3.0 ABUNDANCE AND DENSITY ANALYSES

3.1 DATA PREPARATION AND ASSUMPTIONS

3.1.1 Sightings Data Criteria

The sightings included in the density/abundance analyses had to meet the following criteria:

1) Sightings were recorded by on-duty observers while the team was searching in on-effort mode. Sightings data recorded from other surveys not associated with this baseline study could not be combined with the sightings data collected during this baseline study to generate abundance/ density estimates because we could not assume that the detection function remained constant throughout the different surveys simply due to different weather conditions, observer teams, survey platforms, and protocols. In addition, opportunistic sightings could not be included in the analyses because they were not collected under line transect protocols.

2) Perpendicular sighting distances had to be calculated for each of the on-effort sightings included in the abundance/density analyses.

3) Sightings and effort recorded during a BSS≤5 were included in the density/abundance modeling of all species or groups except the harbor porpoise (*Phocoena phocoena*). The abundance/density analysis for the harbor porpoise was based only on effort conducted in the best survey conditions (BSS ≤2). Harbor porpoises are difficult to detect in a higher BSS because their sighting cue (small, dark dorsal fin) is hard to see, particularly as the distance from the vessel increases (Polacheck 1995). In addition, harbor porpoises typically do not spend much time at the surface and often occur singly or in very small groups which adds to the difficulty in detecting this species (Polacheck 1995).

Conservative modeling of data with high precision (low variance) requires an adequate sample size (n). Generally, as sample size increases, variance decreases and precision improves. A sample size of at least 60 sightings is typically recommended for estimating a detection function (Buckland et al. 2001), and 15 sightings may be the absolute minimum number of sightings that can be used to fit a detection function (Barlow et al. 2006). Due to some of the low number of sightings during this baseline study, we specified a minimum sample size of around 20 sightings in order to model a detection function. Species with fewer than 20 sightings were pooled into taxonomic groups with species of similar sighting characteristics when possible, and modeling of a group detection function was then conducted. The data were then stratified by species to estimate abundance/density of individual species using the pooled detection function. In one case, a minimum of 18 sightings was used to fit a detection function.

Aerial and shipboard survey data could not be combined for density/abundance estimation because of the differences in survey techniques and perception bias (animals were at the surface but were not seen). Therefore, separate analyses were conducted for the aerial and shipboard sightings data. The Conventional Distance Sampling (CDS) method was used to generate abundance/density estimates for the overall Study Area, and the Density Surface Modeling (DSM) method was used to generate surface maps of predicted density at a finer spatial resolution using various environmental covariates as predictors of density. All analyses were carried out using Distance 6.0 release 2 (Thomas et al. 2010)[2] and the statistical program R.[3] Note that the PAM results could not be used to generate density/abundance estimates since these results only provided information on the presence of certain species and did not meet the criteria mentioned above.

3.1.2 Modifications to Sightings Data

We estimated detection functions after filtering the data based on the above criteria. During the exploratory data analysis phase, it is important to identify any “spikes” in the data and what the cause may be since different models will give very different abundance/density estimates for spiked data (Thomas et al. 2010). We plotted histograms of the perpendicular distance data, and selected various

cutpoints to identify suitable truncation points (for removal of spurious data and outliers) for perpendicular distances in order to conform to the conditions of the "ideal" probability detection function. Buckland et al. (2001) recommend truncation of the most distant 5 to 10% of sightings from the right-hand tail of the detection function to remove outliers and improve the ability to fit the detection function; however, due to our small sample sizes, a 5 to 10% right truncation may remove too many sightings and hinder our ability to fit a detection function. Thus, when considering truncation of some of the data, trade-offs must be weighed between the benefits of removing spurious data (which can reduce variance) and the costs of a reduced sample size (which can increase variance). Instead of truncating 5 to 10% of sightings, right truncations were based on specific distances from the trackline which were determined on a case-by-case basis for the different species/group analyses by assessing the Q-Q plots and histogram plots using various truncation lengths. In some cases, spurious data can cause spikes of detections near the trackline. These spikes often arise when animals (e.g., dolphins) are attracted to the survey vessel and detections were not made before any responsive movement occurred (Thomas et al. 2010). Spikes can also be caused by inaccurate estimation of sighting angles for detections ahead of the vessel (often rounding of perpendicular sighting distances to zero; Thomas et al. 2010). For the shipboard survey analyses, the spiked data were not removed with a left truncation because we did not want to eliminate data with a near-100% detection probability at short distances. A left truncation was used for the aerial survey data collected in 2009 not because of a spike near the trackline but because of the limited visibility of the trackline due to the lack of bubble and belly windows on the survey plane. In this case, a left truncation position was chosen where detection was certain.

Distance data are either recorded as exact measurements or are grouped ("binned") into distance categories (Buckland et al. 2001). During the shipboard surveys, sighting distances and angles were recorded as exact measurements and were transformed to perpendicular sighting distances for analysis. Therefore, the shipboard sightings data could be analyzed as exact data in Distance; however, the aerial survey data were collected as both exact data and binned data. During the 2008 aerial surveys, the declination angle of each sighting from the plane was recorded either as an exact distance (measured with an inclinometer) or as a bin number which corresponded to a range of declination angles. During the 2009 aerial surveys, the GPS locations of the plane on the trackline and of the sighting were used to calculate the exact perpendicular sighting distance for each sighting. For some analyses it was necessary to combine the aerial survey data from both years to have an acceptable sample size to use for the density/abundance analyses. Therefore, we had to combine the survey data that was collected in bins and the data that was collected as exact distances. To do so, we had to analyze all the data as though it were collected as binned data. When we analyzed data from only the 2009 aerial surveys, we were able to use the exact distance data (unbinned) in our abundance/density analyses.

3.1.3 Assumptions

The key assumptions for line transect surveys are as follows (Buckland et al. 2001):

1) The detection function (see **Section 3.2.1**) was the same for all animals/detections.

2) Animals were detected at their initial location. Marine mammals and sea turtles are highly mobile; therefore, it can be difficult to determine initial locations. For example, some species, such as harbor porpoises, tend to move away from vessels (Barlow 1988; Polacheck and Thorpe 1990; Palka and Hammond 2001) and other species, such as short-beaked common dolphins (*Delphinus delphis*), are attracted to vessels and often approach ships to bow ride (Palka et al. 2005). To minimize the potential bias of responsive behavior of animals to the ship, the observers used high-powered (bigeye) binoculars so that they could see a great distance from the trackline and detect animals before they reacted (positively or negatively) to the presence of the ship.

3) All measurements recorded during the surveys are exact and not subject to rounding (heaping), measurement errors, or recording errors. For grouped or binned data, the measurements are assumed to be assigned to the correct category (or bin). No measurements are likely to be exact on the moving platforms of the plane and ship; however, we attempted to minimize error in our measurements by using the azimuth rings, reticle scales, and inclinometers. Every effort was

made to avoid rounding any measurements. In regards to group size estimates, we were not able to compare our observer estimates with aerial photographs of sightings; however, we did obtain group size estimates from as many of the observers as possible and used the average of the best group size estimates for each sighting.

4) Animals on the trackline (at zero distance) were detected with certainty such that $g(0)=1$. At zero perpendicular distance y=0 (i.e., when the animal is on the trackline), the detection probability should be at or near 100% (i.e., all or nearly all animals on the trackline should be detected). Over a moderate range of short distances, the detection probability should be ideal (100%) or near ideal (i.e., a broad shoulder in the detection function), meaning that all animals that are actually present are detected by the observer for some distance from the trackline. Instruments that aid in detection at short distances (such as high-power binoculars) can increase the distance range of the "broad shoulder". Naturally, as sighting distance increases over longer distances, the number of sightings/detections should begin to decrease, and at a given distance, large animals and animal clusters are more likely to be detected than smaller animals and animal clusters. Assumption of $g(0)=1$ can lead to bias and underestimation of abundance and density (since density is inversely related to $g[0]$). This assumption rarely holds true during marine mammal and sea turtle surveys due to availability bias and perception bias. Perception bias results when an observer fails to detect an animal on the trackline when the animal is actually at the surface on the trackline. Factors that can influence perception bias include viewing conditions (e.g., BSS, glare, swell height), observer condition (e.g., experience, fatigue), and platform characteristics (e.g., pitch, roll, yaw, altitude). Availability bias results when an animal is submerged or otherwise hidden from view while on the trackline and, hence, is unable to be detected. Factors that can affect availability bias include species-specific behavior, group size, blow and dive characteristics, and dive intervals. Availability bias is particularly a problem for long divers, such as sperm whales (*Physeter macrocephalus*) and beaked whales (family Ziphiidae), and is not as much of a problem for species that have shorter dive times, such as common dolphins.

   A discussion of $g(0)$, factors affecting animal detectability, and methods of accounting for detection bias are discussed in Thomsen et al. (2005). Estimates of $g(0)$ for shipboard and aerial surveys are used to calculate less biased estimates of population size. To estimate $g(0)$ for shipboard surveys, there are two methods that can be used. The first uses a two-team approach (double platform) where two independent teams of observers scan the same trackline simultaneously (see Borchers et al. 1998). This approach is very costly in personnel and equipment and requires a ship large enough to accommodate the two platforms of the observer teams. The second method utilizes a survey aircraft to survey the ship's trackline three to four times during a single day. The NMFS has used this method successfully (e.g., Palka 2005). This technique involves simultaneous ship and aerial surveys that cover the same spatial and temporal area. The sightings from the ship and aircraft are then compared to estimate the number sightings missed by the ship but seen from the aircraft. This method is normally used to estimate $g(0)$ for aircraft but can be used to approximate this metric in reverse. Although this was the most cost effective method that could be used for our baseline study due to budgetary constraints, this method of conducting ship-plane experiments for every species was not practical for our study due to the relatively low encounter rate recorded from the ship and plane. This method is more conducive to an area with a relatively high density of marine mammals and high encounter rate which would allow for sufficient simultaneous recordings of sightings from the plane and ship. Otherwise, the costs of this method greatly increase due to the amount of simultaneous effort the ship and plane would need to run in order to record enough sightings of each species for calculating $g(0)$.

   To estimate $g(0)$ for aerial surveys, one of the following three methods is typically used. The Hiby circle-back data collection method uses a double-platform approach in which the aircraft periodically circles back on itself, and thus acts as both platform 1 (on the first pass) and platform 2 (when it circles back; Hiby 1999). Therefore, once a group of animals is sighted, the aircraft will continue to fly the trackline for 30 s, break trackline and fly the reciprocal heading past the sighting for another 30 s, and then rejoin the trackline. This trackline segment is then repeated

and the presence/absence of the animals is recorded. The ratio of initial sightings to resighting events provides an estimate of *g*(0). This method is unbiased in low density areas but requires a large sample size; therefore, it is not applicable to most species (Palka et al. 2005). Although this method was originally proposed for the baseline aerial surveys, estimates *g*(0) were not obtained due to several factors. The method was originally designed to be used with at least two observers and a data recorder onboard the aircraft. For safety reasons a co-pilot was added to the crew for the 2009 flights. Because of the seating limitations of a Skymaster aircraft (four seats), the seat for the data recorder was eliminated. During initial attempts to consistently implement the Hiby circle-back method, the additional data recording requirements and the circle-back protocol resulted in unconfirmed or loss of sightings due to the multi-tasking of observers. A second method for estimating *g*(0) for aerial surveys uses two independent observer teams; this method is similar to the double platform approach discussed above for ship surveys; however, this method requires an airplane that can accommodate two teams of observers. Our aircraft did not meet this requirement. The third approach involves the use of the ship as mentioned above. These ship-plane experiments are not conducive to our Study Area due to the relatively low encounter rates.

For the purposes of this report, we assumed a *g*(0) of 1 because we were not able to calculate estimates of *g*(0) due to the limitations discussed above. We chose not to use *g*(0) estimates that have been calculated from other similar surveys since detection probability has been shown to vary substantially among observers, platforms, weather conditions, etc. (Borchers 2005). Therefore, the density and abundance estimates calculated for this report should be considered underestimated due to both perception and availability bias.

3.2 CONVENTIONAL DISTANCE SAMPLING

CDS is a design-based approach in which the abundance/density estimates that are generated are based on the survey design which is assumed to provide a representative sample of the entire Study Area. Therefore, we used this method to extrapolate from the sampled strips in our line transect sampling. More information about the CDS approach is discussed below. Additional information can be found in Buckland et al. (2001; 2004) and Thomas et al. (2010).

3.2.1 Detection Function

The CDS engine in Distance uses a flexible semi-parametric detection function modeling framework (Thomas et al. 2010). Sightings data were modeled as a probability detection function g(y), a plot of sightings versus distance between the sighting and the perpendicular distance from the sighting to the trackline on which the ship/plane is traveling. Estimates of density and abundance were based on estimates of encounter rate, detection probability, and mean cluster (group) size.

3.2.1.1 Detection Probability Estimation

An ideal probability detection function has the following characteristics (Buckland et al. 2001):

1) An intercept of *g*(0) = 1.0 (100% probability of detection) at zero perpendicular distance y=0 (where *g*[0] is the probability of detecting an animal on the trackline),
2) A broad shoulder over a range of short distances before beginning to taper off,
3) A monotonically decreasing function *g*(y) with increasing perpendicular distance y, and
4) An upward shift in the detection function *g*(y) as animal/cluster size increases (when animal size or cluster size is included in the modeling).

The decrease in detection probability as a function of increasing perpendicular distance from the transect line was modeled using a half-normal or hazard-rate key function along with cosine series expansion terms as required. This model optimization analysis was conducted for each species/group in which there were around 20 sightings that met the criteria described in **Section 3.1.1**. During the model optimization analysis, the detection functions for each species/group were modeled using different combinations of the

half-normal and hazard-rate key functions with the expansion terms. In most cases, the optimal model was chosen as that model which yielded the smallest value of the Akaike's Information Criterion (AIC) index (Buckland et al. 2001, 2004), given by: AIC = -2*ln(L) + 2*q, where ln(L) is the log-likelihood function evaluated at the maximum likelihood estimates of the model parameters and q = number of estimated model parameters. AIC quantifies the bias-variance trade-off. The first term quantifies how well the model fits the data, which can also be quantified via the chi-square (X^2) goodness-of-fit (GOF) test. The second term quantifies the penalty (increased variance) associated with addition of model parameters. Model parameter addition (increase in q) improves model fit and reduces bias at a cost of increasing variance and model complexity. To aid in model selection, the model with the lowest AIC is identified as the optimal model, which has the best combination of a good fit to the data without too many parameters (parsimony principle). In some cases where the behavioral observations indicated a problem with avoidance or attraction to the survey platform, the optimal model was subjectively chosen. For example, when a spike near the trackline was thought to be caused by the attraction of the animals to the platform, the optimal model chosen was the one that did not fit the detection function to the whole spike. Fitting the spike near the trackline results in inflated abundance/density estimates.

3.2.1.2 Mean Group Size Estimation

We are estimating the density/abundance of animals which often occur in groups or clusters. Therefore, the mean group size of the sightings may be subject to size bias. Large groups are often detected at greater distances from the trackline than small groups which can lead to positively-biased estimates of the mean size of detected groups. In general, the arithmetic mean group size may be an overestimate of the true mean group size and could lead to positively-biased density and abundance estimates. To account for group-size bias, the size-bias regression approach was used to estimate an expected mean group size using Distance. In this approach, the expected mean group size of the population is estimated by using a regression method in which the logarithm of cluster size of observation "i", log(si), is regressed against the estimated detection probability, g(yi), where yi = perpendicular distance of object "i" from the trackline: log(si) = a + b*g(yi), where "a" (intercept) and "b" (slope) are regression coefficients. Mean cluster size in the population is estimated from the predicted mean size of detected clusters in the region where the detection probability is at or near 100% (i.e., g[yi=0] = 1.0, at zero perpendicular distance from the trackline). Thus, from the above regression equation, mean cluster size is approximated by s(mean) = a + b, where g(yi) is set equal to 1. This regression method corrects for size-biased detections and for the underestimation of size of detected groups (Buckland et al. 2001). A statistical hypothesis test was applied to the regression of group size on distance, and the expected mean group size was only used in the analysis if it was significantly ($P<0.15$) smaller than the arithmetic mean group size. If it was not significantly smaller, then the observed mean group size was used.

3.2.1.3 Density, Abundance, and Variance Estimation

According to line transect theory (Buckland et al. 2001), density (abundance per unit area) is estimated as a function of:

1) Encounter rate n/L (where n = sample size or number of sightings and L = line transect length or effort),
2) Probability density function at zero perpendicular distance *f*(0),
3) Mean group or cluster size E(s), and
4) Probability detection function at zero perpendicular distance (*g*[0]).

The estimated density (D) is given by the following equation:

$$D = N/A = n*E(s)*f(0)/2L*g(0)$$

where N = abundance, A = Study Area, E(s) = mean group size, and the other parameters are as defined previously. The term *g*(0) (the availability bias) is assumed to be 1.0. Assuming *g*(0) is constant, the sources of variance associated with abundance/density estimation in the CDS method include contributions of encounter rate (n/L), detection probability *f*(0), and mean group size E(s). Encounter rate

(n/L) is defined as the ratio of the number of animals observed (n) to the effort L (i.e., transect length) associated with those sightings. Group size (E[s]) is effectively the ratio of the total number of individual animals observed (abundance) to the number of observations (see **Section 3.2.1.2**). Density (i.e., ratio of abundance N to Study Area A) is estimated as the ratio of the number of animals sighted (n) to the survey coverage area (a), where a = 2wL, w = strip half-width (truncation distance), and L = transect length. The effective strip half-width (ESW), μ, is defined as the sighting distance such that the number of animals at distances less than μ that were missed by the observer is equal to the number of animals at distances greater than μ that were detected by the observer. The ESW μ is equal to $1/f(0)$. Using the parameters $f(0)$ and μ derived from the optimal detection function and assuming $g(0)=1$, the above density equation can be simplified to the following:

$$D = n*E(s)/2\mu L$$

The error or uncertainty associated with each estimated parameter (D, n/L, $f[0]$, E[s]) can be quantified by the variance (Var), coefficient of variation (CV), and the 95% confidence interval (CI). The CV is the ratio of the square root of variance to the value of the parameter estimate. For example, $CV(x) = Var(x)^{0.5}/x$, where x = D, n/L, $f(0)$, or E(s). The CDS engine in Distance uses the delta method to estimate the analytical variance of a density or abundance estimate. According to the delta method, the squared coefficient of variation for density (D) is equal to the sum of the squared CVs for encounter rate (n/L), detection probability $f(0)$, and mean group size E(s):

$$CV(D)^2 = CV(n/L)^2 + CV(f[0])^2 + CV(E[s])^2$$

After CV(D) is calculated, then, for the 100*(1 - 2*a) CI, the lower (D_l) and upper (D_u) confidence limits for estimated density D are given by $D_l = D/C$ and $D_u = D*C$, where $C = \exp(z_a*[\ln\{1 + CV(D)^2\}])$ and where z_a is the critical z value of the Gaussian normal distribution for the "a" confidence level (Buckland et al. 2001, 2004). For example, for the 95% CI, a = 0.025 and $z_a = 1.96$.

In addition to the estimates of density (D) and abundance (N), the model reports the CV, degrees of freedom (DF), and the 95% CI statistics associated with each density and abundance estimate. In addition, the optimal model parameters (of the optimal model used in the density estimates) are reported along with associated variances. Statistics on the components of density (i.e., n/L and $f[0]$) are also reported. In addition, model output includes the percentages of the variance associated with the global density estimate that is attributed to the encounter rate (n/L), density function $f(0)$, and mean group size E(s).

3.3 DENSITY SURFACE MODELING

The CDS method provides robust estimates of abundance/density of species or groups but cannot give any information about the potential influences on those estimates. The DSM method provides additional information on distribution and abundance/density of marine species in the Study Area at a finer spatial resolution. DSM is a model-based approach in which animal abundance/density can be modeled as a function of spatially-indexed environmental covariates. This method is also known as spatial modeling or habitat modeling (Thomas et al. 2010). The key step in the first phase of DSM is partitioning the survey effort (tracklines) into segments. The DSM analysis engine in Distance utilizes the "count method" in which segment counts (sightings/detections) are modeled as a function of covariates (Hedley and Buckland 2004). The sightings within each segment are converted into an abundance estimate for each segment. The area of the segment (based on chosen segment length and the truncation distance) serves as an offset (Thomas et al. 2010). Generalized additive models (GAMs; Wood 2006) are used to estimate the spatial distribution of abundance/density or counts (the response variable) as a function of numerous geographical, physical, and environmental covariates (explanatory variables), such as longitude, latitude, water depth, distance from shore, bathymetry, SST, and surface chlorophyll concentration. After fitting GAMs to the survey data, the resulting DSM (the chosen model) is applied to a prediction grid superimposed upon the Study Area so that animal abundance/density can be predicted for any portion of the Study Area and related to specific covariates. The variance of the predicted abundance/density is

estimated using the bootstrapping resampling technique (Hedley and Buckland 2004). A brief description of these methods is included below. For more information, please see Hedley and Buckland (2004).

3.3.1 Data Preparation

3.3.1.1 Segmentation Process

The DSM analysis engine in Distance requires all tracklines to be divided into segments (Thomas et al. 2010). There is no objective way to choose the length of segments; however, they should be sufficiently small so that habitat does not vary much within the segments, and expected density is not likely to vary much within the segments (Hedley and Buckland 2004). Due to gaps in search effort along the tracklines (e.g., when the survey team would switch to off-effort mode to approach a sighting to get group size estimates), effort cannot always be split into equal segment lengths. Therefore, the size of each segment may vary.

A variety of segment lengths were assessed for each species/group analysis. We set a goal to have 15% of the segments contain sightings; this goal has been used in other marine mammal DSM analyses (e.g., DoN 2007). The segment length for each analysis was selected to minimize the number of segments with zero sightings and to minimize the variation in habitat within each segment. Due to gaps in search effort along transects, effort could not always be split into segments of the desired length. Therefore, the size of each segment varied, and the model was weighted by segment area. Most segments were around 7 km in length.

For each trackline with and without sightings, effort (transect length L) was calculated as the spatial distance between the starting and ending longitude and latitude coordinates. Efforts were calculated for all tracklines in the survey and summed over the number of tracklines to obtain a total effort (Ltot). Marine mammal sightings were summed to obtain the total number of sightings (Ntot). Total overall encounter rate ER = Ntot/Ltot, and segment length (l) was calculated as:

$$l = 0.15*Etot/Ntot = 0.15/ER$$

where the coefficient "0.15" was chosen so that approximately 15% of the segments would contain a sighting. Each trackline of length L was divided up into equal-sized segments of length "l", where the number of segments in the trackline is Nseg = L/l (i.e., ratio of transect length L to segment length l). The longitude/latitude coordinates of the midpoint of each segment in each trackline were selected based on the length of the segment, and sightings were assigned to the segment whose midpoint was closest to the location of the sightings. The static and dynamic covariates included in the model (see below) were matched to each segment based on the covariate values for the midpoint of each segment.

3.3.1.2 Selection of Covariates (Predictor Variables)

The estimated number of individual animals per segment can be related to environmental covariates by fitting a GAM (**Section 3.3.3**; Wood 2006). A variety of oceanographic and topographic variables can be included in the model as potential predictors of abundance/density; however, the covariate data must be available for the entire Study Area (i.e., not just the segmented tracklines) and the entire study period. Suitable environmental data that meet the criteria can be difficult to obtain. Biological variables, such as the distribution and abundance of prey species, are known factors that influence the distribution of marine mammals but such data are difficult to obtain over a large area (Payne et al. 1986; Kenney et al. 1996). Therefore, remotely sensed data, such as SST and surface chlorophyll *a* (chl *a*) concentrations, and static variables, such as bathymetry and distance from shore, are often the only type of covariates that are available to be included in marine mammal models (e.g., Hamazaki 2002; Cañadas et al. 2005; Ferguson 2005; Redfern et al. 2006; Paxton et al. 2009). Physical oceanographic data are often used as proxies for prey abundance which is thought to directly influence marine mammal distributions (Redfern et al. 2006).

Marine mammal distribution patterns are complex and affected by various demographic, evolutionary, ecological, habitat-related, and anthropogenic factors (Forcada 2002). Prey distribution is one of the main

influences of marine mammal distribution; marine mammals are usually found in areas with high densities of principal prey species (Payne et al. 1986; Kenney et al. 1996; Forcada 2002). Fine resolution spatial information on the distribution and abundance of prey species is often unavailable over large areas and time periods. Therefore, indirect indicators of prey distribution (SST, topography, chl *a*, etc.) are often used to study potential influences on the distribution and abundance of marine mammal species (Fiedler 2002; Ferguson 2005; Redfern et al. 2006). Important oceanographic variables that influence the distribution of prey and characterize marine mammal habitats include SST and chl *a*. In addition, ocean floor topography and bathymetry are often associated with oceanographic phenomena that influence marine mammal distribution (Forcada 2002). We chose a variety of static and dynamic habitat covariates to include in our abundance/density prediction models. Static covariates included water depth, distance from shore, slope of the seafloor, latitude, and longitude while dynamic covariates included SST and chl *a* (**Table 3-1**).

Static Covariates

Latitude, longitude, distance from shore, depth, and slope of the seafloor are static variables which may influence marine mammal distribution and abundance. There are known variations in geographic distributions based on seasonal migrations and movement patterns. For instance, North Atlantic right whales and humpback whales are known for their well-defined seasonal migratory patterns between feeding grounds off the northeast U.S. and breeding/calving grounds off the southeast U.S. (right whales) and in the Caribbean (humpback whales; Dawbin 1966; Winn et al. 1986; Clapham and Mead 1999; Kenney et al. 2001; Clapham 2009). Smaller-scale migratory movements are also evident in other cetacean species such as the bottlenose dolphin which spends the summer and fall months off New Jersey and higher latitudes and moves southward to Virginia and North Carolina during the winter and spring months (CETAP 1982; Kenney 1990; Garrison et al. 2003; Hohn and Hansen 2009; Waring et al. 2009; Toth et al. in press). Fine-scale movements of this species within the Study Area are also documented based on specific distances from shore (Toth-Brown et al. 2007). Topography (slope and depth) is also a critical factor in marine mammal distribution. Bottom topography can influence the abundance of prey; a change in depth on the shelf is often associated with higher concentrations of zooplankton. Baleen whales are known to be associated with shallow waters with high topographic variation in which prey accumulates at frontal interfaces between mixed and stratified waters (Forcada 2002). Humpback whales, for example, are known to base their foraging strategies on areas with high topographic variation (Payne et al. 1986). The static covariates included in **Table 3-1** were attached to each segment by using the covariate value that is closest to the midpoint of each segment.

Dynamic Covariates

SST and chl *a* are two types of dynamic variables known to influence marine mammal distribution and abundance (Smith et al. 1986; Baumgartner et al. 2001; Kaschner et al. 2006; Redfern et al. 2006). Several marine mammal species have temperature-limited distributions. For instance, harbor porpoises occur in sub-polar to cool-temperate waters and are seldom found in waters warmer than 17ºC (63ºF) (Read 1999). In addition, nearshore bottlenose dolphins shift their distribution in response to changes in water temperatures (Barco et al. 1999). Therefore, SST may help to predict abundance/density of certain species in the Study Area. Chl *a* concentrations may also influence marine mammal abundance/density in the Study Area. High chl *a* values are associated with upwelling centers located offshore of the Hudson-Raritan estuary, Barnegat Inlet, the Mullica River estuary, and Townsend/Hereford Inlet (Glenn et al. 2004). Primary production concentrates within upwelled waters and may attract prey species.

The dynamic covariates SST and chl *a* were evaluated for each segment by first generating 1 km by 1 km (1.9 NM by 1.9 NM) spatial grid maps of seasonal average SST and chl *a* using the same seasons as defined in **Section 2.3.1** and then comparing the longitude and latitude coordinates of each segment's midpoint with the longitude and latitude coordinates of each pixel in the gridmap corresponding to the season associated with the segment. The pixel with valid SST and chl *a* values that is in closest proximity to the segment midpoint was identified, and the seasonal average SST and chl *a* values associated with that pixel were assigned to the given segment. If no data were available for the closest pixel (due to cloud cover, etc.), then the next-closest pixels were assessed until a pixel with valid SST and chl *a* was found.

Table 3-1. Environmental covariates included in the DSM analyses.

Covariate	Description	Source
Depth	Average depth of water in meters	NOAA geophysical data system for gridded bathymetric data, National Geophysical Data Center (NOAA 1999)
Offshore Distance	Distance, in meters, from the shoreline	Calculated with the Point Distance Geoprocessing tool available in ESRI's Arc/Info® Toolbox 9.3 using NOAA bathymetric data
Slope	Slope, in degrees, of the sea floor	Calculated with the Surface Analyst function from ArcGIS® 9.3 Spatial Analyst Extension using NOAA bathymetric data
SST	Seasonal and annual averages of SST (in degrees Celsius [°C]) for the Study Area derived from remotely-sensed data from 01 January 2007 through 31 December 2009	Sensor: Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua. Resolution: 1.0 km. (NASA 2010)
chl *a*	Seasonal and annual averages of surface chl *a* concentrations (in milligrams per cubic meter [mg/m^3]) for the Study Area derived from remotely-sensed data from 01 January 2007 through 31 December 2009	Sensor: Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua. Resolution: 1.0 km. (NASA 2010)
Latitude	Latitude in decimal degrees	
Longitude	Longitude in decimal degrees	

3.3.1.3 Construction of Prediction Grid

After fitting GAMs to the survey data (see **Section 3.3.3**), the resulting DSM was applied to a prediction grid superimposed upon the Study Area. Therefore, animal abundance/density could be predicted for the entire Study Area for each season of interest. To construct the prediction grid, a spatial grid of 1-km by 1-km (1.9-NM by 1.9-NM) cells was created using ArcGIS® and overlaid onto the Study Area. The cells were evenly distributed throughout the Study Area, and a point shapefile of the grid was generated using ArcGIS®. This point file was comprised of 5,000 points which were the centroids of the 1-km by 1-km (1.9-NM by 1.9-NM) grid cells. The centroids of each cell were matched to their corresponding latitude and longitude and their values for the following static covariates: water depth, distance from shore, and the slope of the sea floor (**Table 3-1**).

The dynamic covariates SST and chl *a* were generated for each for the 5,000 centroids in the prediction grid by comparing the longitude and latitude coordinates of the given centroid with the longitude and latitude coordinates of each pixel in the gridmap corresponding to the season associated with the prediction grid. The pixel with valid SST and chl *a* values that is in closest proximity to the centroid was identified, and the seasonal average SST and chl *a* values associated with that pixel were assigned to the given centroid. If no data were available for the closest pixel (due to cloud cover, etc.), then the next closest pixels were assessed until a pixel with valid SST and chl *a* was found.

Separate prediction grids were developed for each seasonal analysis of abundance/density of the species or groups. The values for the static covariates remained the same for each prediction grid. The values for

the dynamic covariates (SST and chl *a*) were averaged across each grid cell for the season in question. A total of four prediction grids were developed, one for each of the following seasons: year-round (SST and chl *a* averaged across all seasons), winter (SST and chl *a* averaged across only the winter season), spring (SST and chl *a* averaged across only the spring season), and summer (SST and chl *a* averaged across only the summer season). Note that there were not enough sightings data to model marine mammal abundance/density for the fall season alone.

3.3.2 Fitting a Density Surface

The estimated probabilities of detection were obtained from the fitted models for the detection functions chosen from the CDS analyses (see **Section 3.2**) using the Mark-Recapture Distance Sampling (MRDS) engine in Distance. When using a GAM to model the relationship between the response variable and various covariates, it is ideally desirable to detect all objects (animals) in the segment; however, this is rarely the case, as reflected in monotonically decreasing probability detection functions. Two methods are available to account for objects not detected in a segment: The first method involves estimating the total number (as opposed to the detected number) of objects in the segment, n_i, via the Horvitz-Thompson estimator: n_i = SUM ($1/p_{ij}$), where p_{ij} = probability of detection of object j in segment i, and the summation is conducted from j = 1 to n_i. This method is useful if the detectability is different for objects within the same segment. The second method involves decreasing the segment area (a_i) to reflect the effective area surveyed rather than the covered area (count), using the effective strip half-width μ (or ESW = 1/f[0]): a_i = 2 μ_i l_i, where l_i = length of segment i. The μ value is defined as the sighting distance such that the number of undetected animals at distances less than μ is equal to the number of detected animals at distances greater than μ. After using these two methods to account for undetected objects in a segment in the n_i and a_i terms, density in segment i is then calculated as $D_i = n_i/a_i$.

In the CDS/MRDS methods, the statistical criterion for model selection (i.e., probability detection function) is AIC minimization. In the subsequent DSM analysis, the criteria for selection of the optimal GAM model is (among other criteria) GCV/UBRE minimization (GCV = generalized cross-validation; UBRE = unbiased risk estimator). To fit a GAM to the observed data, we specified the following information: 1) the explanatory variables (covariates) to include in the model (see **Section 3.3.1.2**); 2) the dimension of the smooth functions (univariate which includes one covariate versus bivariate which includes two covariates); 3) the degree of smoothness of the functions (controlled by the number of knots (k): DF = # knots-1; 4) error distribution (quasipoisson); and 5) the logarithmic link function. A small number of knots increases smoothness while suppressing the expression of small-scale variability; this is desired if the function exhibits sharp gradients (i.e., high sensitivity of the response variable to changes in the given covariate) over small scales. Conversely, a large number of knots decreases smoothness while enhancing small-scale variability; this is desired if the functional dependence of the response variable on the given covariate exhibits very low sensitivity. We chose to limit the number of knots used in the analyses to k=7 for univariate smooth functions and k=14 for bivariate smooth functions in order to allow moderate flexibility while reducing the likelihood of fitting unnecessarily complicated functions.

Identification of the "optimal" GAM was aided with the following information and model output: 1) minimization of the GCV/UBRE score; 2) maximization of the % of deviance explained by the model; 3) inspection of the diagnostic plots of the residuals (e.g., normal Q-Q plot, residuals versus linear predictor, frequency histogram of residuals, response versus the fitted values); 4) inspection of the plots of smooth functions (increase or decrease the maximum number of knots, include as a linear term, etc.); 5) assessment of the response surface summary (sensitivity of the density surface/abundance to different models); and 6) assessment of the significance of the covariates in the GAM.

Different GAMs that incorporate various combinations of smooth functions of covariates were tested and compared to each other using the above criteria for ideal model selection. The total number of different combinations of covariates is quite large. In addition to univariate (1-dimensional) functions of the individual covariates, bivariate (2-dimensional) functions were applied to various pairwise combinations of covariates (e.g., longitude and latitude, depth and offshore distance, SST and chl *a*). Generally, for N covariates there are N(N-1)/2 pairwise combinations (i.e., 21 pairs for the total seven covariates). It was not necessary to test every possible combination of smooth functions of covariates. As GAMs were

formulated on a trial-and-error basis, we were able to discern which covariates were more significant than others. The decision to include a given covariate in the model was made based on a tradeoff between model fit (using the above statistical criteria characterizing an "optimal" model) and model complexity. A given covariate was excluded from the GAM if: 1) the estimated DF for the covariate were close to 1; 2) the plotted confidence band for the covariate included zero everywhere; and 3) the GCV/UBRE score decreased when the covariate was omitted from the GAM. After the excluded covariates were identified, a (significantly smaller) list of potential GAMs that include combinations of the remaining covariates were developed, while the combinations involving the excluded covariate(s) were eliminated from further consideration. From this restricted list of potential GAMs, optimal model selection was based on the above statistical criteria (e.g., GCV/UBRE minimization, maximization of % deviance explained, etc.).

3.3.3 Predictions of Density and Abundance

In the DSM analysis, GAM models were developed and an optimal model was chosen based on numerous selection criteria. This optimal GAM was chosen as the best fit to the observations of the response variable (density, abundance) as a function of smooth functions of the various covariates at the available sampled sites, and was used to generate predictions of density and abundance at unsampled sites (i.e., sites where estimates of the covariates are available but where the response variable has not been observed or measured) on a prediction grid that encompasses the entire Study Area.

Caution should be exercised when extrapolating model predictions from regions with observational data to regions far removed from observational data, particularly in situations where sharp spatial gradients in density/abundance and covariates occur. It is probable that the GAM will be applied to regions within the Study Area that are not sufficiently surveyed (i.e., areas with little or no survey effort). In this case, it is imperative that covariate data be collected at these unsampled sites (rather than interpolated from sampled sites) if possible, so that the GAM can be adequately applied to obtain predictions (estimates) of density and abundance. For example, the GIS database stores an abundance of data on static covariates (depth, offshore distance, bathymetric slope) at every conceivable offshore longitude and latitude location. Given the availability and time-invariance of these covariate data, it is more accurate to obtain values of these covariates at the exact locations of the unsampled sites (rather than interpolating from values at sampled sites) and using these exact values in the GAM to generate estimates of abundance and density at these unsampled sites. Using this procedure of covariate data collection at unsampled sites (i.e., every grid cell in the prediction grid), the GAM is applied to estimate density and abundance and extrapolate these predictions to each grid cell, thus generating a density surface (spatial map of density) covering the entire Study Area.

Generally, the accuracy and validity of predictions (of abundance or density) in regions with no observational data (or in regions far removed from observations) depends on model robustness and reliability (model-based analysis) and on the availability of measurements of covariates that are included in the model. At the smallest spatial scale (i.e., within each cell of the prediction grid), the GAM is used to estimate density. Estimated density in each cell is calculated as the ratio of estimated abundance to the cell area. In specified regions of larger spatial scale (i.e., containing several cells of the prediction grid), abundance and density are estimated by density surface integration in which the predicted abundances of all cells in the given region are summed, and density is estimated as the ratio of the summed abundances to the summed areas of all cells in the given region.

3.3.4 Variance Estimation

The variance associated with the prediction grid estimates of density and abundance was estimated using bootstrapping, a technique involving random resampling with replacement (Efron and Tibshirani 1993). Bootstrapping is advantageous in that it is a robust method of variance estimation when variance cannot be calculated analytically. A large number of bootstrap samples are typically generated (to ensure an adequate sample size). The minimum number of resamples should be no less than 200, and 400 to 1,000 resamples are preferred to generate reliable confidence intervals (Buckland et al. 2001). Abundance is estimated from each bootstrap estimate, and these bootstrap abundance estimates are ranked from highest to lowest. The mean of these bootstrap estimates is calculated, and the 95% CI is calculated such

that it is bounded by the 2.5% quantile and the 97.5% quantile. Because of this nonparametric measure of uncertainty, the bootstrapping method is not affected by a few extreme outliers.

Different types of bootstrapping methods include nonparametric, parametric, and moving block. The nonparametric method requires no distributional assumptions, whereas the parametric and moving block methods are based on a fitted model (GAM) that incorporates some distributional assumptions and estimated model parameters. The Distance software is currently able to run only the parametric moving block bootstrap in its variance estimation method. The following is a discussion of the advantages and disadvantages of each method, leading to justification of the choice of method used in Distance.

Nonparametric bootstrapping involves random resampling with replacement of some independent sampling unit whose spatial/temporal scale is sufficiently large that autocorrelation between adjacent sampling units is negligible. Sampling units should be numerous, with a sufficiently fine spatial/temporal scale to capture small-scale variability in the data structure; however, adjacent sampling units should also be spatially/temporally independent of each other, and too fine a scale poses the risk of significant autocorrelation between adjacent sampling units since the degree of correlation generally increases with a decrease in spatial/temporal scale (e.g., decrease in separation distance between adjacent units on a spatial scale or decrease in time difference on a temporal scale). Thus, the sampling units should be constructed on a scale sufficiently fine as to be numerous while also sufficiently coarse as to be independent from each other. The transect (which is spatially finer than the Study Area and region levels and coarser than the segment level) is typically chosen as the independent sampling unit since its data structure is both sufficiently fine to be numerous while also sufficient coarse to be independent. The segment data structure is numerous (since transects are divided up into segments) but may not be independent since its relatively smaller spatial scale renders it susceptible to spatial autocorrelation between adjacent segments (e.g., positively correlated objects in adjacent segments). The data structure at the larger levels of Study Area and region are not sufficiently numerous due to their relatively coarser spatial scales. Nonparametric bootstrapping is advantageous in that it preserves spatial correlation, but it does not preserve spatial coverage and can lead to extreme bootstrap abundance estimates.

Parametric bootstrapping uses a model (e.g., a GAM in the DSM analysis) fitted to the observed data to generate new data values which are then used to generate the bootstrap sample. A GAM uses smooth functions (with model parameters) relating the response variable (i.e., abundance or density) to a number of covariates or explanatory variables (e.g., longitude, latitude, depth, offshore distance, bathymetric slope, SST, chl *a*). The residuals (defined as the difference between the observed value and model-estimated value of the response variable) are selected randomly and with replacement in parametric bootstrapping. Whereas nonparametric bootstrapping preserves spatial correlation but not spatial coverage, parametric bootstrapping preserves spatial coverage but not spatial correlation.

In seeking to address the shortcomings of the nonparametric and parametric methods, the moving block method (which is the method of choice in Distance) preserves both spatial correlation and spatial coverage. This method uses a moving block comprised of a number of sampling units (e.g., segments). Block size m (number of segments in a block) should be sufficiently large so that segments more than m units apart (i.e., in different blocks) are independent (i.e., no spatial correlation between blocks), yet also sufficiently small to retain spatial correlation and structure among the segments within a given block (i.e., spatial correlation within blocks). Information on optimal block size can be obtained from a semivariogram of residuals. Semivariance between a pair of points increases (i.e., autocorrelation decreases) asymptotically with increasing separation distance, reflecting decreased similarity until the points become independent (spatially uncorrelated) at a sufficiently large separation distance.

The moving block is selected randomly and with replacement and is then randomly placed back together to generate the bootstrap sample. The original response variable values cannot be moved since they are connected to spatial location and other explanatory variables; however, the residuals can be moved, thus generating bootstrap samples via random resampling with replacement using a moving block as the sampling unit. Then, many bootstrap samples are randomly generated, a mean value of these samples is calculated, and the 95% CI is estimated to obtain the variance estimate associated with the prediction grid estimate of the response variable (abundance).

Due to its inherent advantages and ease of application, the parametric moving block method is currently the method of choice in Distance for variance estimation. Required user-specified parameters include block size m (typically 3), number of bootstraps (10, 99, 199, 499, or 999), confidence interval desired (0.95, 0.90, 0.85, or 0.80) and inter-quartile range for outlier detection (1.5, 2.0, 2.5, or 3.0). Effects of outliers on variance estimation is generally insignificant, especially if a large number of bootstraps are used, since the relatively rare occurrences of anomalously low and high values will be concentrated in the lower and upper tailings, respectively, which are cut off at the 2.5 and 97.5 quantiles in the estimation of the 95% CI. To balance the tradeoff between spatial detail and time constraints, 499 bootstraps is typically optimal. Using a larger number of bootstraps requires more computation time, whereas using fewer bootstraps runs the risk of an inadequate sample size and renders the method increasingly susceptible to outliers. We chose a block size of 2 or 3 and desired confidence interval of 95% and ran 999 bootstraps for each of our DSM analyses.

This page intentionally left blank