Predictive Habitat Distribution Modeling of Sperm Whale (*Physeter macrocephalus)* within the Central Gulf of Alaska utilizing Passive Acoustic Monitoring

by

Nathan Patrick Novak

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

May 2016

DEDICATION

I dedicate this document to my late grandfather Dr. Gerhard Novak, who provided continuing inspiration and revelation for my most challenging endeavors. His pursuit of knowledge and scientific advancement proved that there are no limits in what one man can achieve.

ACKNOWLEDGMENTS

I am grateful to my mentors, SSI Director and Professor John Wilson and Dr. Tina Yack of Bio-Waves, Inc. Their guidance contributed highly to the development of this manuscript. I would like to acknowledge NAVFAC-Atlantic for funding the survey, and HDR Inc., for coordinating project logistics. With the support from my family and friends, particularly my father Walter Nowak, I have been able to maintain sharp focus and steady motivation. I appreciate all of the assistance you all have provided me as I persevered through the many challenges that this project entailed. Although rewarding and enjoyable, this project never would have made it to its final stages without the many amazing people behind me. Thank you to my family friend Tom Norris, who granted me access to field data, documentation, and expertise required for this spatial and statistical analysis. I would also like to thank Kurt Nelson and Anthony Rodgers at Orca Maritime, for providing me with time off work. With the tremendous help of Martha Rodgers, I have been able to advance my skills in producing underwater GIS. A special thanks to Michelle Kinzel, as she has provided many of the important remote sensing resources I have utilized throughout the duration of the various spatial analyses. It has been wonderful to work with my thesis committee members, Dr. Jennifer Swift and Dr. Daniel Warshawsky, as they have been invaluable in providing feedback for the final revisions. With the culmination of subject matter experts in marine ecology and GIS, I have been allowed a truly unique opportunity to study these fascinating marine mammals utilizing cutting-edge computing technology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC	Akaike's Information Criterion
ANOVA	Analysis of Variance
ASP	Aspect
BTH	Distance to 2000 m isobath
CSV	Comma Separated Value
DAAC	Distributed Active Archive Center
DEM	Digital Elevation Model
DEP	Depth
DF	Degrees of Freedom
DIR	Current direction
GAM	Generalized Additive Model
GLM	Generalized Linear Model
GIS	Geographic Information System
GIST	Geographic Information Science and Technology
GOA	Gulf of Alaska
GOALS-II	Gulf of Alaska Line-transect Survey-II
GP	Geoprocessing
HYCOM	Hybrid Coordinate Ocean Model
IUCN	International Union for Conservation of Nature
MATLAB	Matrix Laboratory
MGET	Marine Geospatial Ecology Tools
MNP	Moonphase

MODIS	Moderate Resolution Imaging Spectroradiometer
MS	Microsoft
NAD	North American Datum
NASA	National Aeronautical and Space Agency
NOAA	National Oceanographic and Atmospheric Administration
PAM	Passive Acoustic Monitoring
PSU	Practical Salinity Units
OSCAR	Ocean Surface Current Analyses - Realtime
SC	Surface Chlorophyll-a
SLP	Slope
SMI	Standard Mapped Image
SSI	Spatial Sciences Institute
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
TKE	Total Kinetic Energy
TMAA	Temporary Maritime Activities Area
USC	University of Southern California

ABSTRACT

This research involves a habitat model of sperm whale (*Physeter macrocephalus)* distribution utilizing Passive Acoustic Monitoring (PAM) within the central Gulf of Alaska. The main goal of this project was to explore the relationship between distinct occurrences of sperm whale vocalizations and environmental variables within a 144,560 km^2 Temporary Maritime Activities Area (TMAA) during the Summer of 2013 (Rone et al., 2014). A total of 6,304 km of trackline was utilized to produce 426 hours of 'standard' real-time monitoring to detect vocally active cetaceans. Acoustic activity, along with nine static geophysical and dynamic oceanographic variables were used to produce empirical statistical models in order to express correlations and spatially represent their probable habitat range. The application of customized GIS-based components has allowed the performance of iterative geoprocessing, and a precision-based spatial approach to habitat distribution modeling. Various Generalized Additive Models (GAMs) were developed with discrete trackline acoustic encounters and combinations of habitat variables to offer a comparison of encounter rate differences across the study area, as well as demonstrate the habitat variables' ability to predict sperm whale presence. Modeling efforts indicated that the most important explanatory variables for sperm whale habitat within the spatial and temporal scale of this study were depth, slope, proximity to the 2,000 m isobath, sea surface temperature, chlorophyll-a concentration, and magnitude of oceanic currents. This work demonstrated that acoustically detected sperm whales found within the central Gulf of Alaska follow predictable foraging patterns and demonstrate consistent preferences for specific oceanographic conditions.

CHAPTER 1 - INTRODUCTION

Sperm whales (*Physeter macrocephalus)* are deep-diving toothed marine mammals that communicate and forage using echolocation click vocalizations. During the early eighteenth century through the late twentieth century, this vulnerable cetacean remained a prime target of whalers. Their bulbous head contains a liquid wax called spermaceti, which was frequently sought after for use in lubricants, candles, and oil lamps. As a result of extensive whaling, population sizes decreased until the recent whaling moratorium, which now protects them from global exploitation (Barlow and Taylor, 2005). Recent knowledge has proved to be limited; there is a need to more fully comprehend population trends and conditional preferences.

1.1 Habitat Modeling

Habitat modeling has recently been adopted by ecologists as a tool to quantify species-environment relationships. In this thesis project, an examination of northeast Pacific sperm whale abundance within the Gulf of Alaska will provide valuable information to improve marine mammal (cetacean) conservation and sustainable habitat management. This is especially pertinent because the International Union for Conservation of Nature (IUCN) Red List defines *P. macrocephalus'* conservation status as Vulnerable A1d. Current data is available on the abundance and distribution of the species, although current modelling methods and the pooling of data have obscured geographic migration patterns (Notarbartolo di Sciara et al., 2014).

This thesis addresses the efforts being made to develop estimates for regional habitats with insufficient sampling in the central Gulf of Alaska. There is limited amount of global data on sperm whale abundances, although study estimates for currently sampled areas infer a worldwide population of about 100,000 individuals (Notarbartolo di Sciara et al., 2014).

Advanced population studies are fundamental so that vulnerable populations are moved to more protected designations. It is likely that numerous cetaceans are eligible for a threatened category and steps such as habitat modeling and analyses should proceed to assess their conservation status (Notarbartolo di Sciara et al., 2014). Although cetacean habitat modeling is still a relatively new tool for cetacean investigations, it has proven to be highly powerful and flexible for the purpose of understanding cetacean-habitat relationships (Redfern et al., 2006).

Surveys on population structure, abundance and life history are needed for most regions. Habitat deterioration and over-exploitation exert significant pressure on *P. macrocephalus*, especially as pods can be small and highly specialized. Combined effects of pollutants, as well as depletion of prey species are expected to potentially reduce global populations by up to 30% over the next three generations of sperm whales (Notarbartolo di Sciara et al., 2014). The proposed habitat model for data collection performed may produce a greater knowledge-based understanding of northeast Pacific sperm whale abundances and distributions in the nutrient-rich region.

The Gulf of Alaska Line-transect Survey II (GOALS-II) acquired data to analyze the abundance, density, and spatial distribution of marine mammals within the TMAA in 2013, employing visual line-transect and bioacoustic surveys using a towed-hydrophone array and sonobuoys (Rone et al., 2014). This combination of methods aims to enhance knowledge of the vocal repertoire of cetaceans by relating visual sightings to sounds recorded from vocally active whales. Sperm whales use underwater acoustics in the ocean environment for such purposes as navigation, communication, locating prey, mating and courtship. In addition, past research also links certain military-related sonar activity in TMAAs to disturbances of sperm whales and other marine mammals (Batchelor and D'Spain, 2005). For these reasons, the development of a sperm whale habitat model using acoustic monitoring provides further spatial insight into their vocal habits during these behavioral activities.

This project represents the first biogeographic study to investigate the relationships between environmental variables and sperm whale acoustic encounters to predict habitat use areas. These applications are particularly efficient because they do not rely on ineffective visual line-transect surveys to confirm a marine mammal's presence (Thompson, Brooks, and Cordes et al. 2015). Although previous studies have investigated cetacean occurrences in correlation with environmental variables (see Burtenshaw et al., 2004; Bush, 2006; Forney et al., 2012), none have primarily focused on using these methods to predict suitable sperm whale habitat. The project exhibits a unique combination of methods in the application of both PAM-based data collection techniques and spatial analysis using the Marine Geospatial Ecology Toolset (MGET). Cetacean habitat modeling has been performed to determine spatial patterns of occurrence; however, the potential of this technique has not yet been fully explored.

1.2 Thesis Organization

The remainder of this thesis is organized into four chapters: Chapter 2 - Background and Literature Review; Chapter 3 – Methods; Chapter 4 – Results; and Chapter 5 – Discussion and Conclusions. The next chapter will deliver a general description of key topics and techniques involved in this study. The third chapter details methods and applications used to collect field data, process raw data, and develop models. Chapter 4, Results consists of a report detailing various aspects of the final outputs with figures. Descriptive statistics of each environmental covariate used in the model are also found in this chapter. Chapter 5, Discussion and Conclusions reports the project findings and how they influence existing research on developing effective predictive habitat distribution models.

CHAPTER 2 - BACKGROUND AND LITERATURE REVIEW

This chapter reviews literature related to the underlying concepts behind bioacoustic systems, habitat models, data collection and habitat data. The goal of this section is to provide background information that supports an understanding of processes used within this study. Comprehending the dynamics of a multi-faceted habitat modeling approach requires a wide range of practical insight into the mechanics of bioacoustics and habitat modeling.

2.1 Passive Acoustic Monitoring

Cetaceans use sound for sensing their surroundings. Most marine mammals are considered acoustic specialists that rely on sounds for behavioral and navigational purposes. Scientists and engineers have successfully developed acoustic-based technologies to record underwater sounds produced by marine mammals. The use of Passive Acoustic Monitoring (PAM) for the detection of marine mammal vocalizations is an emerging tool for characterizing the presence and geographic range of cetaceans in a variety of pelagic environments. Incorporating PAM recordings into a marine GIS, gives spatial scientists applicable information for comprehending ocean habitat characteristics and describing acoustic interactions between marine mammals and their environment (Moore et al., 2006).

Acoustic sensors typically consist of multiple underwater acoustic sensors, or hydrophones. These acoustic sensors are combined with advanced signal and array processing methods, allowing studies which demonstrate the locations of specific sound sources as a function of time and recording position. Acoustic data can be utilized for a variety of purposes, including habitat classification, behavioral studies, and identifying areas of anthropogenic sources of ocean noise (Batchelor and D’Spain 2006).

Hydrophones are rated by frequency, with mid- and high-frequency instruments detecting a majority of vocalizing cetaceans. A primary signal processing and recording system was utilized for the duration of the survey, solely for the detection of sperm whales. It comprised a mid-frequency system, which was used in the offshore, seamount, and slope strata. This system was targeted mainly toward sperm, killer, and beaked whales within these regions.

The strength of PAM-based techniques lies in its ability to detect vocalizing odontocete and mysticete individuals using reliable and non-invasive methods. With the onset of satellite telemetry tracking tags, invasive population studies have recently become more common (Rone et al., 2014). The use of PAM has provided a more sustainable and less destructive method of studying spatial population dynamics of cetaceans. In comparison with more common visual surveys, PAM-integrated surveys allow for effort during adverse weather and lighting which would otherwise be preventative. (Thompson et al., 2010; Rayment et al., 2011; Teilmann and Carstensen, 2012).

Typical cetacean population surveys rely primarily on visual-based detection techniques. The application of PAM-based technologies provides researchers with a higher detection rate, larger spatial range, and fewer weather restrictions than standard visual-based surveys. Environmental variables that must be taken into account during visual surveys include sea state, swell height, precipitation, glare, and visibility (Munger et al., 2009). During a sighting, observers are required to record subjective information such as best, high, and low estimates of group size (Sousa-Lima et al., 2013). Acoustic systems present advantages in the form of functioning independently of human operators, therefore reducing human-error in the field (Mellinger and Barlow, 2003). The use of PAM-based methods such as towed-hydrophone arrays represents a prevailing alternative over antiquated visually-reliant surveys.

Although ship-based PAM surveys can be effective, they are more expensive when compared with visual-based cetacean detecting methods (D'Spain, 2013). It is not uncommon for PAM applications to have a very high cost per detection (Mellinger and Barlow, 2003). Increased expenses may be attributed to additional hardware, software and personnel training. These tradeoffs must be considered when deciding which approach is most cost-effective (Sousa-Lima et al., 2013).

2.2 Cetacean Habitat Modeling

A central issue in ecology has been centered on the species-environment relationship. Some of the earliest of these studies of this type have used climate-related anomalies to explain animal and plant distribution (von Humboldt and Bonpland, 1807; de Candolle, 1855). In addition, many researchers have strived to take into account a variety of environmental factors to describe influences in main vegetation patters around the world (e.g. Salisbury, 1926; Cain, 1944; Good, 1953; Holdridge, 1967; MacArthur, 1972; Box, 1981; Stott, 1981; Woodward, 1987; Ellenberg, 1988). Such species-environment relationships have been quantified to represent the core of predictive geographical modeling in ecology. These models conceptually employ various hypotheses to understand how environmental factors control the distribution of species and communities (Yack, 2013).

Habitat modeling is a process of accurately describing and understanding the spatial and temporal presence of organisms and their relationship to habitat properties (Redfern et al., 2006). With the arrival of effective methods in statistics and computing power, a comprehensive series of quantitative procedures can be used to perform complex and iterative calculations that integrate non-linear relationships and a multitude of explanatory variables (Diaconis and Efron

1983; Efron and Tibshirani, 1991; Guisan and Zimmermann, 2000; Manly, 2006). An appropriate marine species habitat model includes carefully-selected habitat variables that define physical processes and biological responses such as upwelling currents and the mechanisms associated with planktonic transport (Redfern et al., 2006). Although cetacean habitat modeling is still a relatively new tool for cetacean investigations, it has proven to be highly powerful and flexible for the purpose of understanding cetacean-habitat relationships.

A variety of determinants may go into the development of the appropriate model. Primary factors that should be contemplated include cetacean-habitat model purpose, scale, data considerations, and modeling techniques (Barlow and Taylor 2005). Defining the true purpose of a model is one of the most critical steps in the modeling process. A model's purpose is largely influenced by how much is currently understood about the ecology of the species. When limited biogeographic research has been performed on a specific species, models are designed to explore relationships between cetacean distributions and other features of the study area (Redfern et al., 2006). An underlying understanding of dominant oceanographic features within the ecological area of interest may be utilized in selecting habitat variables for the analysis.

Very few cetaceans have been studied using sufficient modeling efforts that would give specific hypotheses of ecological processes determining distributions; therefore, it is crucial to first construct models that describe associations between living resources and oceanographic variables (Barlow and Taylor 2005). As more is known about a species' distribution, the model may shift from merely observing cetacean-habitat correlation to predicting cetacean distribution patterns. Descriptive statistical techniques are applied in this stage during an iterative process in which every successive observation aids to refine the model and advance long-term predictive capabilities (Redfern et al., 2006).

2.3 Scale

One of the most important selection criteria of a cetacean-habitat model is spatial and temporal scale. The final model output will depend on the influences of scale at which the data are collected and studied (Wiens, 1989). In order to understand cetacean-habitat relationships, knowledge of prey distribution and seasonality must be examined. Cetaceans are apex marine predators which primarily feed on small pelagic schooling fish, crustaceans and plankton. The scale of the study area should be directly related to prey abundance, as well as any other factors that may contribute to areas rich in biomass. Aggregating prey species rely on many components, many times reflecting water masses and currents, with relationship to spawning and feeding distributions. Scaled models developed using known prey densities can examine environmental variables using surface data, water column data, and oceanographic features. Fine-scale models describe localized knowledge of marine species, while larger-scale models may generate hypotheses about global cetacean population structure and distribution (Redfern et al., 2006).

2.4 Data Collection

A reliable and accurate model relies on the methods of cetacean and habitat data collection. There are numerous methods of collecting cetacean data including ship, aerial, and acoustic survey techniques. In order to ensure equal sampling probabilities within the area of interest, line-transect sampling methods are typically followed to make quantitative approximations. Although designed to address proper sampling theory within the region, logistics such as fueling limitations and weather are likely to affect the actual track line path. If bathymetric strata types are varied within the study area, transects are allocated among strata with respect to expected cetacean densities or total size of the strata surveyed areas (Redfern et al., 2006).

During both ship and aerial surveys, visual detections of animals are missed due to two types of bias. Animals are potentially missed due to perception bias (animals are at surface but are visually undetected), as well as availability bias (animals are submerged) (Marsh and Sinclair, 1989). Many times perception bias is caused by factors associated with animal behavior and group sizes, in addition to survey conditions such as swell height and visibility (Barlow et al., 2001). Availability bias is mostly affected by dive durations and relative time spent surfacing (Redfern et al., 2006). Common visual detection methods are susceptible to missing animals due to both of these biases. Acoustic methods such as towed hydrophone arrays on ships are effective techniques to reduce perception or availability bias by increasing the range of detection and allowing nighttime surveys (Thompson et al., 2015).

2.5 Habitat Data

Habitat data utilized to model cetacean distributions may derive from a variety of sources including *in-situ* data, remote sensing data, and bathymetric data. Many ship-based surveys measure *in-situ* data in which habitat variables are measured along the cruise path to describe water column properties, surface water conditions and ecological dynamics such as densities of prey, and predator species. Properties taken from the water column often include total chlorophyll concentration within the euphotic zone depth and strength of the thermocline, as well euphotic zone and mixed layer depths. Surface condition measurements include temperature, salinity, chlorophyll-a, dissolved oxygen content, water transparency and coloration. It must be noted that a majority of physical oceanographic data represent proxies for prey abundance or availability, and are imperative because these values may indirectly influence cetacean presence (Redfern et al., 2006).

Remotely sensed data comprises satellite-derived images, with the ability to provide such dynamic information as sea surface temperature, chlorophyll-a, sea surface salinity, and currents. Bathymetric information is accessible for most regions, including the Gulf of Alaska, providing the capability to include topographic features within the cetacean-habitat model. Bathymetric datasets allow model variables such as bottom depth, bottom slope, and distance to fixed depth contour lines. Statistically significant relationships exist between bathymetric variables and cetacean population distributions including bottlenose dolphin *Tursiops truncates* ecotypes in the northwest Atlantic (Torres et al., 2003), northern bottlenose whales *Hyperoodon ampullatus* in Nova Scotia (Hooker et al., 2002), as well as harbor porpoises *Phocoena phocoena* in northern California (Carretta et al., 2001). Although high spatial resolution has been a focus of many multi- and hyper-spectral passive sensors, temporal resolution may often times be restrictive. The finest temporal resolution possible is often daily or more, and may contain temporal lags between cetacean and habitat data collection (Redfern et al., 2006).

Static geophysical properties such as depth, slope, and aspect have been found to be important predictors of whale distribution in other regions, particularly the Bahamas and North Atlantic (MacLeod and Zuur, 2005; MacLeod, 2000). Cephalopods are a basis of the sperm whale diet and it has been suggested that cephalopod species may be associated with seamounts and other steeply sloping ocean features because they are carried on slopes by oceanic currents (Nesis, 1993; Clarke and Pascoe, 1997).

2.6 Satellite-based Remote Sensing Background

Environmental variable data for input into associated models are extracted through a customized geospatial module utilizing imagery collected via satellite remote sensing. In order to gain a greater notion of the process in which these environmental data are acquired, a broad background of a typical satellite-based sensor and associated oceanographic model is described in this section.

2.6.1 MODIS

Acquisition of environmental variables such as sea surface temperature and chlorophyll concentration data is made possible via MODIS (or Moderate Resolution Imaging Spectroradiometer). This sensor is a main instrument aboard the Aqua (originally known as EOS PM-1) and Terra (originally known as EOS AM-1) satellites, both operated and managed by NASA. Aqua's asynchronous orbit around the Earth is precisely timed so that it passes from south to north across the equator in the afternoon, while Terra passes north to south over the equator in the morning. Aqua and Terra MODIS provide swaths of the entire Earth's surface every 24 to 48 hours, acquiring a multitude of data in 36 spectral bands, or ranges of wavelengths. These data have improved global understanding of ongoing dynamics and processes occurring on land, oceans, and lower portions of the atmosphere. The application of MODIS in this study and many others is contributing to the development of ecosystem models. Furthermore, these sensors have produced data allowing for the prediction of global patterns. Observations of trends have been used for the provision of evidence-based detection and more informed policy (Maccharone, 2015).

In addition to using short- and long-wave bands to determine the sea surface temperature for the top one millimeter of the ocean surface, MODIS uses near infrared bands to provide a useful estimate of live phytoplankton biomass, or green pigment chlorophyll-a within the top portion of the surface waters. This is gained by measuring reflectance, representing a proxy for water attenuation, and ultimately, total primary productivity in euphotic zones (Feldman, 2008). Indeed, MODIS-derived data provides an accurate source of remotely-sensed variables for the purpose of quantifying species-habitat relationships.

2.6.2 HYCOM

Accurate temporal and spatial dynamics of geophysical properties such as sea surface salinity and currents must be understood to create species-habitat predictive models. Therefore, a repertoire of data-driven models have been accessed for the purpose of relating behavioral interactions. Data representing values extracted from the study area during the survey were available from the HYCOM (HYbrid Coordinate Ocean Model). HYCOM is a database representing three-dimensional models depicting the ocean state at fine resolution in real time, providing the basis for coastal and regional models. HYCOM is a multi-institutional effort to develop and evaluate a variety of data-assimilative hybrid isopycnal-sigma-pressure (generalized) coordinate ocean models in pursuit of maintaining a global coupled ocean-atmosphere prediction model available to the international research community (Chassignet, 2008).

2.6.3 OSCAR

The incorporation of ocean surface velocity calculations was required for the multi-variable modeling efforts, necessitating ocean current magnitude and direction data. Ocean Surface Current Analyses Real-time (OSCAR) provides near-surface ocean current estimates, which are derived from quasi-linear and steady flow momentum equations. The model computes horizontal current velocity with data comprising sea surface height, surface vector wind and sea surface temperature. Data sources include various satellites and *in situ* instruments. The model incorporates a quasi-steady model, which combines geostrophic, Ekman, and Stommel shear dynamics with a complementary term from the surface buoyancy gradient. Data are on a 1/3 degree global grid with 5 day resolution averaged over the top 30 m of the upper ocean (Dohan, 2012).

2.7 Modeling Techniques

There are many important considerations when determining the best modeling techniques for quantifying large-scale associations between cetacean distributions and habitat variables. Usually overlays of sightings and maps of habitat variables are used to produce subjective outlines of species ranges, presenting substantial variation. There are three main types of techniques that will generate objective and reproducible results with transparent and customizable assumptions. The first is environmental envelope modeling, where minimum and maximum values of the habitat variables are calculated to define an envelope, encompassing a pre-determined percentage of the cetacean occurrences. Environmental envelope models are ideal because they do not require large sample sizes and may be applied to datasets with no associated information of effort-status. The simple conceptual framework of environmental envelope models allows for the testing of ecological hypotheses but further interpolation to fine spatial scales may obscure important cetacean-habitat relationships. These models perform best for answering broad questions about large-scale species abundance and distributions (Redfern et al., 2006).

Another commonly used technique to model the relationship between cetacean data and habitat data is the use of regression models. Regression comprises an assortment of distinct modeling techniques that differ in their assumptions about the distribution of habitat variables and the functional form of the relationship. One of the simple techniques is linear regression, relating variability in observed values to a sum of linear functions of predictor variables. Linear models typically produce models that are relatively simple in their application and understanding. Another type of more sophisticated regression model may be used to deal with discrete response variables with non-normal error distributions. In this case, generalized linear models (GLMs) may be appropriate as they incorporate a link function to induce linearity between response and predictor variables, use non-constant variances directly into the analysis, with the final response constrained within a specific range (i.e. binary response such as 0 to 1). A weakness of both linear regression and GLM is their assumption that the relationship between the response variable and the predictor variables is parametric (i.e. linear or quadratic relationship), which is potentially an unrealistic assumption for cetacean-habitat relationships (Redfern et al., 2006).

Generalized additive models (GAMs) are non-parametric extensions of GLMs, replacing the linear function of the predictor variables with a smoothing function (Hastie and Tibshirani, 1990). Various smoothing functions include moving averages, smoothing splines (Eubank, 1988; Wood 2000; Wood and Augustin 2002), running medians (Goodall, 1990), and kernel smoothers (Hardle and Marron, 1991). GAMs are appropriately used when the response variable is binary (i.e. presence/absence data), discrete (e.g. count data), or continuous. The most significant benefit of using GAMs rests in their flexibility in capturing non-linear cetacean-habitat relationships. Regression is currently the most common technique for modeling cetacean-habitat models, but may be limited by the characteristics of the dataset as well as the span of spatial and temporal availability of cetacean and habitat data (Redfern et al., 2006).

CHAPTER 3 - METHODOLOGY

This chapter describes the application of a multifaceted approach, utilizing novel methods of passive acoustic detection of sperm whales in the Gulf of Alaska and predicting habitat using best-fit GAMs. The various sections detail the methods and applications used to collect field data, process raw data, and develop non-linear models that show functional relationships between localized encounters and a suite of environmental variables. Figure 1 summarizes the work flow and tasks that were crucial for the completion of this project. The numbers on the left side show the sections used to describe each task.

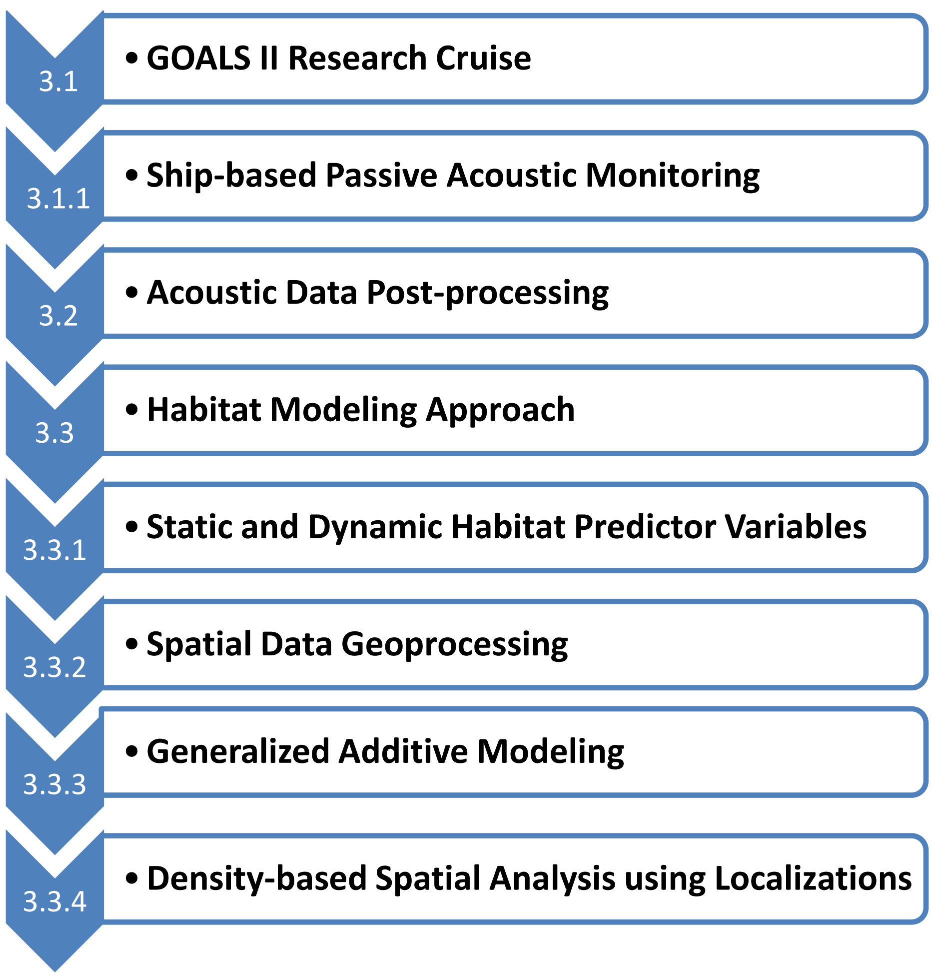


Figure 1. Workflow diagram showing overall project methodology.

3.1 GOALS-II Research Cruise

The GOALS-II line-transect survey was performed from 23 June to 18 July 2013. The entire survey effort was conducted aboard the *R/V Aquila*, a 50 m charter crab-fishing vessel. Survey cruises were implemented using custom designed line-transect track line patterns within the 144,560 km^2 Temporary Maritime Activities Area (TMAA) (Figure 2) (Rone et al., 2014). The customized line-transect tracklines were designed to offer uniform sampling coverage by following an equal-spaced zigzag pattern (Strindberg et al., 2004). Estimating abundance of marine mammals with line-transect surveys is particularly well-developed (Holt, 1987; Kinzey et al., 2000; Barlow et al., 2009).

In the course of the line-transect survey, researchers used towed hydrophones to conduct acoustic efforts during 23 of 26 survey days, turning out a total of 426 hours of 'standard' real-time monitoring along 6,304 km of cruise trackline. This survey yielded an average of 18.5 hours and 274 km of acoustic monitoring per day. In total, GOALS-II produced approximately 522 hours of mid-frequency recordings, as well as over 130 hours high-frequency recordings (Rone et al., 2014).

The TMAA was designed to survey cetaceans across four distinct areas representing sea bed stratum habitat types ('inshore', 'offshore', 'slope' and 'seamount') (Figure 3). The 'inshore' stratum was included to provide information on animals occurring within the continental shelf area. The 'offshore' stratum was sampled in order to gain knowledge of pelagic cetacean presence within the TMAA. The 'slope' stratum was established to sample the cetacean occurring on sloping regions off the continental shelf, and the seamount strata was included as a way to investigate cetaceans near seamounts (Rone et al., 2014). This survey plan is specifically designed to give an equal and accurate sampling of the area of interest.

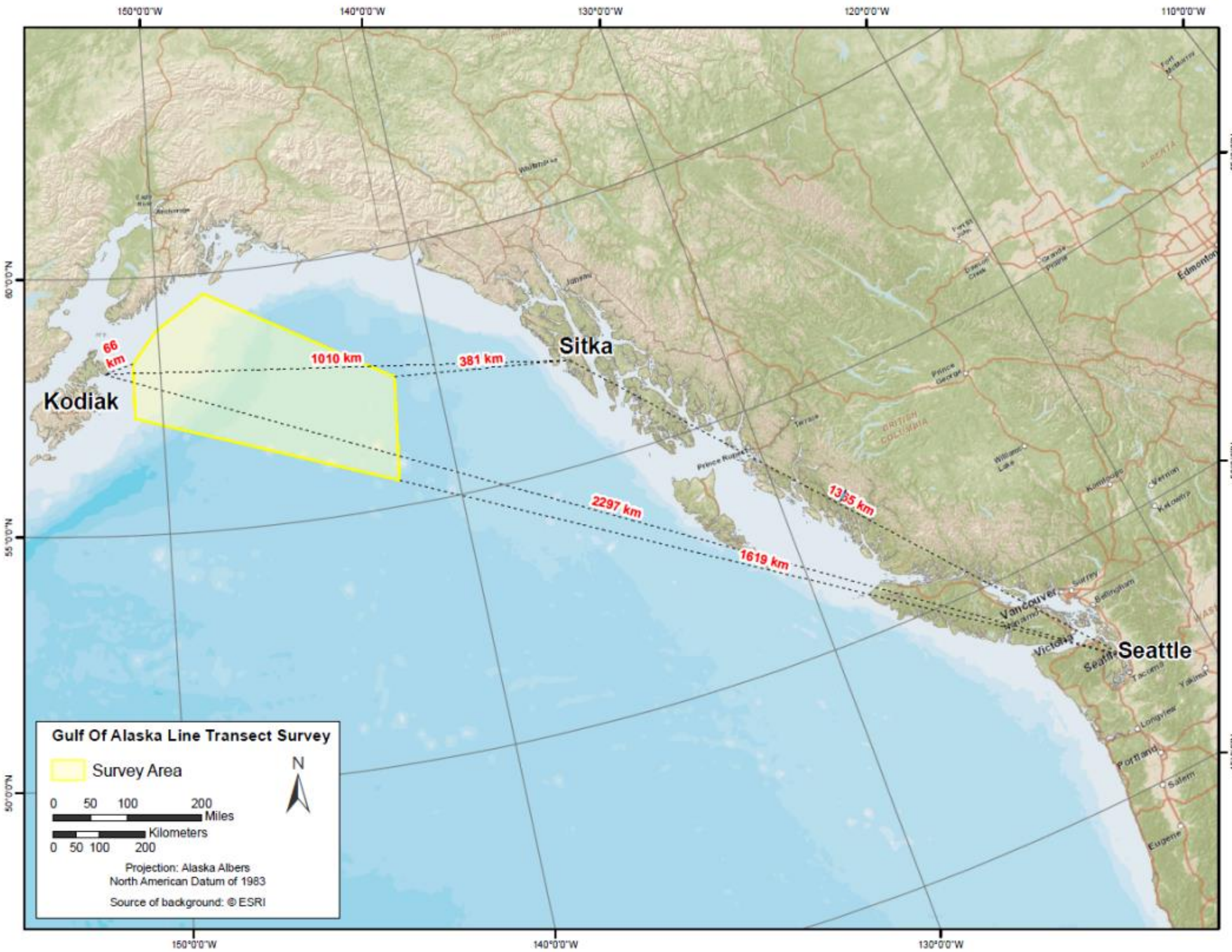


Figure 2. Full-extent map of study area of Gulf of Alaska Line-transect Survey – II Temporary Maritime Activities Area (Rone et al., 2014).

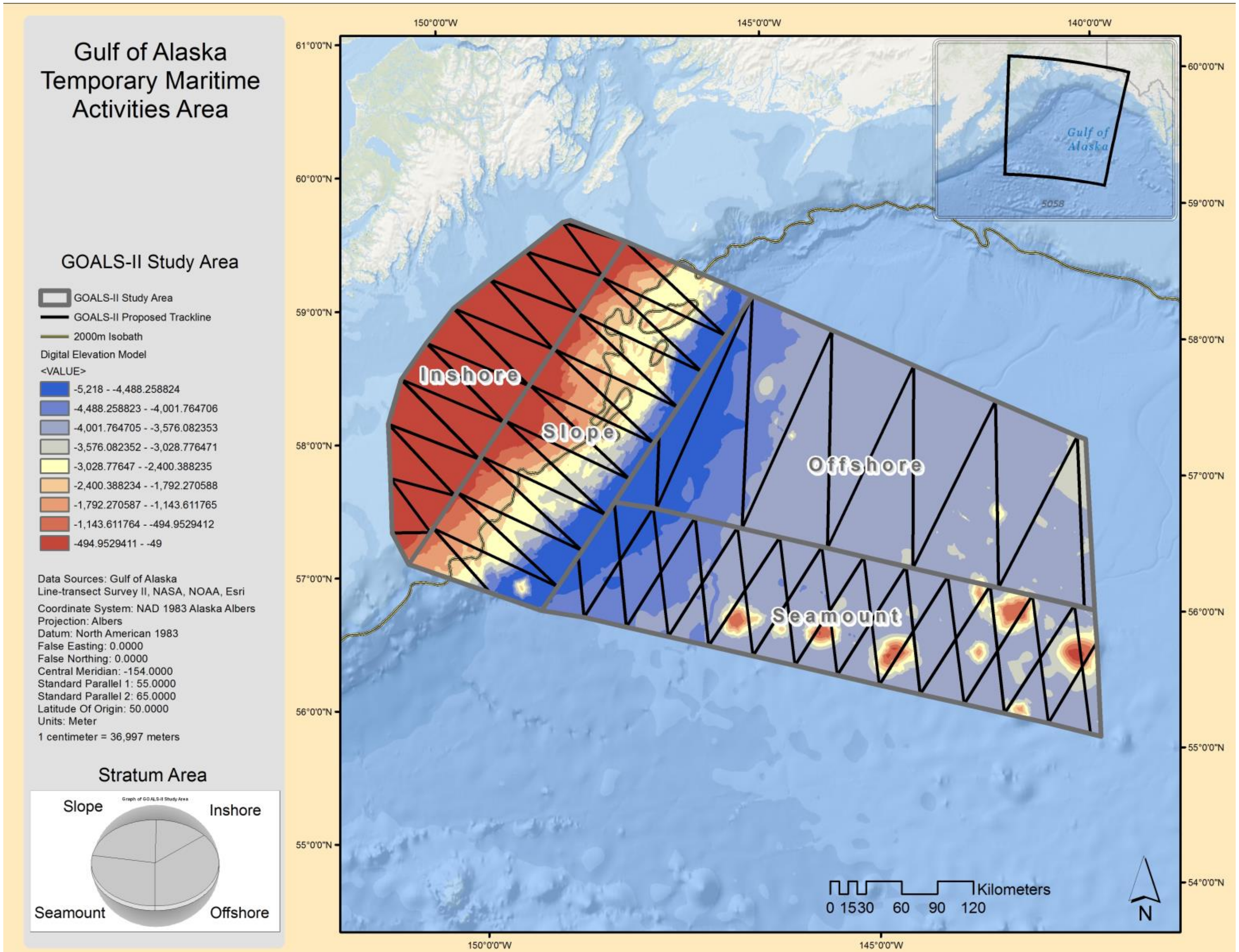


Figure 3. Map showing the GOALS-II TMAA proposed trackline, seabed relief, and relative area of survey strata.

3.1.1 Ship-based Passive Acoustic Monitoring

The ship-based acoustics surveying effort was conducted by performing real-time monitoring and recording efforts using the deployment of a towed-hydrophone array and sonobuoys. A total of 6,304 km of trackline was utilized to produce 426 hours of 'standard' real-time monitoring to detect vocally active cetaceans. PAM was managed by personnel consisting of four bioacousticians, with two on watch at all times. Personnel worked 12-hr shift rotations that were divided up between towed-array and sonobuoy monitoring (Rone et al., 2014).

The towed hydrophone array incorporates five hydrophones. The array includes two high-frequency Reson (R) hydrophones (flat frequency response (± 1.5 decibels [dB]) from 1 to 180 kilohertz [kHz]; 35 dB gain), as well as three mid-frequency hydrophones (APC International, Inc.; flat frequency response [± 1.5 dB] from 1 to 100 kHz; 36 dB gain). Two signal processing and recording systems were operated during the course of the survey. The first comprised a mid-frequency system utilized for the offshore, seamount, and slope strata and the second, a mid- and high-frequency system utilized in the inshore stratum. The various systems allowed researchers to target species of interest within each habitat type. The mid-frequency system detected sperm whales, while the high-frequency system targeted other surveyed species such as porpoise (Rone et al., 2014).

While on-effort, perpendicular distances from vessel to animals were measured in order to produce a georeferenced location of the animal (Figure 4), this is known as the localization technique. Localized acoustic encounters are utilized in an effort to estimate a detection function, and are particularly well-suited for sperm whales (Leaper et al., 1992, 2000; Barlow and Taylor, 2005).

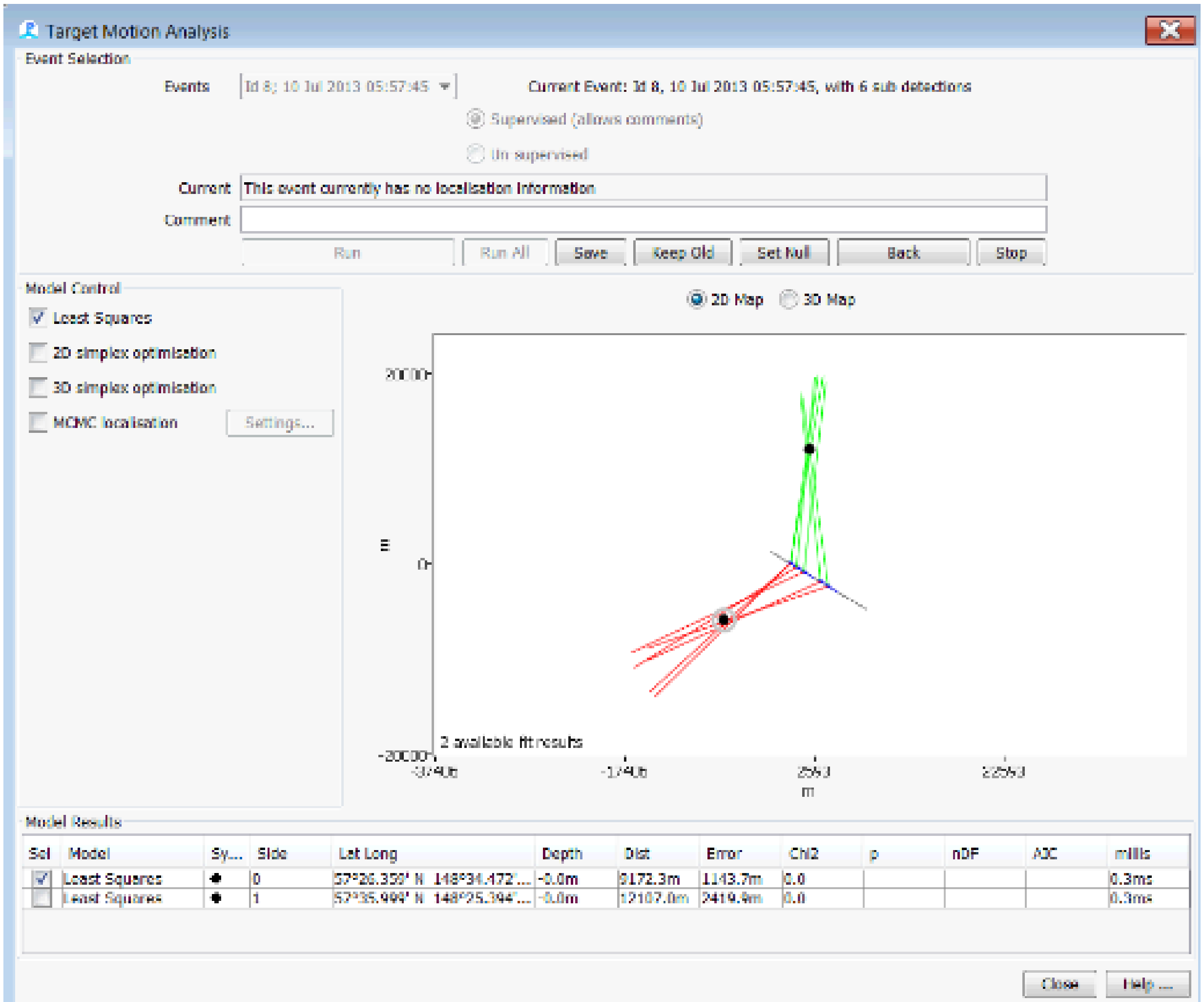


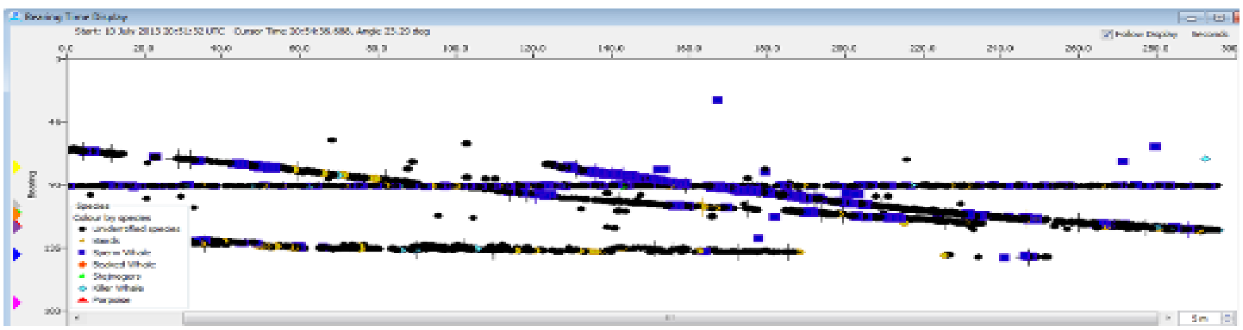
Figure 4. Example of target motion analysis feature of PAMGuard's ViewerMode software. As indicated by the model results panel, a least-squares-fit is used to obtain a port and starboard localization to the sperm whale event and PAMGuard indicates which of these two localizations has the best fit (Rone et al., 2014).

3.2 Acoustic data post-processing

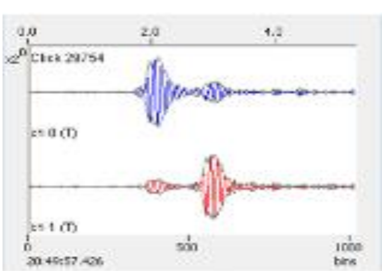
Signal processing, localization, recording, and documentation was executed using a combination of software programs including *Ishmael* (Mellinger, 2001), *Whaletrak II* (developed by Glem Gailey at Texas A&M University), and *PAMGuard* (http://www.pamguard.org). In order to record acoustic data and obtain bearings to user-selected vocalizations, *Ishmael* was utilized via target motion analysis (Mellinger, 2001). Two-channel recordings were continuously sampled at

192 kHz, with .wav formatted files saved at 10-minute intervals. Acoustic data were backed up to internal and external hard drives every 24 hours. *Whaletrak II* was utilized for the purpose of manually localizing and plotting whistling cetaceans and compact groups. This software suite provided a means of recording the vessel and array position, heading and speed, as well as form data entry via *MS Access*. *PAMGuard* was utilized for its powerful capability as an automatic click and whistle detector. Sperm whales were detected via configuring *PAMGuard* with an automatic click classification module (Figure 5).

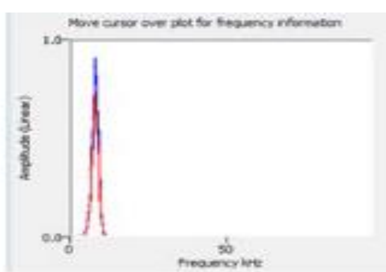
(a) Bearing Time Display



(b) Waveform Display



(c) Spectrum Display



(d) Wigner Plot

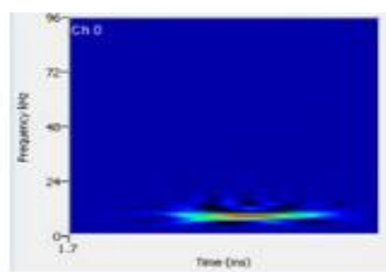


Figure 5. PAMGuard semi-automated click classification and tracking of sperm whales. Panel (a) displays bearing angles from individual animals as they pass the beam (~90°) of the vessel. The x-axis shows time and the y-axis shows the bearing angle. Panel (b) shows a typical waveform of a sperm whale click with amplitude displayed along the y-axis over sample bins displayed on the x-axis. Panel (c) is the spectrum display indicating the peak frequencies of the incoming echolocation signals. Panel (d) shows a Wigner plot of the sperm whale click (Rone et al., 2014).

Independent acoustic encounters are those considered to be isolated from the previous or next encounter, and are each assigned a unique encounter number. Acoustic encounters were characterized by the bio-acoustician on watch by using bearings, time-frequency signatures, and patterns of the signals detected.

Bio-acousticians identified acoustic encounters to the lowest taxonomic level possible based on descriptions of calls from literature and past experience. Whale calls and echolocation clicks were discerned between species based on various characteristics such as duration, spectrum peaks, and Wigner-Ville transform plots (Rone et al., 2014). A Wigner-Ville transform plot, or more commonly known as a Wigner Plot, is a quadratic time-frequency representation used to visually represent the time frequency structure of short duration, broadband cetacean clicks (Papandreou-Suppappola and Antonelli, 2001).

3.3 Habitat Modeling Approach

A habitat is defined not only by the geographic region where an organism may reside but also by biotic and abiotic environmental factors and their interaction (MacArthur and Wilson, 1967). Behavioral, physiological, and environmental factors govern an organism's distribution, movement, dispersal ability, tolerance of environmental conditions, with respect to inter- and intra-specific interactions. A culmination of these factors, in concert with their interaction effects, defines a species' niche within its habitat range. The niche that a species can occupy consists of all locations where sufficient conditions exist to support it, whereas inter and intra-specific competition defines the realized distribution of actual occurrence (Hutchinson, 1959, 1961).

In order to confidently predict habitat-species associations, researchers have worked to develop models that rely on spatially explicit multivariate techniques using a suite of environmental predictor variables. Variables within models often serve as proxies for broader scale conditions, such as mixed layer depth and frontal zones (Ferguson et al., 2006; Becker et al., 2012). An underlying assumption of these models is that animal density is a good predictor of habitat preference (Yack, 2013).

Through recent advancements in GIS technologies and highly-functional statistical techniques, there has also been an increase in predictive habitat models to relate geographical distributions of cetaceans to many environmental variables with the ability to model trends in abundance (Guisan and Zimmermann, 2000). Studies have shown the habitat modeling approach to be successful, although different methods of exploring cetacean habitat relationships can be advantageous in an analysis of habitat prediction outputs. The type of datasets available within the specified area of interest was considered foremost when determining techniques to be used for modeling. Due to the fact that many encounters (localized and non-localized) were recorded during the span of the approximately five-week study, the sample size was deemed large enough to perform regression modeling as opposed to a less indicative environmental envelope modeling.

An abundance of satellite-based predictor variables with a wide range of available spatial and temporal data allowed the potential for regression-based and spatial analysis techniques. A lack of *in situ* oceanographic variables during data collection restricted temporal scaling of the models to pre-determined satellite remotely sensed dataset averages. In this robust analysis design, trackline segments with associated acoustic data are prepared and utilized to inform Generalized Additive Models (GAMs) of encounter rate for sperm whales utilizing a multi-stage process. Localized acoustic encounters were modeled for best fit using an array of fixed spatial features and dynamic oceanographic variables to predict species distribution and habitat preferences.

3.3.1 Static and Dynamic Habitat Predictor Variables

In order to understand the relationship between recorded animal locations and covariates, appropriate static and dynamic habitat variables were determined to develop regression models. A variety of environmental variables were evaluated to define their fit for the purpose of modeling sperm whale habitat. Non-linear relationships are expected for cetaceans, which exist in a three-dimensional, highly variable environment with exceptionally dynamic trends, often characterized by a multitude of physical and biological variables (Yack, 2013). With the complex cetacean-habitat relationship that was expected, two categories of variables were developed to ensure individual covariates would fit the functional form of the GAM. For the purpose of explaining declines in abundance that are due to environmental variability or movement in response to variable ocean conditions, predictive habitat variables were chosen based on current knowledge of cetacean behavior and physiology.

Visual-based studies indicate that many whale species prefer habitat associated with dramatic topographic features, such as canyons, escarpments, shelf-edges and steep slopes (MacLeod and Zuur, 2005). More recently, GAMs of surface detections in the eastern tropical Pacific suggest that previously proposed definitions of whale habitat may be too narrow and/or not applicable to all geographic regions (Ferguson et al., 2006). In order to explore general habitat preferences of sperm whale in higher latitudes, nine predictor variables were used in the construction of models.

Environmental predictor variables utilized in the GAM models included static, or fixed variables such as depth, slope, aspect and distance to the 2,000 m isobath (Table 1). Dynamic predictor variables used within the models include sea surface temperature, salinity, chlorophyll-a, moon phase, current direction, and current magnitude. These data variables were acquired using custom geoprocessing tools that directly download data values from servers which contain databases of various remote sensing missions. These custom tools were configured using ModelBuilder, described in Section 3.3.3 and shown in Figure 6.

Table 1. List of static and dynamic explanatory variables used in predictive habitat models for sperm whales.

Static habitat predictor variables	**Dynamic habitat predictor variables**
• Depth (m) • Slope (degrees grade) • Aspect (degrees) • 2000m Bathy Dist (km)	• SST (° Celsius) • Salinity (PSU) • Moon Phase (0-0.999) • Current Direction (degree bearing) • Current Magnitude (m^2/s^2) • Chlorophyll-a (mg/m^3)

3.3.2 Cruise Trackline Data Conversion

Density of projected sperm whale within the study area was modeled using discrete data based on acoustic encounter rate along the cruise trackline. In order to create samples for modeling, survey data was divided into transects of approximately 5 km. This transect segment length was chosen to match the scale of response variables in a similar study (Yack, 2013). Lengths of continuous sections of survey effort could not be evenly divided by 5 km, consequently remaining segments were between 0.1 and 4.9 km. The resulting acoustic dataset contained 1,184 segments, with 81% of the segments equal to 5 km. Acoustic samples were utilized from trackline for segments from Beaufort sea states 0-6 (because acoustic detections are not affected by sea conditions and will not result in bias) and during transit periods when the acoustic team was 'on-effort' (Rone et al., 2014).

In order to prepare survey trackline for modeling efforts, the total cruise path during acoustic efforts were evenly divided into 5 km segment samples using the R programming language for statistical and quantitative data analysis. A segment chopping script was prepared to process the raw cruise data and produce segments with this specified length, created by Tina Yack and Aly Flemming (Appendix B). The script imported raw trackline, in addition to calculating the positions of 5 km segments and averaging the Beaufort sea state. The script was instructed to write a matrix with values for start, middle and end positions with timestamps, as well as segment length in kilometers. The final output was a comma separated value (CSV) file which contained the entire matrix. The final step was to manually associate sperm whale localizations with their respected trackline segments. The populated CSV was then imported into ArcGIS so trackline segment midpoints could be used for modeling.

3.3.3 Spatial Data Geoprocessing

ArcGIS ModelBuilder was utilized as a geoprocessing framework to link data input, customized tools, and data output. A geoprocessing model was constructed to describe the relationship between acoustic data in vector format to covariates in raster format using specified parameters. ModelBuilder was utilized as a core data workflow in order to calculate fixed and dynamic habitat variables for interchangeable input vector datasets. In this study, the following datasets are used as inputs for this geoprocessing model:

1) Trackline segment midpoint samples;
2) Localized acoustic encounters of sperm whale; and

3) Study area fishnet grid midpoints.

To ensure an accurate spatial analysis, associated input and study area data were properly imported and projected as NAD 1983 Alaska Albers Equal Area. In order to acquire data describing depth within the study area, a 30 arc second DEM of Alaska was downloaded in a geotiff format from the NOAA National Geospatial Data Center. This coastal relief model raster was clipped to the GOALS-II study area.

The next step was to prepare the final input dataset by dividing the study area into a 5x5 km (25 km^2) fishnet grid to produce smoothed encounter rate plots during the final stages of the model. Using the Create Fishnet (Data Management) tool, a grid of rectangular polygons and related midpoints was created. Polygon cell size width and height was set to equal 5,000 m, with no rows or columns. Tool environment parameters were set up for the study area, but had to be clipped with the GOALS-II boundary again to retain only cells and midpoints within the survey area.

After the DEM and all three input datasets were prepared to be geoprocessed, they were iterated through the spatial analysis geoprocessing model. Multiple attribute fields were created within a single feature class to represent specified environmental variable values, thereby iterating a multitude of geoprocessing functions (Tables 19 and 20), and outputting one feature class. Data values are saved into existing feature classes based on date fields. Each tool's parameters were defined to most accurately represent actual environmental conditions for each predictor variable. All required fixed and dynamic oceanographic variables were calculated within the model (Figure 6).

Figure 6. Customized geoprocessing model using MGET to extract desired environmental variables at points (Further explanation found in Figures 33 and 34 of Appendix A).

With the appropriately-sized DEM of the study area, the Slope Tool (Spatial Analyst) was utilized to create a slope raster in degrees rise. The clipped coastal relief model raster was again used to run the Aspect Tool (Spatial Analyst), in degree bearing. After the rasters were created, this allowed the ability to run all point features through Extract Values to Point (Spatial Analysis). This tool added the values of depth (DEP), slope (SLP), and aspect (ASP) to the attribute table of the output feature class. The bathymetric contour shapefile was acquired from Esri ArcGIS Online, with the correct isobath depth selected using the Definition Query of "DEPTH" = 2,000. The last step is to calculate the distance of the animals from the 2,000 m isobath (BTH) contour line using Near (Analysis). All of the tools used to obtain fixed environmental variables can be further referenced in Appendix A.

Extraction of dynamic predictor variables incorporates the novel use of Duke University's Marine Geospatial Ecology Tools (MGET), an extensible powerful open-source geoprocessing toolbox for Esri ArcMap. MGET provides marine ecologists capabilities of specialized platforms such as Python, R, MATLAB, and C++ (Roberts and Best, 2010). Six MGET tools were used to extract values to points via direct access to online geophysical data servers provided by NOAA, NASA, MODIS, HYCOM, and OSCAR. The custom geoprocessing model was designed to extract satellite measurement values for the input point feature class. Before acquiring values from the respected servers, labeled fields were created using a float data format to ensure compatibility with tool scripts. After these fields were created, they were populated by customized MGET tools using the sampling date field created in R. Customized geoprocessing tools used to extract values for each environmental predictor variable are detailed in Table 2, along with other information associated with environmental variables such as units, data sources, and object-oriented modeling abbreviations.

Table 2. Environmental predictor variables with units, data sources, variable name abbreviation used in S-S-Plus/R scripts, and geoprocessing tools used to extract values.

Environmental Predictor Variable	**Units**	**Data Source**	**Variable Name for S-Plus/R**	**Raster Attribute Extraction GP Tools**
Depth	meters	NOAA NGDC	DEP	Extract Values to Points
Slope	degrees of grade	Slope (Spatial Analyst) run on NOAA NGDC DEM	SLP	Extract Values to Points
Aspect	degrees bearing	Aspect (Spatial Analyst) run on NOAA NGDC DEM	ASP	Extract Values to Points
2000m Isobath Distance	kilometers	Near (Analyst) run on occurrences to 2000m isobath	BTH	Near
Sea Surface Temperature	degrees Celsius	PO.DAAC MODIS L3	SST	Interpolate PO.DAAC MODIS L3 SST at Points
Sea Surface Salinity	PSU (practical salinity unit)	HYCOM GLBa0.08 Equatorial 4D	**SSS**	Interpolate HYCOM GLBa0.08 Equatorial 4D Variables at Points
Chlorophyll (surface concentration)	milligrams/meter^3	NASA OceanColor L3 SMI	SC	Interpolate NASA OceanColor L3 SMI Product at Points
Moon Phase	moon phase (0-0.999)	GeoEco Module from MGET	MNP	Calculates moon phase for dates in field
Current (Direction of Water)	degrees bearing	OSCAR	DIR	Interpolate OSCAR Currents at Points
Current (Total Kinetic Energy)	meters^2/second^2	OSCAR	TKE	Interpolate OSCAR Currents at Points

In order to obtain values for sea surface temperature (SST), the acoustic dataset was inputted into PO.DAAC MODIS L3 SST at Points (Spatial Analyst), using the Aqua satellite with a temporal resolution of 8 days and a spatial resolution of 9 km for cruise segment midpoints and acoustic encounters. Sea surface salinity (SSS) was calculated using the Interpolate HYCOM GLBa0.08 Equatorial 4D Variables at Points in which estimates were taken using a daily average. Chlorophyll (SC) was downloaded using the Interpolate NASA OceanColor L3 SMI Product at Points tool, with a temporal resolution of 8 days and a spatial resolution of 9 km for cruise segment midpoints and acoustic encounters. Next, numeric moon phase was extracted from the MGET GeoEco Module using the Calculates moon phase for dates in the field. Current magnitude (TKE) and direction (DIR) were obtained using Interpolate OSCAR Currents at Points, the temporal resolution was set to an 8-day average. After all of the tools completed updating the attribute table, a single feature class was created to consolidate fixed and dynamic variable outputs. Ultimately, two columns of attributes were created for each variable. This included the creation of attribute columns for all trackline midpoints with associated localizations representing presence-absence data and another for localizations only. Statistics of all variables were exhibited in Tables 5 through 15.

For the fishnet grid midpoints, feature classes were created for each explanatory variable separated by date. When possible, temporal resolutions were set to monthly in order to produce averages for the approximately one-month long survey period. In the case of current magnitude and direction, OSCAR does not have an option for monthly temporal resolution and only provides 8-day averages. To overcome this, date fields were created for each week of the survey, in which attribute tables are joined. The field calculator function was then used to create average values between fields.

3.3.4 Generalized Additive Modeling

Modeling efforts include a multifaceted approach that uses GAMs with discrete trackline acoustic encounters and habitat variables to first offer a comparison of encounter rate differences across the study area, as well as demonstrate the ability of habitat variables to predict sperm whale presence. GAMs are data-driven models that can accommodate non-parametric relationships between examined variables and are therefore particularly effective for modeling complex ecological relationships (Hastie and Tibshirani, 1990). A GAM may be represented as:

$$g(\mu) = \alpha + \sum_{j=1}^{p} f_j\left(x_j\right) \tag{1}$$

Where g(μ) is the link function, which relates the mean of the response variable given the predictor variables $=\Sigma(Y|x_i, + \ldots + x_p)$ to the additive predictor $f_j(x_j)$.

Model building was achieved through a multi-step progression of scripts via the object oriented S-Plus statistical language. The null model assumes a uniform distribution for sperm whales that initially includes none of the predictor variables. After building a null model, habitat variables were added to determine if they statistically improved the fit of the null model. Typically, GAMs are utilized for visual surface observation-based data to model distribution (Ferguson et al., 2006; Becker et al., 2012). Within this study, acoustic-based habitat models for the Gulf of Alaska were generated following methods developed for the first beaked whale acoustic based models by Becker (2007) and Yack (2013).

Trackline segment midpoint samples had been prepared by associating fixed spatial features and dynamic oceanographic variables for input into GAMs. After data was entered into S-Plus, sample size summaries, and oceanographic correlations were calculated. These matrices were then written to CSVs. This provided further information on the relationship between variables. The oceanographic correlation matrix was analyzed for values that would indicate linear relationships, at which point sea surface salinity was removed. Values within this matrix also provided information about useful covariate combinations for each model.

Four GAMs were independently created, utilizing different combinations of habitat variables (Table 3) with a forward/backward stepwise selection. For each model, predictions are made with dispersion and deviance stored in a matrix. Functions were brought into stored formulas and plots representing functional relationships, further allowing the assessment of explanatory variables using Akaike’s Information Criterion (AIC) (Akaike, 1973). The resulting models of habitat predictors provide a baseline to allow for careful inspection of relative sperm whale encounter rate differences across the study area and to illustrate which habitat variables are most predictive of sperm whale presence.

The best-fit GAM contained the variables which would be used to predict back on to the study area grid (Table 3). The script used the formulas and explanatory variables from Model 4 with grid midpoint values extracted from the MGET tools to produce encounter rate predictions for each grid cell output. The encounter rate model predictions were then imported back into ArcGIS to create weighted density plots interpolated using Kriging (Spatial Analyst). Kriging is an advanced geostatistical procedure that generates an estimated autocorrelation surface from a scattered set of points with z-values (Roberts and Best, 2013). Kriging was set to interpolate fishnet labeled midpoints that indicate a spatial representation of preferred habitat of sperm whale within the study area.

Table 3. Models showing predictor variable combinations and resulting AIC values.

Model Name	AIC	Variables Included
Model A	363.0302	DEP, SLP, ASP, SC, MNP, SST, BTH, DIR, TKE
Model B	405.5429	DEP, SLP, ASP, SST, BTH, DIR, TKE
Model C	419.6675	DEP, SST, BTH
Model D	403.7779	DEP, BTH, SLP, SST, SC, TKE

3.3.5 Density-based Spatial Analysis using Localizations

In order to fully explore localized acoustic encounters, a hotspot analysis was performed. A similar approach to using animal locations as an input for the MGET geoprocessing model, the goal of this section is to create a visual demonstration. This side analysis shows predicted density values without influences from environmental covariates, essentially a distribution of samples (Figure 32).

The Kernel Density (Spatial Analysis) tool calculates a magnitude-per-unit area from point features with a kernel function to fit a smoothly tapered surface to each point (Roberts and Best, 2013). Conceptually, each point produces a surface value which is highest at the epicenter and diminishes with increasing distance from the point, only to reach zero at the search distance. The kernel probability density function is based on a quadratic function where inferences are made about the distribution without producing statistical confidence levels (Silverman, 1986). Parameters were set to create a surface raster of single acoustic observations, defining no population values to ensure the performance of a non-weighted analysis.

CHAPTER 4 - RESULTS

This chapter describes the final results generated with each component of the methodology. These final products explore a sample of localized acoustic sperm whale encounters and their modeled relationship to nine environmental variables. Variable relationships include supporting figures and maps in order to visually represent possible correlations. Non-linear relationships were obtained using a multi-stage process, primarily integrating Generalized Additive Modeling of encounter rate for the purpose of predicting sperm whale habitat within the study area.

During the GOALS-II survey, a total of 241 sperm whale acoustic encounters were recorded while 176 (73 percent) were localized in real time (Table 4). An additional 11 sperm whale encounters were localized during post-processing effort, representing an increase of 6 percent over the real-time localization total. 54 encounters were unable to be localized due to an inability to use a least-squares-fit to obtain beam distances to the sperm whale event. Of the 241 acoustic encounters, 187 were used as the basis for this analysis for the reason that perpendicular distances were obtained for these samples and represent the highest-quality samples. Sperm whales were encountered primarily in the slope stratum (63 percent), followed by the offshore (20 percent), and seamount (16 percent) strata (Figure 7). An additional sperm whale was heard while on 'non-standard' effort, since the encounter occurred on transit between transect lines (Rone et al., 2014).

Table 4. During the GOALS-II survey, 241 sperm whales were acoustically encountered although only 176 were localized during field operations, and 11 localized using post-processing. A total of 187 localized encounters were utilized for modeling efforts.

Real-time Localizations	Post-Processed Localizations	Total Localizations	Non-Localized Encounters	Total Acoustic Encounters
176	11	187	54	241

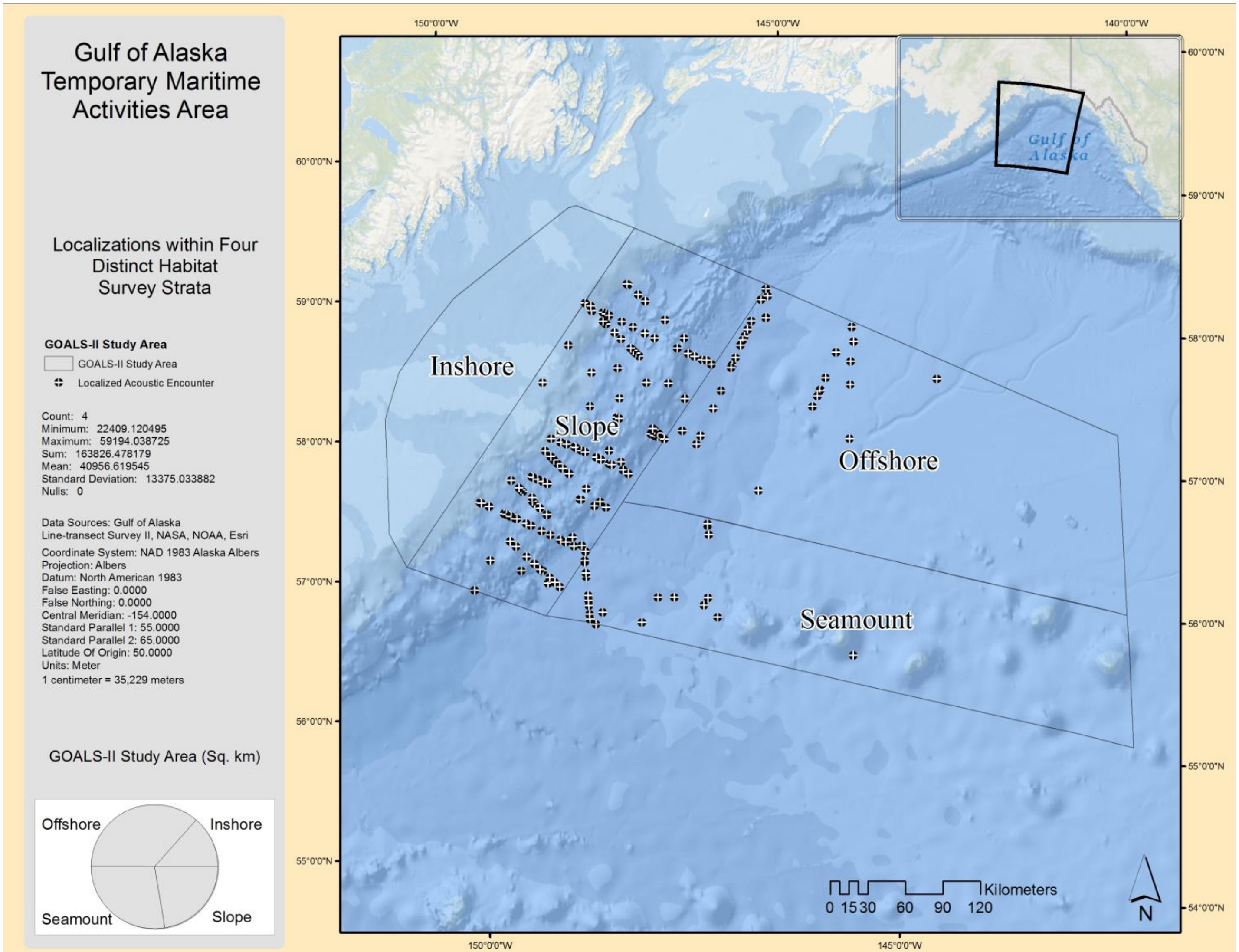


Figure 7. Map of GOALS-II TMAA showing georeferenced locations of localized acoustic encounters within four survey strata. There were 129 samples found within the ‘Slope’ stratum, 28 samples found within the ‘Offshore’ stratum, 17 within the ‘Seamount’ stratum, and no localized samples found within the ‘Inshore’ stratum.

There were 1,183 original segments, although only 839 had values for all covariates and were included in the models. The remaining 839 segment features remained after the MGET extraction-to-point tools yielded 344 null values, due to the presence of cloud cover during the survey dates. From the GOALS-II survey, most of the segments yielded no encounters (n=1,039), with the remainder having one encounter (n=119), two encounters (n=22), three encounters (n=5), and four encounters (n=1) (Table5). Of the original segments, there were 147 segments with associations between 0-4 occurrences. Most common were single encounter (79.7 percent), with the remainder containing multiple encounters such as two (14.1 percent), three (4.1 percent), and four (0.7 percent).

Table 5 Summary of GOALS-II trackline segments with associated localized encounters.

Number of Localized Encounters	Number of Segments	Percent of Total Trackline Segments	Percent of Total Segments with Localized Encounters
0	1,039	87.82%	0.00%
1	119	10.06%	80.95%
2	22	1.86%	14.97%
3	5	0.42%	3.40%
4	1	0.08%	0.68%

In order to spatially represent discrete encounter rate of trackline segments within the survey, Figure 8 displays the 5 km segments created in R. The number of localizations associated with segments are symbolized using a green to red color scheme. A simple frequency distribution plot of localization-associated segments was produced in the bottom left corner of the figure.

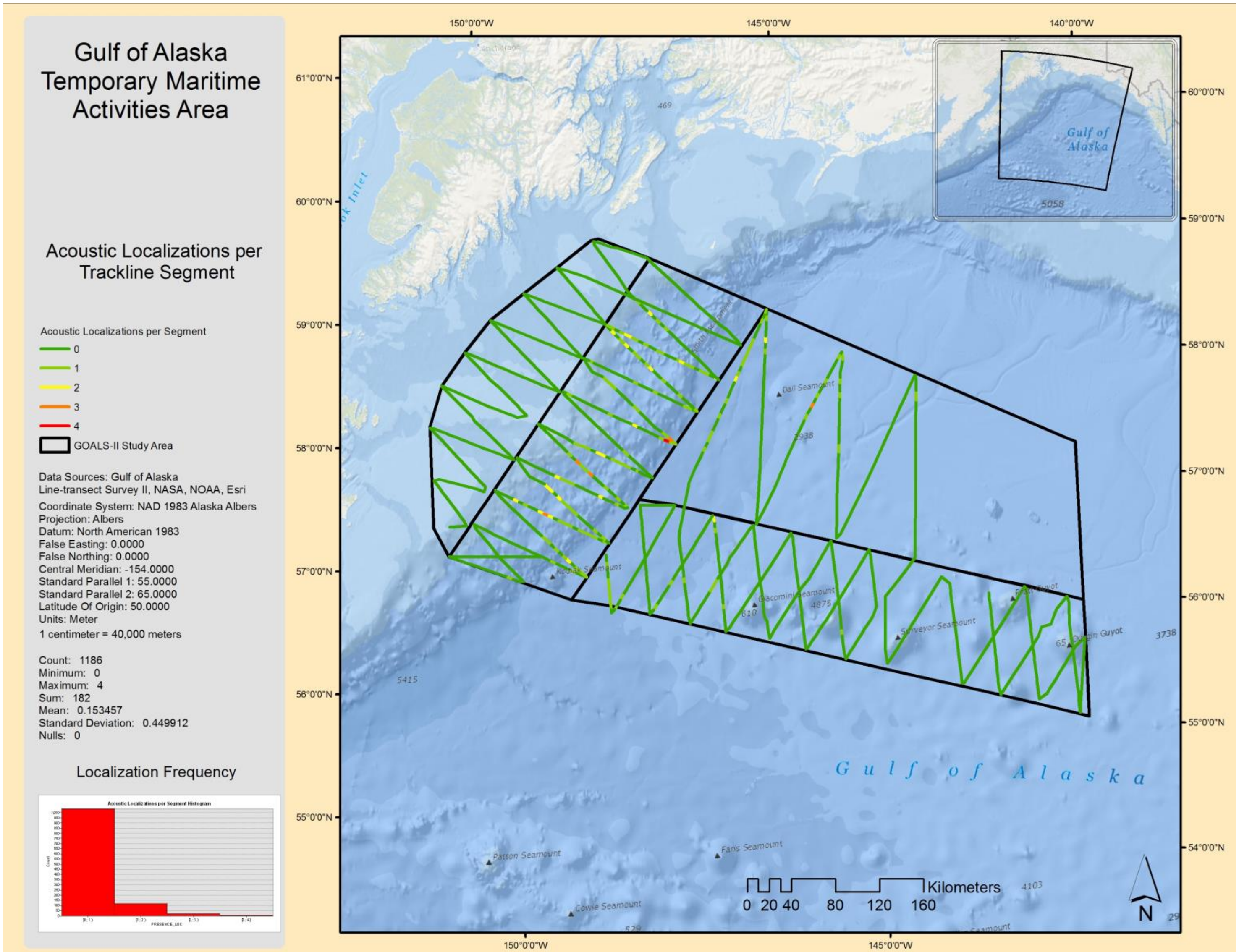


Figure 8. Map displaying how many acoustic localizations associated per 5 km trackline segment.

4.1 Encounter Relationship to Habitat Variables

Species distributions were modeled utilizing a combination of fixed spatial features and dynamic oceanographic habitat predictor variables. Fixed spatial features include depth, slope, aspect and distance away from the 2000 m depth contour. Dynamic oceanographic variables include sea surface temperature, sea surface salinity, moon phase, current direction, current magnitude, and chlorophyll concentration.

During the GAM development efforts, frequency distribution plots were created for each of the model covariates (Figure 9). These are important for the purpose of analyzing the distribution of values present for each of the environmental variables. This study shows that localized encounters were found to be related to static spatial features. Localizations are commonly located in areas with depths in ranges between 0-500 and 2,500-3,000 meters. As expected, localizations were also found to be related to seabed slope. Sperm whales were frequently vocalizing in areas with seabed slope ranging between 0-5 degrees grade. Localized encounters were found to be related to other fixed spatial geological properties such as seabed aspect. Animals tended to be located in areas featuring a southwest facing seabed aspect. Sperm whale localizations were also strongly linked to other static spatial features such as the 2,000 m isobath and displayed a preference for areas within 50 km of this contour line.

Dynamic oceanographic variables were also found to have an influence on sperm whale habitat. Modeling efforts showed a strong correlation with sea surface temperatures, with a high frequency of localizations present in areas with surface temperatures between 11-12 degrees Celsius. Dynamic oceanographic variables such as moon phase show a weak functional relationship with frequencies showing the numeric moon phase to skew away from 0.5, likely due to the occurrence of GOALS-II survey dates. Localized acoustic encounters have a robust correlation with current magnitude. Total kinetic energy for the majority of localized encounter samples is between 0.00-0.01 m^2/s^2. In addition to current magnitude, functional relationships to current direction were also explored. The highest number of localizations occurred with the current direction of flow being between 200-250 degrees bearing. Lastly, localized acoustic encounters were observed to have a moderate correlation with chlorophyll-a concentrations in the range between 0.5 and 2.5 mg/m^3.

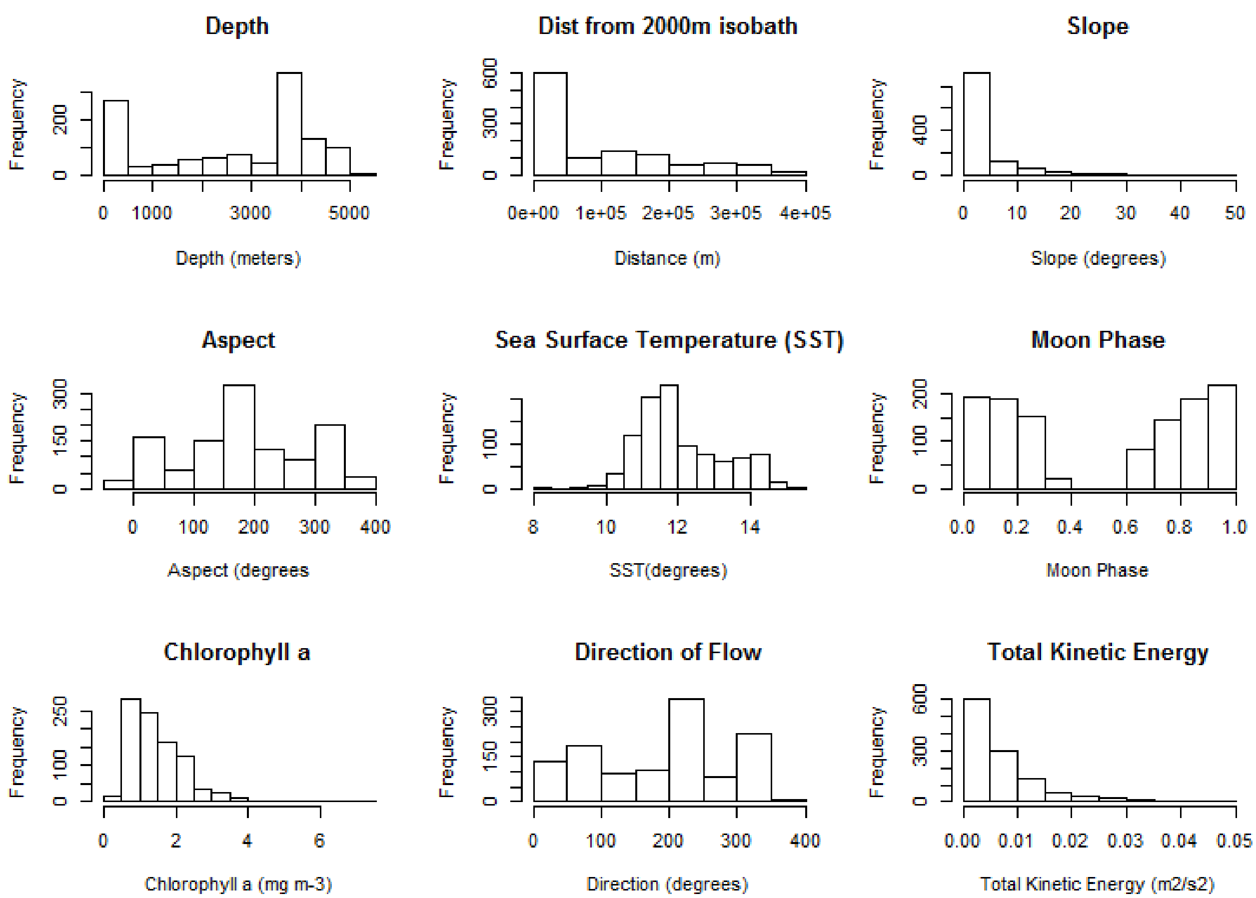


Figure 9. Frequency distribution plots showing the localized acoustic sperm whale encounters observed for every possible model covariate.

4.1.1 Static Spatial Features

Static spatial features include environmental variables for depth, slope, aspect and distance away from the 2000 m isobath.

4.1.1.1 Depth

Seabed relief is considered to be a primary fixed environmental variable because it has an effect on prey abundance. This section compares the difference between depth values obtained for localized encounters and all trackline segment midpoints, shown in Figures 10 and 11, with reference to Table 6. 186 samples were utilized to calculate statistics for localized encounters, and 1,183 samples were used to calculate statistics for segment midpoints. The minimum values were similar for both localizations and trackline segments, at 5,045 m and 5,053 m, respectively. The maximum values of localizations and midpoints were also similar, at -155 m and -56 m, respectively. The two columns differ in their mean values, showing a drastic difference in presence and presence/absence data. Localized encounter locations showed a mean value of 3,293.18 m while segment midpoints yielded a mean of 2,697.06 m. This shows that sperm whales were found about 596 m deeper than sampled survey trackline. Localizations had a standard deviation of 1,273.77 m, while trackline segments had a standard deviation of 1,655.70 m from the mean. This means that localization values for depth are clustered with less deviation from the mode. No null values were generated. Localized acoustic encounters were found to have a strong correlation with depth in the ranges of 0-500 and 2,500-3,000 m.

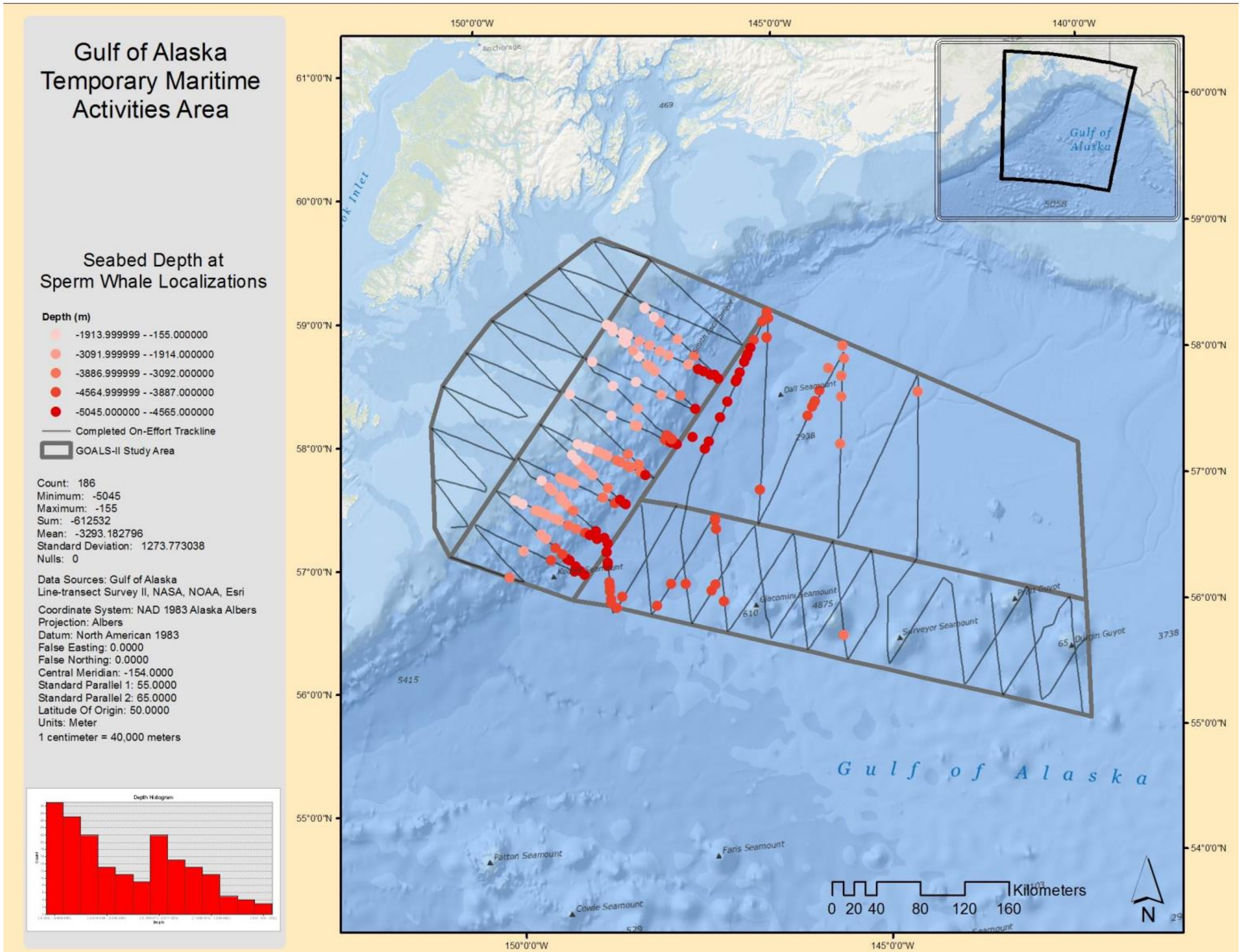


Figure 10. Map displaying seabed depth found at the location of each sperm whale localization. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

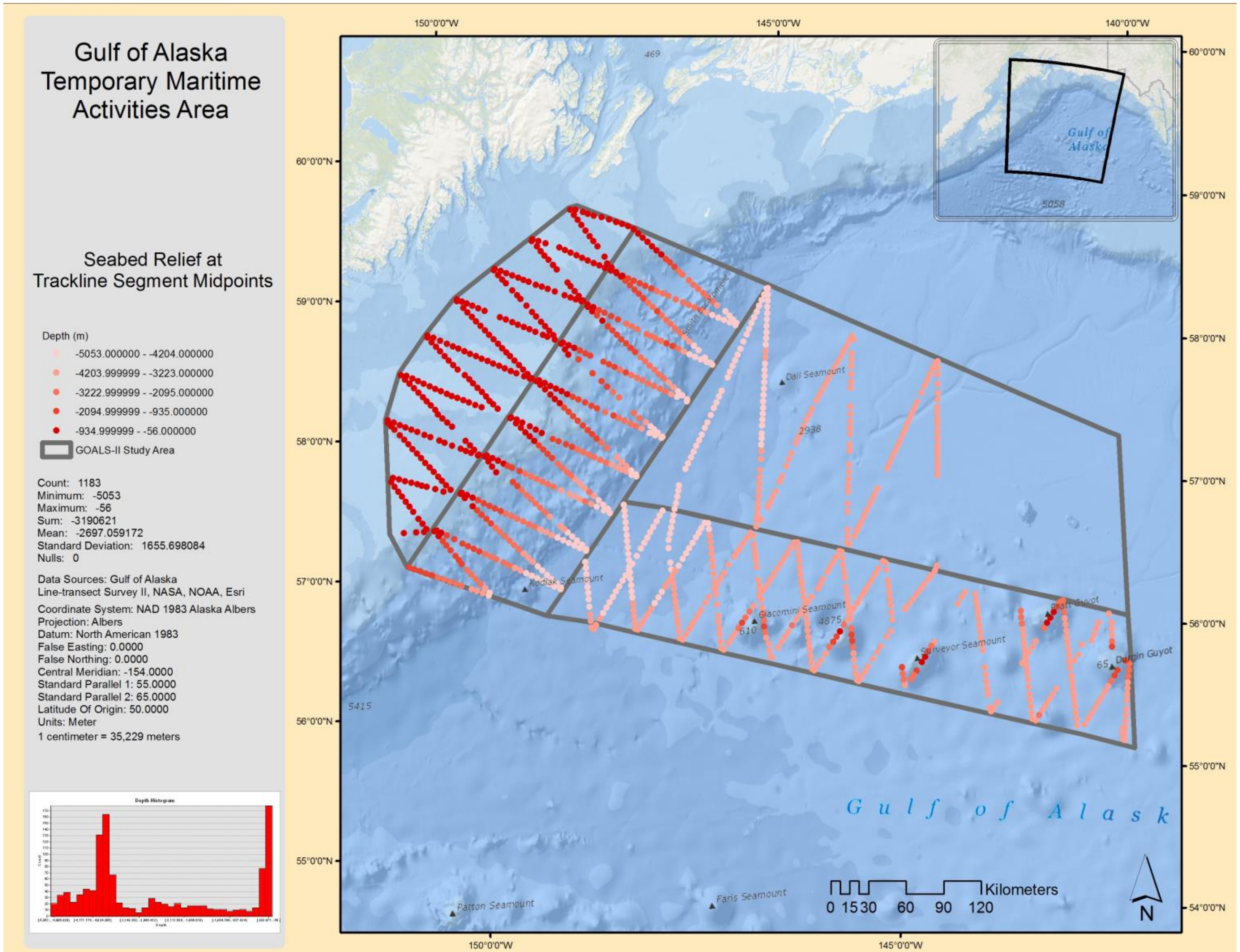


Figure 11. Map displaying seabed depth found at the location of each trackline segment midpoint. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

Table 6. Depth statistics of acoustic localizations compared to all trackline segment midpoints.

	Localized Encounters	Segment Midpoints
Count:	186	1,183
Minimum:	-5045.00	-5053.00
Maximum:	-155.00	-56.00
Mean:	-3293.18	-2697.06
Standard Deviation:	1273.77	1655.70
Nulls:	0	0

4.1.1.2 Slope

Slope can indicate the rate at which ocean currents push up cold, deep, and dense nutrient-rich water to the surface, stimulating the growth and abundance of prey. This section compares the difference between gradient values obtained for localized encounters and all trackline segment midpoints, shown in Figures 12 and 13, with reference to Table 7. 186 samples were utilized to calculate statistics for localized encounters, and 1,183 samples were used to calculate statistics for segment midpoints. The minimum values were similar for both localizations and trackline segments, at 0.11 degrees rise and 0 degrees rise, respectively. As expected, the maximum values of localizations and midpoints showed a much higher maximum slope for midpoints. Maximum slope values for localizations and trackline midpoints were 36.00 degrees rise and 43.48 degrees rise, respectively. The characterization of the two datasets is first demonstrated in their mean values, which shows a statistically insignificant difference between presence and presence/absence data. Localized encounter locations showed a mean value of 6.01 degrees rise while segment midpoints yielded a mean of 3.90 degrees rise.

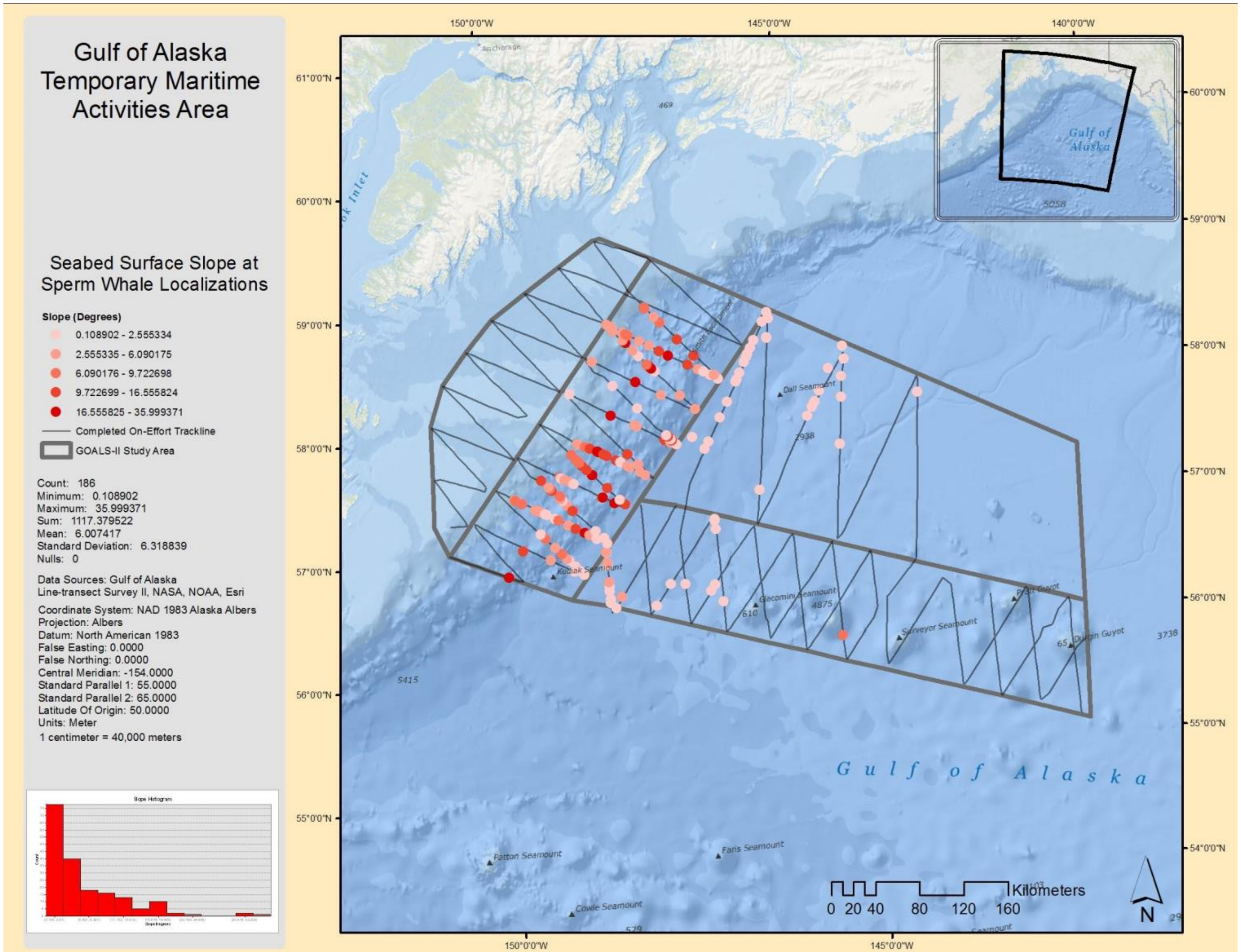


Figure 12. Map displaying seabed slope found at the location of each sperm whale localization. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

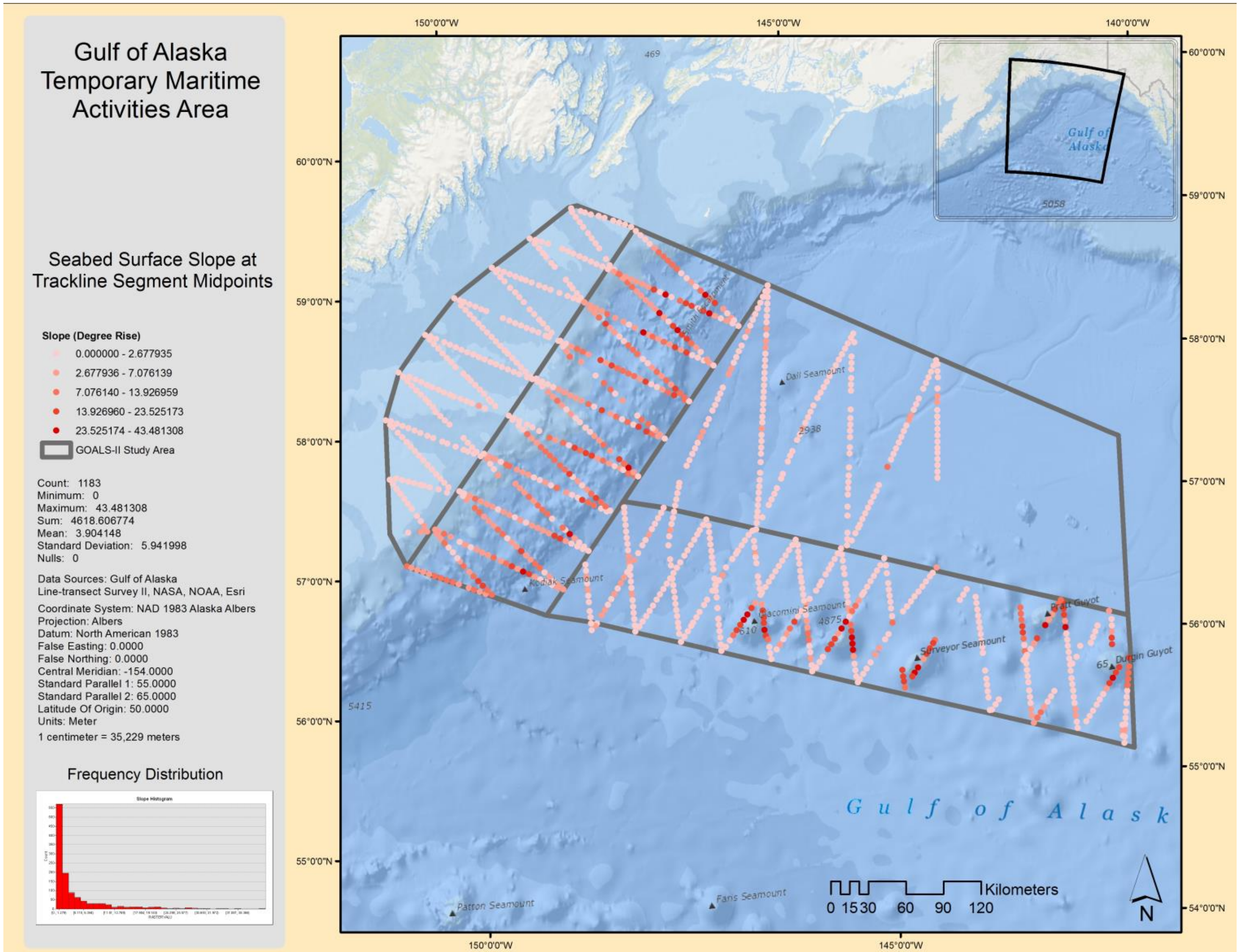


Figure 13. Map displaying seabed slope found at the location of each trackline segment midpoint. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

Sperm whales were found in regions with about 2.10 degrees more gradient than average sampled survey trackline. Localizations had a standard deviation of 6.32 degrees while trackline segments had a standard deviation of 5.94 degrees from the mean. This means that localization values for slope are less clustered with more deviation from the mode. No null values were generated. Localized acoustic encounters were found to be common with seabed slope values between 0 and 5 degrees gradient.

Table 7. Slope statistics of acoustic localizations versus all trackline segment midpoints.

	Localized Encounters	Segment Midpoints
Count:	186	1183
Minimum:	0.11	0
Maximum:	35.99	43.48
Mean:	6.01	3.90
Standard Deviation:	6.32	5.94
Nulls:	0	0

4.1.1.3 Aspect

Seabed aspect may serve as a proxy for the direction at which upwelling occurs, and could be important for regions that are susceptible to certain wind-driven current directions. This section compares the difference between aspect values obtained for localized encounters and all trackline segment midpoints, shown in Figures 14 and 15, with reference to Table 8. 186 samples were utilized to calculate statistics for localized encounters, and 1,183 samples were used to calculate statistics for segment midpoints. The minimum values were similar for both localizations and trackline segments, at 0 degrees and - 1 degrees, respectively, although represent very different aspects. In this analysis, 0 represents an aspect facing true north, meanwhile -1 equates to a completely flat surface. The maximum values of localizations and midpoints were similar, at 358.03 degrees bearing and 359.24 degrees bearing, respectively. The two columns were also quite similar in their mean values, showing a statistically insignificant difference between presence and presence/absence data. Localized encounter locations showed a mean value of 181.90 degrees bearing while segment midpoints yielded a mean of 186.76 degrees bearing. This shows that sperm whales were found near aspects facing about 5 degrees southward than sampled survey trackline. Localizations had a standard deviation of 96.20 degrees, while trackline segments had a standard deviation of 103.05 degrees from the mean. This means that localization values for aspect are clustered with less deviation from the mode. No null values were generated. Localized acoustic encounters were mainly found to occur over the south- and south-west facing aspects.

Table 8. Aspect statistics of acoustic localizations versus all trackline segment midpoints.

	Localized Encounters	Segment Midpoints
Count:	186	1183
Minimum:	0	-1
Maximum:	358.03	359.24
Mean:	181.90	186.76
Standard Deviation:	96.20	103.046
Nulls:	0	0

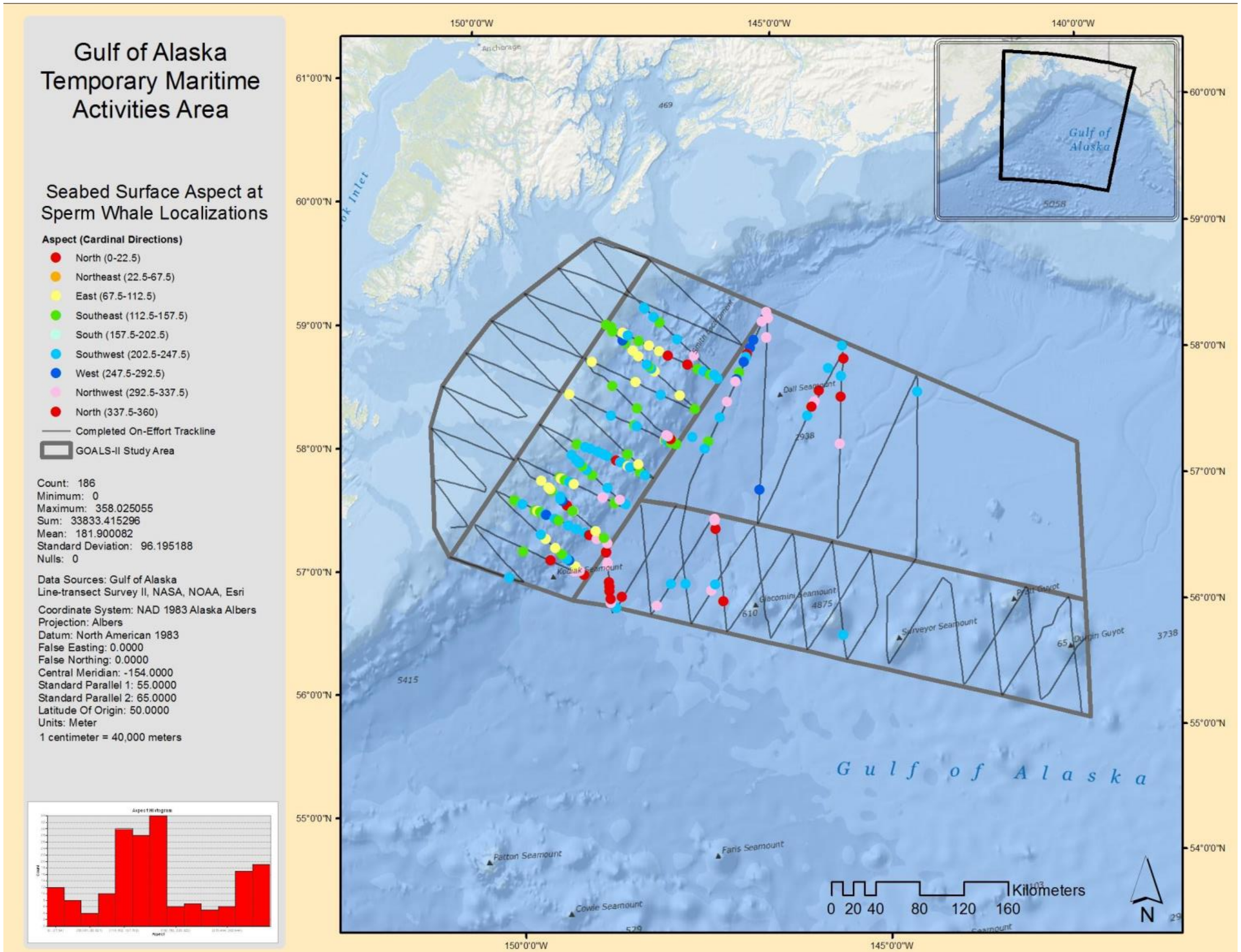


Figure 14. Map displaying seabed aspect found at the location of each sperm whale localization. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

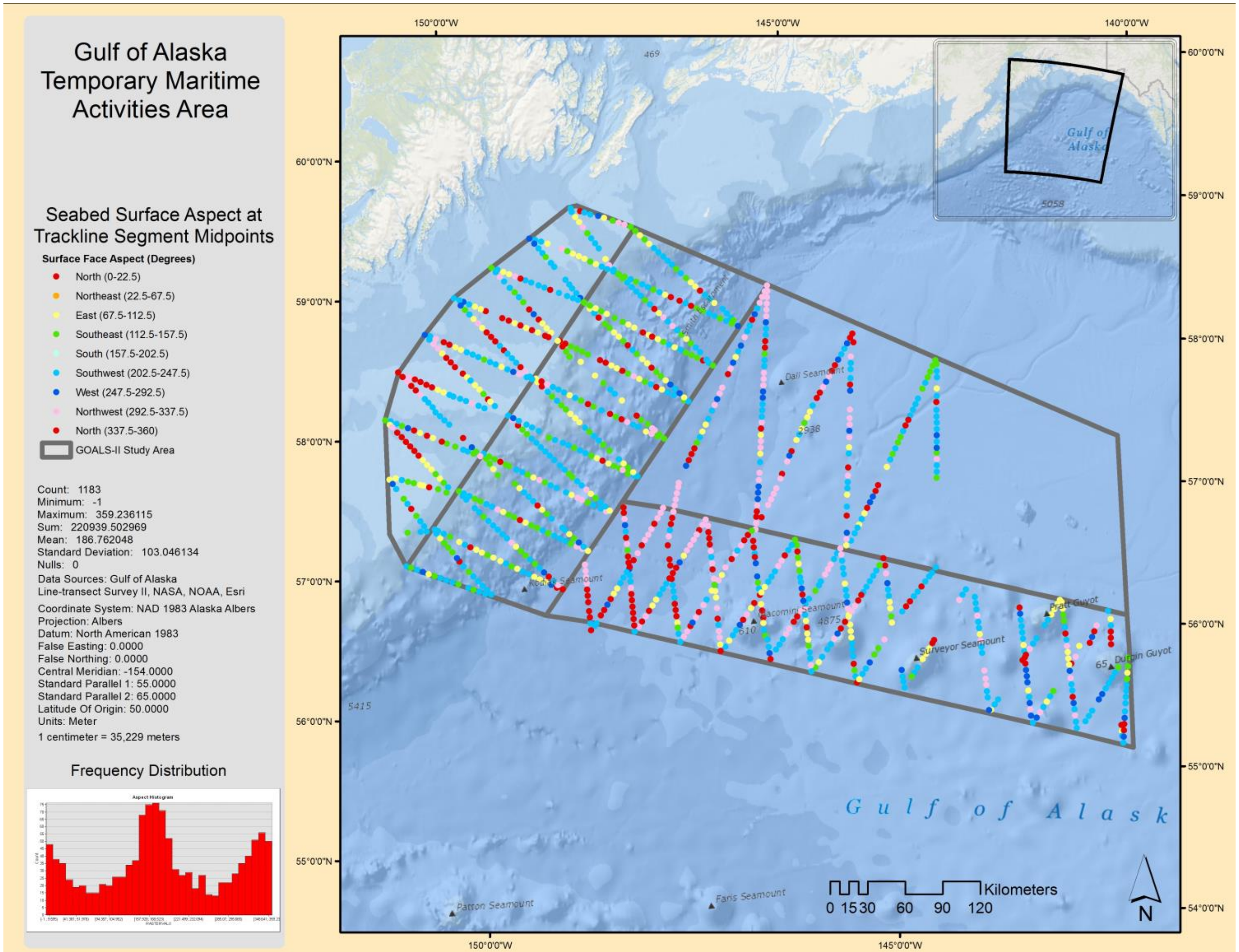


Figure 15. Map displaying seabed aspect found at the location of each trackline segment midpoint. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

4.1.1.4 Distance to 2000 m isobath

Encounters were found to be related to fixed spatial features such as distance to the 2000 m isobath. The 2000 m isobath is considered to be a primary fixed environmental variable because it has shown to be associated with upwelling and biomass density (Forney et al., 2012). This section compares the difference between isobath distance values obtained for localized encounters and all trackline segment midpoints, shown in Figures 16 and 17, with reference to Table 9.

186 samples were utilized to calculate statistics for localized encounters, and 1,183 samples were used to calculate statistics for segment midpoints. The minimum and maximum values were similar for both localizations and trackline segments, at 0.062 km and 0.023 km, respectively. The maximum values of localizations and midpoints were insignificant, at 298.50 km and 367.03 km, respectively.

The two columns differ in their mean values, showing a statistically significant difference between presence and presence/absence data. Localized encounter locations showed a mean value of 44.05 km while segment midpoints yielded a mean of 99.54 km. This shows that sperm whales were found about 55 km closer to the 2000 m isobath than the sampled survey trackline.

Localizations had a standard deviation of 44.72 km, while trackline segments had a standard deviation of 106.74 km from the mean. This means that localization values for depth are clustered with less deviation from the mode. No null values were generated. Localized acoustic encounters were found to be most common within 50 km of the 2000 m contour line. Figures 16 and 17 demonstrate frequency in the histogram located in the bottom left-hand corner of each figure.

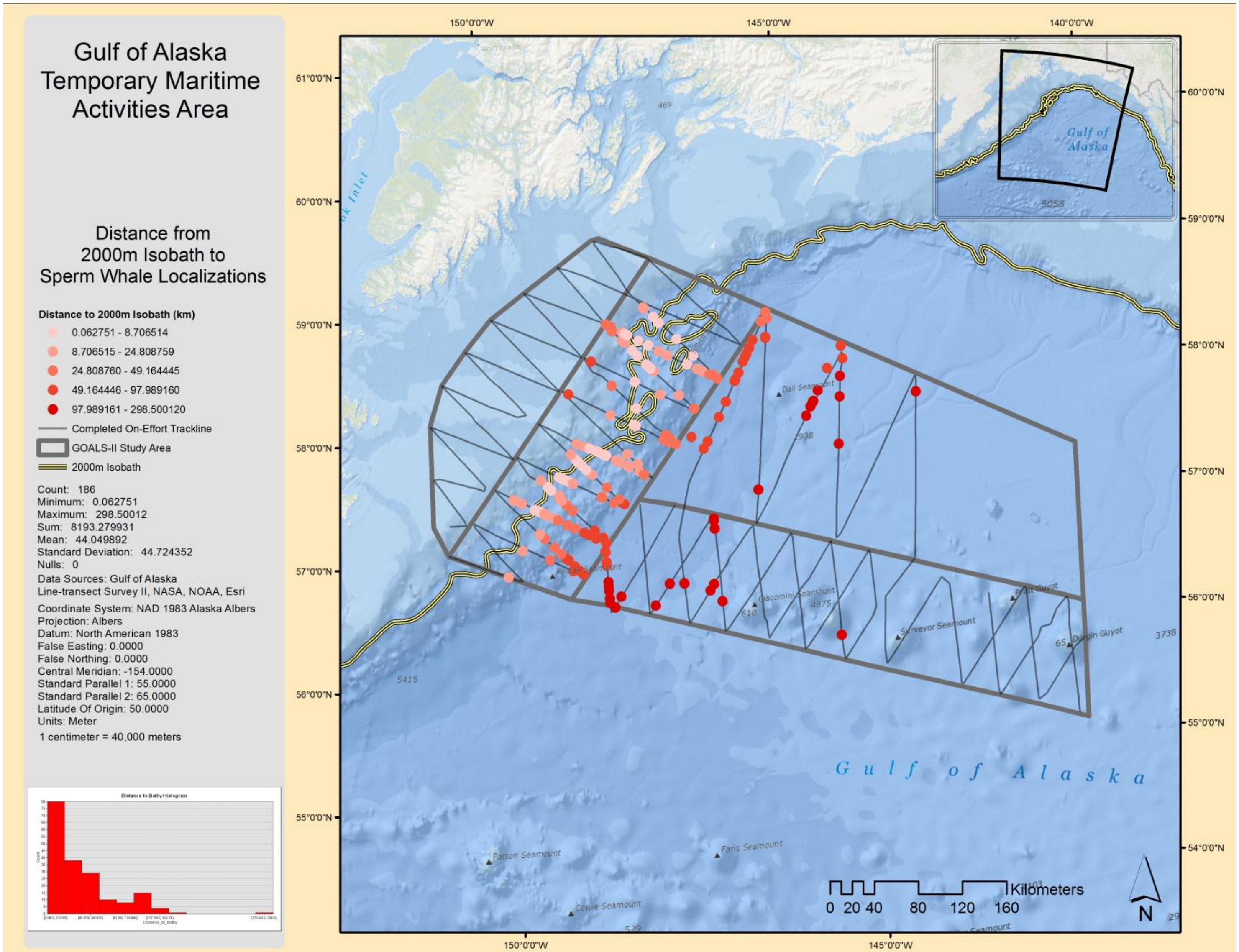


Figure 16. Map displaying the distance from 2000 m isobath for each location of each sperm whale localization. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

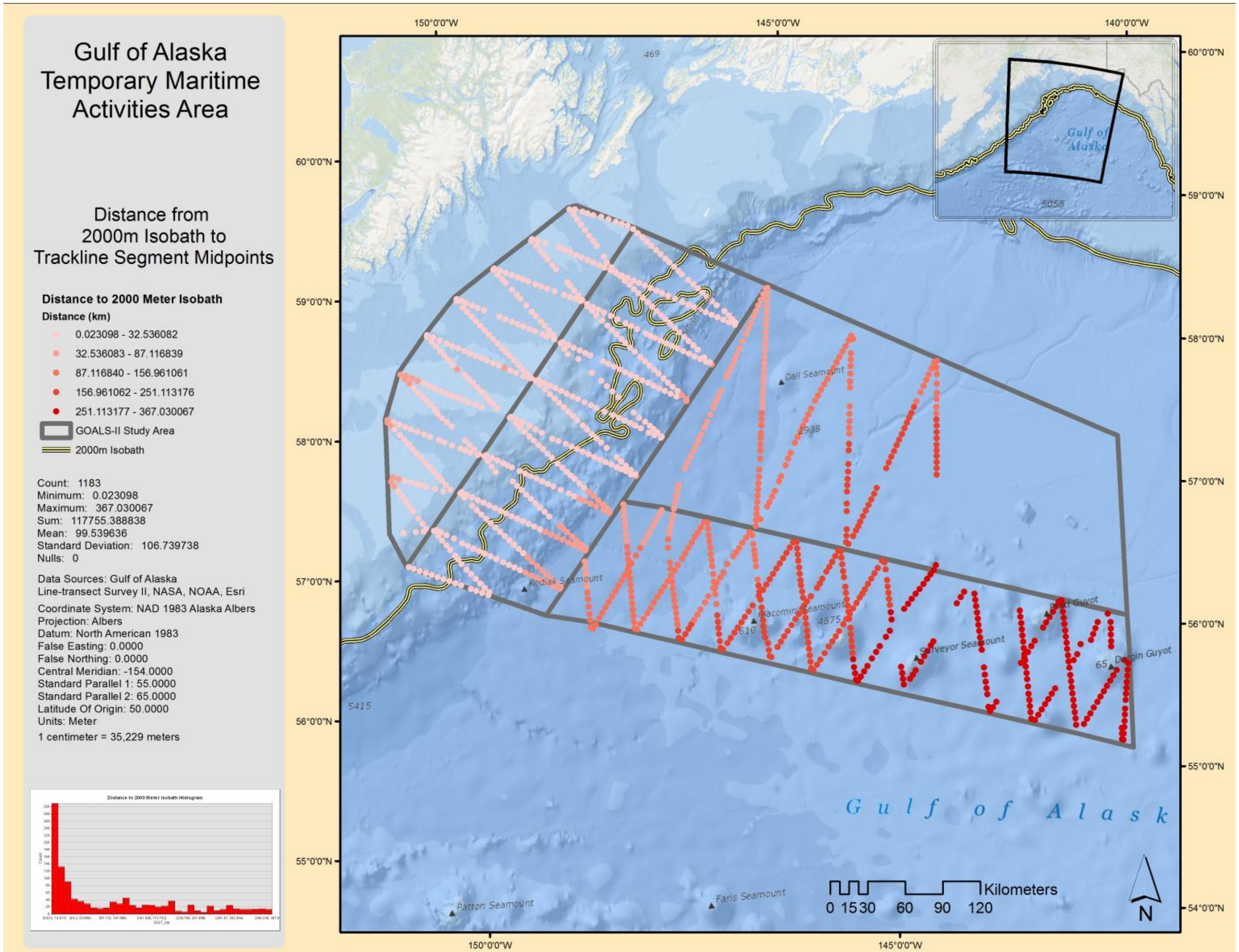


Figure 17. Map displaying distance from 2000 m isobath for each trackline segment midpoint. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

Table 9. Distance statistics of acoustic localizations versus all trackline segment midpoints.

	Localized Encounters	Segment Midpoints
Count:	186	1183
Minimum:	0.06	0.02
Maximum:	298.50	367.03
Mean:	44.05	99.54
Standard Deviation:	44.72	106.74
Nulls:	0	0

4.1.2 Dynamic Oceanographic Variables

Dynamic oceanographic variables include sea surface temperature, sea surface salinity, moon phase, current direction, current magnitude, and chlorophyll concentration.

4.1.2.1 Sea Surface Temperature

Sea surface temperature has been linked to proxies representing prey density, chlorophyll-a concentration, and mixed layer depth (Barlow and Taylor, 2005). This section compares the difference between SST values obtained for localized encounters and all trackline segment midpoints, shown in Figures 18 and 19, with reference to Table 10.

165 samples were utilized to calculate statistics for localized encounters, and 993 samples were used to calculate statistics for segment midpoints. The minimum were similar for both localizations and trackline segments, at 10.08 degrees C and 8.05 C, respectively. The maximum values of localizations and midpoints were also similar, at 14.29 degrees C and 15.03 degrees C, respectively.

Their mean values show a difference between presence and presence/absence data. Localized encounter locations showed a mean value of 12.09 degrees C while segment midpoints yielded a mean of 12.03 degrees C.

Sperm whales were found in waters about 0.06 warmer than those in the sampled survey trackline dataset. Localizations had a standard deviation of 1.15 degrees C, while trackline segments had a standard deviation of 1.17 degrees C from the mean. This means that localization values for depth are clustered with less deviation from the mode.

21 null values were generated for the localized encounter and 190 were generated for the segment midpoints. Localized acoustic encounters were most correlated with sea surface temperatures between 11-12 degrees. This can be attributed to their preference for deep-water coastal regions typically found with warmer sea surface temperatures.

Table 10. SST statistics of acoustic localizations versus all trackline segment midpoints.

	Localized Encounters	Segment Midpoints
Count:	165	993
Minimum:	10.08	8.05
Maximum:	14.29	15.034
Mean:	12.09	12.03
Standard Deviation:	1.15	1.17
Nulls:	21	190

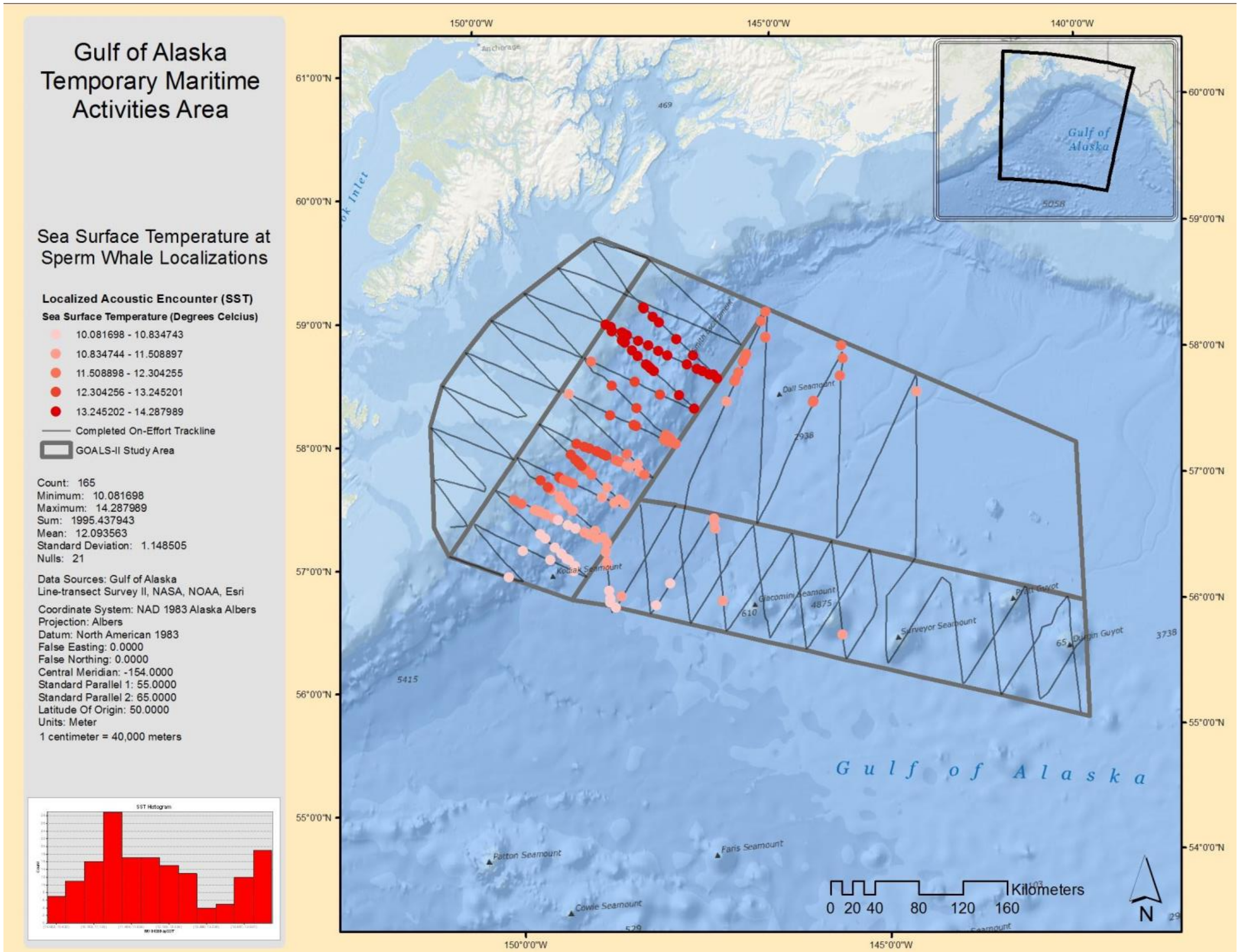


Figure 18. Map displaying sea surface temperature found at the locations of each sperm whale localization. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

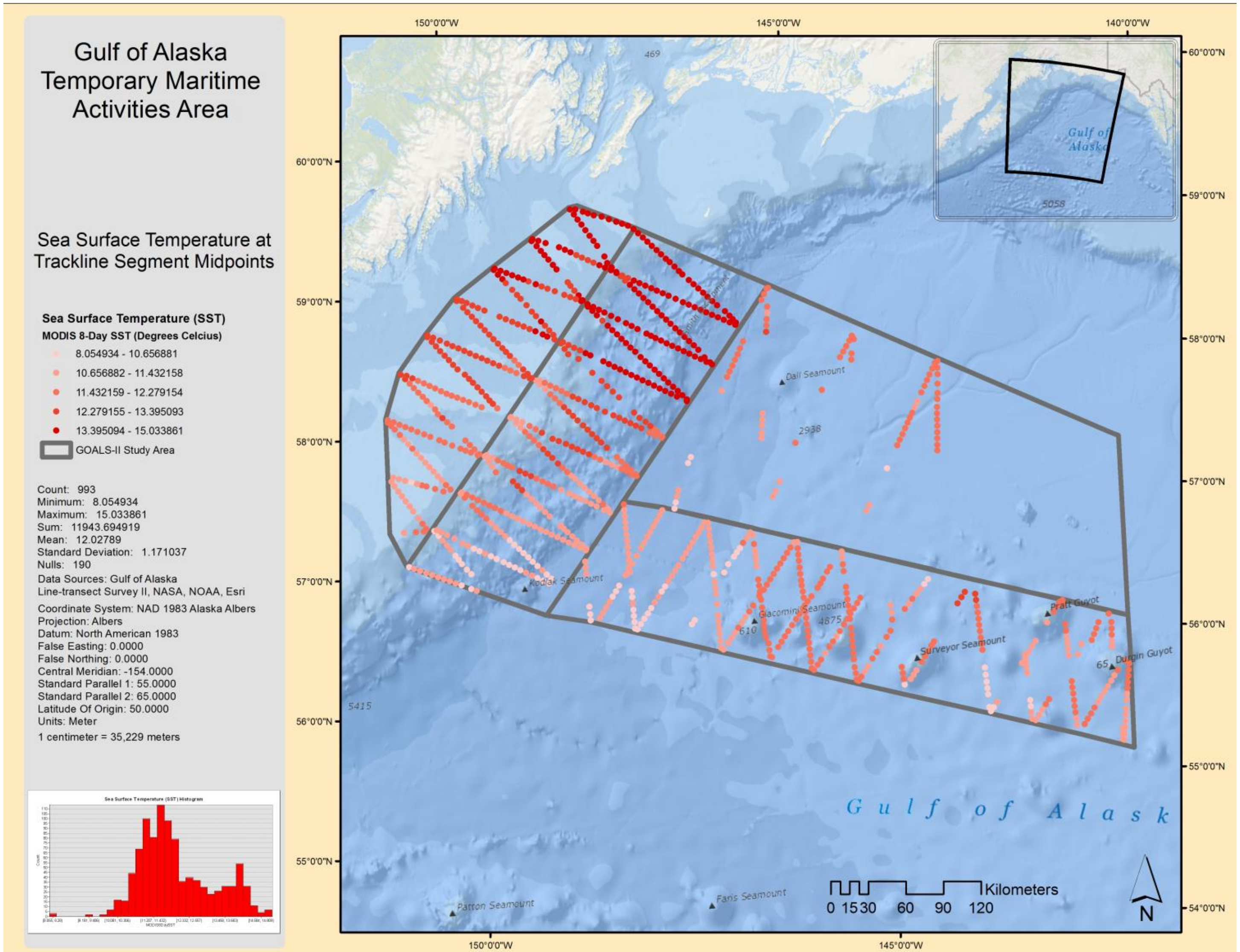


Figure 19. Map displaying sea surface temperature found at the locations of each trackline segment midpoint. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

4.1.2.2 Sea Surface Salinity

The relationship between localized encounters and sea surface salinity was explored in this section. Sea surface salinity is usually higher in deep, offshore waters while it remains lower in shallow onshore regions. This section compares the difference between SSS values obtained for localized encounters and all trackline segment midpoints, shown in Figures 20 and 21, with reference to Table 11.

186 samples were utilized to calculate statistics for localized encounters, and 1,149 samples were used to calculate statistics for segment midpoints. The minimum values were similar for both localizations and trackline segments, at 31.86 PSU and 31.56 PSU, respectively. The maximum values of localizations and midpoints were also similar, at 32.60 PSU and 32.67 PSU, respectively.

The two columns differ slightly in their mean values, showing a small difference between presence and presence/absence data. Localized encounter locations showed a mean value of 32.25 PSU while segment midpoints yielded a mean of 32.27 PSU.

This shows that sperm whales were found in waters about 0.02 PSU less saline than the sampled survey trackline. Localizations had a standard deviation of 0.18 PSU while trackline segments had a standard deviation of 0.26 PSU from the mean. This means that localization values for SSS are clustered with less deviation from the mode. A total of 34 null values were generated for the segment midpoints. Localized acoustic encounters were found to have a weak correlation with sea surface salinity around 32.25 PSU. Due to the weak correlation found between the georeferenced locations of acoustic localizations and sea surface salinity, this predictor variable was not inputted into the GAM.

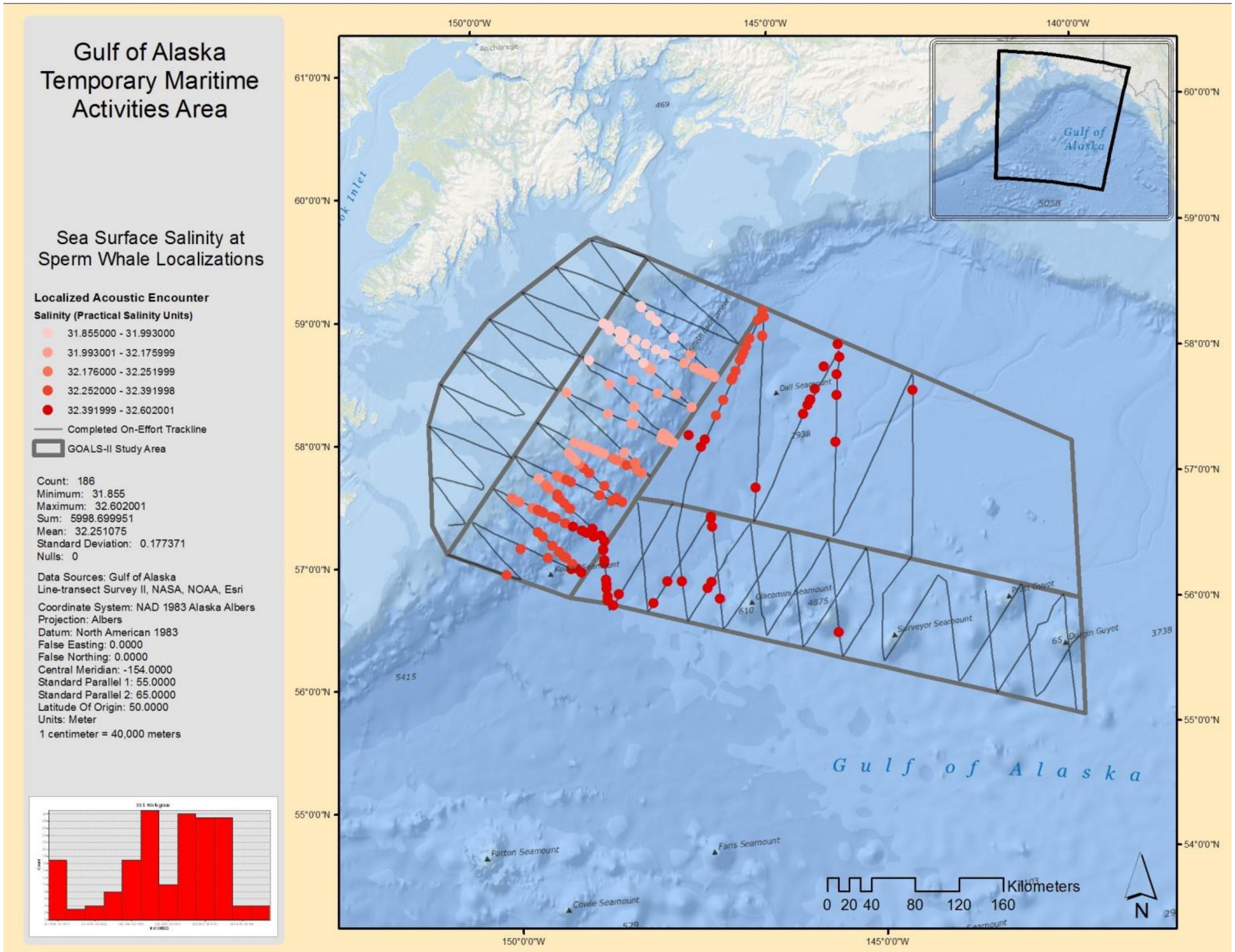


Figure 20. Map displaying sea surface salinity found at the locations of each sperm whale localization. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

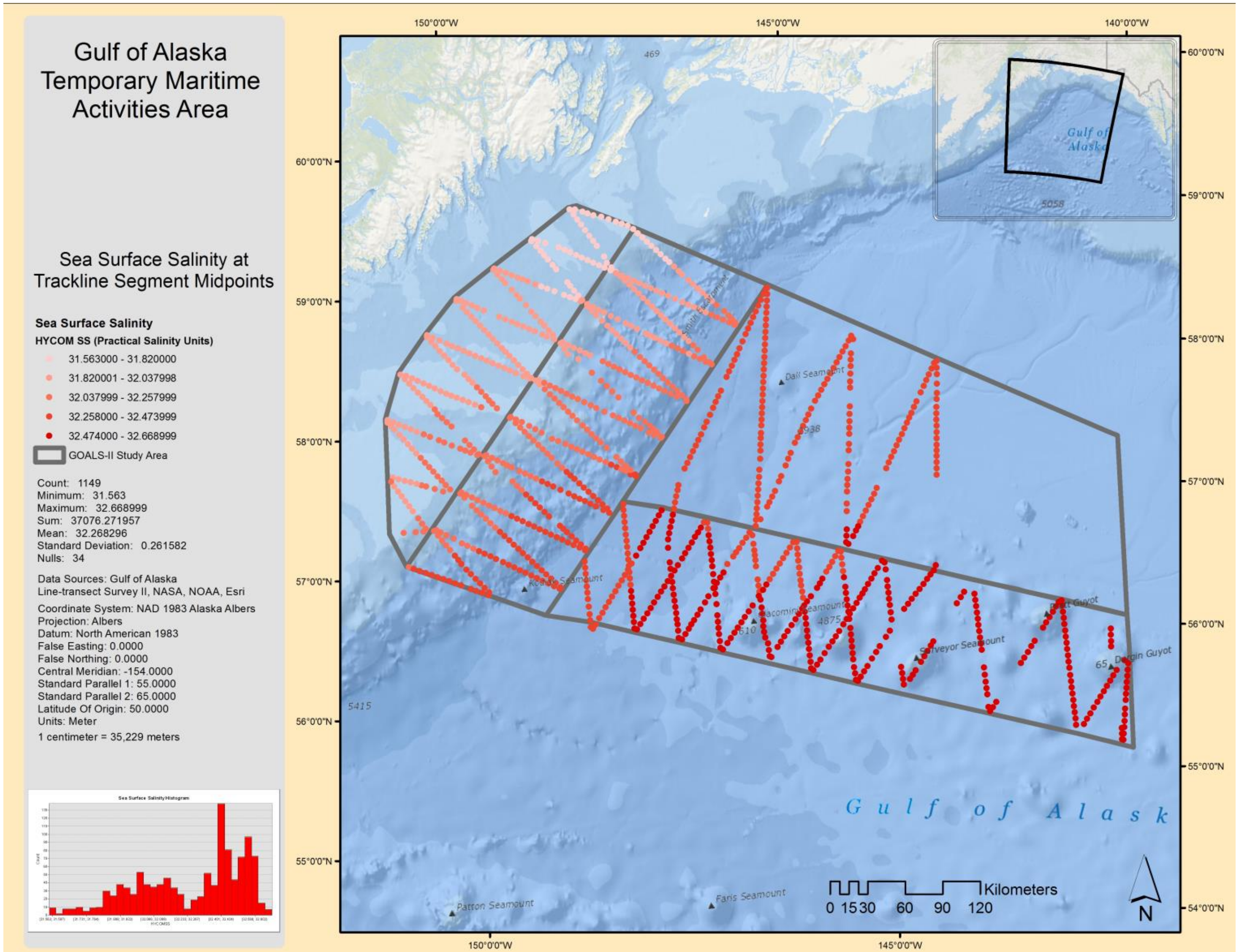


Figure 21. Map displaying sea surface salinity found at the locations of each trackline segment midpoint. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

Table 11. SSS statistics of acoustic localizations versus all trackline segment midpoints.

	Localized Encounters	Segment Midpoints
Count:	186	1149
Minimum:	31.86	31.56
Maximum:	32.60	32.67
Mean:	32.25	32.27
Standard Deviation:	0.18	0.26
Nulls:	0	34

4.1.2.3 Moon phase

Encounters were found to have a weak correlation to certain dynamic oceanographic variables such as moon phase. This section compares the difference between depth values obtained for localized encounters and all trackline segment midpoints, shown in Figures 22 and 23, with reference to Table 12.

A total of 186 samples were utilized to calculate statistics for localized encounters, and 1,183 samples were used to calculate statistics for segment midpoints. Numeric moon phase of the moon ranges from 0 to 0.999999, where 0 is a new moon, 0.25 is the first quarter, 0.5 is full, and 0.75 is the third quarter. The minimum and maximum values were identical for both localizations and trackline segments, at 0.02 and 0.02, respectively.

The maximum values of localizations and midpoints were also identical, at 0.99 and 0.99, respectively. The two columns differ in their mean values, showing a drastic difference in presence and presence/absence data. Localized encounter locations showed a mean value of 0.35, while segment midpoints yielded a mean of 0.51.

This shows that sperm whales were found closer to a new moon than survey trackline samples. Localizations had a standard deviation of 0.39, while trackline segments had a standard deviation of 0.35 from the mean. This means that localization values for depth are less clustered with more deviation from the mode. No null values were generated. Localized acoustic encounters were found to have a weak correlation with numeric moon phase, with encounters occurring earlier in the moon cycle than trackline values of the survey.

Table 12. Moon phase statistics of acoustic localizations versus all trackline segment midpoints.

	Localized Encounters	Segment Midpoints
Count:	186	1183
Minimum:	0.02	0.02
Maximum:	0.99	0.99
Mean:	0.35	0.51
Standard Deviation:	0.39	0.35
Nulls:	0	0

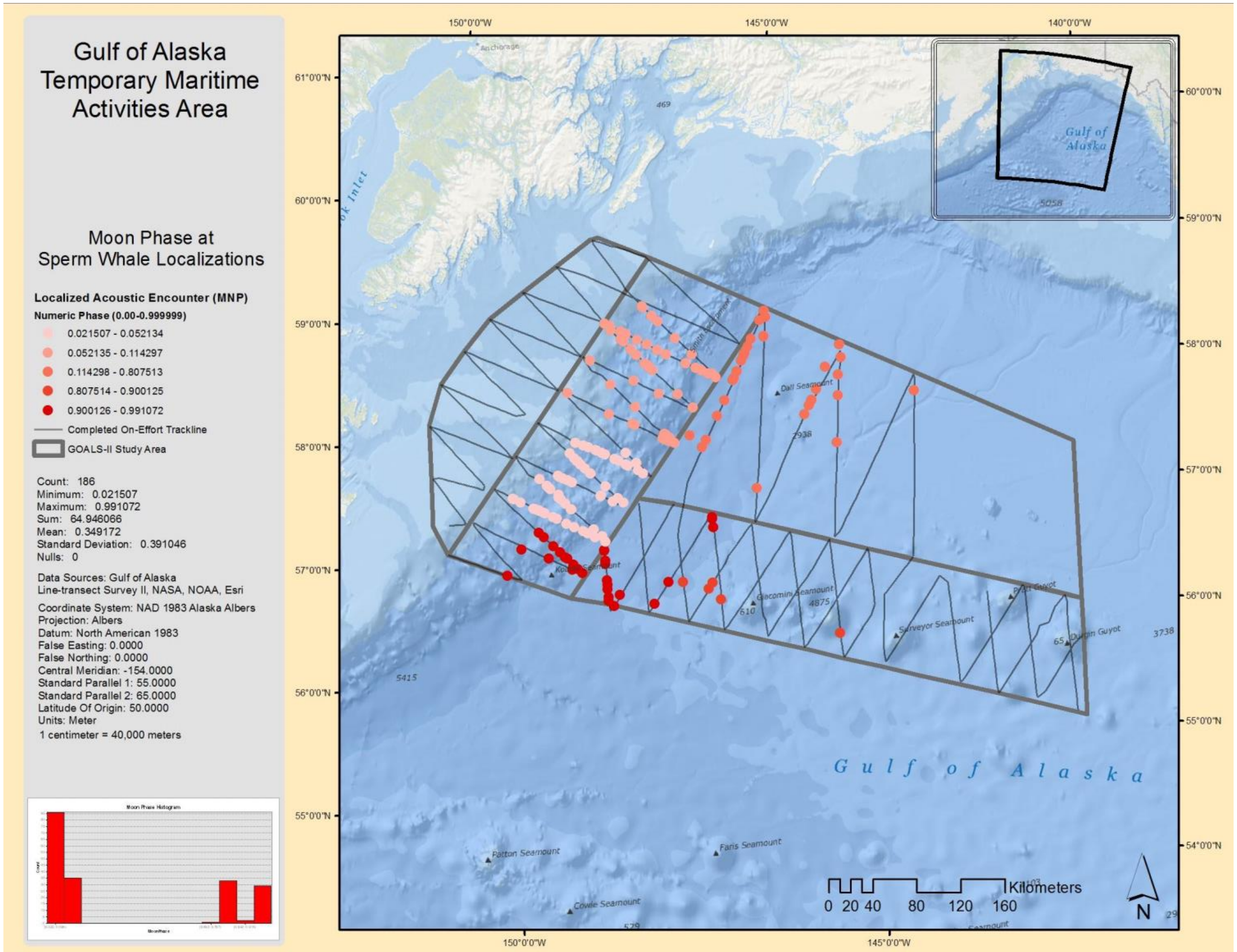


Figure 22. Map displaying moon phase on the survey date of each sperm whale localization. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

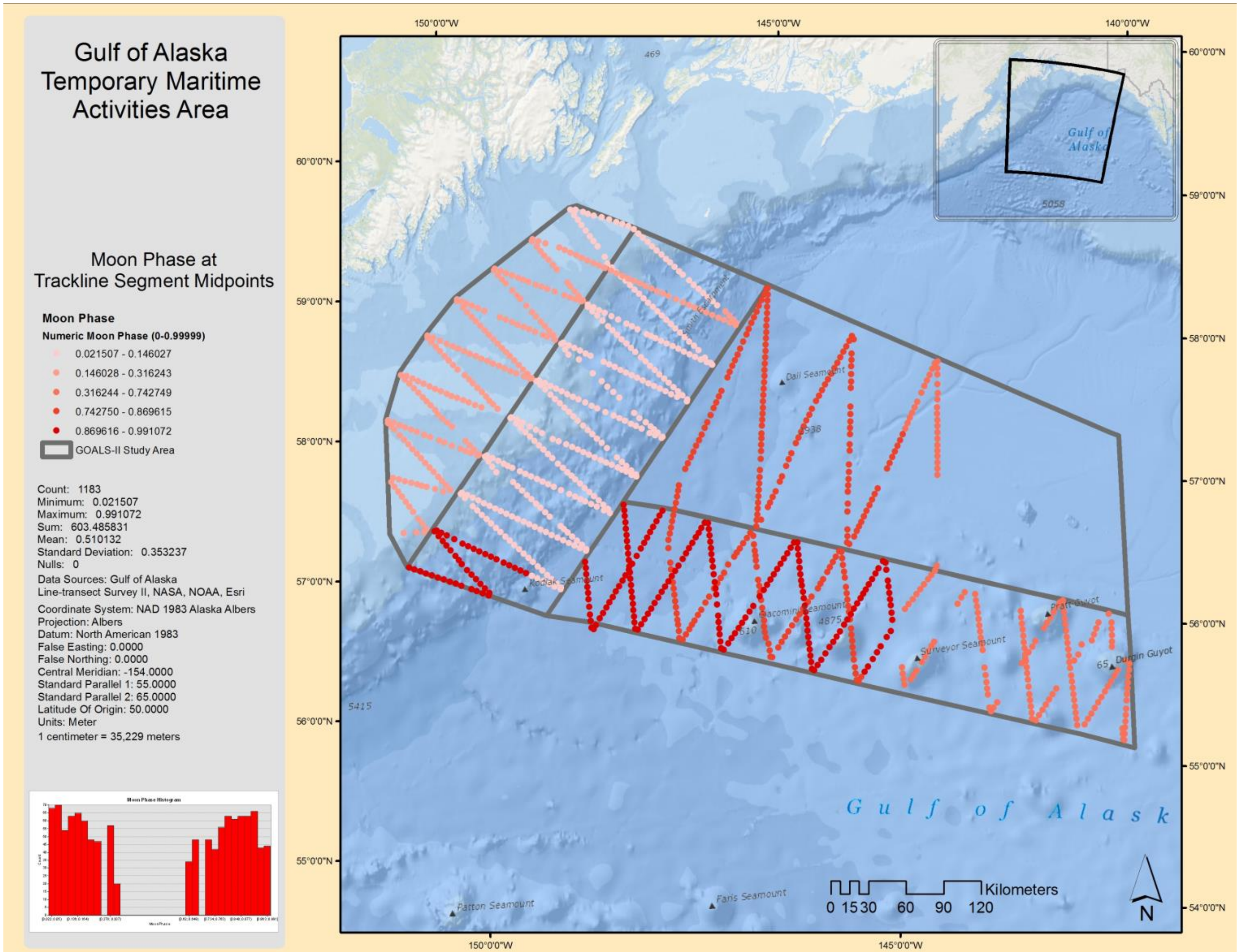


Figure 23. Map displaying moon phase on the survey date of each trackline segment midpoint. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

4.1.2.4 Current magnitude

Encounters were found to be related to dynamic oceanographic variables such as current magnitude measured in TKE. Current magnitude was calculated using the OSCAR model vector. This section compares the difference between TKE values obtained for localized encounters and all trackline segment midpoints, shown in Figures 24 and 25, with reference to Table 13. From the map figures, it is possible to see that TKE is typically more robust in deeper, offshore regions.

186 samples were utilized to calculate statistics for localized encounters, and 1,183 samples were used to calculate statistics for segment midpoints. The minimum and maximum values were similar for both localizations and trackline segments, at 1.65 x 10^{-4} m^2/s^2 and 1.7 x 10^{-5} m^2/s^2, respectively.

The maximum values of localizations and midpoints were also similar, at 0.042 m^2/s^2 and 0.047 m^2/s^2, respectively. The two columns differ in their mean values, showing a drastic difference in presence and presence/absence data.

Localized encounter locations showed a mean value of 5.8 x 10^{-3} m^2/s^2 while segment midpoints yielded a mean of 7.9 x 10^{-3} m^2/s^2. This shows that sperm whales were found in waters with about 2.1 x 10^{-3} m^2/s^2 less TKE than standard survey trackline. Localizations had a standard deviation of 6.6 x 10^{-3} m^2/s^2, while trackline segments had a standard deviation of 8.3 x 10^{-3} $m^2/s^2$from the mean.

This means that localization values for TKE are clustered with less deviation from the mode. No null values were generated. Total kinetic energy for the majority of localized encounter samples is between 0.00-0.01 m^2/s^2. The best-fit model chose current magnitude as one of the predictor variables for sperm whale habitat within the TMAA.

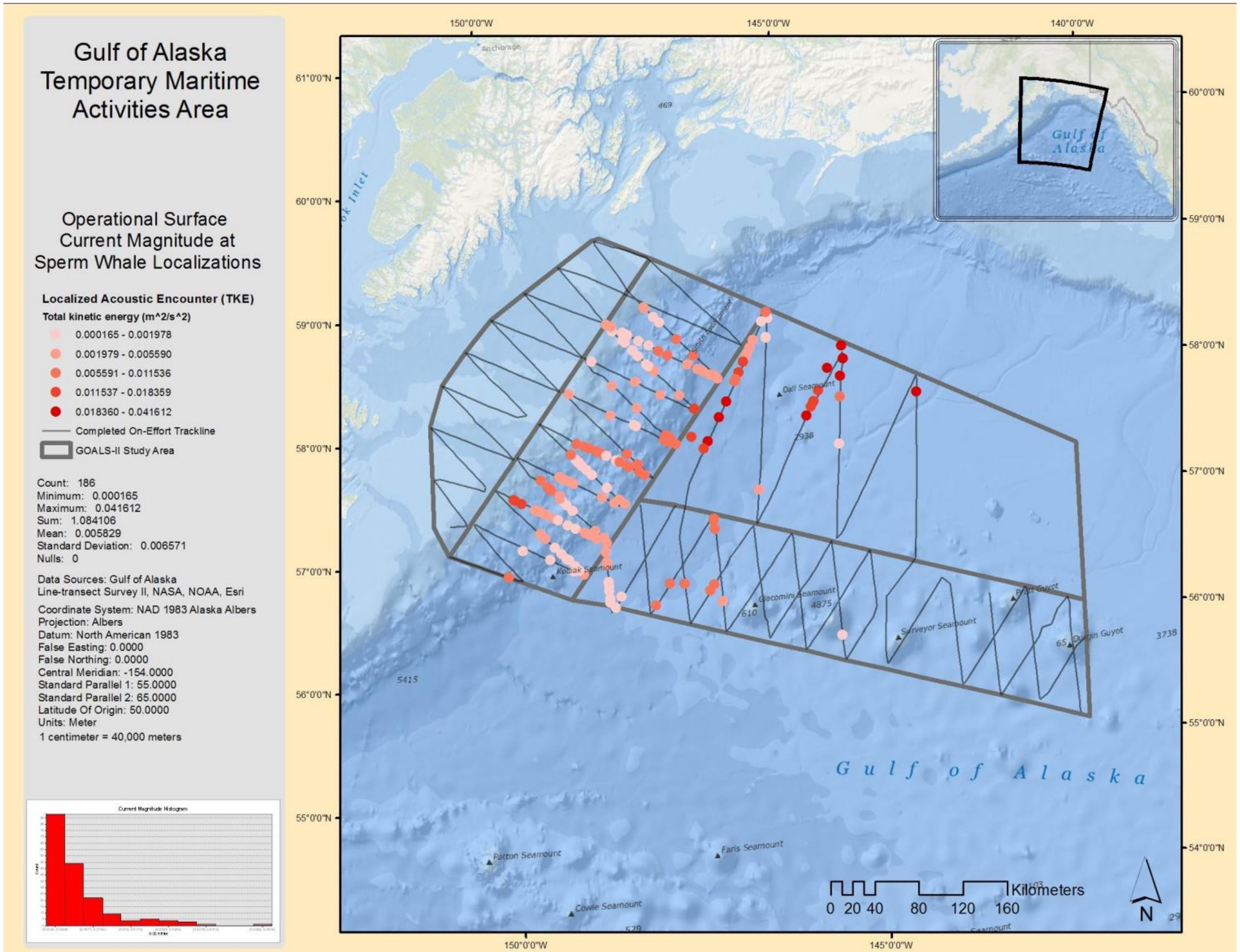


Figure 24. Map displaying ocean current magnitude found at the location of each sperm whale localization. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

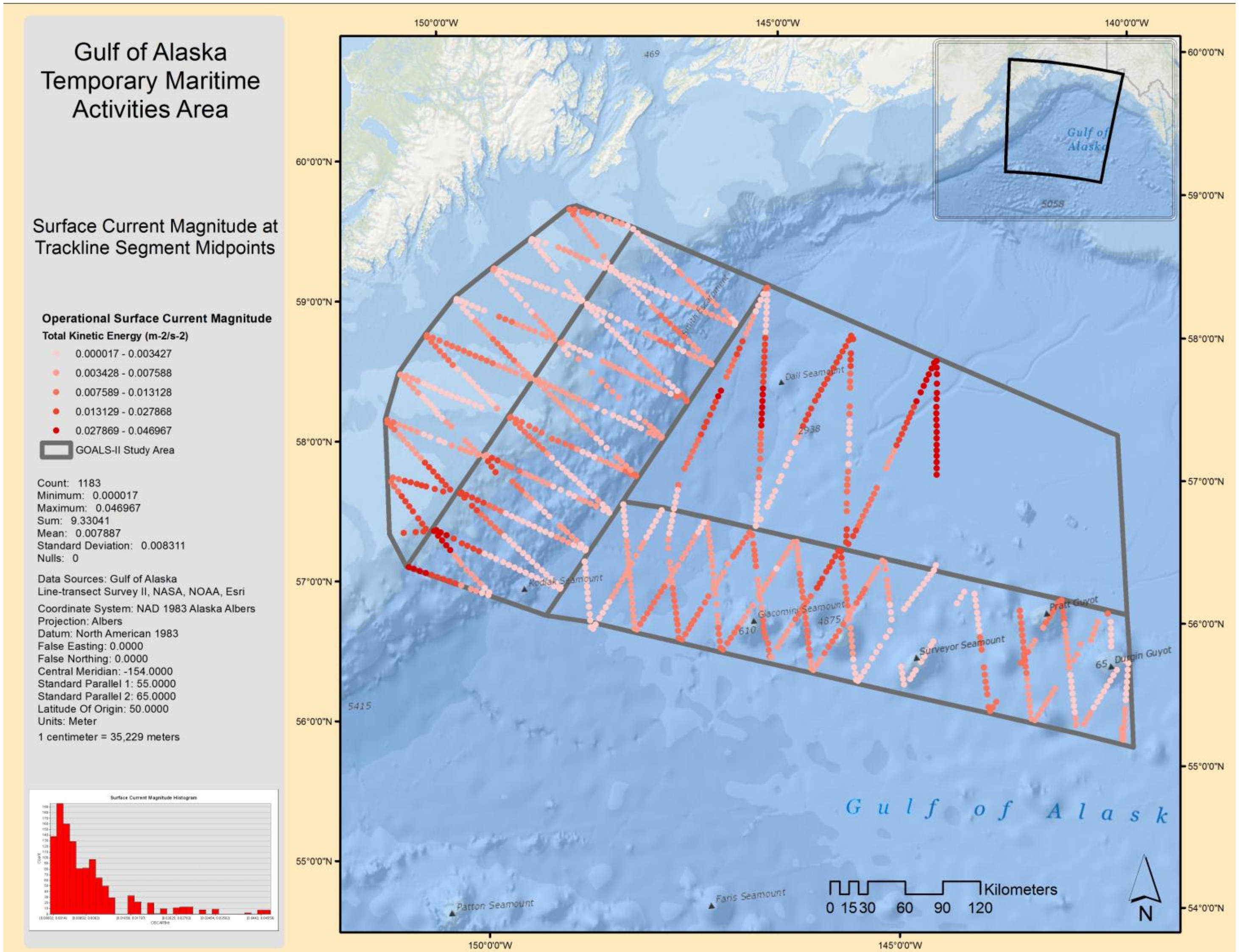


Figure 25. Map displaying ocean current magnitude found at the locations of each trackline segment midpoint. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

Table 13. Current magnitude statistics of acoustic localizations versus all trackline segment midpoints.

	Localized Encounters	Segment Midpoints
Count:	186	1183
Minimum:	0.00017	0.000017
Maximum:	0.042	0.047
Mean:	0.0058	0.0079
Standard Deviation:	0.0066	0.0083
Nulls:	0	0

4.1.2.5 Current direction

Current direction, a dynamic oceanographic variable, was explored to predict sperm whale habitat. This section compares the difference between degrees bearing values obtained for localized encounters and all trackline segment midpoints, where 0 degrees indicates an ocean current directed toward true north, and so forth.

A total of 186 samples were utilized to calculate statistics for localized encounters, and 1,183 samples were used to calculate statistics for segment midpoints, shown in Figures 26 and 27, with reference to Table 14. The minimum values were identical for both localizations and trackline segments, at 2.69 degrees and 2.69 degrees, respectively. The maximum values of localizations and midpoints were also similar, at 343.32 degrees and 359.41 degrees, respectively.

The two columns differ in their mean values, showing a drastic difference in presence and presence/absence data. Localized encounter locations showed a mean value of 167.57 degrees while segment midpoints yielded a mean of 188.82 degrees.

This shows that sperm whales were found about in areas with currents that flow about 21.25 degrees less than sampled survey trackline, more northwest. Localizations had a standard deviation of 103.61 degrees, while trackline segments had a standard deviation of 101.25 degrees from the mean. This means that localization values for depth are less clustered with more deviation from the mode. No null values were generated. Localized acoustic encounters were found to peak with current direction of flow between 200-250 degrees.

Table 14. Current direction statistics of acoustic localizations versus all trackline segment midpoints.

	Localized Encounters	Segment Midpoints
Count:	186	1183
Minimum:	2.69	2.69
Maximum:	343.32	359.41
Mean:	167.57	188.82
Standard Deviation:	103.61	101.25
Nulls:	0	0

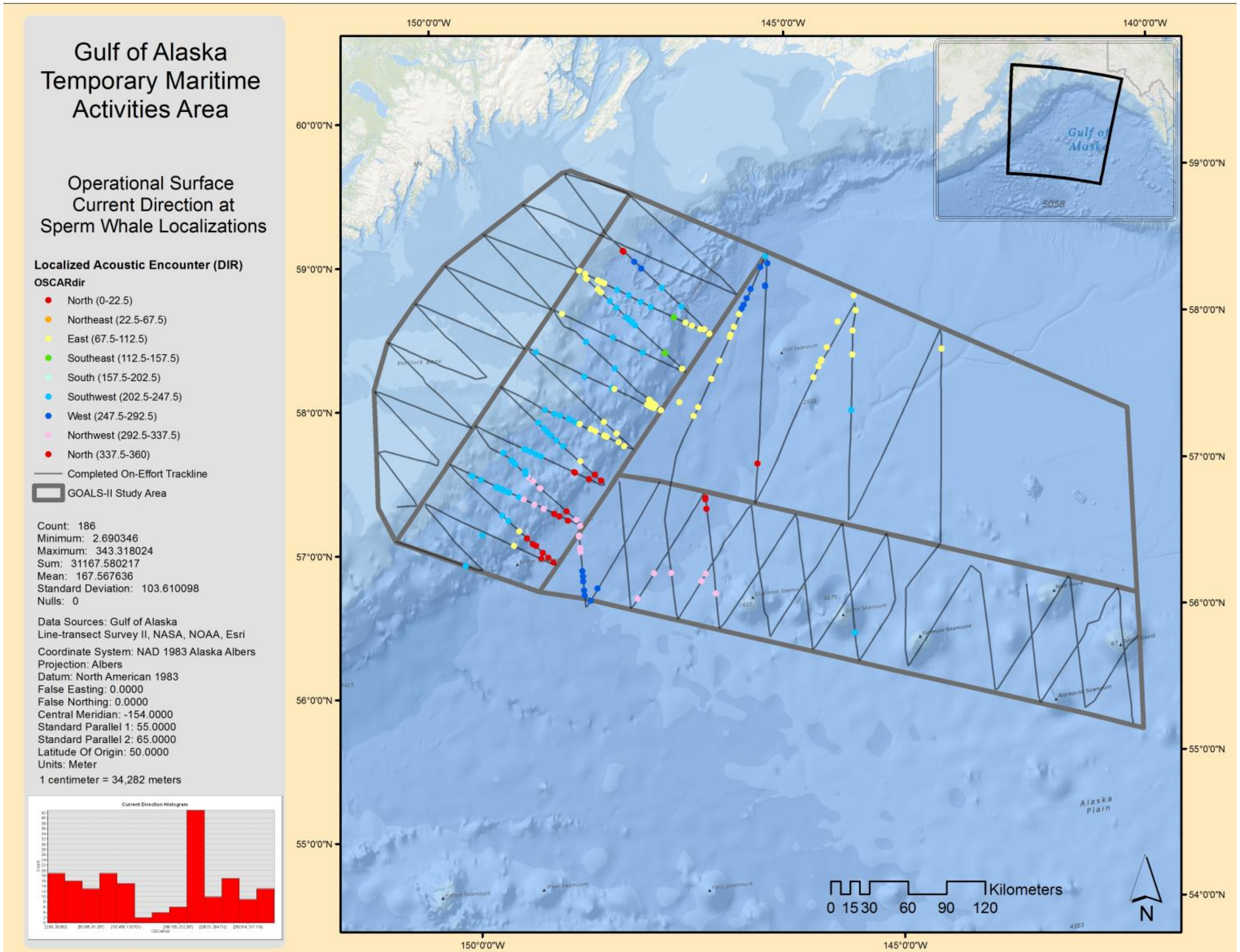


Figure 26. Map displaying the current direction found at the location of each sperm whale localization. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

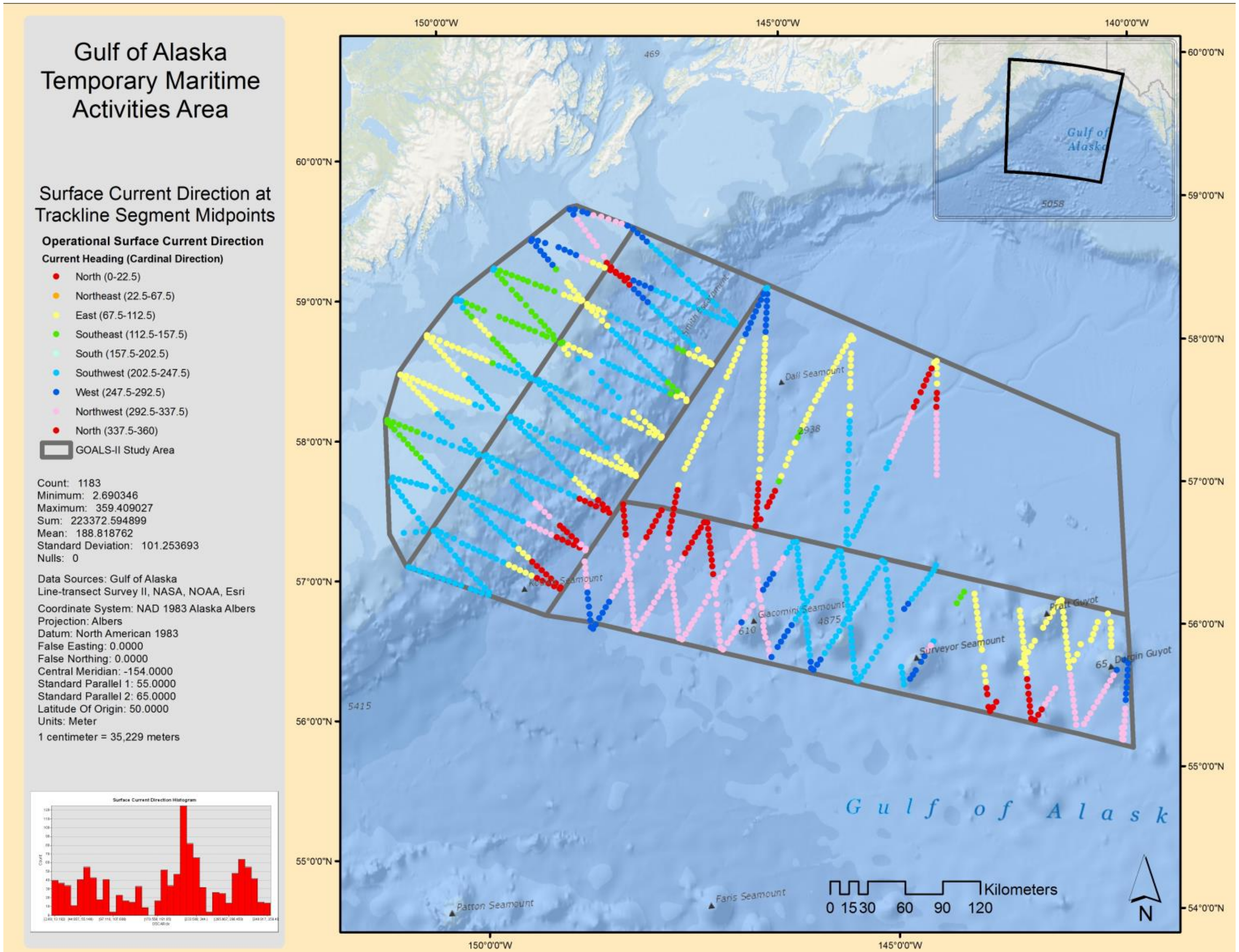


Figure 27. Map displaying the current direction found at the location of each trackline segment midpoint. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

4.1.2.6 Chlorophyll-a concentration

Encounters were found to be related to dynamic oceanographic variables such as chlorophyll-a concentration. Chlorophyll-a is considered to be a primary fixed environmental variable because of the effect that phytoplankton has on attracting cephalopods and other sperm whale prey species (Clarke and Pascoe, 1997). This section compares the difference between remotely sensed chlorophyll values obtained for localized encounters and all trackline segment midpoints, shown in Figures 28 and 29, with reference to Table 15.

A total of 147 samples were utilized to calculate statistics for localized encounters, and 907 samples were used to calculate statistics for segment midpoints. The minimum values were similar for both localizations and trackline segments, at 0.42 mg/m^3 and 0.27 mg/m^3, respectively. The maximum values of localizations and midpoints were also similar, at 4.03 mg/m^3 and 7.06 mg/m^3, respectively. The two columns diverge in their mean values, showing a noticeable difference in presence and presence/absence data. Localized encounter locations showed a mean value of 1.72 mg/m^3 while segment midpoints yielded a mean of 1.49 mg/m^3.

This shows that sperm whales were found in waters with about 0.22 mg/m^3 more chlorophyll-a concentration than survey trackline samples. Localizations had a standard deviation of 0.86 mg/m^3 while trackline segments had a standard deviation of 0.82 mg/m^3 from the mean. This means that localization values for depth are less clustered with more deviation from the mode. 39 null values were generated for localized encounters, and 276 null values were assigned for segment midpoints. Localized acoustic encounters were found to have a strong correlation with chlorophyll-a concentrations around 0.5-2.5 mg/m^3.

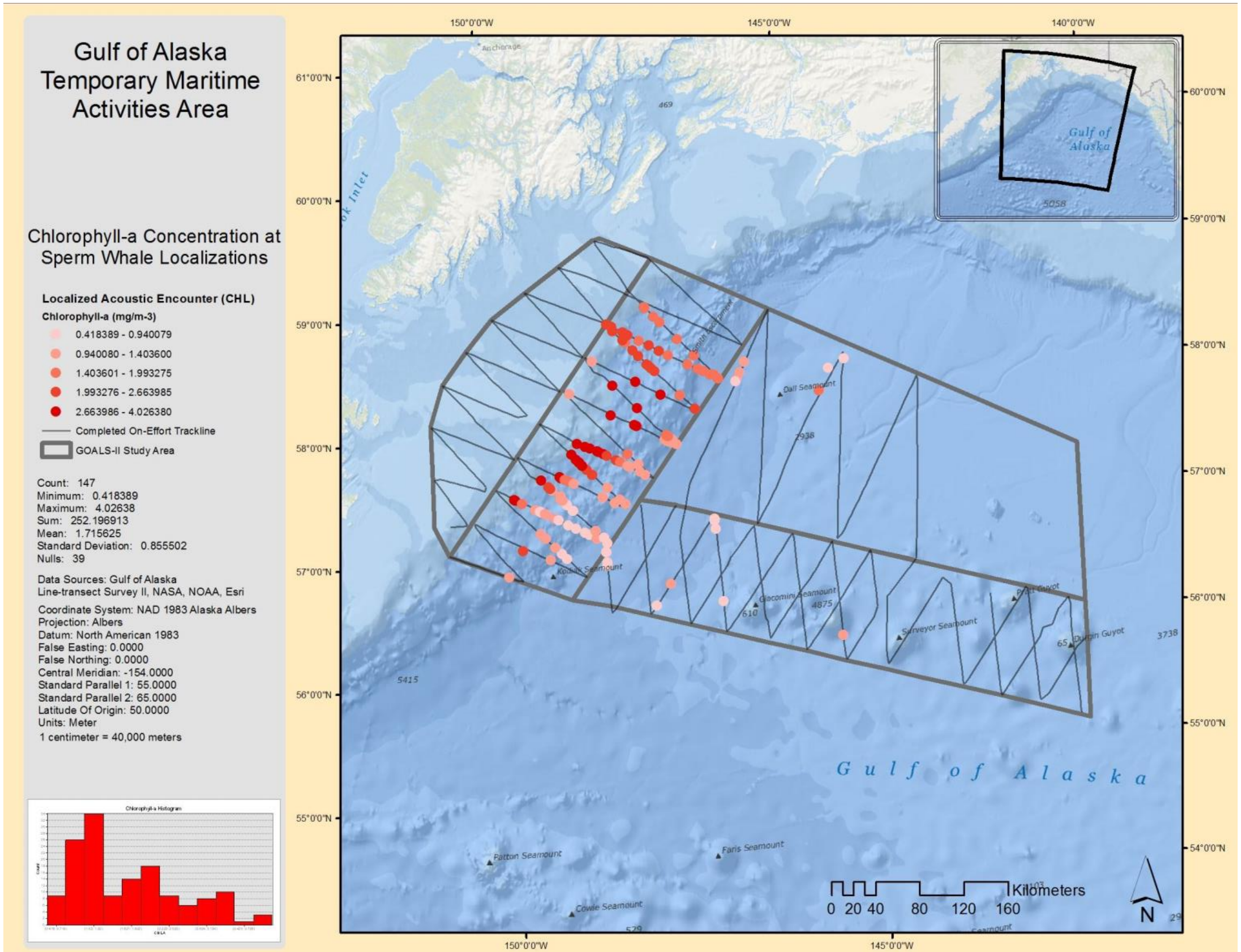


Figure 28. Map displaying chlorophyll-a concentration found at the location of each sperm whale localization. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

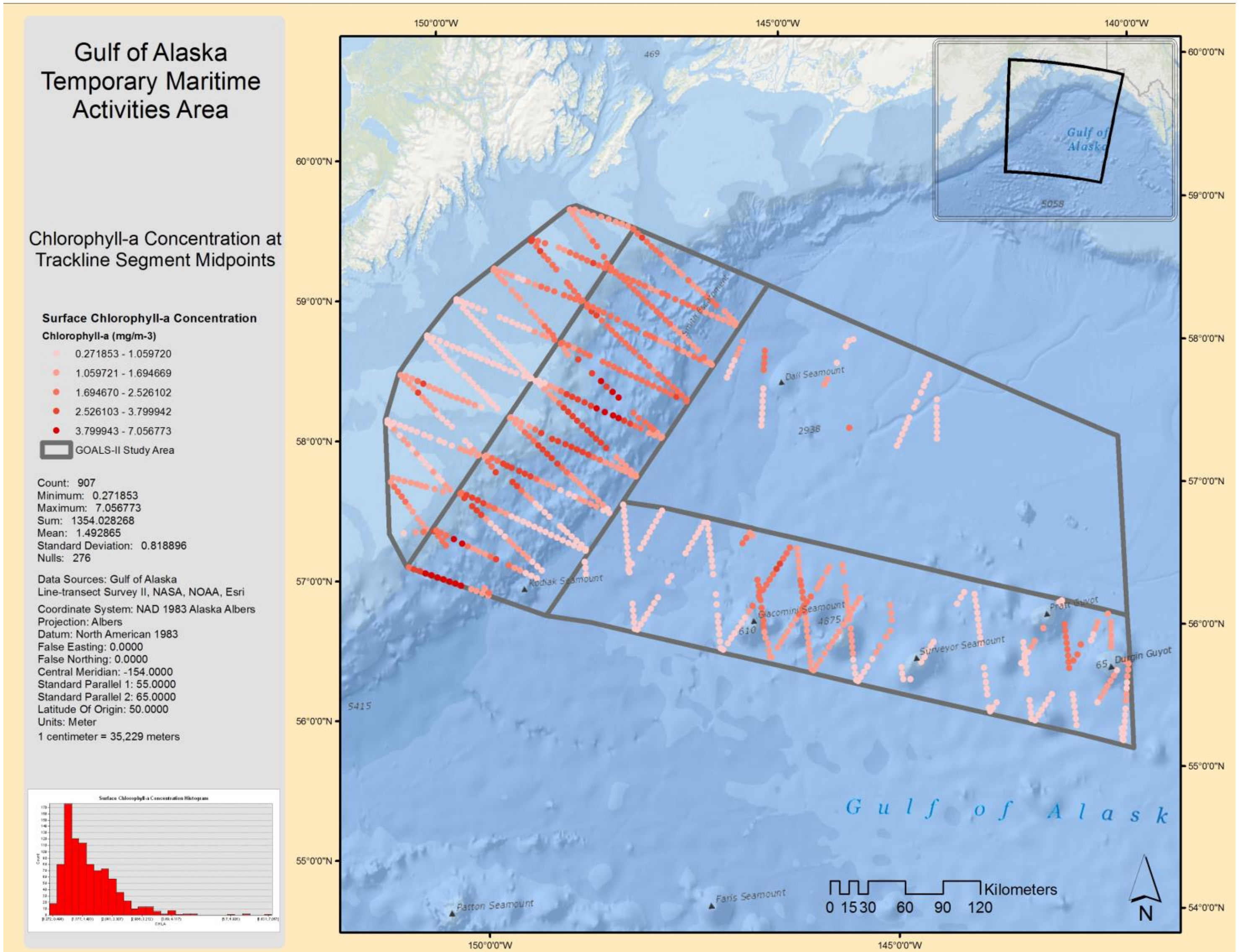


Figure 29. Map displaying chlorophyll-a concentration found at the location of each trackline segment midpoint. An abbreviated frequency distribution plot is located on the lower-left corner of the figure.

Table 15. Chlorophyll-a concentration statistics of acoustic localizations versus all trackline segment midpoints.

	Localized Encounters	Segment Midpoints
Count:	147	907
Minimum:	0.42	0.27
Maximum:	4.03	7.06
Mean:	1.72	1.49
Standard Deviation:	0.86	0.82
Nulls:	39	276

4.2 Modeling Results

Four GAMs were initially constructed from varied combinations of predictive model covariates and then were assessed to produce the best-fit model for sperm whale localizations within the study area (Table 16). The final model was then evaluated to understand its predictive power, in addition to covariate properties including the approximate significance of each smoothed term.

Acoustic Model D was determined to be the best-fit model for sperm whale localizations within the study area. The AIC value for this model was 403.7779. This showed that depth, slope, surface chlorophyll, sea surface temperature, distance to the 2,000 m isobaths, and oceanic current magnitude were the most important explanatory variables for sperm whale habitat. A correlation matrix was produced to confirm variable-to-variable relationships indicated no redundancies (Table 17).

Table 16. Statistics of acoustic localizations versus all trackline segment midpoints.

Model	Dispersion	Null	Deviance	Explained Deviance	AIC Score
A	0.709964	612.81	332.99	45.66%	363.03
B	0.793492	619.64	380.91	38.53%	405.54
C	0.86174	619.64	391.14	36.88%	419.67
D	0.792334	619.64	374.98	39.48%	403.78

Table 17. Correlation matrix produced from best-fit GAM showing the variable-to-variable relationship using a range of unit-less values, indicating covariance. Values range from -1 to 1. Positive 1 equates to a perfect correlation, while negative 1 represents disassociation.

	Depth	Slope	Chlorophyll-a	Sea Surface Temperature	Distance to 2000 m	Current Magnitude
Depth	1	0.1419	-0.2191	-0.3647	0.5152	-0.0110
Slope	0	1	0.1044	0.0340	0.08164	-0.1037
Chlorophyll-a	0	0	1	0.2852	-0.3796	0.0427
Sea Surface Temperature	0	0	0	1	-0.4605	-0.2276
Distance to 2000 m	0	0	0	0	1	-0.0041
Current Magnitude	0	0	0	0	0	1

In addition, plots were created to display the functional relationship between localized encounters and predictor variables within the best-fit model (Figure 30). Depth entered the model as a smoothing spline with 2 degrees of freedom (df), displaying maximum encounter rates for depths greater than 1,000 m (Figure 30(a)). Distance to the 2,000 m isobath entered the model as a linear term showing encounter rate decreasing with distance away from the 2,000 m isobath, as is expected (Figure 30(b)). Slope was included in the model as a smoothing spline with 3df showing maximum encounter rates for slopes peaking at about 10 degrees (Figure 30(c)). Sea surface temperature was included in the model as a smoothing spline with 3df, displaying a peak in encounter rate for SST around 12°C (Figure 30(d)). Surface chlorophyll concentration was entered into the model as a linear term and showed decreasing encounter rate as a function of increasing chlorophyll concentration (Figure 30(e)). Current magnitude was also included in the model as a linear term and displayed decreasing encounter rate as a function of increasing magnitude (Figure 30(f)).

Each model covariate was analyzed to understand the effect that environmental variables had on localizations (Table 18). Depth ($F_2 = 8.106$, $p < 0.001$) was found to be the most influencing and correlated environmental variable of the model with a confidence level of 99.9%. Distance to the 2,000 m isobath ($F = 4.139$, $p < 0.001$) was found to be the second most statistically significant variable on the presence of localizations with a confidence level of 99.9%. Sea surface temperature ($F_3 = 2.320$, $p < 0.001$) was found to be the third-ranking influential variable on preferred sperm whale habitat, also with a confidence level of 99.9%. The model suggested that chlorophyll-a ($F = 0.610$, *ns*) had the fourth most important influence on sperm whale localizations although was not significant due to F being less than 1. Slope ($F_3 = 0.590$, *ns*) was ranked the fifth most influential environmental variable although was also not significant due to *F* being less than 1. Current magnitude (F = 0.512, *ns*) was found to have the sixth most influential impact on sperm whale habitat with no significance due to *F* being less than 1. Although chlorophyll-a, slope, and current magnitude were not significant, including them increased the fit of the model as measured by the AIC.

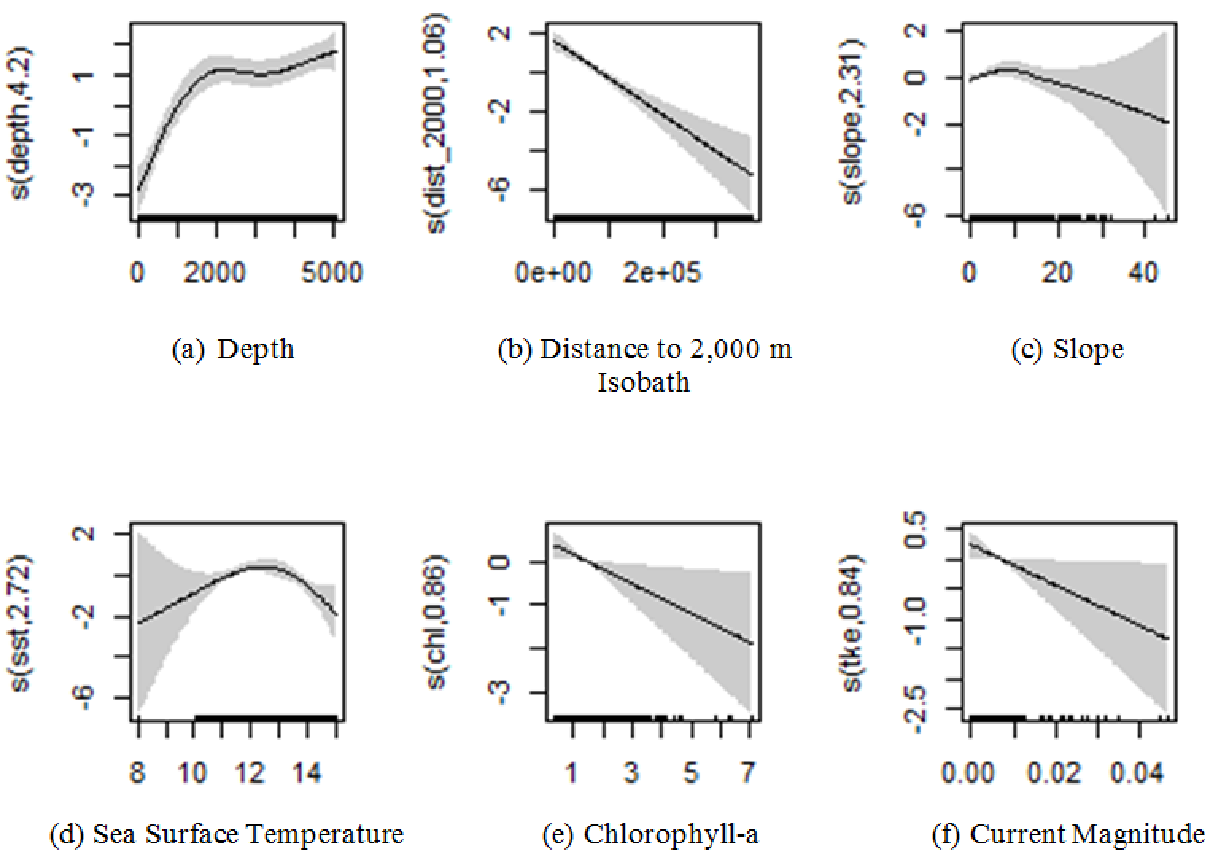


Figure 30. Encounter rate model functions for localized acoustic encounters of sperm whales (Model D). The y-axes represents the term's (linear, spline or polynomial) function. Zero on the y-axes corresponds to no effect of a predictor variable on the estimated response variable (encounter rate). Grey areas represent the amount of statistical uncertainty. Tick marks above the x-axis indicate the distribution of observations in all segments. Variables shown include (a) Depth, (b) Distance to 2,000 m isobath, (c) Slope, (d) Sea Surface Temperature (e) Chlorophyll-a, and (f) Current Magnitude.

The results from this regression analysis indicate that model covariates were in fact contributing to the preference of sperm whale habitat. As a matter of scientific practice, a well-accepted significance level is at least 0.05. Three of the six covariates were found to confidently show a statistically significant relationship by this standard and, therefore, rejecting the null hypothesis. All six variables considered were shown to have a sufficiently low co-variance to be included in the model as separate terms. After determining the effect that covariates had on fitting the model, final prediction maps were produced by selecting sperm whale habitat preferences within the study area.

Table 18. Results showing statistical significance of each covariate within the best-fit model (n=864). The predicting ability rank shows the relative predicting power of each covariate. The mean and standard deviation describe the distribution of the data with respect to each covariate. The F-statistic defines how much the model explains variability in the response. The p-value is the probability of obtaining an equal or more extreme mean, assuming the null hypothesis. The significance levels are related to confidence levels and show the probability of the statistic being incorrect, assuming the null hypothesis.

	Prediction Ability Rank	Mean	Standard Deviation	F-statistic	p-value	Significance level
Depth (m)	1	3293.182796	1273.773038	8.106	< 2e-16	0.001
Distance to 2000 m (km)	2	44.049892	44.724352	4.139	3.82e-10	0.001
Sea Surface Temperature (°C)	3	32.251075	0.177371	2.320	1.57e-05	0.001
Chlorophyll-a (mg/m^3)	4	1.715625	0.855502	0.610	0.00617	0.01
Slope (° gradient)	5	6.007417	6.318839	0.590	0.06851	0.1
Current Magnitude (m^2/s^2)	6	0.005829	0.006571	0.512	0.01605	0.05

When analyzing the final encounter rate distribution map, it is important to understand how the data affects the final output. Unlike presence-only studies, in which only observations are recorded and used to predict habitat, presence and absence were used to map predicted encounter rate using non-linear GAMs. Due to this, the map does not automatically show density around encounters and interpolate like presence-only studies. As an example to show a distribution of samples without model covariates, a kernel density interpolation was performed (Figure 31). This shows predicted encounter rate based on presence-only data for the same dataset and allows a comparison between simple and complex approaches, thereby demonstrating the usefulness of the GAM approach used in this study.

The GAM-generated habitat suitability map shows predicted sperm whale habitat within the GOALS-II study area utilizing the most relevant response variables (Figure 32). GAMs used correlated parameters from the best predictor variables to calculate preferred habitat, without the direct input of encounter rates or georeferenced localizations. The model produced high density regions on the continental shelf within the TMAA. As seen in Figure 16, this phenomenon may be attributed to their preference for environmental proxies near the 2,000 m isobath. In addition, this area comprises of high sloping features that support nutrient upwelling (Munger et al., 2009). Sea surface temperatures ramp up toward the shore line, possibly explaining habitat suitability. Therefore, this region is attractive due to its elevated surface chlorophyll-a concentration, indicating regional biomass. Predicted encounter rate density also extended southwest off the shelf, indicative of their preference for deep waters. Seamounts were also marked as areas of medium density with respect to their sloping faces and high surface chlorophyll concentrations. The eastern region of the TMAA displayed density due to the presence of strong current magnitude.

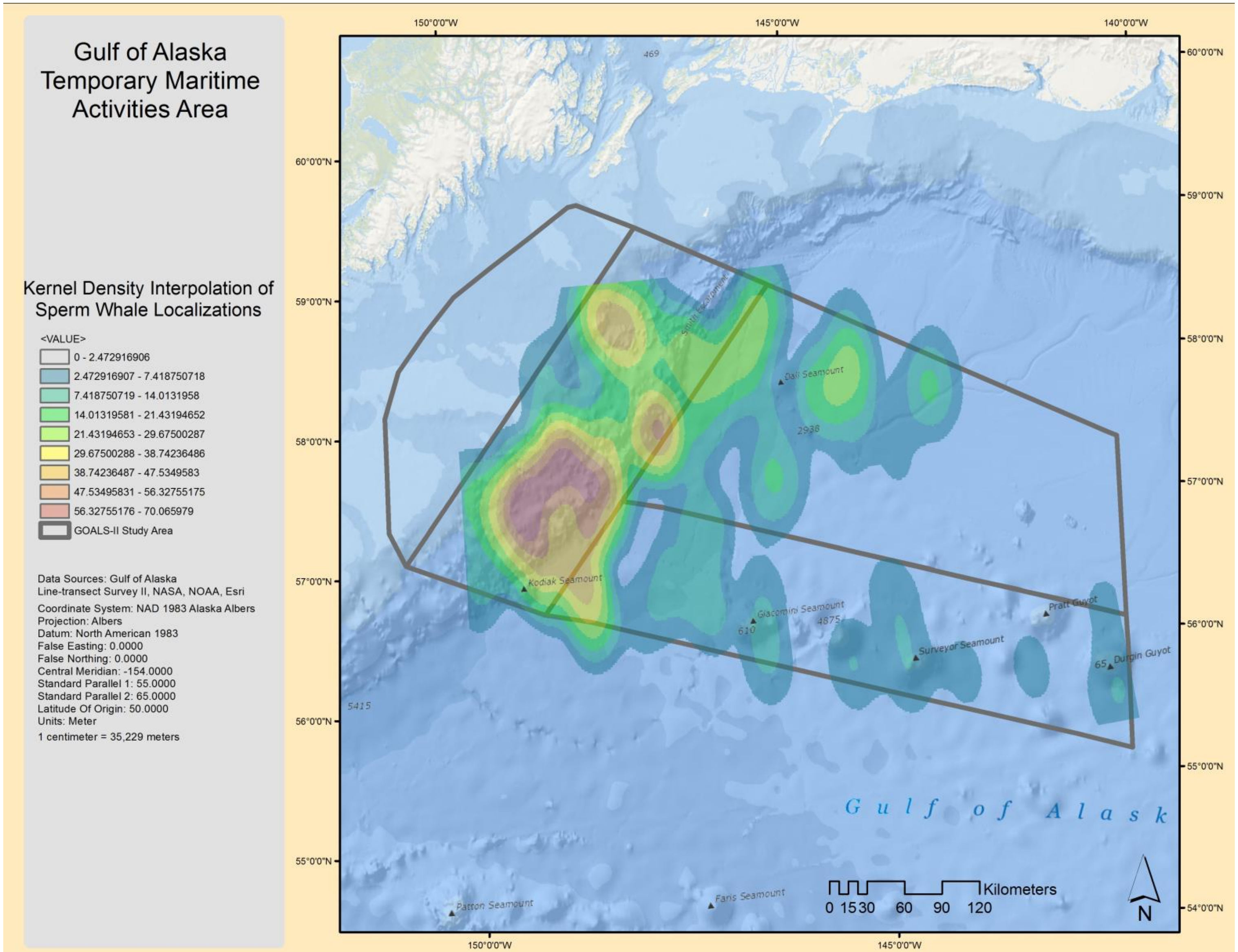


Figure 31. Density-based spatial analysis showing non-weighted kernel interpolation of localization distributions within the TMAA.

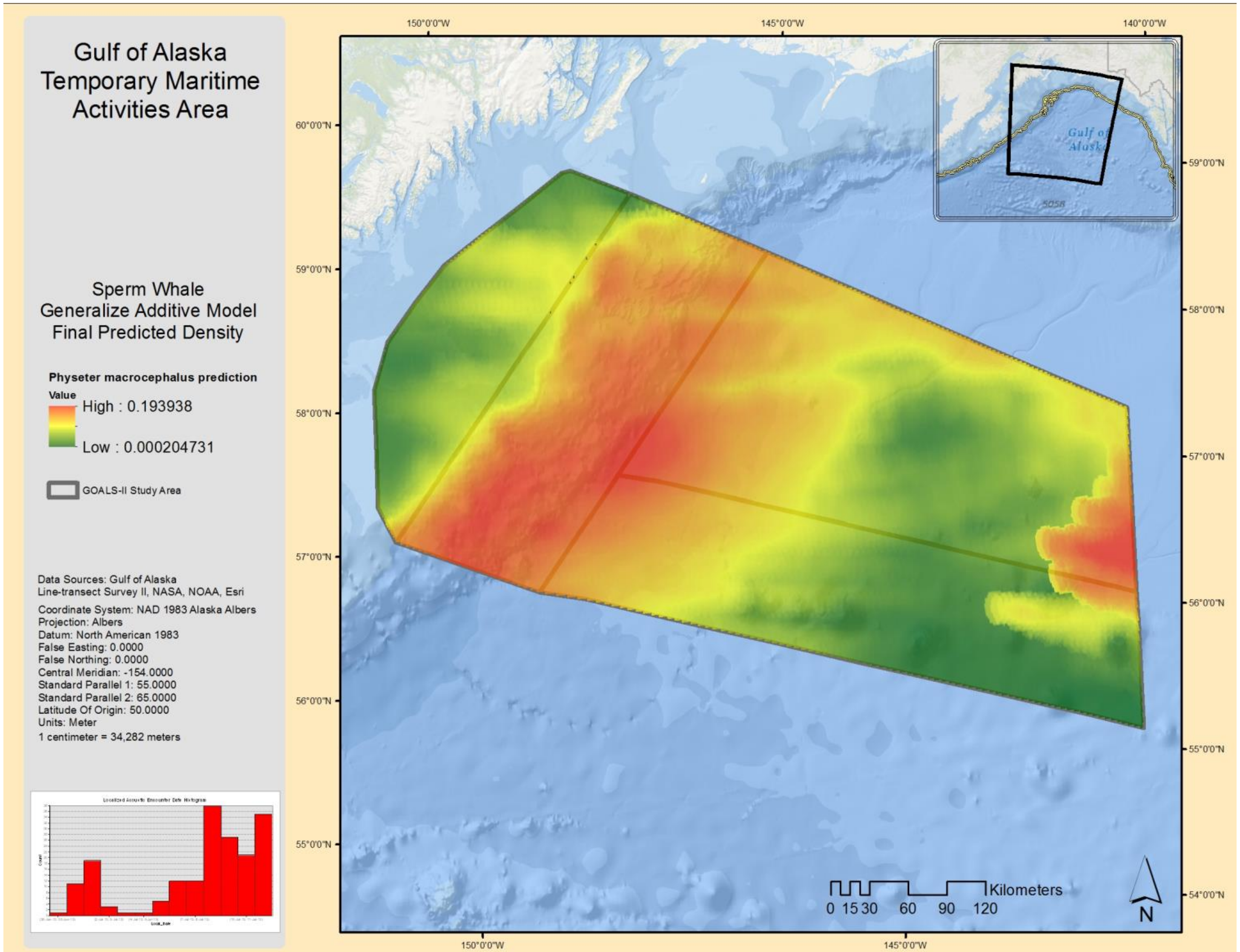


Figure 32. Final encounter rate prediction map displaying predicted density from GAM results. Histogram indicates frequency of localized encounter per survey date.

CHAPTER 5 - DISCUSSION AND CONCLUSIONS

This chapter discusses the impact of the project findings and how they influence existing research on developing GIS-based species habitat models, along with their relevance to similar biogeographic studies and how acoustic data affects the modeling results. Shortcomings of the evaluation are discussed, and recommendations for future habitat modeling efforts are made.

Species distributions were modeled using a combination of fixed spatial features (depth, slope, aspect, and distance to the 2,000 m isobaths) and dynamic oceanographic variables (sea surface temperature, sea surface salinity, moon phase, current magnitude, current direction, and chlorophyll-a concentration). After the creation of four different models, they were then evaluated using AIC scores. Models A and D produced the lowest AICs, indicating best fit, although Model D was ultimately chosen due to a smaller range of co-variance between terms. Moon phase was excluded as a possible predictor, deemed redundant due to its strong relationship with other variables within the GAM. Model D showed that both fixed and dynamic features (i.e., depth, slope, surface chlorophyll, sea surface temperature, distance to 2,000 m isobath, and current magnitude) as explanatory variables within the best fit model, with depth, distance to 2,000 m isobath and sea surface temperature being statistically significant.

Fixed spatial features have been found to be essential predictors of deep-diving cetacean habitat in other regions, particularly the Bahamas and North Atlantic (MacLeod and Zuur, 2005; MacLeod, 2000). This is most likely attributed to sperm whale's inclination toward prey in these regions. It is a well-known fact that cephalopods are a key component of sperm whale diet and existing research suggests that cephalopod species may be associated with seamounts and other steeply sloping ocean features because they are carried on to slopes by oceanic currents (Nesis, 1993).

A caveat for the PAM-based data is the assumption that no animals were missed on the trackline and that sperm whales present were vocalizing frequently enough to be within detection range along the surveyed trackline. The basis of the study relies on an estimate of acoustically active animals in the TMAA. Due to the fact that animals are not vocalizing 100 percent of the time, the PAM-based estimate may be an underestimate of the true density and encounter rate. Conversely, unless sperm whales within range are vocally inactive for extended periods of time, it is likely they will be acoustically detected within the limits of system capabilities (Rone et al., 2014).

Hydrophone arrays are subject to negative bias in localization ability near the trackline. Physical properties of the acoustic sensitivity of hydrophones limit their ability to 'look' directly forward or backward. The region monitored by hydrophone arrays during the survey is three-dimensional, in comparison to visual observation methods where two-dimensional surfaces are monitored. Foraging sperm whale in this region are engaging in deep-diving to depths of no more than approximately 1,600 m (Madsen et al., 2002) and can be detected as far as 20 km, so the vertical component of PAM-based negative bias is more common than the horizontal. Although missed encounters appear to be infrequent, slight negative responses to the research vessel (i.e., movement away) would produce fewer detections because animals are only detected for a brief period of time adjacent to the trackline (Rone et al., 2014).

The model used for habitat prediction appeared to under-fit the data (i.e. was not tightly correlated to acoustic encounters), although prediction areas generally overlapped areas of high encounter rate. A few anomalies were noted, particularly in areas of high predicted density within the eastern regions of the TMAA study area. The model suggested that seamounts and guyots located on the southeast portion of the surveyed study area were not predicted to be sperm whale habitats, despite rising seafloor synonymous with their prey (Nesis, 1993).

The use of acoustic data to inform habitat models for sperm whales may also ultimately allow more accurate models because acoustic encounters are obtained when the animals are foraging at depth in the desired habitat while animals sighted visually may be transiting between foraging areas. This study demonstrated that acoustic data offers a valuable contribution, providing a potential alternative to visually-based surveys.

GOALS-II comprises an approximately six-week long PAM-based survey with no follow-up efforts. Larger studies have applied cross-validation techniques to obtain more robust models featuring multi-year prediction capabilities (Becker et al., 2012; Becker, 2007; Forney, 2000). Although GOALS-II represented the second iteration of the GOALS marine mammal survey, the first survey lacked temporal and locational variables, and, therefore, did not allow for this type of cross-validation technique. An increase in the adoption of long-term population surveys using acoustic data would allow researchers to conduct these types of evaluative techniques to confirm predictions and re-evaluate models.

Utilizing the S-Plus platform as the framework of the step-wise GAM development process was demonstrated to be effective; however, a major drawback of S+ lies in its use of the Poisson distribution, which falsely relies on the assumption that variance equals the mean of the data. Practical studies have shown that Poisson regression hardly ever works for ecological modeling because most ecological datasets include a variance that is larger than the mean, also known as overdispersion. Negative binomial and tweedie distributions within regression models have become increasingly popular in handling overdispersion (Zuur et al., 2009), and this would be possible by performing modeling functions within the R environment. R is a currently supported statistical platform with continual community support in source code repositories such as GitHub. For these reasons, it would be beneficial to use R in place of S+ for future work for the reason that it offers extended functionality appropriate for this study.

The addition of explanatory variables that describe prey indices should be considered to aid habitat predictions. These include mixed layer depth, deep scattering layer depth, and squid density; unfortunately, these were not available for this effort. The inclusion of *in situ* data collection techniques would also prove useful, particularly to increase spatial resolution and provide real-time values of oceanographic variables.

The spatial component of this analysis presented a novel approach to understanding behavioral patterns by examining functional relationships to fixed and dynamic environmental variables. An array of technologies used was infused into a coordinated methodological workflow, demonstrating the performance and versatility of applied geographic information in achieving significant results in non-linear regression modeling effort. In addition, increased capabilities in geoprocessing and visualizing georeferenced data is facilitating a focus on the spatial aspects of population distributions. An emergence of recent GIS technologies has afforded researchers the ability to investigate and communicate the complex ecological interactions and tendencies that occur within these populations.

The capability to successfully detect, localize, and classify sperm whales and their locations using a towed hydrophone array denotes a significant contribution to the field of marine mammal science and thus provided the foundation for this study. With the addition of customized GIS-based components, the work has presented a unique and powerful analysis that has not been possible until recent years. This effort upholds existing understanding of sperm whale distribution and habitat preferences, thereby supporting the role of behavioral and physiological processes in habitat selection. This information exhibits the first in-depth analysis of sperm whale habitat within the central Gulf of Alaska, further advancing sperm whale management and conservation efforts. Future work will benefit from utilizing bioacoustic data to inform multi-variable habitat models for sperm whales and other cetacean species as well. This study has validated the feasibility of using PAM-based localizations and geospatial applications to enable higher levels of precision for predictive habitat distribution models.

REFERENCES

Akaike, Htrotugu. "Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models." *Biometrika* 60, no. 2 (1973): 255-265.

Barlow, Jay, Megan C. Ferguson, Elizabeth A. Becker, Jessica V. Redfern, Karin A. Forney, Ignacio L. Vilchis, Paul C. Fiedler, Tim Gerrodette, and Lisa T. Balance. *Predictive Modeling of Marine Mammal Density from Existing Survey Data and Model Validation using Upcoming Surveys*. No. SERDP-SI-1391. National Marine Fisheries Southwest Fisheries Center, 2009.

Barlow, Jay, Tim Gerodette, and Jaume Forcada. "Factors Affecting Perpendicular Sighting Distances on Shipboard Line-Transect Surveys for Cetaceans." *Journal of Cetacean Research and Management* 3, no. 2 (2001): 201-212.

Barlow, Jay, and Barbara L. Taylor. "Estimates of Sperm Whale Abundance in the Northeastern Temperate Pacific from a Combined Acoustic and Visual Survey." *Marine Mammal Science* 21, no. 3 (2005): 429-445.

Batchelor, Heidi H., and G. L. D'Spain. "Passively Recorded Acoustic Data as a Component of Marine GIS." Paper presented at 25th ESRI Conference, San Diego, California, July 25-29, 2005.

Becker, Elizabeth Ann. “Predicting Seasonal Patterns of California Cetacean Density based on Remotely Sensed Environmental Data” PhD diss., University of California- Santa Barbara, 2007.

Becker, Elizabeth. A., David. G. Foley, Karin. A. Forney, Jay Barlow, Jessica V. Redfern, and Chelle L. Gentemann. "Forecasting Cetacean Abundance Patterns to Enhance Management Decisions." *Endangered Species Research* 16, no. 2 (2012): 97-112.

Becker, Elizabeth A., Karin A. Forney, Megan C. Ferguson, Jay Barlow, and Jessica V. Redfern. "Predictive Modeling of Cetacean Densities in the California Current Ecosystem Based on Summer/Fall Ship Surveys in 1991-2008." *NOAA Technical Memorandum, NMFS-SWFSC* 499 (2012): 1-45.

Box, Elgene O. "Predicting Physiognomic Vegetation Types with Climate Variables." *Vegetation* 45, no. 2 (1981): 127-139.

Burtenshaw, Jessica C., Erin M. Oleson, John A. Hildebrand, Mark A. McDonald, Rex K. Andrew, Bruce M. Howe, and James A. Mercer. "Acoustic and Satellite Remote Sensing of Blue Whale Seasonality and Habitat in the Northeast Pacific." *Deep Sea Research Part II: Topical Studies in Oceanography* 51, no. 10 (2004): 967-986.

Bush, Natalie. "Spatio-Temporal Comparisons between Acoustic and Visual Detections of the Short-Beaked Common Dolphin (Delphinus delphis) in the St George's Channel" Master of Science Thesis, University of Wales - Bangor, 2006.

Cain, Stanley Adair. *Foundations of Plant Geography.* New York: Harper, 1944.

Carretta, James. V., Barbara. L. Taylor, and Susan. J. Chivers. "Abundance and Depth Distribution of Harbor Porpoise (Phocoena phocoena) in Northern California Determined from a 1995 Ship Survey." *Fishery Bulletin* 99, no. 1 (2001): 29-39.

Chassignet, Eric. "*HYCOM*: Center for Ocean-Atmospheric Prediction Studies (COAPS)." Last modified January 1, 2016. http://hycom.org/.

Clarke, Malcolm. R., and Sean L. Pascoe. "Cephalopod species in the Diet of a Sperm Whale (Physeter catodon) Stranded at Penzance, Cornwall." *Journal of the Marine Biological Association of the United Kingdom* 77, no. 04 (1997): 1255-1258.

Diaconis, Persi, and Efron Bradley (1983) Computer intensive methods in statistics. Scientific American 248:116–130.

D'Spain, Gerald L. *Analysis of High Spatial, Temporal, and Directional Resolution Recordings of Biological Sounds in the Southern California Bight*. Scripps Institute of Oceanography, 2013.

De Candolle, Alphonse. *Géographie Botanique Raisonnée: ou, Exposition des Faits Principaux et des Lois Concernant la Distribution Géographique des Plantes de L'époque Actuelle*. Paris: Nabu Press, 1855.

Dohan, Kathleen. "*OSCAR*." Last modified January 1, 2012. http://www.esr.org/oscar_index.html.

Efron, Bradley, and Robert Tibshirani. “Statistical Data Analysis in the Computer Age.” *Science* 253, no. 2018 (1991): 390-395.

Ellenberg, Heinz. *Vegetation Ecology of Central Europe*. Cambridge University Press, 1988.

Esri. "ArcGIS 10.3 Help." Last modified January 1, 2015. http://resources.arcgis.com/en/help/main/10.2/.

Feldman, Gene. "Ocean Color and Temperature Scanner." Last modified November 10, 2008. http://oceancolor.gsfc.nasa.gov/OCTS/.

Ferguson, Megan C., Jay Barlow, Paul Fiedler, Stephen B. Reilly, and Tim Gerrodette. "Spatial Models of Delphinid (family Delphinidae) Encounter Rate and Group Size in the Eastern Tropical Pacific Ocean." *Ecological Modelling* 193, no. 3 (2006): 645-662.

Forney, Karin A. "Environmental Models of Cetacean Abundance: Reducing Uncertainty in Population Trends." *Conservation Biology* 14, no. 5 (2000): 1271-1286.

Forney, Karin A., Megan Ferguson, Elizabeth Becker, Paul Fiedler, Jessica Redfern, Jay Barlow, Ignacio Vilchis, and Lisa Ballance. "Habitat-based Spatial Models of Cetacean Density in the Eastern Pacific Ocean." *Endangered Species Research* 16, no.2 (2012): 113-133.

Good, Irving J. "The Population Frequencies of Species and the Estimation of Population Parameters." *Biometrika* 40, no. 3-4 (1953): 237-264.

Goodall, Colin. *A Survey of Smoothing Techniques*. Newbury Park: Sage Publications, 1990.

Guisan, Antoine, and Niklaus E. Zimmermann. "Predictive Habitat Distribution Models in Ecology." *Ecological modelling* 135, no. 2 (2000): 147-186.

Hardle, Wolfgang., and John. S. Marron. "Bootstrap Simultaneous Error Bars for Nonparametric Regression." *The Annals of Statistics* 19, no. 2 (1991): 778-796.

Hastie, Trever. J., and Robert. J. Tibshirani. *Generalized Additive Models.* Chapman & Hall/CRC Press, 1990.

Holdridge, Leslie R. "The Determination of Atmospheric Water Movements."*Ecology* 43, no. 1 (1962): 1-9.

Holt, Rennie S. "Estimating Density of Dolphin Schools in the Eastern Tropical Pacific Ocean by Line Transect Methods." *Fishery Bulletin* 85, no. 3 (1987): 419-434.

Hooker, Sascha K., Hal Whitehead, and Shannon Gowans. "Ecosystem Consideration in Conservation Planning: Energy Demand of Foraging Bottlenose Whales (Hyperoodon ampullatus) in a Marine Protected Area." *Biological Conservation* 104, no. 1 (2002): 51-58.

Hutchinson, G. Evelyn. "Homage to Santa Rosalia or Why Are There So Many Kinds of Animals?." *American Naturalist* 93, no. 870 (1959): 145-159.

Hutchinson, G. Evelyn. "The Paradox of the Plankton." *American Naturalist 95, no. 882* (1961): 137-145.

Kinzey, Douglas, Paula Olson, and Tim Gerrodette. "Marine Mammal Data Collection Procedures on Research Ship Line-transect Surveys by the Southwest Fisheries Science Center." *NOAA, SWFSC Administrative Report LJ-00-08*, 2000.

Leaper, Russell, Jonathan Chappell, and Jonathan Gordon. *The Development of Practical Techniques for Surveying Sperm Whale Populations Acoustically*. Oxford: Oxford University Press, 1992.

Leaper, Russell, Douglas Gillespie, and Vassili Papastavrou. "Results of Passive Acoustic Surveys for Odontocetes in the Southern Ocean." *Journal of Cetacean Research and Management* 2, no. 3 (2000): 187-196.

MacArthur, Robert, and Richard Levins. "The Limiting Similarity, Convergence, and Divergence of Coexisting Species." *American Naturalist* 101, no. 921 (1967): 377-385.

Maccharone, Brandan. "*MODIS: Moderate Resolution Imaging Spectroradiometer*." Last modified November 1, 2015. http://modis.gsfc.nasa.gov/.

MacArthur, Robert H. *Geographical Ecology: Patterns in the Distribution of Species*. Princeton: Princeton University Press, 1972.

MacLeod, Colin D., and Alain F. Zuur. "Habitat Utilization by Blainville's Beaked Whales Off Great Abaco, Northern Bahamas, in Relation to Seabed Topography." *Marine Biology* 147, no. 1 (2005): 1-11.

Macleod, Colin D. "Review of the Distribution of Mesoplodon Species (order Cetacea, family Ziphiidae) in the North Atlantic." *Mammal Review* 30, no. 1 (2000): 1-8.

Madsen, Peter, Magnus Wahlberg, and Boris Møhl. "Male Sperm Whale (Physeter macrocephalus) Acoustics in a High-Latitude Habitat: Implications for Echolocation and Communication." *Behavioral Ecology and Sociobiology* 53, no. 1 (2002): 31-41.

Manly, Bryan F. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. London: Chapman and Hall, 2006.

Marsh, Helene., and Dennis F. Sinclair. "Correcting for Visibility Bias in Strip Transect Aerial Surveys of Aquatic Fauna." *Journal of Wildlife Management* (1989): 1017-1024.

Mellinger, David. K. "Ishmael 1.0 User's Guide. NOAA Technical Report OARPMEL-120." *NOAA Pacific Marine Environmental Laboratory, Seattle* (2001): 1-26.

Pacific Marine Environment Laboratory (US), David Mellinger, and Jay Barlow. *Future Directions for Acoustic Marine Mammal Surveys: Stock Assessment and Habitat Use*. Presentation at the Southwest Fisheries Science Center in La Jolla, California, November 20-22, 2002.

Moore, Sue E., Kathleen M. Stafford, David K. Mellinger, and John A. Hildebrand. "Listening for Large Whales in the Offshore Waters of Alaska."*BioScience* 56, no. 1 (2006): 49-55.

Munger, Lisa M., Dominique, Camacho, Andrea Havron, Greg Campbell, John Calambokidis, Annie Douglas, and John Hildebrand. "Baleen Whale Distribution Relative to Surface Temperature and Zooplankton Abundance Off Southern California, 2004–2008." *California Cooperative Oceanic Fisheries Investigations Reports* 50 (2009): 155-168.

Nesis, Kir N. "Population Structure of Oceanic Ommastrephids, with Particular Reference to Sthenoteuthis Oualaniensis: A Review." In *Recent Advances in Cephalopod Fisheries Biology*, edited by John D. Kelly, 375-383. Tokyo: Takai University Press, (1993).

Notarbartolo di Sciara, Giuseppe., Alexandros Frantzis, Giovanni Bearzi, and Randall Reeves. "Sperm Whales, Physeter Macrocephalus, in the Mediterranean Sea: a Summary of

Status, Threats, and Conservation Recommendations." *Aquatic Conservation: Marine and Freshwater Ecosystems* 24, no. 1 (2014): 4-10.

Rayment, William, Steve Dawson, Silvia Scali, and Liz Slooten. "Listening for a Needle in a Haystack: Passive Acoustic Detection of Dolphins at Very Low Densities." *Endangered Species Research* 14 (2011): 149-156.

Redfern, Jessica. V., Megan. C. Ferguson, Elizabeth. A. Becker, Kurt. D. Hyrenbach, Caroline P. Good, Jay Barlow, and Kristin Kaschner. "Techniques for Cetacean–Habitat Modeling." Marine Ecology Progress Series 310 (2006): 271-295.

Roberts, Jason, and Benjamin Best. "Marine Geospatial Ecology Tools: An Integrated Framework for Ecological Geoprocessing with ArcGIS, Python, R, MATLAB, and C##." *Environmental Modeling and Software* 25, no. 2010 (2010): 1197-207.

Rone, Brenda. K., Annie. B. Douglas, Tina M. Yack, Alexandre N. Zerbini, Tom. N. Norris, Elizabeth Ferguson, and John Calambokidis. "Report for the Gulf of Alaska Line-Transect Survey (GOALS) II: Marine Mammal Occurrence in the Temporary Maritime Activities Area (TMAA)." Report prepared by Cascadia Research Collective, Alaska Fisheries Science Center, and Bio-Waves, Inc. Naval Facilities Engineering Command Pacific, Honolulu, Hawaii, 2014.

Salisbury, Edward. J. "The Geographical Distribution of Plants in Relation to Climatic Factors." *Geographical Journal* 67, no.4 (1926): 312-335.

Silverman, Bernard. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall/CRC, 1986.

Sousa-Lima, Renata S., Thomas F. Norris, Julie N. Oswald, and Deborah P. Fernandes. "A Review and Inventory of Fixed Autonomous Recorders for Passive Acoustic Monitoring of Marine Mammals." *Aquatic Mammals* 39, no. 1 (2013): 23-53.

Stott, Gary. "What is Animal Stress and How is it Measured?" *Journal of Animal Science* 52, no. 1 (1981): 150-153.

Strindberg, Samantha, and Stephen T. Buckland. "Zigzag Survey Designs in Line Transect Sampling." *Journal of Agricultural, Biological, and Environmental Statistics* 9, no. 4 (2004): 443-461.

Teilmann, Jonas, and Jacob Carstensen. "Negative Long Term Effects on Harbour Porpoises from a Large Scale Offshore Wind Farm in the Baltic—Evidence of Slow Recovery." *Environmental Research Letters* 7, no. 4 (2012): 1-10.

Torres, Leigh G., Patricia E. Rosel, Caterina D'Agrosa, and Andrew J. Read. "Improving Management of Overlapping Bottlenose Dolphin Ecotypes through Spatial Analysis and Genetics." *Marine Mammal Science* 19, no. 3 (2003): 502-514.

Thompson, Paul M., Kate L. Brookes, and Line S. Cordes. "Integrating Passive Acoustic and Visual Data to Model Spatial Patterns of Occurrence in Coastal Dolphins." *ICES Journal of Marine Science: Journal du Conseil* 72, no. 2 (2015): 651-660.

Thompson, Paul M., David Lusseau, Tim Barton, Dave Simmons, Jan Rusin, and Helen Bailey. "Assessing the Responses of Coastal Cetaceans to the Construction of Offshore Wind Turbines." *Marine Pollution Bulletin* 60, no. 8 (2010): 1200-1208.

von Humboldt, Alexander, and Aimé Bonpland. *Essai Sur la Géographie des Plantes*. Paris: Maxtor France,1807.

Wiens, John A. “Spatial Scaling in Ecology.” *Functional Ecology* 3, no. 4 (1989): 385–97.

Wood, Simon N., and Nicole H. Augustin. "GAMs with Integrated Model Selection using Penalized Regression Splines and Applications to Environmental Modelling." *Ecological Modelling* 157, no. 2 (2002): 157-177.

Wood, Simon N. "Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 62, no.2 (2000): 413-428.

Woodward, Frank Ian. *Climate and Plant Distribution*. Cambridge: Cambridge University Press, 1987.

Yack, Tina M. "The Use of Passive Acoustic Data to Predict Beaked Whale Habitat in the California Current Ecosystem." PhD diss., University of California – Davis & San Diego State University, 2013.

Zuur, Alain, Elena N. Ieno, Neil Walker, Anatoly A. Saveliev, and Graham M. Smith. *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer Science & Business Media, 2009.

APPENDIX A: Tables and Figures

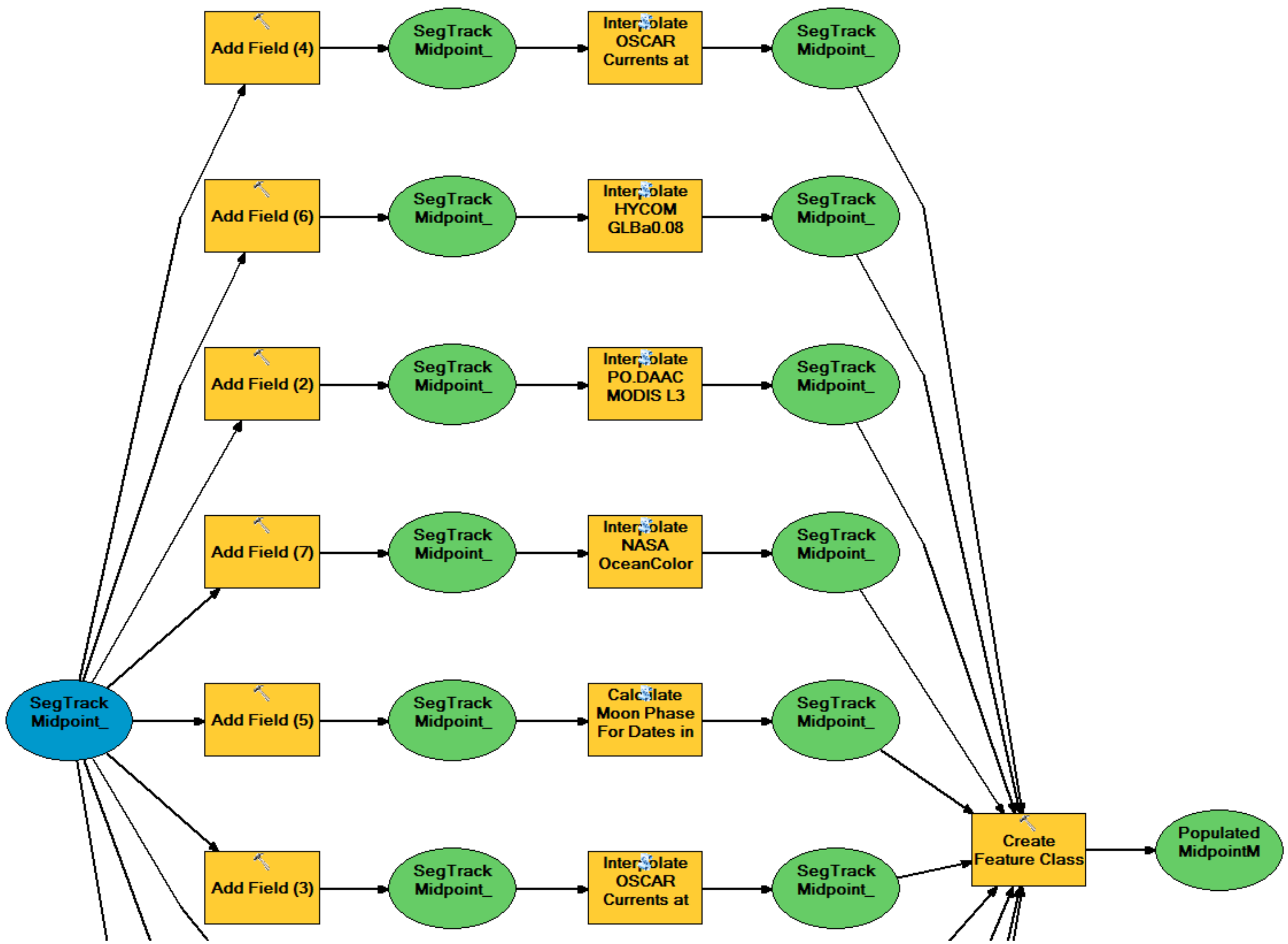


Figure 33. Upper portion of MGET-based geoprocessing model showing extraction of six environmental variables. For the input feature class, fields with appropriate data types were created, data fields were then populated and used to create a final feature class product.

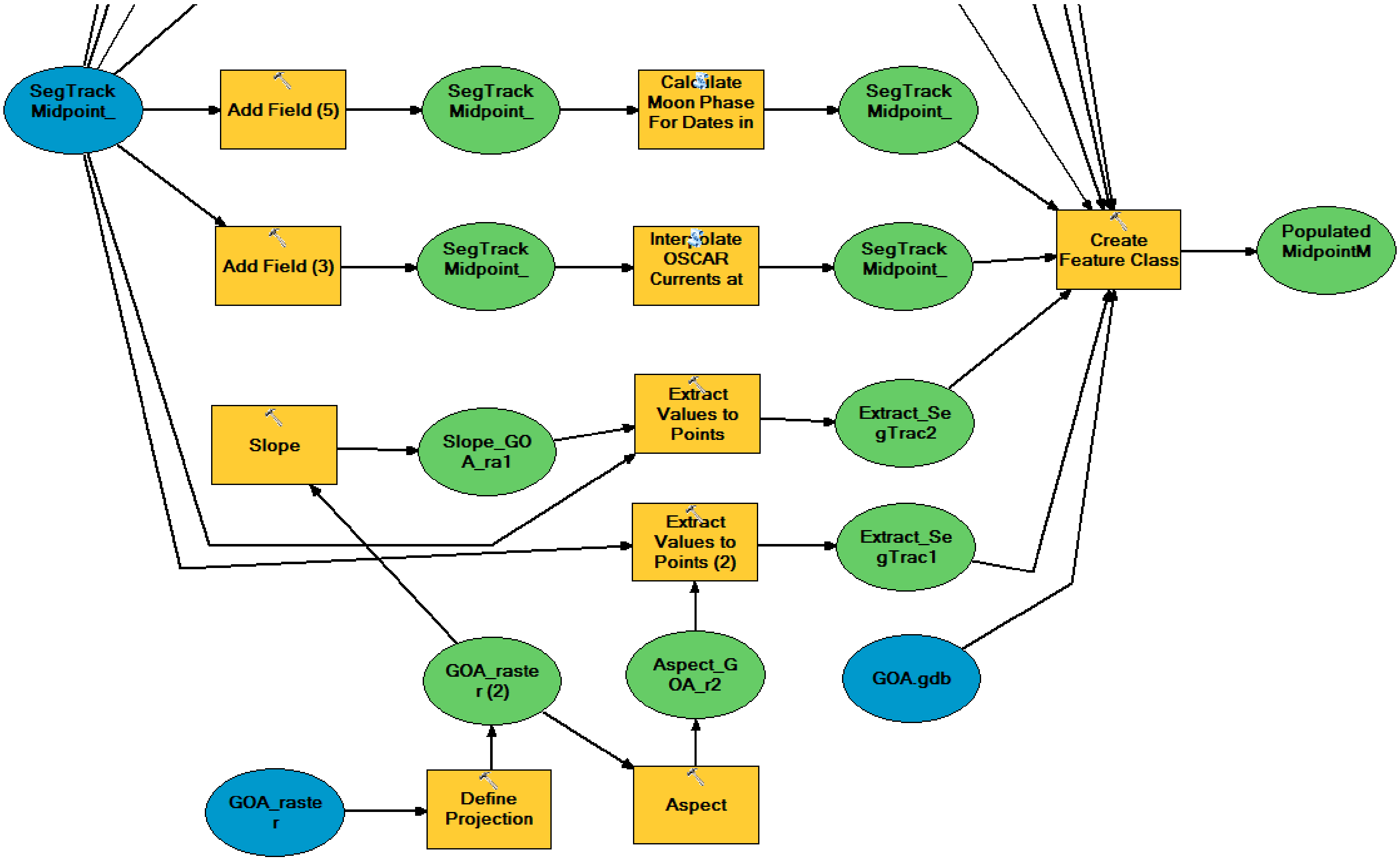


Figure 34. Lower portion of MGET-based geoprocessing model. The Gulf of Alaska DEM was defined with a Alaska Albers Equal Area Conic projection. This input was used to produce Slope and Aspect, and this was used to extract variables from the raster and populate fields within the input feature class. The Near tool was used to calculate distance of input point features to the 2,000 m isobath. The outputs from these tools were then used to create a feature class within the "GOA" file geodatabase.

Table 19. Fixed variables and associated tools used in habitat predictive model for sperm whales (Esri, 2015).

<table>
<tr><th>Environmental Variable</th><th>Tool and Description</th></tr>
<tr><td>Depth (m)</td><td>

- Extract Values to Points (Spatial Analyst) extracts the cell values of a raster based on a set of point features and records the values in the attribute table of an output feature class.

</td></tr>
<tr><td>Slope (degrees grade)</td><td>

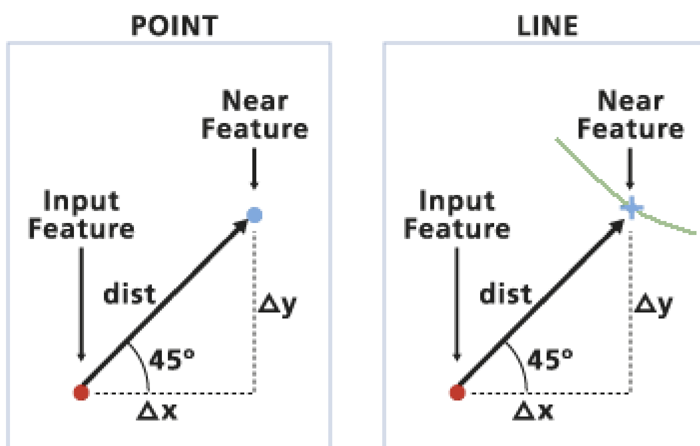
- Extract Values to Points (Spatial Analyst) extracts the cell values of a raster based on a set of point features and records the values in the attribute table of an output feature class.

</td></tr>
<tr><td>Aspect (degrees bearing)</td><td>

- Extract Values to Points (Spatial Analyst) extracts the cell values of a raster based on a set of point features and records the values in the attribute table of an output feature class.

</td></tr>
<tr><td>2000m Bathy Dist. (km)</td><td>

- Near (Analysis) determines the distance from each point in the Input Features to the nearest point or polyline in the Near Features, within the Search Radius



- The results are recorded in the Input Features attribute table. Fields for distance and feature ID of the closest feature are added or updated. The field names are NEAR_DIST and NEAR_FID.
- The values for NEAR_DIST will be zero if no match is found within the specified Search Radius.
- The values for NEAR_FID will be -1 if no match is found within the specified Search Radius.
- If the fields NEAR_DIST and NEAR_FID already exist, the values will be recalculated.
- If no Search Radius is specified, a radius large enough to calculate a distance from each point in the Input Features to the closest point or polyline in the Near Features will be used.
- The calculated distance from a point to a line will be from the point to the nearest location along the line.

</td></tr>
</table>

	• The distances calculated by Near are determined by the units of the Input Features. If the linear units of the input feature class are in Meters, the search radius will default to Meters. The units of the Search Radius can be changed. Specifying one kilometer is the same as entering one thousand meters. • Fields for x and y coordinates are added when Location is checked, and a field for ANGLE is added when Angle is checked. • Angles are measured in degrees, where one degree represents 1/360 of a circle, and fractions of a degree are represented as decimal points. Angles are measured from 180° to -180° ; 0° to the east, 90° to the north, 180° (-180°) to the west, and -90° to the south. • Near is useful for assigning attributes to nearest lines or points. For example, when assigning address ranges to lines or searching for the closest sewer line in a sewage network for a specific property.

Table 20. Dynamic variables and associated tools used in habitat predictive model for sperm whales (Esri, 2015).

Environmental Variable	Tool and Description
SST (° Celsius)	• Interpolate PO.DAAC MODIS L3 SST at Points (Spatial Analyst) interpolates PO.DAAC MODIS Level 3 SST values at points, similarly to the Extract Values to Points (Spatial Analyst) tool • The NASA Jet Propulsion Laboratory (JPL) Physical Oceanography Distributed Active Archive Center (PO.DAAC) publishes collections of sea surface temperature (SST) images gathered by the Moderate Resolution Imaging Spectroradiometer (MODIS) carried by the Terra and Aqua satellites. This tool accesses the PO.DAAC SST datasets that begin with the following names: o MODIS Terra Level 3 Thermal IR o MODIS Terra Level 3 Mid-IR o MODIS Aqua Level 3 Thermal IR o MODIS Aqua Level 3 Mid-IR • Given a satellite name, temporal resolution, spatial resolution, and desired geophysical parameter, this tool interpolates the value of that parameter at the given points. This tool performs the same basic operation as the ArcGIS Spatial Analyst's Extract Values to Points tool, but it reads the MODIS data directly from NASA's servers using theOPeNDAP protocol, rather than reading rasters stored on your machine.
Sea Surface Salinity (PSU)	• Interpolate HYCOM GLBa0.08 Equatorial 4D Variables at Points (Spatial Analyst) interpolates salinity using 4D variables of the HYCOM GLBu0.08 dataset at points. • This tool accesses a concatenation of several sequential HYCOM + NCODA Global 1/12 Degree "uniform" (GLBu0.08) datasets, treating them as a continuous virtual dataset running from late 1992 to the present day.

	• The dataset consists of a collection of 3D and 4D gridded variables. The 3D variables represent conditions at the surface of the ocean and have dimensions of x, y, and time. The 4D variables represent conditions at depth and have dimensions of x, y, depth, and time.

Moon Phase (0-0.999)	• Calculates moon phase for dates
	  ○ Given a table with a date field, calculates the phase of the moon and writes it to another field.
	  ○ Moon phase property to calculate from the date field. One of:
	  ○ Phase - Numeric phase of the moon ranging from 0 to 0.999999, where 0 is new moon, 0.25 is first quarter, 0.5 is full, and 0.75 is third quarter.

Current Direction (Degrees Bearing)	• Interpolate HYCOM GLBa0.08 Equatorial 4D Variables at Points (Spatial Analyst) interpolates salinity using 4D variables of the HYCOM GLBu0.08 dataset at points. • This tool accesses a concatenation of several sequential HYCOM + NCODA Global 1/12 Degree "uniform" (GLBu0.08) datasets, treating them as a continuous virtual dataset running from late 1992 to the present day. • The dataset consists of a collection of 3D and 4D gridded variables. The 3D variables represent conditions at the surface of the ocean and have dimensions of x, y, and time. The 4D variables represent conditions at depth and have dimensions of x, y, depth, and time.
Current Magnitude (TKE)	• Interpolate HYCOM GLBa0.08 Equatorial 4D Variables at Points (Spatial Analyst) interpolates salinity using 4D variables of the HYCOM GLBu0.08 dataset at points. • This tool accesses a concatenation of several sequential HYCOM + NCODA Global 1/12 Degree "uniform" (GLBu0.08) datasets, treating them as a continuous virtual dataset running from late 1992 to the present day. • The dataset consists of a collection of 3D and 4D gridded variables. The 3D variables represent conditions at the surface of the ocean and have dimensions of x, y, and time. The 4D variables represent conditions at depth and have dimensions of x, y, depth, and time.
Chlorophyll-a Concentration (mg/m^3)	• Interpolate NASA OceanColor L3 SMI Product at Points (Spatial Analyst) is an MGET customized tool that interpolates the values of a Level 3 Standard Mapped Image (SMI) product published by the NASA GSFC OceanColor Group at points. • Given a sensor name, temporal resolution, spatial resolution, and desired Level 3 SMI product, this tool interpolates the value of that product at the given points. This tool performs the same basic operation as the ArcGIS Spatial Analyst's Extract Values to Points tool, but it downloads and reads HDF files from NASA's servers rather than reading rasters stored on the local machine. • The NASA Goddard Space Flight Center (GSFC) OceanColor Group publishes a variety of satellite image products derived from ocean color observations made by polar-orbiting sensors such as MODIS, SeaWiFS, OCTS, and CZCS. The most popular product is an estimate of chlorophyll-a concentration. • This tool accesses the Level 3 Standard Mapped Image (SMI) products, which have global spatial extent, use a geographic coordinate system with the WGS 1984 datum, and have square cells with either 1/12 or 1/24 degree

	resolution (about 9.3 km or 4.6 km at the equator).
	
	• Product code of the NASA Level 3 Standard Mapped Image (SMI) product to use, such as CHL_chlor_a for chlorophyll concentration.
	
	• NASA publishes the SMI products as collections of compressed HDF version 4 files that are downloadable from the OceanColor web site. This tool automatically downloads, decompresses, and reads HDF files as they are needed. Unless you specify a directory to cache the files, they will be stored in your user TEMP directory and deleted when processing is finished.
	
	• Data for this particular use is taken from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor carried by the Aqua satellite. Aqua datasets started in July 2002 and are still being collected.
	
	• This tool provides L3 SMI products at 4 km and 9 km
	  o 4 km was used in this instance
	
	• Interpolation method used:
	  o Nearest - nearest neighbor interpolation. The interpolated value will simply be the value of the cell that contains the point. This is the default.

APPENDIX B: Modeling Code

Script 1. Segment conversion code. Written in R, this code was used to convert complete GOALS-II trackline from GPS archive to 5 km segments. As discussed from section 3.3, the script was instructed to write a matrix with values for start, middle and end positions with timestamps, as well as segment length in kilometers.

```
rm(list=ls())
options (digits=5)
GOALSIItracks<-read.csv("GOALSII_Trackline.csv", header=TRUE, stringsAsFactors = F, na.strings = c(NA,
"NA", "NaN"))

date.int <- GOALSIItracks$Date
date.char <- as.character(date.int)
date.char <- ifelse(nchar(date.char) < 6, paste(0, date.char, sep = ""), date.char)
month <- as.integer(substr(date.char, 1, 2))
day <- as.integer(substr(date.char, 3, 4))
yr <- as.integer(substr(date.char, 5, 6))

time.int <- GOALSIItracks$Time
time.char <- as.character(time.int)
time.char <- ifelse(nchar(time.char) < 6, paste(0, time.char, sep = ""), time.char)
hr <- as.integer(substr(time.char, 1, 2))
min <- as.integer(substr(time.char, 3, 4))
sec <- as.integer(substr(time.char, 5, 6))

GOALSIItracks$datetime <- ISOdatetime(2000 + yr, month, day, hr, min, sec)


str(GOALSIItracks)

# Find the start positions and end positions for the Overall straight line segment
find.start.point <- function(df, end) {
        if(is.na(end)) return(NA)
        start.rows <- which(df$Trackline.Status == "S") ##pick the lines that have standard effort
        start.rows <- start.rows[start.rows > end] ## only take the rows that are greater than your start point that are
standard
        ifelse(length(start.rows) == 0, NA, start.rows[1]) ## if there are no standard values, return NA and only
start at the first value that meets standard effort conditions
}

find.end.point <- function(df, start) {
        if(is.na(start)) return(NA)
        effort.end <- df$Visual.Effort.Status == "off"
    effort.off <- df$Trackline.Status == "NS"
        end.rows <- which(effort.end | effort.off)### IS THIS CORRECT???
        end.rows <- end.rows[end.rows > start]
        ifelse(length(end.rows) == 0, NA, end.rows[1])
}
```

```
############
# Get Great circle distance
grtcircl <- function(start.pos, end.pos){
	rlat1 <- start.pos[1] * pi/180  ### converting lats and longs to radians
	rlat2 <- end.pos[1] * pi/180
	rlong1 <- start.pos[2] * pi/180
	rlong2 <- end.pos[2] * pi/180
	R <- 6371 # Earth mean radius [km]

	threshold <- 0.0000001 ### to prevent NAs due to numerical precision errors in R
	disc <- sin(rlat1)*sin(rlat2) + cos(rlat1)*cos(rlat2) * cos(rlong2-rlong1) ### calculating distnce between
start and end but removes risk of precission errors
	if(disc > 1 && disc < 1+threshold) ### restrains discriminants between 1 and -1
	{return(0)}
	if(disc < -1 && disc > -1-threshold)
	{return(pi*R)}
	d <- acos(disc) * R
	return(d) # Distance in km  ### great circle distance between start and end positions
}

# Given 5 km distance from start pos get the lat/long of the endposition
	get.seg.end <- function(start.pos, end.pos, dist) {
	rlat1 <- start.pos[1] * pi/180
	rlat2 <- end.pos[1] * pi/180
	rlong1 <- start.pos[2] * pi/180
	rlong2 <- end.pos[2] * pi/180

	R <- 6371 # Earth mean radius [km]

	bearing <- atan2(sin(rlong2-rlong1), cos(rlat1)*tan(rlat2)-sin(rlat1)*cos(rlong2-rlong1))

	dest.lat <- asin(sin(rlat1)*cos(dist/R)+cos(rlat1)*sin(dist/R)*cos(bearing))
	dest.long <- rlong1 + atan2(sin(bearing)*sin(dist/R)*cos(rlat1), cos(dist/R) - sin(rlat1)*sin(rlat2))

	return(c(dest.lat * 180/pi, dest.long * 180/pi))

	# Calculate bearing to endpos returns bearing as tc1, must do this for every new endpoint
	# if (sin(rlong2-rlong1)<0 )
	# (tc1<-acos((sin(rlat2)-sin(rlat1)*cos(d))/(sin(d)*cos(rlat1))))
	# ifelse
	# (tc1<-2*pi-acos((sin(rlat2)-sin(rlat1)*cos(d))/(sin(d)*cos(rlat1))))
	# Calculate lat and long positions for endpoint given bearing tc1
    #   rlat<-asin(sin(rlat1)*cos(d)+cos(rlat1)*sin(d)*cos(tc1))
    #   rlong<-mod(rlong1-asin(sin(tc1)*sin(d)/cos(rlat1))+pi,2*pi)-pi
	# lat <- rlat * (180/pi)
	# long<- rlong * (180/pi)
	}

#Find midpoint of start pos and end seg

	get.mid <- function(start.pos, end.pos) {
	dist = grtcircl(start.pos, end.pos)
	get.seg.end(start.pos, end.pos, dist/2)
	}
#Find the average beaufort value for the 5km segment
```

```
	get.beaufort.avg <- function(seg.start, seg.end, df) {
	current.distance <- grtcircl (seg.start, c(Lat =df$Lat[1], Long = df$Long[1])) ### will calc dist from seg start to 1st geographic location point in data frame
	seg.length <- grtcircl(seg.start, seg.end) ### calculate distance between seg. start and end and set as seg length
	if (current.distance > seg.length)  ### if the dist between seg start and first lat long is greater than seg length, return NA (not a valid segment per our qualifiers so dont want to use beaufort data in it)
	{return (c(NA, 1))}

	for(i in 2:nrow(df))
	{
	current.distance <- grtcircl (seg.start, c(Lat =df$Lat[i], Long = df$Long[i]))### calc dist from seg start to lat long for each subsequent position
	if(current.distance > seg.length) ### if that dist is greater than seg length, stop there and take the mean of the beauforts in that segment (cut it off and take the mean up to that point)
	{return (c(mean(df$Bft[1: (i-1)], na.rm = TRUE),i))} ### na.rm allows NAs in data to be taken out and the rest averaged - won't return NA if there is data to average from
}
return (c(mean(df$Bft[], na.rm = TRUE),i))### for the end of the database in case have a segment that isn't greater than 5km
}


###########################


get.end.time <- function(time.start, dist, rate.of.travel) time.start + as.difftime(dist / rate.of.travel, units = "secs")


#Using above functions that calculated starts, ends, midpoints, etc, put all that info into a matrix
get.segment.positions <- function(df, start, end) {
  start.pos <- c(Lat = df$Lat[start], Long = df$Long[start])
  end.pos <-  c(Lat = df$Lat[end], Long = df$Long[end])
bf.search.df <- df[start:end,] # this data frame contains the coordinates to search through (aka the coordinates along the overall straight line segment)
  start.time <- df$datetime[start]
  end.time <- df$datetime[end]
	tot.dist <- grtcircl(start.pos, end.pos)
  rate.of.travel <- tot.dist / as.integer(abs(difftime(start.time, end.time, units = "secs")))
  num.segments <- floor(tot.dist / 5) ## only takes whole 5 km long segs
  remainder <- tot.dist %% 5
  add.in <- remainder < 2.5
  if(!add.in) num.segments <- num.segments + 1
  ran.segment <- sample(1:num.segments, 1)
  seg.end <- start.pos
  seg.time.end <- start.time
	if(num.segments == 0) return(NA)

# initialize data structures
  segment.mat <- matrix(NA, num.segments, 8)
  time.df <- data.frame()
  beaufort.df<-data.frame()

# compute values for each segment
  for(i in 1:num.segments)
```

```
            {
                        seg.start <- seg.end
       seg.time.start <- seg.time.end
                        dist <- ifelse(i == ran.segment, ifelse(add.in, 5 + remainder, remainder), 5)
                        seg.end <- get.seg.end(seg.start, end.pos, dist)
       seg.time.end <- get.end.time(seg.time.start, dist, rate.of.travel)
                        seg.mid <- get.mid(seg.start, seg.end)
       seg.time.mid <- get.end.time(seg.time.start, grtcircl(seg.start, seg.mid), rate.of.travel)
                        segment.mat[i,] <- c(i, seg.start[1], seg.start[2],seg.end[1], seg.end[2], seg.mid[1], seg.mid[2],
dist)
       beaufort.return <- get.beaufort.avg(seg.start, seg.end, bf.search.df)
seg.beaufort.avg <- beaufort.return[1]
if (beaufort.return[2] > 0)
{
            bf.search.df <- bf.search.df[beaufort.return[2]:nrow(bf.search.df),] ### beaufort.return[2] is bc need 2
values, beaufort and line number
}
             time.df <- rbind(time.df, data.frame(seg.time.start, seg.time.end, seg.time.mid))
             beaufort.df <-rbind(beaufort.df, data.frame(seg.beaufort.avg))
            }

# return data structures
 return(data.frame(segment.mat, time.df, beaufort.df))
}


  # segment.mat <- sapply(1:num.segments, function(i) {
  #        seg.start <- seg.end
  #        dist <- ifelse(i == ran.segment, ifelse(add.in, 5 + remainder, remainder), 5)
  #        seg.end <- get.seg.end(seg.start, end.pos, dist)
  #        seg.mid <- get.mid(seg.start, seg.end)
  #        c(sub.segment = i, start.lat = seg.start[1], start.long = seg.start[2],
  #          end.lat = seg.end[1], end.long = seg.end[2],
  #          mid.lat = seg.mid[1], mid.long = seg.mid[2],
  #          dist = dist
  #        )
  # })



# Run code through data and write csv of segment positions

start <- find.start.point(GOALSIItracks, 1)
end <- find.end.point(GOALSIItracks, start)
segment.list <- list()
seg.num <- 1
segment <- data.frame()

while(!is.na(end)) {
            this.segment <- get.segment.positions(GOALSIItracks, start, end)
            #this.segment <- cbind(main.segment = rep(seg.num, nrow(this.segment)), this.segment)
            #segment.list <- c(segment.list, this.segment)
    if(is.data.frame(this.segment)) segment <- rbind(segment, this.segment)
            start <- find.start.point(GOALSIItracks, end)
            end <- find.end.point(GOALSIItracks, start)
            seg.num <- seg.num + 1
```

```
}
colnames(segment) <- c("sub.segment index", "start.lat", "start.lon", "end.lat", "end.lon", "mid.lat",
  "mid.lon", "dist","start.time", "end.time", "mid.time", "avg.beaufort")
segment$secs <- abs(difftime(segment$start.time, segment$end.time, units = "secs"))
#segment.mat <- do.call(rbind, segment.list)
#write.csv(segment.mat, file = "segment matrix.csv", row.names = F)
write.csv(segment, file = "segment matrix_GOALSII_Final.csv", row.names = F)
```

Script 2. Enter data into modeling console. This S-Plus code imported acoustic localization data into TIBCO Spotfire S+ Workbench. This step prepares the data by only accepting positive covariate values, applies a natural log function to surface chlorophyll to normalize values, and calculates oceanographic correlations. The script then generates matrices, such as those to indicate sample size and variable-to-variable correlations.

```
data.restore("E:\\DataFromNathanPC\\Data\\S-Plus\\Acoustic B\\ScriptsToUse\\Jessica Code\\step_gam_off.txt")

# Bring in the data.
GOA.all <- read.table("E:\\DataFromNathanPC\\Data\\S-Plus\\Acoustic B\\GOALS_II.csv",sep=",",header=T)

# Exclude segments with Beaufort > 5.
#GOA.bft5 <- GOA.all[GOA.all$beauf<= 5,]
# Delete all rows with -9999 values for in situ.
#GOA.nona <- GOA.bft5[GOA.bft5$SSS > 0 & GOA.bft5$MNP > 0 & GOA.bft5$CHL > 0,]
GOA.nona <- GOA.all[GOA.all$SST > 0 & GOA.all$SSS > 0 & GOA.all$DEP > 0 & GOA.all$SLP > 0 &
GOA.all$ASP > 0 & GOA.all$BTH > 0 & GOA.all$MNP > 0 & GOA.all$CUR > 0 & GOA.all$DIR > 0 &
GOA.all$TKE > 0 & GOA.all$SC > 0,]
#Get values for log of chlorofill and bind into dataframe
logC <- log(GOA.nona$SC)
GOA.f <- cbind(GOA.nona,logC=logC)


# Calculate sample size summaries
ACSummary <- matrix(0,1,3)
ACSummary[1,1] <- nrow(GOA.f[GOA.f$ACPM>0,])
ACSummary[1,2] <- sum(GOA.f$ACPM)
ACSummary[1,3] <- length(GOA.f$ACPM)



write.table(ACSummary, file="E:\\DataFromNathanPC\\Data\\S-Plus\\Acoustic B\\ACSummary.csv", sep=",")


# Calculate oceanographic correlations
cor.oceo <- matrix(0,9,9)
cor.oceo[1,1] <- cor(GOA.f$DEP,GOA.f$DEP)
cor.oceo[1,2] <- cor(GOA.f$DEP,GOA.f$SLP)
cor.oceo[1,3] <- cor(GOA.f$DEP,GOA.f$ASP)
cor.oceo[1,4] <- cor(GOA.f$DEP,GOA.f$logC)
cor.oceo[1,5] <- cor(GOA.f$DEP,GOA.f$MNP)
cor.oceo[1,6] <- cor(GOA.f$DEP,GOA.f$SST)
cor.oceo[1,7] <- cor(GOA.f$DEP,GOA.f$BTH)
cor.oceo[1,8] <- cor(GOA.f$DEP,GOA.f$DIR)
cor.oceo[1,9] <- cor(GOA.f$DEP,GOA.f$TKE)

cor.oceo[2,2] <- cor(GOA.f$SLP, GOA.f$SLP)
cor.oceo[2,3] <- cor(GOA.f$SLP, GOA.f$ASP)
cor.oceo[2,4] <- cor(GOA.f$SLP, GOA.f$logC)
cor.oceo[2,5] <- cor(GOA.f$SLP, GOA.f$MNP)
cor.oceo[2,6] <- cor(GOA.f$SLP, GOA.f$SST)
cor.oceo[2,7] <- cor(GOA.f$SLP, GOA.f$BTH)
```

```
cor.oceo[2,8] <- cor(GOA.f$SLP, GOA.f$DIR)
cor.oceo[2,9] <- cor(GOA.f$SLP, GOA.f$TKE)

cor.oceo[3,3] <- cor(GOA.f$ASP, GOA.f$ASP)
cor.oceo[3,4] <- cor(GOA.f$ASP, GOA.f$logC)
cor.oceo[3,5] <- cor(GOA.f$ASP, GOA.f$MNP)
cor.oceo[3,6] <- cor(GOA.f$ASP, GOA.f$SST)
cor.oceo[3,7] <- cor(GOA.f$ASP, GOA.f$BTH)
cor.oceo[3,8] <- cor(GOA.f$ASP, GOA.f$DIR)
cor.oceo[3,9] <- cor(GOA.f$ASP, GOA.f$TKE)

cor.oceo[4,4] <- cor(GOA.f$logC, GOA.f$logC)
cor.oceo[4,5] <- cor(GOA.f$logC, GOA.f$MNP)
cor.oceo[4,6] <- cor(GOA.f$logC, GOA.f$SST)
cor.oceo[4,7] <- cor(GOA.f$logC, GOA.f$BTH)
cor.oceo[4,8] <- cor(GOA.f$logC, GOA.f$DIR)
cor.oceo[4,9] <- cor(GOA.f$logC, GOA.f$TKE)

cor.oceo[5,5] <- cor(GOA.f$MNP, GOA.f$MNP)
cor.oceo[5,6] <- cor(GOA.f$MNP, GOA.f$SST)
cor.oceo[5,7] <- cor(GOA.f$MNP, GOA.f$BTH)
cor.oceo[5,8] <- cor(GOA.f$MNP, GOA.f$DIR)
cor.oceo[5,9] <- cor(GOA.f$MNP, GOA.f$TKE)

cor.oceo[6,6] <- cor(GOA.f$SST, GOA.f$SST)
cor.oceo[6,7] <- cor(GOA.f$SST, GOA.f$BTH)
cor.oceo[6,8] <- cor(GOA.f$SST, GOA.f$DIR)
cor.oceo[6,9] <- cor(GOA.f$SST, GOA.f$TKE)

cor.oceo[7,7] <- cor(GOA.f$BTH, GOA.f$BTH)
cor.oceo[7,8] <- cor(GOA.f$BTH, GOA.f$DIR)
cor.oceo[7,9] <- cor(GOA.f$BTH, GOA.f$TKE)

cor.oceo[8,8] <- cor(GOA.f$DIR, GOA.f$DIR)
cor.oceo[8,9] <- cor(GOA.f$DIR, GOA.f$TKE)

cor.oceo[9,9] <- cor(GOA.f$TKE, GOA.f$TKE)


write.table(cor.oceo, file="E:\\DataFromNathanPC\\Data\\S-Plus\\Acoustic B\\Cor.Oceo.csv", sep=",")
```

Script 3. Build model. This is the S-Plus code used for developing habitat models with an entire set of survey data and predictions. A forward and backward stepwise selection of variables was utilized. Observed/Predicted ratios (number of individuals), AIC values, dispersions and percent explained deviances were then written to matrices.

```
# Bring in function to store formula and AIC values from each trial in second call to step.gam
models.keep <- function(object, AIC)
list(term = as.character(object$formula)[3], AIC = AIC)

# ER GAMs All AC Gen BW
null.gam.ACPM <- gam(formula = ACPM~offset(log(dist)), family = quasi(link = log,variance = "mu"), data =
GOA.f)

null.gam.ACPM.1 <- step.gam.off(null.gam.ACPM, scope=list(

"DEP"=~1 + s(DEP,2) + s(DEP,3),
"logC"=~1 + s(logC,2) + s(logC,3),
"SLP"=~1 + s(SLP,2) + s(SLP,3),
"ASP"=~1 + s(ASP,2) + s(ASP,3),
"SST"=~1 + s(SST,2) + s(SST,3),
"SSS"=~1 + s(SSS,2) + s(SSS,3),
"MNP"=~1 + s(MNP,2) + s(MNP,3),
"DIR"=~1 + s(DIR,2) + s(DIR,3),
"TKE"=~1 + s(TKE,2) + s(TKE,3),
"BTH"=~1 + s(BTH,2) + s(BTH,3)))

null.gam.ACPM.2<- step.gam.off(null.gam.ACPM.1, scope=list(
"DEP"=~1 + DEP + s(DEP,2) + s(DEP,3),
"logC"=~1 + logC + s(logC,2) + s(logC,3),
"SLP"=~1 + SLP + s(SLP,2) + s(SLP,3),
"ASP"=~1 + ASP + s(ASP,2) + s(ASP,3),
"SST"=~1 + SST + s(SST,2) + s(SST,3),
"SSS"=~1 + SSS + s(SSS,2) + s(SSS,3),
"MNP"=~1 + MNP + s(MNP,2) + s(MNP,3),
"DIR"=~1 + DIR + s(DIR,2) + s(DIR,3),
"TKE"=~1 + TKE + s(TKE,2) + s(TKE,3),
"BTH"=~1 + BTH + s(BTH,2) + s(BTH,3)), keep = models.keep)


# AIC

# Store GAM AIC values from each step of second call to step.gam

x <- null.gam.ACPM.2
null.gam.ACPM.2.aic. <- data.frame(c(numerical.matrix(data.frame(x$keep[2,])),as.character(x$keep[1,])))

# DISPERSION AND DEVIANCE

# Store dispersion parameter and % explained deviance in a matrix

disdev.AC.gam <- matrix(0,1,4)

disdev.AC.gam[1,1] <- summary(null.gam.ACPM.2)$dispersion
```

```
disdev.AC.gam[1,2] <- null.gam.ACPM.2$null
disdev.AC.gam[1,3] <- summary(null.gam.ACPM.2)$deviance
disdev.AC.gam[1,4] <- (null.gam.ACPM.2$null - summary(null.gam.ACPM.2)$deviance)/null.gam.ACPM.2$null


# PREDICTION AND CALCULATION OF OBSERVED/PREDICTED SPATIAL RATIOS BY REGION AND
TOTAL AREA- ERGAM.2

# Predict number of individuals

p.null.gam.ACPM.2 <- predict.gam(null.gam.ACPM.2, GOA.f, type="response")


#Add predicted values to original segment database
Acoustic.segs.predictions <- cbind.data.frame(GOA.f, "predict.ACPM" = p.null.gam.ACPM.2)

write.table(Acoustic.segs.predictions, file="E:\\DataFromNathanPC\\Data\\S-Plus\\Acoustic.segs.predictions.csv",
sep=",")
```

Script 4.Plot functional relationships. S-Plus code used for plotting GAMs in different scales. These graphs were used to analyze the functional relationship of modeling covariates.

```
#graphsheet(page="every graph")

par(mfrow=c(2,2))
plot.gam(null.gam.Model4.2)



# Scaled

#graphsheet(page="every graph")

par(mfrow=c(2,2))
plot.gam(null.gam.Model4.2,scale=16)


#graphsheet(page="every graph")

par(mfrow=c(2,2))
plot.gam(null.gam.Model4.PD.2,scale=16)

#graphsheet(page="every graph")

par(mfrow=c(2,2))
plot.gam(null.gam.BW.2,scale=16)

#graphsheet(page="every graph")

par(mfrow=c(2,2))
plot.gam(null.gam.ACBb.2,scale=16)


#graphsheet(page="every graph")

par(mfrow=c(2,2))
plot.gam(null.gam.ACBb.PD.2,scale=16)

#graphsheet(page="every graph")

par(mfrow=c(2,2))
plot.gam(null.gam.Bb.2,scale=16)
```

Script 5.Associate predictions with cell midpoints. S-Plus code used for relating GAM predictions to the grid database developed in section 3.3.3. The script brings in the prediction and grid data, then prepares the data using similar techniques as Script 2. The data prediction data is associated with the grid cell midpoints using CellIDs, the grid is then exported out with Lat/Lon attributes written to a database for import into the GIS.

```
#rm(list=ls())

##########################################################################################
##############
# Bring in the data
# Delete all rows with -9999 values for in situ.  I am assuming that all cells have depth and isobath values.
# Get ln of chlorophyll, ln of depth, add year, and add effort = 1.
p.data.2013 <- read.table("E:\\DataFromNathanPC\\Data\\GridPredict.csv",sep=",",header=T)
p.data.2013.nona <- p.data.2013[p.data.2013$SST > 0 & p.data.2013$DEP > 0 & p.data.2013$SLP > 0 & 
p.data.2013$BTH > 0 & p.data.2013$TKE > 0 & p.data.2013$SC > 0,]
# p.data.2013.nona <- p.data.2013[p.data.2013$SST > 0 & p.data.2013$SSS > 0 & p.data.2013$SC > 0 & 
p.data.2013$MLD > 0,]
#logC <- log(p.data.2013.nona$SC)
#p.data.2013.effort1 <- cbind(p.data.2013.nona,effort=1)
p.data.2013.effort1 <- cbind(p.data.2013.nona,effort=1,dist=1,year = 2013)
write.table(p.data.2013.effort1, file="E:\\DataFromNathanPC\\Data\\S-Plus\\Acoustic 
B\\Model_4\\p.data.2013.effort1.csv", sep=",")

# d2isoneg <- p.data.1991.effort1$DEP
# d2isoneg[d2isoneg>=-200] <- -1
# d2isoneg[d2isoneg<(200*-1)] <- 1 # This line uses (200*-1) instead of -200 because the <-200 kept being 
confused with an assignment.
# p.data.1991.effort1 <- cbind(p.data.1991.effort1,d2isoneg=d2isoneg*p.data.1991.effort1$dist2iso200m)

# Predict on the grids using the isobath models and effort1

# Make predictions
p.2013.grid.predict.ACPM <- predict.gam(null.gam.Model4.2, p.data.2013.effort1, type="response")


#MAKE SURE CELL ID IS IN YOUR GRID, LAT AND LON NEED TO BE NAMED HOWEVER THEY ARE 
NAMED IN THE GRID DATABASE
grid.d2isoneg.effort1.pred.2013 <- cbind.data.frame(CELLID = p.data.2013.effort1$CELLID, LAT = 
p.data.2013.effort1$LAT, LON = p.data.2013.effort1$LON, denACPM = p.2013.grid.predict.ACPM)

# Output the data

write.table(grid.d2isoneg.effort1.pred.2013,file="E:\\DataFromNathanPC\\Data\\S-
Plus\\GOALS_II__MODEL4_PM_PREDICTIONS.csv",sep=",")
```