d| < 0.2$).

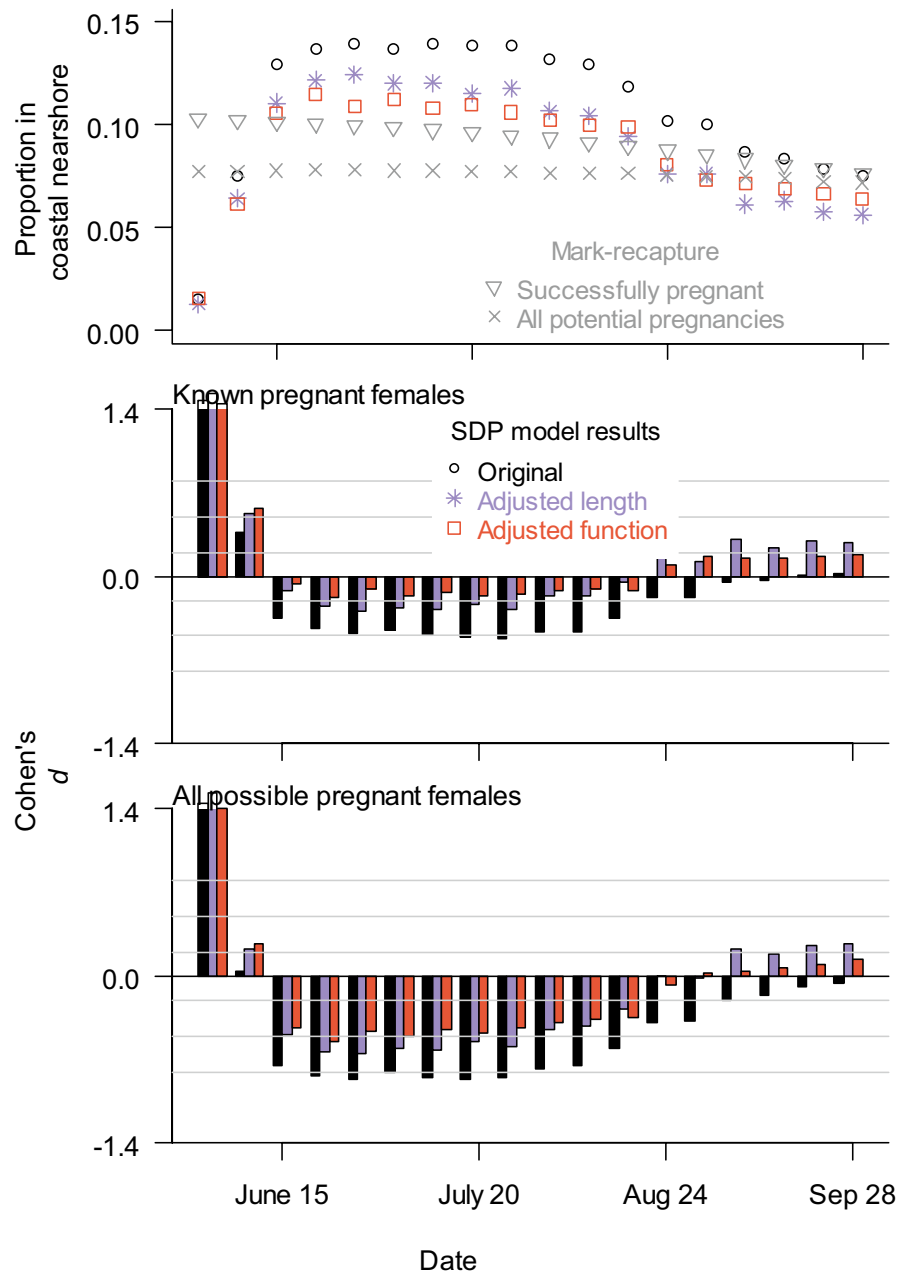
Discussion

Objective 1: Application of SDP model with different reproductive female lengths and fitness functions

Reproductive fitness functions are one of the greatest sources of uncertainty for SDP models, particularly for large whales, because there is little-to-no empirical data to determine their shape or limits. Therefore, it was important to explore various functions to determine their effect on habitat use, probability of disturbance, and probability of reproductive success. Overall, our modifications of the reproductive functions used in McHuron et al. (2021) did not result in large differences in nearshore-offshore habitat use patterns or proportion of animals disturbed. Predictions from all of the SDP models indicated a general decline over the season in nearshore habitat use that coincided with an increase in use of offshore Cell 11.

Increased use of the most prey energy-dense area (Cell 11) over the season aligned with the females'

Fig. 7 Mean proportion of pregnant females in coastal nearshore area (Cells 1, 3, 5, and 7), comparing predictions from three SDP models with regression estimates from photo-identification data (top). Cohen's *d* values comparing SDP models with the proportion of known pregnant females nearshore (middle) and the proportion of all potentially pregnant females nearshore (bottom). Cohen's *d* > 0 indicates higher estimates from empirical data, while Cohen's *d* < 0 indicates higher predictions from SDP models. Gray lines in Cohen's *d* graphs indicate thresholds between no-to-little effect ($|d| < 0.2$), small ($0.2 \leq |d| < 0.5$), intermediate ($0.5 \leq |d| < 0.8$), and large effect ($|d| \geq 0.8$)



higher energetic needs as the season progressed. When the pregnant female population consisted of smaller animals ($L_{\min} = 11.0$ m), a lower proportion of females utilized Cell 11 because their energetic needs were lower. Future analysis of offshore whale density from line-transect vessel surveys could further our understanding of those links.

Objective 2: Comparison of modeled habitat use and reproductive success without disturbance and with acoustic disturbance data from 2015

Using measured air gun sound levels as the only source of disturbance, SDP models predicted a moderate increase in use of offshore Cells 8 and 9, with

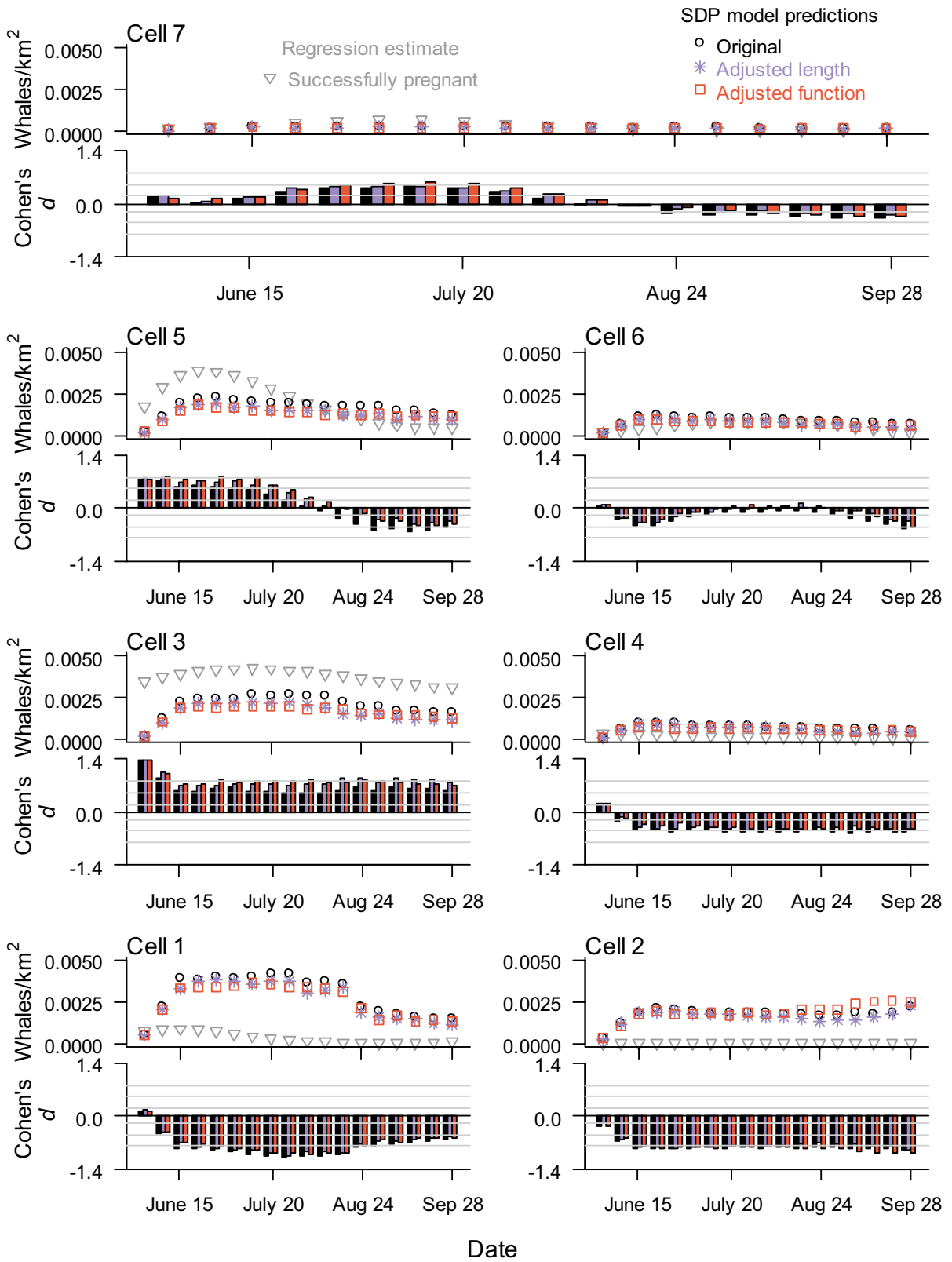


Fig. 8 Density of known pregnant females in the nearshore area (Cells 1–7) over the foraging season, comparing SDP model predictions with regression estimates from empirical data, assuming ideal sighting conditions (visibility=1, Beaufort=0). Cohen’s *d* values >0 indicate a higher regression estimate than SDP model predictions, while Cohen’s *d* values <0 indicate a lower regression estimate than SDP model predictions. Gray lines in Cohen’s *d* graphs indicate thresholds between no-to-little effect ($|d| < 0.2$), small ($0.2 \leq |d| < 0.5$), intermediate ($0.5 \leq |d| < 0.8$), and large effect ($|d| \geq 0.8$). For density of all possibly pregnant females, see Figure S3

small changes in other cells, including nearshore. In the SDP models, disturbed whales moved to the nearest undisturbed cell, and the SDP model Cells 8 and 9 were defined as each other’s closest cell. When disturbance was high in Cell 8, it was comparatively lower in Cell 9 and vice versa. The predicted pattern of higher proportions in Cells 8 and 9 during disturbance indicates movement between the cells. While the model indicated offshore females may have lost foraging time when moving out of a disturbed cell, they moved to cells with equal or more prey availability, effectively compensating for lost energy later in the season.

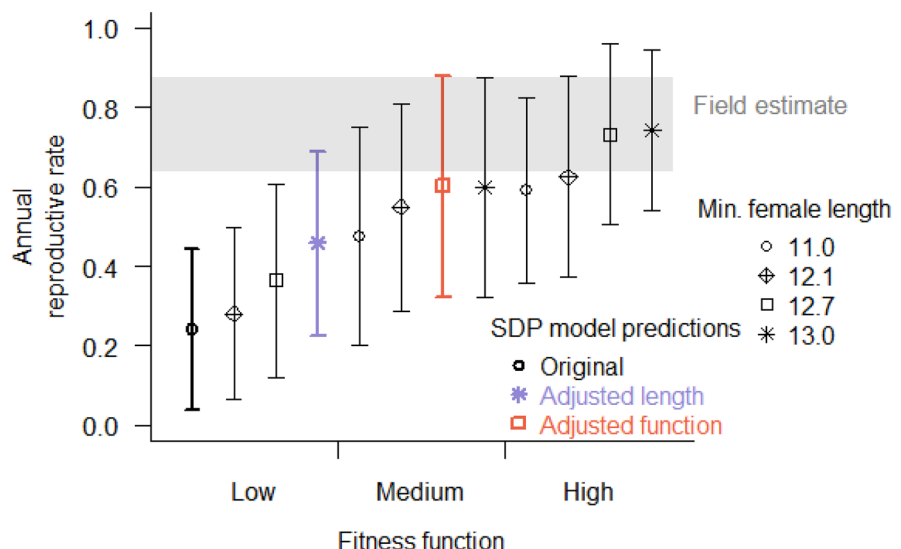
However, the assumption that whales would move to the closest undisturbed cell was not based on empirical data, since the magnitude of the behavioral response (how far an animal moves and how much foraging is lost) has yet to be quantified. Whales could show displacement farther from disturbed cells, which could lead to different predicted habitat use

patterns and energetic outcomes. While disturbance had the most impact on use of Cells 8 and 9 (which have lower prey energy than Cell 11), the location of the highest offshore prey resources may not be consistent (Blanchard pers. com.). Thus, future seismic surveys with similar spatial and temporal patterns may not have the same impact on whale foraging or distribution.

Objective 3: Comparison of model results with observed gray whale distribution and photo-identification data from 2015

SDP model predictions of the proportion of pregnant females utilizing the coastal nearshore area in 2015 were similar to estimates from photo-identification data. However, the data indicated the animals were present in the area before data collection began. Although arrival time of gray whales to Sakhalin Island is unknown, very few animals are successfully identified at the beginning of the season, and discovery rates were relatively constant throughout 2015 (Schwarz et al., 2022). That information led us to believe that whales were still arriving after photo-identification data collection began. Therefore, we assumed an arrival time of $June\ 15 \pm 5$ days in the SDP models but did not account for demographic-specific differences in arrival time. For example, the identified pregnant and potentially pregnant females in 2015 were first seen on June 2nd and June 7th, respectively. The first calf was identified on June

Fig. 9 Predicted probability of successful reproduction from 12 SDP simulations with varying reproductive fitness functions and minimum maternal length compared to the possible range of observed reproductive rates in 2015 (gray-shaded area). The three SDP models used for habitat use comparisons are highlighted for reference. Points are means and whiskers are ± 2 sd



24th, and the first mother was identified on July 7th, indicating a much earlier arrival time for pregnant females. Model sensitivity analysis from the initial SDP effort indicated arrival date had no effect on successful reproduction or maternal survival (McHuron et al., 2021) nor on the timing of the shift between nearshore and offshore cells. So, a change in the SDP arrival date would only change whale proportions in the nearshore and offshore area during the first few weeks of the season. Studies to determine arrival times earlier than June 1st would be limited due to ice and fog, but future model assumptions about arrival time should be developed with demographic-specific information in mind.

Within the nearshore area, SDP models predicted a higher density of animals in Cells 1, 2 (large effect, $|d| \geq 0.8$), and 4 (small-to-intermediate effect, $0.2 \leq |d| < 0.8$) and a lower density in Cell 3 (intermediate-to-large effect, $0.5 \leq |d|$) compared to density estimates from scan surveys. In fact, density estimates from scan surveys indicated an almost complete lack of whales in the southeastern part of the nearshore foraging area (Cell 2). Inter-annual whale densities in the southern part of the nearshore area (Cells 1 and 2) have been highly variable since density data have been collected (Exxon Neftegas, 2018), with Cell 1 usually an area of lower prey biomass (Blanchard et al., 2019, 2022a, b). Little is known about whale prey energy in Cell 2. Amphipods are considered the whales' prey of choice (Budnikova & Blokhin, 2012; Dunham & Duffus, 2001; Oliver et al., 1983), and we adjusted habitat use probabilities in the SDP model as a function of the proportion of energy from amphipods. We also assumed in SDP models that Cell 2 prey energy had the same mean prey energy value as other offshore cells (Cells 4 and 6) with lower variability (McHuron et al., 2021). Clearly the adjustment and assumption relative to prey availability need to be reassessed.

Cell 3 includes the mouth of Piltun Bay, and more-detailed analysis of the distribution data indicated whale density in that area was particularly high (Gailey et al., 2022b). Empirical density estimates in Cell 3 may be a reflection of high densities of mother-calf pairs and young animals concentrated at the mouth of Piltun Bay. The highest observed densities were within 1 km of shore (Gailey et al., 2022b), while pregnant females are more likely to be seen further from shore

(Sychenko, 2011). Thus, our adjustment to determine the density of pregnant females from the density of all whales may not be accurate for this area since non-pregnant females may make up a higher proportion of animals in the area.

Conversely, if pregnant females are foraging < 1 km from shore, benthic prey data (a key input in the SDP model) may not be an accurate representation of the level of prey energy available to pregnant whales in Cell 3 because nearshore whales are regularly seen foraging in depths considered too shallow to safely or adequately sample benthic prey (< 7 m) (Blanchard et al., 2022b).

Density estimates from scan data indicated highest use of Cell 7 from late June through early July, which was not predicted using the SDP model. Analysis of smaller 1-km² blocks indicated a higher density of gray whales farther than usual from shore in Cell 7 (Gailey et al., 2022b) which could have been related to a seasonal patch of high energy sand lance (*Ammodytes* sp.) (Blanchard et al., 2022a). Gray whales have previously been observed in that area feeding on sand lance (Fadeev, 2011). Because of the limited structure of the distribution of prey energy across space and time, the SDP models were not sensitive enough to capture behavioral differences driven by seasonal prey patches. That is, limitations in temporal and spatial benthic biomass sampling (Blanchard et al., 2022b) restricted us to one overall prey energy distribution for the entire foraging season in Cell 7 (McHuron et al., 2021; Maresh et al., 2022).

While disturbance had no-to-little effect on predictions of reproductive rate from SDP models, predictions were generally lower than the range of reproductive rate observed in 2015. However, of the 12 SDP model simulations that included 2015 disturbance, only three models (low reproductive fitness function and minimum length < 13.0) were inconsistent with observations. Analysis of long-term photo-identification data of Sakhalin Island gray whales estimated age at sexual maturity at 9.0 years old (7.7–11.2 years old) (Bradford et al., 2010; Cooke, 2010), which is consistent with findings in eastern gray whales (Rice & Wolman, 1971). An additional 10.5-year-old female was first seen with a calf in 2014 (Schwarz et al., 2022). Based on gray whale growth rates (Agbayani et al., 2020), we estimated roughly 50% of females would have successfully reproduced by age 9 years (with a mean value of 12.1 m in length) using both the medium and high reproductive

fitness functions (Fig. S3). Thus, both the empirical probability of reproductive success and age/size at sexual maturity support use of the medium and high reproductive fitness functions. Empirically determined age at sexual maturity estimates a minimum length around 11.7 m (mean length at 7.7 years old) and does not support a minimum maternal length > 12.1 m, which would translate to a minimum age of sexual maturity > 10 years old.

The empirical probabilities of successful reproduction would be biased high if the counted number of pregnant females was low. The long-term dataset (2002–2015) shows that the mothers are unknown for 38% of identified calves, likely due to variability in timing and photo-identification data collection effort each year, as calves could have separated from their mothers prior to data collection. Sakhalin Island gray whales have also been sighted off Kamchatka Peninsula, including mother-calf pairs that have separated before arrival at Sakhalin Island (Tyurneva et al., 2013). Therefore, it is likely that some Sakhalin Island reproductive females have not yet been identified. In 2015, 11 individuals of reproductive age (>6 years old) were either female or of unknown gender and could have been unidentified first-time mothers in 2016. An additional 81 identified whales were of unknown reproductive status (two female, 79 unknown gender). Four of those individuals were first identified in 2015. Based on sighting histories, the other 77 animals had a mean minimum age of 11 years, with 71% of them ($N=55$) at least 10 years old, at the higher end of the range for age of sexual maturity (Cooke, 2010; Rice & Wolman, 1971). Overall, several attributes of the photo-identification data lead to the potential for underestimates of the number of pregnant females in 2015.

If SDP models were adjusted to more-closely simulate the habitat use patterns seen from empirical data (no-to-little use of Cells 1 and 2, higher use of Cell 3), SDP predictions of reproductive rates would most likely remain the same, either with or without disturbance. McHuron et al. (2021) tested several theoretical disturbance scenarios that involved high levels of disturbance in all nearshore cells, resulting in no effect on predictions of successful reproduction. Alternatively, given that prey energy in Cell 11 can be up to ten times higher than energy in nearshore cells, exclusion from offshore cells could result in complete reproductive failure and maternal death (McHuron et al., 2021).

Conclusion

To identify the behavioral choices and demographic results of disturbance on pregnant western gray whales during one foraging season, a stochastic dynamic programming (SDP) model has been developed (McHuron et al., 2021). However, important underlying model assumptions about minimum maternal length (L_{\min}) and the reproductive fitness functions (R_{fit}) remained untested. Coincidentally, multiple seismic surveys were conducted in 2015 off the northeast coast of Sakhalin Island, Russia, in and near an important foraging habitat, which could lead to decreased gray whale foraging activity and reduced population growth via lower survival or reproduction. We varied assumptions about L_{\min} and R_{fit} to determine their effects on SDP predictions of habitat use, proportion of animals disturbed (p_{dist}), and reproductive success (ϕ), as well as the effects of disturbance from air gun sounds on predictions of habitat use and ϕ .

Benthic prey surveys, photo-identification studies, and distribution sampling occurred in conjunction with seismic surveys in 2015, providing a unique opportunity to compare output from SDP models with empirical observations of whale distribution, behavior, and vital rates. Resulting SDP predictions indicated higher nearshore habitat use with shorter L_{\min} , which reflects the expected differences in behavior based on size-specific energetic needs. Changes in L_{\min} and R_{fit} had no-to-little effect on predictions of p_{dist} or ϕ . Disturbance had a small-to-medium effect on predicted cell use in the northern offshore foraging area where disturbance was highest, showing an increased proportion of females in cells with higher probability of disturbance. SDP predictions of large-scale habitat use over the season were generally similar to values and trends seen from photo-identification and distribution data. However, empirical estimates of the proportion of pregnant females nearshore were much higher than SDP model predictions (large effect: |Cohen's d | > 0.8) during the first week, so arrival date was most likely too late in the SDP model. Within their nearshore habitat, the SDP model overestimated whale density in the south and underestimated density around the mouth of Piltun Bay compared to empirical estimates. SDP predictions of ϕ increased with increased L_{\min} and R_{fit} , and measured ϕ was a possible outcome in nine of the 12 SDP models.

Overall, this work, using data from a specific well-studied situation, underscores the potential usefulness of SDP modeling as a means of quantitatively applying the PCOD framework. However, and not surprisingly, SDP model predictive abilities can be limited due to assumptions in the model and parameter inputs. Sensitivity analyses can test the effects of SDP model assumptions on predictions and provide valuable information about where researchers need to focus data collection to improve SDP model accuracy. For example, measured lengths of known-age animals, particularly reproductive females, would aid in predicting reproductive success. Comparisons with field studies also identify SDP model limitations. For example, the SDP model could not capture some detailed gray whale nearshore spatial distribution because the temporal and spatial scales of prey input parameters (derived from data) were too coarse. Notably, the assumptions and limitations in the SDP model did not affect the comparisons between disturbed and undisturbed scenarios. If future SDP models are expanded to determine the effects of disturbance over longer periods, accuracy of vital rates may become more important, and additional and expanded field studies will be needed to improve SDP model inputs and functions and validate SDP model outputs.

Acknowledgements We dedicate this paper to Dr. H. R. Melton who passed away in 2017. Without his leadership, counsel, and technical knowledge, this work would not have been possible. Many people contributed to the collection of data used in this paper, and we thank all researchers and field assistants involved in the acoustic, benthic, distribution, and photo-identification studies in 2015. Ervin Kalinin and Mike Swindoll from Exxon Neftegas Limited provided logistical and safety support for the field teams. Olga Tyurneva, Valery Vertyankin, and Yuri Yakovlev assisted in processing photo-identification data, and Igor Zhmayev (LGL Eco) provided logistical and photo-identification database support. The Zhirmunsky National Science Center of Marine Biology of the Russian Academy of Sciences Far East Branch provided benthic prey data, and Arny Blanchard processed the data for our analyses. Alexander Rutenko and Mikhail Fershalov (V.I. Il'ichev Pacific Oceanological Institute of the Russian Academy of Sciences Far East Branch), Mikhail Zykov and Roberto Racca (JASCO Applied Sciences), and Mike Jenkerson and Vladimir Nechayuk (ExxonMobil) processed acoustic data for use in this paper. We appreciate the feedback from Lisanne Aerts and from the guest editor Bill Streever on this paper. This work originated from discussions within the US Office of Naval Research (ONR), Population Consequences of Disturbance (PCOD) Working Group. Marc Mangel was partially supported by ONR grant N00014-19-2494 during the completion of the manuscript. Exxon Neftegas Limited provided financial support for all authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aerts, L., Jenkerson, M. R., Nechayuk, V. E., Gailey, G., Racca, R., Blanchard, A. L., et al. (2022). Seismic surveys near gray whale feeding areas off Sakhalin Island, Russia: Assessing impact and mitigation effectiveness. *Environmental Monitoring and Assessment*, *194*. <https://doi.org/10.1007/s10661-022-10016-9>
- Agbayani, S., Fortune, S. M. E., & Trites, A. W. (2020). Growth and development of North Pacific gray whales (*Eschrichtius robustus*). *Journal of Mammalogy*, *100*(3), 742–754.
- Blanchard, A. L., Aerts, L., Yazvenko, S. B., Demchenko, N. L., Shcherbakov, I., & Melton, H. R. (2019). Prey biomass dynamics in gray whale feeding areas adjacent to Northeastern Sakhalin (the Sea of Okhotsk), Russia, 2001–2015. *Marine Environmental Research*, *145*, 123–136. <https://doi.org/10.1016/j.marenvres.2019.02.008>
- Blanchard, A. L., Ainsworth, L., Gailey, G., Demchenko, N. L., & Shcherbakov, I. (2022a). Benthic studies adjacent to Sakhalin Island, Russia 2015a III: benthic energy density spatial models in the nearshore gray whale feeding area. *Environmental Monitoring and Assessment*, *194*. <https://doi.org/10.1007/s10661-022-10018-7>
- Blanchard, A. L., Demchenko, N. L., Aerts, L., Yazvenko, S. B., Ivin, V. V., & Shcherbakov, I. (2022b). Benthic studies adjacent to Sakhalin Island, Russia, 2015b I: benthic biomass and community structure in the nearshore gray whale feeding area. *Environmental Monitoring and Assessment*, *194*. <https://doi.org/10.1007/s10661-022-10017-8>
- Bradford, A. L., Weller, D. W., Lang, A. R., Tsidulko, G. A., Burdin, A. M., & Brownell Jr., R. L. (2010). Comparing observations of age at first reproduction in western gray whales to estimate of age at sexual maturity in eastern gray whales. *Paper SC/62/BRG2 submitted to the IWC Scientific Committee*. <https://digitalcommons.unl.edu/usdeptcommercepub/140/>
- Bradford, A. L., Weller, D. W., Punt, A. E., Ivashchenko, Y. V., Burdin, A. M., VanBlaricom, G. R., et al. (2012). Leaner leviathans: Body condition variation in a critically endangered whale population. *Journal of Mammalogy*, *93*(1), 251–266. <https://doi.org/10.1644/11-MAMM-A-091.1>
- Bröker, K., Gailey, G., Tyurneva, O. Y., Yakovlev, Y. M., Sychenko, O., Dupont, J. M., et al. (2020). Site-fidelity and spatial movements of western North Pacific gray

- whales on their summer range off Sakhalin, Russia. *PLoS One*. <https://doi.org/10.1371/journal.pone.0236649>
- Budnikova, L. L., & Blokhin, S. A. (2012). Food contents of the eastern gray whale *Eschrichtius robustus* Lilljeborg, 1861 in the Mechigmentsky Bay of the Bering Sea. *Russian Journal of Marine Biology*, 38, 149–155.
- Bürkner, P. (2017). brms: An R package for Bayesian multi-level models using stan. *Journal of Statistical Software*, 80(1). <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P. (2018). Advanced Bayesian multilevel modeling with the R Package brms. *The R Journal*, 10(1). <https://doi.org/10.32614/RJ-2018-017>
- Clark, C. W., & Mangel, M. (2000). *Dynamic state variable models in ecology: Methods and applications*. Oxford University Press.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Academic Press.
- Cooke, J. G. (2010). Joint population assessment of western gray whales using data from IBM and Russia-US photo-identification teams collected off Sakhalin Island through 2008, Paper GWAP 8/09 revised. *8th Meeting of the Western Gray Whale Advisory Panel*.
- Cooke, J. G., Weller, D. W., Bradford, A. L., Sychenko, O., Burdin, A. M., Lang, A. R., et al. (2017). Updated population assessment of the Sakhalin gray whale aggregation based on a photo-identification study at Piltun, Sakhalin, 1995–2015. Report SC/66b/BRG presented to the IWC Scientific Committee.
- Costa, D. P., Schwarz, L. K., Robinson, P. W., Schick, R. S., Morris, P. A., Condit, R. S., et al. (2016). A bioenergetics approach to understanding population consequences of disturbance: Elephant seals as a model system. *The Effects of Noise on Aquatic Life II. Advances in Experimental Medicine and Biology*, 875, 161–169.
- Dunham, J. S., & Duffus, D. A. (2001). Foraging patterns of gray whales in Central Clayoquot Sound, British Columbia, Canada. *Marine Ecology Progress Series*, 223, 299–310. <https://doi.org/10.3354/meps223299>
- Exxon Neftegas, L. (2018). Report on gray whale monitoring program off northeast Sakhalin Island in 2017. Western gray whale advisory panel, 19th meeting GWAP-19/16-EN 14–16 Nov 2018.
- Fadeev, V. I. (2011). Benthos studies in feeding grounds of western gray whales off the northeast coast of Sakhalin Island (Russia), 2002–2010. Paper SC/63/BRG15 presented to the IWC Scientific Committee. (pp. 14 pp).
- Gabry, J., & Mahr, T. (2021). bayesplot: Plotting for Bayesian Models. *R package version 1.8.0*. <https://mc-stan.org/bayesplot/>
- Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society*, 182, 389–402. <https://doi.org/10.1111/rssa.12378>
- Gailey, G., & Ortega-Ortiz, J. (2002). A note on a computer-based system for theodolite tracking of cetaceans. *Journal of Cetacean Research and Management*, 4, 213–218.
- Gailey, G., Sychenko, O., McDonald, T., Racca, R., Rutenko, A., & Bröker, K. (2016). Behavioural responses of western gray whales to a 4-D seismic survey off northeastern Sakhalin Island, Russia. *Endangered Species Research*, 30, 53–71. <https://doi.org/10.3354/esr00713>
- Gailey, G., Sychenko, O., Tyurneva, O. Y., Yakovlev, Y. M., Vertyankin, V. V., van der Wolf, P., et al. (2020). Effects of sea ice on growth rates of an endangered population of gray whales. *Scientific Reports*, 10, 1553. <https://doi.org/10.1038/s41598-020-58435-3>
- Gailey, G., Sychenko, O., Zykov, M. M., Rutenko, A., Blanchard, A. L., & Melton, H. R. (2022a). Western gray whale behavioural response to seismic surveys during their foraging season. *Environmental Monitoring and Assessment*, 194. <https://doi.org/10.1007/s10661-022-10023-w>
- Gailey, G., Würsig, B., & McDonald, T. (2007). Abundance, behavior, and movement patterns of western gray whales in relation to a 3-D seismic survey, northeast Sakhalin Island, Russia. *Environmental Monitoring and Assessment*, 134, 75–91.
- Gailey, G., Zykov, M. M., Sychenko, O., Rutenko, A., Blanchard, A. L., Aerts, L., et al. (2022b). Gray whale density during seismic surveys near their Sakhalin feeding ground. *Environmental Monitoring and Assessment*, 194. <https://doi.org/10.1007/s10661-022-10025-8>
- Houston, A. I., Stephens, P. A., Boyd, I. L., Harding, K. C., & McNamara, J. M. (2006). Capital or income breeding? A theoretical model of female reproductive strategies. *Behavioral Ecology*, 18, 241–250. <https://doi.org/10.1093/beheco/arl080>
- Leeper, R., & Gordon, J. (2001). Application of photogrammetric methods for locating and tracking cetacean movements at sea. *Journal of Cetacean Research and Management*, 3, 131–142.
- Lerczak, J. A., & Hobbs, R. C. (1998). Calculating sighting distances from angular readings during shipboard, aerial, and shore-based marine mammal surveys. *Marine Mammal Science*, 14, 590–598.
- Malme, C. I., Würsig, B., Bird, J. E., & Tyack, P. (1988). Observations of feeding gray whale responses to controlled industrial noise exposure. In W. M. Sackinger, M. O. Jeffries, J. L. Imm, & S. D. Treacy (Eds.), *Port and ocean engineering under arctic conditions vol. II, symposium on noise and marine mammals* (pp. 55 – 73). Fairbanks, AK: Univ. Alaska Fairbanks.
- Mangel, M., & Clark, C. W. (1988). *Dynamic modeling in behavioral ecology*. Princeton University Press.
- Mangel, M., & Ludwig, D. (1992). Definition and evaluation of the fitness of behavioral and developmental programs. *Annual Review of Ecology and Systematics*, 23, 507–536. <https://doi.org/10.1146/annurev.es.23.110192.002451>
- Maresh, J. L., Blanchard, A. L., Demchenko, N. L., Shcherbakov, I., Aerts, L., & Schwarz, L. K. (2022). Benthic studies adjacent to Sakhalin Island, Russia, 2015 II: Energy content of the zoobenthos in western gray whale feeding grounds. *Environmental Monitoring and Assessment*. <https://doi.org/10.1007/s10661-022-10020-z>
- Mate, B. R., Bradford, A. L., Tsidulko, G. A., Vertyankin, V. V., & Ilyashenko, V. Y. (2011). Late-feeding season movements of a western North Pacific gray whale off Sakhalin Island, Russia and subsequent migration into the Eastern North Pacific. *Paper SC/63/BRG23 submitted to the IWC Scientific Committee*.
- McHuron, E. A., Aerts, L., Gailey, G., Sychenko, O., Costa, D. P., Mangel, M., et al. (2021). Predicting the population consequences of acoustic disturbance, with application to an endangered gray whale population. *Ecological Applications*, 31(8), e02440. <https://doi.org/10.1002/eap.2440>
- McHuron, E. A., Costa, D. P., Schwarz, L. K., & Mangel, M. (2017). State-dependent behavioural theory for assessing the fitness consequences of anthropogenic disturbance on capitial and income breeders. *Methods in Ecology and*

- Evolution*, 8(5), 552–560. <https://doi.org/10.1111/2041-210X.12701>
- McHuron, E. A., Schwarz, L. K., Costa, D. P., & Mangel, M. (2018). A state-dependent model for assessing the population consequences of disturbance on income-breeding mammals. *Ecological Modelling*, 385, 133–144. <https://doi.org/10.1016/j.ecolmodel.2018.07.016>
- Moore, S. E., Grebmeier, J. M., & Davies, J. R. (2003). Gray whale distribution relative to forage habitat in the northern Bering Sea: Current conditions and retrospective summary. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, 81(4), 734–742. <https://doi.org/10.1139/z03-043>
- Muir, J. E., Ainsworth, L., Joy, R., Racca, R., Bychkov, Y., Gailey, G., et al. (2015). Distance from shore as an indicator of disturbance of gray whales during a seismic survey off Sakhalin Island, Russia. *Endangered Species Research*, 29(2), 161–178.
- Muir, J. E., Ainsworth, L., Racca, R., Bychkov, Y., Gailey, G., Vladimirov, V., et al. (2016). Gray whale densities during a seismic survey off Sakhalin Island, Russia. *Endangered Species Research*, 29(3), 211–227. <https://doi.org/10.3354/esr00709>
- Nerini, M. (1984). A review of gray whale feeding ecology. In M. L. Jones, S. L. Swartz, & S. Leatherwood (Eds.), *The gray whale Eschrichtius robustus*. New York: Academic Press.
- New, L. F., Clark, J. S., Costa, D. P., Fleishman, E., Hindell, M. A., Klanjšček, T., et al. (2014). Using short-term measures of behaviour to estimate long-term fitness of southern elephant seals. [Article]. *Marine Ecology Progress Series*, 496, 99–108. <https://doi.org/10.3354/meps10547>
- Oliver, J. S., Slattery, P. N., Silberstein, M. A., & Oconnor, E. F. (1983). A comparison of gray whale, *Eschrichtius robustus*, feeding in the Bering Sea and Baja California. *Fishery Bulletin*, 81(3), 513–522.
- Perryman, W. L., Joyce, T., Weller, D. W., & Durban, J. W. (2020). Environmental factors influencing eastern North Pacific gray whale calf production 1994–2016. *Marine Mammal Science*. <https://doi.org/10.1111/mms.12755>
- Pirotta, E., Mangel, M., Costa, D. P., Goldbogen, J., Harwood, J., Hin, V., et al. (2019). Anthropogenic disturbance in a changing environment: Modelling lifetime reproductive success to predict the consequences of multiple stressors on a migratory population. *Oikos*. <https://doi.org/10.1111/oik.06146>
- Plummer, M. (2019). rjags: Bayesian graphical models using MCMC. <https://CRAN.R-project.org/package=rjags>.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6, 7–11.
- Rice, D. W., & Wolman, A. A. (1971). *The life history and ecology of the gray whale (Eschrichtius robustus)*. American Society of Mammalogists.
- Rutenko, A. N., Zykov, M. M., Gritsenko, V. A., Fershalov, M. Y., Jenkerson, M. R., Manulchev, D. S., et al. (2022). Acoustic monitoring and analyses of air gun, pile driving, vessel, and ambient sounds during the 2015 seismic surveys on the Sakhalin shelf. *Environmental Monitoring and Assessment*, 194. <https://doi.org/10.1007/s10661-022-10021-y>
- Sanchez-Pacheco, J. A., Vazquez-Hanckin, A., & De Silva-Davila, R. (2001). Gray whales' mid-spring feeding at Bahia de los Angeles, Gulf of California, Mexico. *Marine Mammal Science*, 17(1), 186–191. <https://doi.org/10.1111/j.1748-7692.2001.tb00991.x>
- Schwarz, L. K., Gailey, G., Tyurneva, O., Yakovlev, Y., Sychenko, O., van der Wolf, P., et al. (2022). Western gray whales on their summer feeding grounds off Sakhalin Island in 2015: who is foraging where? *Environmental Monitoring and Assessment*, 194. <https://doi.org/10.1007/s10661-022-10022-x>
- Sychenko, O. (2011). *Western gray whale (Eschrichtius robustus) mother and calf ecology off Sakhalin Island (Thesis)*. Texas A&M University.
- Tyurneva, O. Y., Yakovlev, Y. M., & Vertyankin, V. V. (2013). Photo-identification study of western gray whales (*Eschrichtius Robustus*) offshore northeast Sakhalin Island and southeast Kamchatka Peninsula, Russia. Paper SC/65a/BRG08 prepared for the Scientific Committee Meeting of the International whaling Commission. <https://archive.iwc.int/?r=8757&k=9067feccd8>
- Tyurneva, O. Y., Yakovlev, Y. M., Vertyankin, V. V., van der Wolf, P., & Scott, M. J. (2018). Long-term photo identification studies of gray whales (*Eschrichtius robustus*) offshore northeast Sakhalin Island, Russia, 2002–2017. Paper SC/67B/ASI/04 presented to the IWC Scientific Committee, 15 pp.
- Villegas-Amtmann, S., Schwarz, L. K., Gailey, G., Sychenko, O., & Costa, D. (2017). East or West: The energetic cost of being a gray whale and the consequence of losing energy to disturbance. *Endangered Species Research*, 34, 167–183. <https://doi.org/10.3354/esr00843>
- Villegas-Amtmann, S., Schwarz, L. K., Sumich, J. L., & Costa, D. P. (2015). A bioenergetics model to evaluate demographic consequences of disturbance in marine mammals applied to gray whales. *Ecosphere*, 6(10), 183. <https://doi.org/10.1890/ES15-00146.1>
- Weller, D. W., Bradford, A. L., Lang, A. R., Brownell Jr., R. L., & Burdin, A. M. (2009). Birth-intervals and sex composition of western gray whales summering off Sakhalin Island, Russia. *Publications, Agencies and Staff of the U.S. Department of Commerce*, 128. https://digitalcommons.unl.edu/usdeptcommercepub/128?utm_source=digitalcommons.unl.edu%2Fusdeptcommercepub%2F128&utm_medium=PDF&utm_campaign=PDFCoverPages
- White, J. W., Rassweiler, A., Samhoury, J. F., Stier, A. C., & White, C. (2014). Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos*, 123, 385–388.
- Yakovlev, Y. M., Tyurneva, O. M., Vertyankin, V. V., & van der Wolf, P. (2016). Gray whale (*Eschrichtius robustus*) photo identification off north-east coast of Sakhalin in 2015. Western gray whale advisory panel, 17th meeting WGAP-17/8-EN 11–13 November 2016. (pp. 57).
- Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2017). Using stacking to average Bayesian predictive distributions. *Bayesian Analysis*. <https://doi.org/10.1214/17-BA1091>
- Yazvenko, S. B., McDonald, T., Blokhin, S. A., Johnson, S. R., Meier, S. K., Melton, H. R., et al. (2007). Distribution and abundance of western gray whales during a seismic survey near Sakhalin Island, Russia. *Environmental Monitoring and Assessment*, 134, 45–73.
- Youngflesh, C. (2018). MCMCvis: Tools to visualize, manipulate, and summarize MCMC output. *Journal of Open Source Software*, 3(24). <https://doi.org/10.21105/joss.00640>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.