UNIVERSITY OF CALGARY

A marine GIS case study of micro-scale gray whale (*Eschrichtius robustus*) habitat-use off Vancouver Island, British Columbia

by

David Nathan Laskin
B.Sc., University of Victoria, 2000

A DOCUMENT
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF GEOGRAPHIC INFORMATION SYSTEMS

DEPARTMENT OF GEOGRAPHY

CALGARY, ALBERTA
JUNE, 2007

ABSTRACT

GIS based habitat modeling and a modified form of binary logistic regression were used to assess habitat-use of gray whales (*Eschrichtius robustus*) along the south coast of Flores Island, British Columbia. Principle objectives include the derivation of a resource selection function (RSF) to determine relative likelihood of use of available bathymetric depth, slope and complexity features within the study area. Micro-scale pelagic currents were subsequently incorporated to examine their potential impact on gray whale habitat use. 877 whale presence observations were contrasted with bathymetric GIS layers to produce a RSF identifying increased whale occurrence in waters ~10 meters deep, in combination with areas having higher benthic topographical complexity. Acoustic Doppler current profiling data were used to derive continuous, dynamic current surfaces at three separate depths of the water column. The effects of current speed and direction on foraging whales were found to be negligible, but areas with south flowing surface currents consistently predicted increased use.

Keywords: gray whale, resource selection function, GIS, logistic regression, habitat model, marine GIS.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$\vec{V}_x$	Total magnitude in plane of x
$\vec{V}_y$	Total magnitude in plane of y
$\vec{V}$	Total velocity
θ	Current flow direction in radians
β_o	Constant, y intercept
y	Dependent variable, presence-available
χ_k	Independent variable at location χ
β_k	Logistic regression coefficient
β_o'	Constant term in the fitted logistic regression equation
P	Probability
P_a	Probability of sampling an available resource unit, independently of the Selection of any other unit
P_u	Probability of sampling a used resource unit, independently of the Selection of any other unit
$w^*(x_i)$	Resource selection probability function for a single period of selection where the probability that a unit is used as a function of variables $\chi = (\chi_1, \chi_2 .. \chi_k)$ measured on the unit used to describe it
$\tau(x_i)$	Probability of observing resource unit i as used, given that it is one of the samples
$w(x)$	Resource selection function, an estimate of the relative likelihood of use for the entire sample (the exponential model)

ACKNOWLEDGEMENTS

Dr. Dave Duffus for his research guidance and generous contribution of a decades worth of data. Brian Kopach and Sonya Meier for providing the groundwork for this study. Dr. Darren Bender, my supervisor, for refining my perspectives on landscape ecology. Dr. Greg McDermid whose cavalier suggestions provided clarity in times of uncertainty. Dr. Scott Nielsen, Dr. Tak Fung, Dr. Larry Bates, and Dr. Bryan Manly for liberally contributing their time to help solve my statistical and mathematical queries. Dr. William Megill for facilitating my international voyages in search of our enigmatic quarry. Joe Breman for his overall enthusiasm. Michelle Kinzel and Jeff Ardron for providing the large-scale bathymetric imagery. Shawn Mueller and Robin Poitras for their ruthless evaluation of my cartographic skills. Dr. Guillermo Castilla for his critical remarks and positive critiques. My family for their persistent attempts to distract me – and finally the whales, for tolerating my consistent efforts to ride them.

1 INTRODUCTION

Sample data are rarely available to explain species presence at every location on the landscape. Therefore, models are used to extrapolate beyond the locations where species presence is known, by linking presence to underlying habitat variables (Pearce and Boyce 2005). However, accurately predicting species interaction with their environment is a challenging task (Scott *et al.* 2002). The complex relations that define ecological processes are often dynamic and blur the patterns that statistical models attempt to uncover (Huston 2002). This vagueness is proliferated when the habitat under study is submerged, and the processes therein are inherently more obscured (Urbanski and Szymelfenig 2003).

Many statistical techniques have been used to accurately model habitat, with the intent of predicting species occurrence solely from environmental variables (Scott *et al.* 2002; Guisan and Zimmermann 2000). Selecting the most effective quantitative method depends on the amount and quality of species presence data, and landscape information. Additionally, not all of these data suit the popular statistical models, often violating the assumptions that govern hypothesis testing (Lowell *et al.* 2003).

A common modeling technique employs binary logistic regression, a method that assesses how a species uses resource units within an area of interest (MacKenzie *et al.* 2006). Logistic regression utilizes dichotomous data of used and available habitat to produce a resource selection function (RSF). This method is favourable because it accommodates *presence-only* data for which there is no information on locations where a species does not occur on the landscape, a frequent result of species sampling. There are several approaches for constructing models using this type of data (Pearce and Boyce 2005; Manly *et al.* 2002).

The exponential form of logistic regression offers a more representative prediction of species occurrence than other common logistic models (Manly 2002). Species presence data are often paired with an independent sample of available habitat data subsequent to being in the field. These available sites are generally sampled post-hoc from habitat layers represented in a geographic information system (GIS) (Erickson *et al.* 1998). The exponential model is particularly well suited for this sample design, and does not require a measure of species abundance (Manly *et al.* 2002).

This analysis demonstrates the functional application of the exponential model on foraging gray whale (*Eschrichtius robustus*) locations on the west coast of Vancouver Island, British Columbia. The derived RSF surface depicts areas of increased likelihood of gray whale habitat-use, which has potential management applications that include: delineating the spatial extent of marine protected areas (Short 2005), monitoring effects of boat traffic on foraging sites (Duffus 1996), and examining shifts in whale occurrence as a result of changing oceanic climate (Grebmeier *et al*. 2006; Moore *et al*. 2003)[1].

Species presence data were obtained from a continuing, long-term study by the University of Victoria Whale Research Laboratory. A previous investigation of whale habitat use was performed here by Meier (2003). The results were significant at various spatial and temporal scales, showing that whales select for particular depths and bathymetric slopes. Kopach (2004) expanded this investigation in pursuit of linking whale occurrence to dynamic current features in the study area. Although no association was found, a complex portrayal of current dynamics within the study area was attained.

The whale occurrence data are suitable candidates for the exponential model using *Study Design 1* as outlined by Manly *et al.* (2002:6), which utilizes transect-based presence surveys, and a random sample of available habitat extracted from a GIS. The RSF surface was derived while considering the following objectives at a fixed spatial micro-scale across several seasons:

Objective 1: Use logistic regression to confirm the findings of Meier (2003), and identify gray whale preference for specific bathymetric slope and depth.

Objective 2: Construct a RSF model through the introduction of additional static marine landscape variables: Benthic topographical complexity, distance from shore, and a polynomial form of depth.

Objective 3: Examine the influence of dynamic, multi-depth, continuous marine current variables on the model.

[1] Conservation status of gray whales, refer to Appendix C

GIS was used to facilitate the modeling process. This analytic geographic environment permits the integration of multiple habitat layers that represent the natural features negotiated by gray whales in the ocean. This case study offers insight into the effectiveness of habitat-use models within this unique ecological context. Although GIS retains fixed terrestrial origins, it has only recently been adapted for use in aquatic ecosystems (Wright 2000; Li and Saxena 1993).

The marine environment presents additional GIS modeling complexities with the introduction of an intrinsic third dimension, perpetual dynamics, and indistinct boundaries (Valavanis 2002). The ocean obscures regular human observation, making data collection difficult compared to land-based investigations (Wright and Goodchild 1997; Lockwood and Li 1995). Unlike land, which has rigid coordinate systems, few things in the ocean are static. There is a significant lack of control points, datums, and landmarks available to model this space (Wright 2002). This has prompted developments in GIS software to accommodate these ambiguities; the result of the process has been termed *Marine GIS* (Wright and Scholz 2006; Breman 2002). This particular nomenclature does not refer simply to the use of GIS in a marine context, but designates an independently burgeoning branch of GIS itself[2].

2 BACKGROUND

2.1 PATTERNS OF SELECTION AND HABITAT USE

In a marine context, this investigation considers habitat to be defined as the characteristic space occupied by an individual, population or species (Baretta-Bekker *et al*. 1992). RSF modeling provides a controlled methodology that identifies relationships between variables coexisting in this characteristic space. The link between whales and their abiotic habitat is not direct, rather, intermediate factors such as prey influence their distribution. An animal is going to occupy the space that offers the most favorable biological requirements for that individual. Patterns of occurrence are influenced by the availability of food and the energy required to obtain it (Dunham and Duffus 2002).

[2] For a complete review of concepts in *marine GIS*, refer to Appendix B

RSF models isolate these patterns by statistically combining available environmental attributes (extruded from GIS layers) with whale presence sites (Manly *et al.* 2002).

Gray whales are unique because they are the only cetaceans to exploit both the open water and seafloor features of their habitat (Rice and Wolman 1971). As a result, both static (benthic) attributes and dynamic (current) attributes are considered in this study. By utilizing both static and dynamic spatial elements, whales increase their probability of encountering areas of concentrated prey.

A population of approximately 250 summer residents forage along the coast of Vancouver Island (Megill *et al.* 2003; Dunham and Duffus 2002, 2001; Darling *et al.* 1998). Their primary prey species are sub-benthic invertebrates that live in seafloor sediments (Moore 2003). These include tube dwelling ampeliscid amphipods (*Ampelisca* spp.) and ghost shrimp (*Callianassa californiensis*) which occur nearshore in waters up to 35 m deep (Dunham and Duffus 2002, 2001; Darling *et al*. 1998). The whales essentially vacuum soft sediments from benthic substrate into their mouths, then expel the silt while retaining prey with their comb-like baleen (Nerini 1984).

Gray whales also forage on dense groups of swarming plankton. These species include hyper-benthic mysids (*Holmesimysis sculpta*) and pelagic porcelain crab larvae (*Petrolithes eriomerus*) (Darling *et al*. 1998). Mysids consistently swarm above shallow rock reefs up to 15 m deep, while porcelain crab larvae are found in greater concentrations above boulder covered substrate in waters less than 30 m (Megill *et al*. 2003; Dunham and Duffus 2002, 2001). In essence, these prey are directly linked to the surrounding static and dynamic landscape, acting as the spatial connection between whales and the places they occupy.

This investigation parallels previous research linking whale distribution to oceanic landscape variables using GIS (Yen *et al*. 2005; Meier 2003; Moses and Finn 1997). Moore *et al*. (2002) established a connection between northwest Pacific blue whales (*Balaenoptera musculus*) and seafloor bathymetry, sea-surface temperature, surface currents, and chlorophyll-a concentration. Chlorophyll-a acts as an indicator of primary production and is directly related to the principal prey of baleen whales: plankton (Littaye *et al*. 2004). Gregr and Trites (2001) made predictions of critical habitat for five whale species off the coast of British Columbia, by correlating historic whaling location records with seafloor depth, slope, sea-surface temperature, and salinity. Depth was shown to be a strong influence in distribution for most whale species. Moore *et al*. (2003) sampled benthic prey biomass directly to characterize the distribution of gray whales in the northern extent of their range. A variety of statistical models were used in this previous research to predict whale distribution patterns based on physical oceanographic habitat attributes. For this case study, the exponential form of logistic regression has been selected.

2.2 THE EXPONENTIAL MODEL

Originally explained in an ecological context by Manly (1992), and later readdressed by Manly *et al*. (2002), the exponential form of the logistic regression model is used to derive realistic estimates of resource selection by animals. It has been applied in several ecological studies (Nielsen 2005; Boyce and McDonald 1999; Erickson *et al.* 1998) in an attempt to quantify species-landscape associations.

Logistic regression is widely used in biology because it conveniently restricts predicted probability values between 1 and 0, and is similar to linear regression, but forgives many of the assumptions (Manly *et al*. 2002). This restriction is based on the dichotomous nature of species location data, being either present (1) or absent (0), matching the logistically confined probabilistic result of either high (1) or low (0) use. The purpose of using a logistic transformation is to accommodate the binary proportions of the response variables. Unlike a linear model where a single variable (y) is regressed against a set of independent predictor variables ($\chi_1, \chi_2 .. \chi_k$), logistic regression uses a link function to combine the ratio of presence vs. absence in the data to act as a single variable (y). This way, the logistic regression assumes many of the desirable properties of a regular linear model (Hosmer and Lemeshow 2000).

When modeling presence-only data, a sample of non-presence is also required. True absence across the landscape is not known because it was not measured concurrently during presence sampling. Traditional logistic regression assumes that both were sampled simultaneously within the same timeframe (Manly 1992). Absence sites can be obtained after being in the field, by randomly sampling habitat layers in a GIS. This is now in fact *pseudo-absence*, because true absence was still never recorded. The predictive power of the traditional logistic model is slightly weakened by the use of pseudo-absence, because many cases of absence may contain an unknown number of presences, acting to contaminate the sample (Pearce and Boyce 2005).

Alternate forms of logistic regression exist to accommodate the various sampling techniques used to acquire pseudo-absence. Depending on the definition, one of two models can be used: *presence-absence* or *use-available*. The former is utilized to contrast consumed resource units against the characteristic units where use has not been recorded. The latter considers all resource units to be available for use, but some units are used more frequently than others. The distinction between these definitions is small, because their sampling designs are similar, but there are larger conceptual differences between them. This is a main reason for differing logistic modeling techniques (Pearce and Boyce 2005). The exponential model in particular assumes the dichotomous data to be defined as used (1) and available (0) habitat[3]. Therefore this case study defines the RSF model as *use-available* herein.

Traditional logistic regression requires that total species abundance in the study area is known, so that sampling probabilities of use and available can be calculated. However, actual abundance is rarely measured, and the proportion of available habitat is randomly sampled using GIS. The equation constant (β_0) that contains these probabilities is dropped from the logistic regression equation, because it is meaningless without species abundance (Manly *et al.,* 2002). The constant is simply an extra parameter that scales the intercept to reflect how rare or common a species is on the landscape, which is unknown in the first place (Nielsen pers. com. 2006). Equation 2.1 then assumes the exponential form:

$$w(x) = exp(\beta_1\chi_1 + \beta_2\chi_2 + ... + \beta_k\chi_k) \qquad (2.1)$$

The exponential model now only effectively estimates relative likelihood of occurrence rather than calculating true probability. Firstly since the sampling probabilities are removed, and secondly because the proportion of used and available

[3] The derivation of the exponential model is explained in Appendix A

samples has been arbitrarily determined. The results of the model are still approximately equal to the traditional logistic regression model, and reflect the proportional probabilities of habitat use (Manly pers. com. 2006).

3 METHODS

3.1 SCALE

Levin (1992) states that a substantial problem in ecology is to relate ecological phenomena across scales. Scale is considered the continuum through which entities, patterns, and processes can be observed and linked (Marceau 1999). Meier (2003) found that gray whales display differing patterns of selection at separate spatial and temporal scales. This analysis uses a fixed spatial and temporal scale to focus on the effects of how additional habitat variables influence occurrence.

Gray whales have the longest migration of any mammal, traveling over 16,000 km annually from their summer feeding grounds in the Chukchi, Bering, and Beufort Seas, to the shallow lagoons of Baja Mexico to reproduce (Rice and Wolman 1971). The breadth of scale at which they interact is vast, and presents differing criteria for habitat-use across these scales. The summer residents do not complete the journey to the arctic, but preferably consume substantial nutrients in habitual feeding sites along the coast of Vancouver Island. For the purpose of this investigation, and in contrast to the entire range of the summer residents, the study area can be considered a micro-site (Figure 3.1).

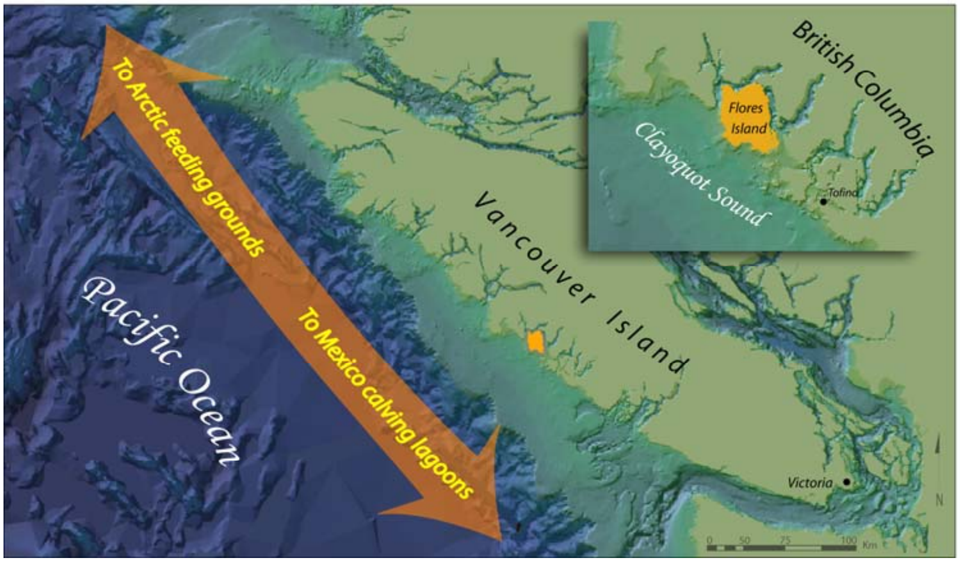

Figure 3.1 Vancouver Island, gray whale migration route, and Flores Island[4].

The spatial extent of the study area is limited by the extent of the University of Victoria whale census transect, and is coincident with the coverage of the GIS layers that characterize available habitat. Although the study area is constrained by these layers, it is recognized that in reality, these boundaries may not coincide with the actual limits of availability (Erickson *et al.* 1998). There is no established method for defining a study area, but these constraining factors ultimately limit the spatial extent of the analysis (Marceau *et al*. 1994). It was determined that the grain of the GIS layers must be finer than the scale at which whales congregate within the micro-site, in order to detect variation in habitat at individual locations (Duffus pers. com. 2006). This spatial resolution was kept consistent between layers at <30 m, which is within the GPS accuracy used to record whale locations.

Temporally, surveys from nine independent, four month-long seasons were available. Given enough time, presence could eventually occur at every point in the study area, possibly obscuring any spatial patterns. Therefore, a subset of these data was used; by only observing occurrence during a brief temporal window from every independent season. Observations during July were chosen because calm weather

[4] Background bathymetry credit: Jeff Ardron, Living Oceans Society.

permits a more consistent survey effort, and the summer resident population has stabilized in the region.

The landscape encompassing the study area is an aggregation of adjacent hierarchical levels of spatial and temporal scales (O'Neill *et al*. 1986). The focus of this case study is on a single level of available scales, with no quantitative effort to investigate how ecological patterns and processes are linked across the hierarchy.

3.2 STUDY AREA

The population of summer residents have displayed long term site fidelity to feeding sites in Clayoquot Sound, BC. (Moore *et al.* 2003; Calambokidis *et al.* 2002). The study area contains several of these feeding sites along the southwest coast of Flores Island, which is flanked by non-productive areas that are rarely used by foraging resident whales (Meier 2003). The area extends along approximately 12 km of shoreline from Dagger Point in the northwest, to the eastern edge of Cow Bay (Red Rocks) (Figure 3.2). The study area is characterized by an assortment of physical substrate types including shallow rocky reefs, boulders, mud and sand bays, with water depths ranging to 35 m (Dunham and Duffus 2001).

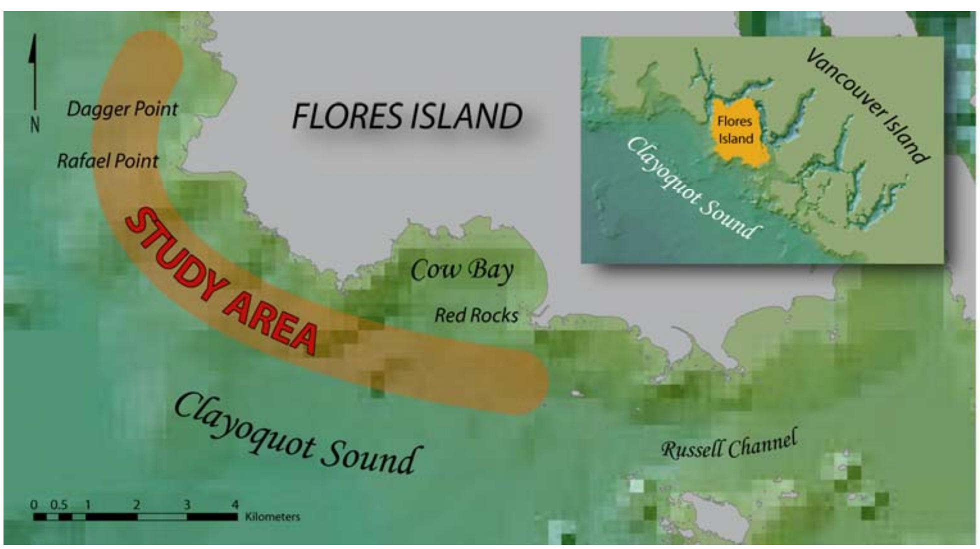

Figure 3.2 Flores Island study area.

3.3 PRESENCE DATA

Whale location data were collected from late May to early September from 1997 to 2005, weather permitting. This analysis used every presence site that occurred within the month of July across nine seasons, totaling 877 points. Vessel based surveys were conducted along transects as described by Kerr (2005) and Meier (2003). An inshore transect was followed ~350 m from shore heading north, then adhering to a return southbound transect ~2 km offshore. Surveys were conducted with a crew of three to seven observers following speeds of ~15 km/hr (8.5 knots), with individuals scanning fore, aft, and on each side of the vessel for whale blows. When a whale was observed, the vessel would deviate from the transect to approach the animal, and collect its position using a GPS, digital flux compass, and a laser range finder. The vessel then returned to the transect and continued the survey. Positional accuracy is assumed to exist in the region of ±100 m.

3.4 AVAILABLE HABITAT: GIS SURFACE DESIGN

Model layers were selected using *a priori* knowledge of gray whale foraging behavior, and their exploitation of benthic and pelagic habitat. Variables were also considered on their use and success in previous resource selection investigations (Kopach 2004; Meier 2003). It is assumed that each candidate variable plays an active role in selection by gray whales, representing both static and dynamic oceanographic entities.

3.4.1 DEPTH

Nearshore depth was derived from multibeam sidescan sonar data provided by Nautical Data International Inc., an associate of the Canadian Hydrographic Service. Sonar data were converted to a grid of discrete points with 50 m intervals, and then interpolated to create a continuous bathymetric seafloor surface for the extent of the study area. Interpolation was executed using kriging in *ArcGIS Geostatistical Analyst* (ESRI 2006). Kriging estimated unknown values of depth for areas between sonar points. Unlike other interpolators, kriging determines the non-arbitrary range where autocorrelation decays, and spatial dependence dissipates. Sonar points beyond this distance were not used for weighted estimates.

The derived bathymetric surface had an overall RMS error of 0.84 m, with a 30 m spatial resolution (Figure 3.3). This surface provided the foundation for the other static landscape features of the same resolution. This layer was used to confirm the significance of depth as a static indicator of whale selection. Where Meier (2003) found preference for shallower depths by whales, possibly indicating better foraging in these areas.

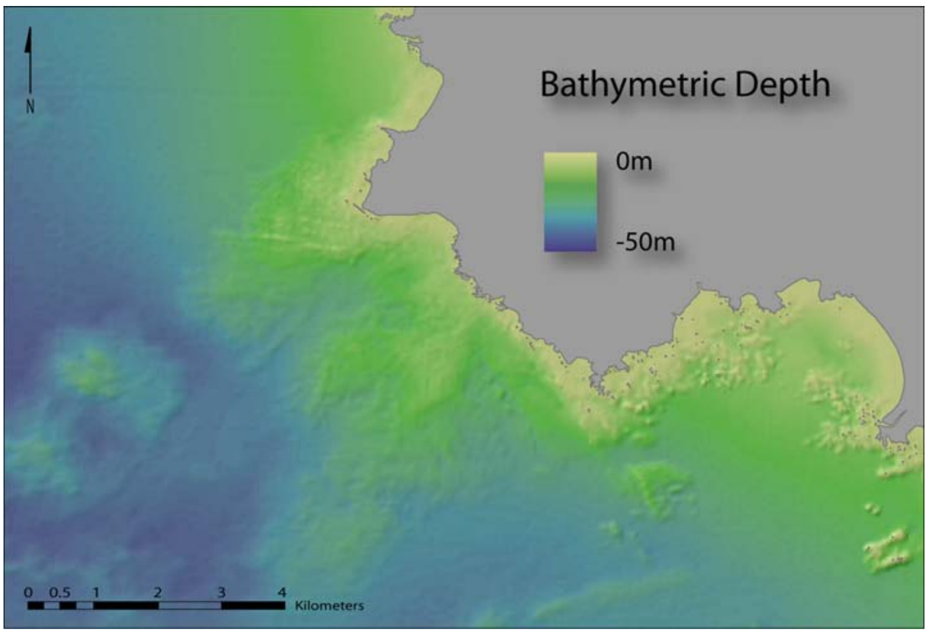

Figure 3.3 Interpolated bathymetric depth surface.

3.4.2 SLOPE

Bathymetric slope was derived from the bathymetric surface using the *Surface Analysis* tools in *ArcGIS Spatial Analyst*. The slope surface is measured in degrees, showing minimal relief in the study area with a maximum slope of 19º (Figure 3.4). Meier (2003) found disproportionate use of available bathymetric slope by whales.

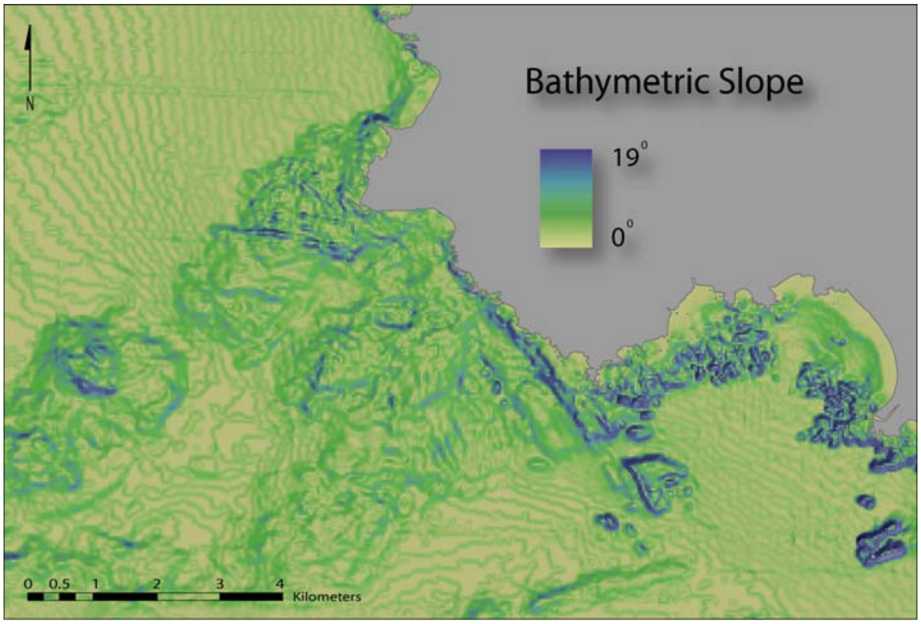

Figure 3.4 Bathymetric slope (degrees).

3.4.3 DISTANCE FROM SHORE

Gray whales forage along the coast, and are rarely found more than a couple kilometers offshore at any point during their migration (Rice and Wolmann 1971). Distance from shore was calculated using the *Euclidean Distance* tool in *ArcToolbox*, with a resolution of 10 m (Figure 3.5).

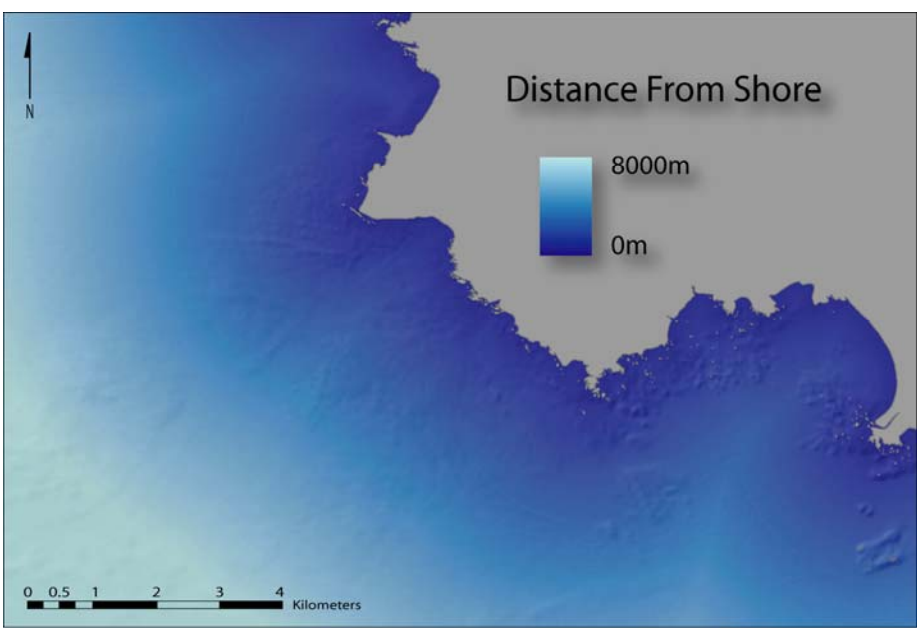

Figure 3.5 Distance to shore (meters).

3.4.4 BENTHIC TOPOGRAPHICAL COMPLEXITY

Benthic topographical complexity is an indicator of heterogeneous areas commonly associated with high species richness. It is often possible to use benthic complexity as a surrogate for species abundance in the absence of comprehensive biological data (Ardron 2002). Species exist in greater diversity in complex areas that provide more available niches. Increased niches support more concurrent life stages, offering greater resiliency when a swarm of plankton is consumed by a passing whale for instance. These areas will frequently attract more intense foraging because they can quickly rebound from heavy use (Patterson 2004).

Benthic complexity measures how frequently the slope of the seafloor changes, it is a measure of variability. Slope represents steepness and relief looks at the maximum change in depth; benthic complexity examines the intensity of convolutions along the bottom. A similar measurement is rugosity, where bottom surface distance is divided by the total planar distance. This can be strongly influenced by single large change in depth, where complexity considers all surface variations more equally (Ardron 2002).

The derivation of this marine specific surface was adapted form Ardron (2002). The bathymetry layer was exaggerated in *ArcMap Raster Calculator* to expose smaller shifts in depth that may have been overlooked. The exaggeration also groups very steep features together as they generally dominate the results. The slope of the exaggerated depth was calculated, and then again taking the slope of this surface identified dense areas of seafloor variation. The technique creates a final surface that diffuses complexity around abrupt linear features such as ridges, and displays convoluted regions with increasing values of higher benthic complexity (Figure 3.6). This is an enhanced representation of the marine environment, an improvement over the discrete categories common to most GIS habitat layers.

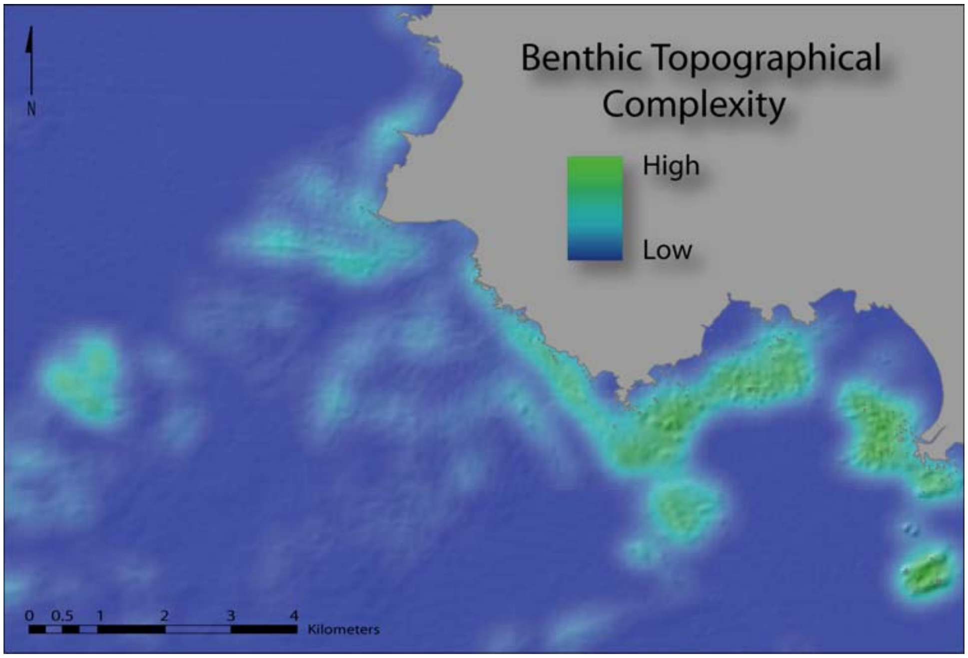

Figure 3.6 Benthic topographical complexity.

3.4.5 CURRENT VELOCITY

The interpolated dynamic variables originate from discrete data collected by Kopach (2004) using an acoustic Doppler current profiler (ADCP). The ADCP measures the Doppler shift of sonar pulses reflected off particles suspended in the water column (RD Instruments 1996). The ADCP collects both current velocity (cm/s) and flow direction (360º) in 50 cm bins extending from the surface to the seafloor. Bins were averaged into three layers of the water column: top, middle, and bottom (Kopach 2004). A series of vessel based transect ensembles were run both perpendicular and parallel to shore throughout the study area. Each transect was sampled four times in total, twice during ebb (receding), and flood (advancing) tides.

Kopach (2004) determined that tides are the major driving force of current dynamics in the study area. It is believed that currents physically concentrate swarming mysids and provide refuge in slower waters, while faster currents would sweep them away. Areas of slower velocity may produce detritus stalls and fallout nutrients would supply both pelagic and benthic prey (Kopach 2004). Gray whale summer residents have an increased reliance on prey suspended in the water column, but Kopach (2004) found that whales did not seem to select for any particular current speed. This analysis was limited by the difficulty of synchronizing whale presence sites with discrete ADCP ensemble locations.

Interpolated current GIS surfaces permitted the use of all whale occurrences, regardless of their proximity to the discrete ADCP measurements. Unlike other phenomena such as temperature or elevation, the interpolation of velocity was much more complex. By nature, any object with velocity has both speed and direction components acting on the whole. For example, the interpolated velocity in between two currents depends if they are traveling on a collision course, which will ultimately affect their velocities (Figure 3.7). Interpolation of velocity must account for the influence of directional momentum impacting the final values of the interpolated surface.

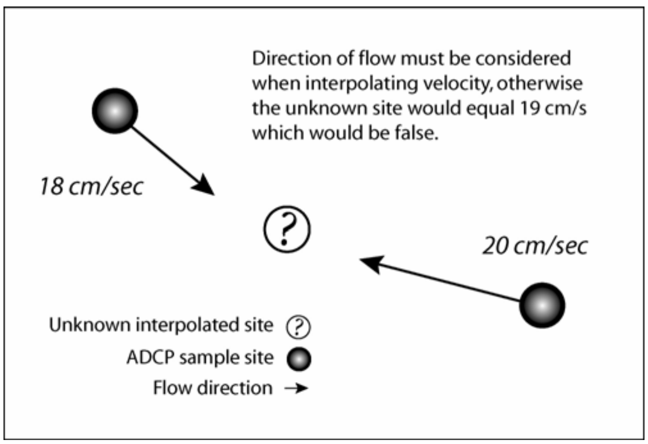

Figure 3.7 Impact of flow direction on the interpolation of velocity.

A solution was to decompose velocity into directional components of *x* and *y*. interpolate them separately, and then recombine to form the final velocity surface. The direction of travel was converted from degrees to radians. Since the azimuth originates at 0º, the total velocity $\vec{V}$ is multiplied by the cosine and sine of θ to produce component magnitudes $\vec{V}_x$ and $\vec{V}_y$ respectively (Equations 3.1 & 3.2).

$$\vec{V}_x = \left|\vec{V}\right| sin\theta \tag{3.1}$$

$$\vec{V}_y = \left|\vec{V}\right| cos\theta \tag{3.2}$$

Each vector component was interpolated using *Geostatistical Analyst*. Kriging was initially attempted, but a lack of spatial autocorrelation in the vector components resulted in no suitable semivariogram. Most likely a result of fine scale current turbulence such as rapid shifts in speed and direction. Inverse distance weighting (IDW) was used instead, and is analogous to kriging a surface with no spatial trend, because an arbitrary range is employed. IDW estimates unknown locations by receiving more influence from nearby points, and inversely reducing the influence from distant points. Shifts in speed and direction occur very quickly in the ADCP measurements, a shorter range was used as not to include weights from distant dissimilar points. The power value of the IDW was increased to preserve the abrupt spatial variation in a regionalized structure similar to Thiessen polygons (Figure 3.8).

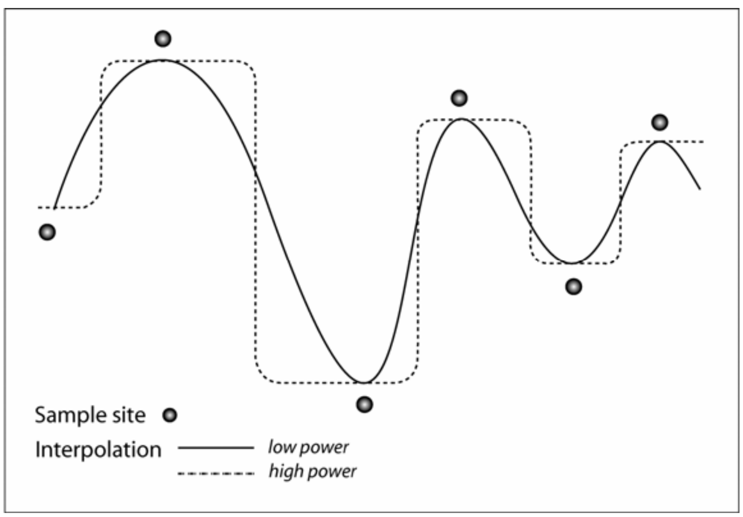

Figure 3.8 Conceptual diagram of regionalization as an effect of increasing IDW power.

This heavily localized structure on the final surface implicitly generated regions of confidence surrounding each ADCP point. Following the interpolation of each vector component, both surfaces were recombined by applying the root sum of squares to each surface (Equation 3.3) using *Raster Calculator* in *Spatial Analyst*.

$$\vec{V} = \sqrt{\vec{V}_x^2 + \vec{V}_y^2} \tag{3.3}$$

The result was final velocity surfaces for the top, middle and bottom of the water column during both ebb and flood tides (Figure 3.9).

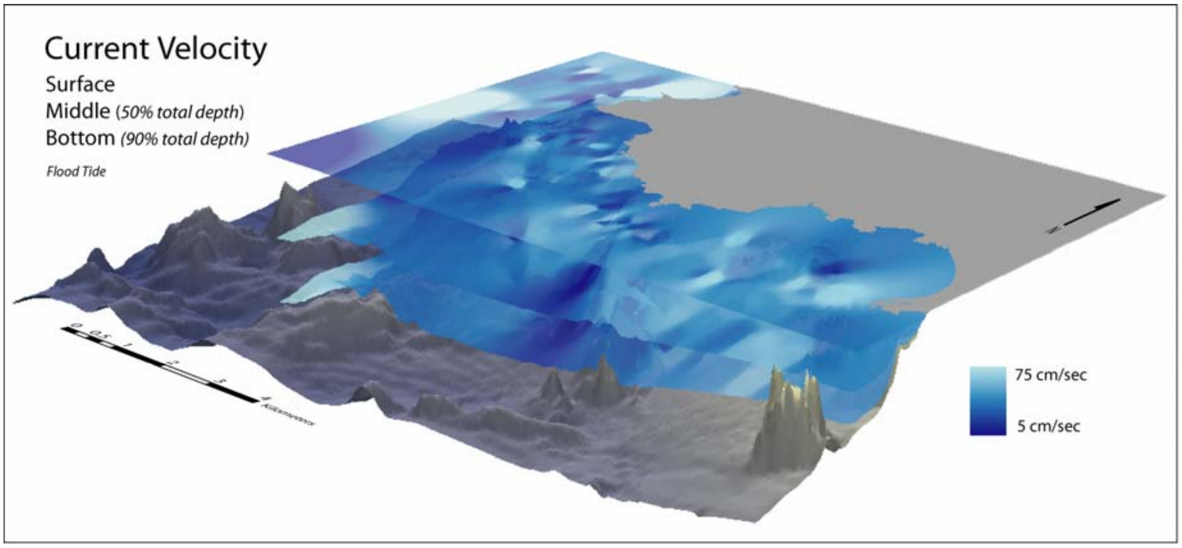

Figure 3.9 Interpolated current velocity (cm/sec) at the surface, middle and bottom of the water column.

3.4.6 CURRENT DIRECTION

The ADCP provides current flow direction at the surface, middle, and bottom of the water column. A circular transformation was performed to avoid problems interpolating similar directions represented by very different values (i.e., 5º and 355º). The cosine of their radians converted north and south into values of +1, and -1, while the sine of their radians changed east to +1 and west to -1. As a result, two separate direction surfaces were used in the model, *north-south* and *east-west* flow (Figure 3.10).

Similar to the interpolation of velocity, direction was found to be intrinsically linked with speed. Therefore, current momentum was incorporated as an inertial component acting on the interpolated values of direction. North-south corresponds with values of $\vec{V}_y$ and east-west is represented by $\vec{V}_x$ for both ebb and flood tides (Equations 3.1 & 3.2). Prevailing current direction is assumed to be analogous to the effects of prevailing wind in a terrestrial habitat, an influence on gray whale occurrence that has not been previously explored.

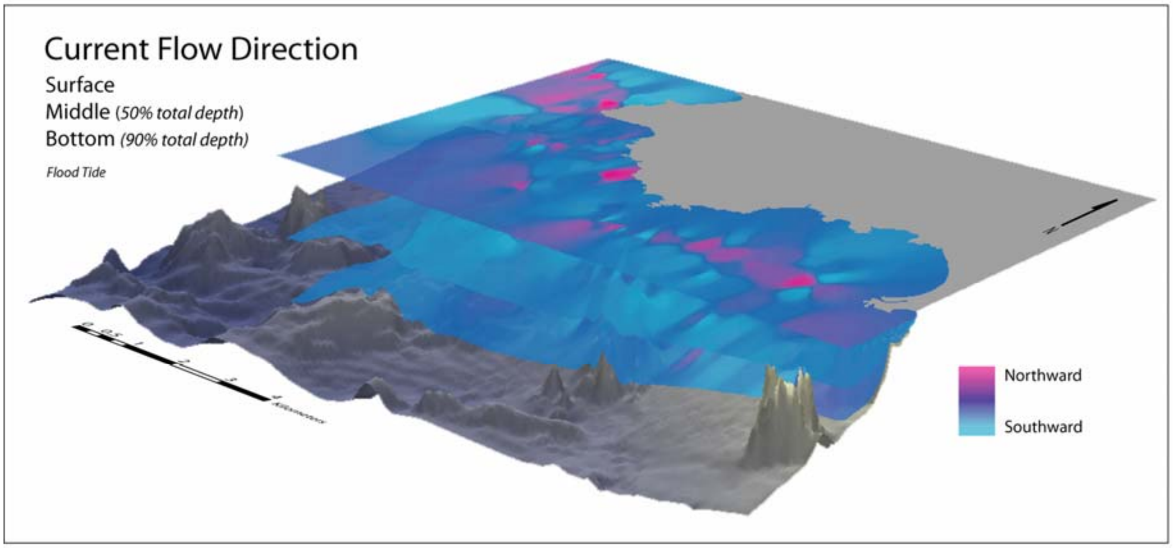

Figure 3.10 Interpolated current flow direction at the surface, middle and bottom of the water column.

3.5 STATISTICAL ANALYSIS

The RSF model considers all habitat units available for use, but some are more frequently used than others (Pearce and Boyce 2005). This definition follows that outlined in Manly *et al.* (2002) from *Study Design 1* where presence measurements are made at the population level and GIS layers are used to randomly census habitat availability. The exponential model produces a RSF that predicts relative likelihood, not probability of use, therefore the ratio of presence to available points needs only to reflect that the available points are actually representative of what is 'available' (Manly *et al.* 2002).

There are no strict means of determining the number of random available habitat units. However, the higher the sampling intensity the better, this ensures that the RSF coefficients are stable by effectively sampling what is available, and solely reflect the result of animal use. A drawback of a larger available habitat sample size is the reduced effectiveness of model selection techniques such as Akaikes Information Criterion (AIC) (Nielsen pers. com. 2006).

Increasing the number of available units will eventually exceed the point where available resources are so well defined that further increase is unnecessary (Nielson *et al.* 2003). For this model, a magnitude of three times the number of presence points was randomly sampled to derive the available sample. This was determined by increasing random sample sizes until the *ArcMap Spatial Analyst* histogram distribution closely matched the distribution of the raster layer from which they were sampled. Nielson *et al.* (2003) found that exceeding a ratio of three times the number of presence units provided minimal change in coefficients, or a reduction in standard error. A ten percent, random equal proportion of presence and available points were withheld for model validation.

In this model there was no adjustment for autocorrelation from individual overrepresentation, because no individuals were identified; measurements were made at the population level. All of the independent variables were tested for multicollinearity using Spearman's rank correlation in *SPSS 13* (SPSS Inc. 2006). Any relationship between two or more habitat covariates above $R^2 = 0.7$ were deemed too analogous for use in the same model (Suring *et al.* 2003).

3.5.1 POLYNOMIALS

Logistic regression does not require, but generally assumes linear relationships between locations of presence, and underlying habitat variables (Hosmer and Lemeshow 2000). In all reality, this association may not necessarily be linear. For that reason, polynomials can be applied to enhance the fit of the function curve. The quadratic function form of bathymetric depth was used as a surface because it is obvious that whales do not continue linearly into the abyss during foraging. No other functions were included to avoid forcing model fit by using increasingly complex polynomials which could potentially stray from simplicity.

3.5.2 MODEL ASSUMPTIONS

This model considers presence as an event that occurs solely at the ocean surface. Presence below the surface is only assumed to extend vertically to the bottom of the water column. This may not be the case, as whales may diverge laterally from this location while out of sight; a consequence of research in a relatively opaque, three dimensional ocean (Hooker and Baird 2001). However, it was assumed that over time, presence points cluster in areas where whales consistently forage.

Much effort was employed in the creation of the current speed and direction surfaces, but this did not prevent measurement error. The discrete ADCP data were found to be very spatially and temporally coarse considering the extent of the study area (Kopach 2004). In practice, an ADCP must collect data over an extended duration to resolve any enduring trends. Insufficient measurements would be directly transferred to the interpolated surfaces. IDW partially mitigated uncertainty by restraining the ADCP values to local regions around the measurement sites. The use of a 95% confidence interval may have been too strict for the stochastic nature of these variables. On the contrary, any detectable signal emerging above the noise would indicate a very strong natural relationship worthy of further examination.

All dynamic variables occur at three separate depths, and then combined with the top ranking static variables to detect any model improvement. The resulting number of possible model combinations would be too vast for individual testing at discrete depths. To contend with the number of dynamic variables, it was assumed that there are two main biological influences of ocean currents: velocity and direction. Therefore, these variables were entered into the candidate models at all depths simultaneously. The intent was to detect any obvious relationships, and then identify a particular depth where these relationships have the greatest effect. Significant variables were then reassessed after the initial modeling, because it was assumed that the large number of dynamic variables may produce spurious relationships that confound the models, overestimating goodness-of-fit through over-fit (Hosmer and Lemeshow 2000). Through the introduction of many novel dynamic variables, this portion of the analysis should be considered exploratory because it was observational rather than a manipulative experiment (Manly *et al*. 2002).

3.5.3 AKAIKES INFORMATION CRITERION (AIC)

All candidate models were ranked using AIC. This technique is regularly used for biological research, and derives candidate models that contain variables based on both statistical significance and biologically plausible contributions (Hosmer and Lemeshow 2000). The AIC process ranks the best model as the one with the lowest AIC score, a compromise between reducing deviance in model fit (log likelihood), and reducing the total number of parameters (Burnham and Anderson 2002). Models with fewer variables tend to be more robust as they seek parsimony, or overall simplicity. The explanation of any phenomenon should hold as few assumptions as possible, and eliminate those that make no difference in observable predictions (Young *et al*. 1996). Therefore, penalties are added to the AIC scores of models having more parameters than necessary (MacKenzie 2006).

$$AIC = -2\{\log_e(L_M)\} + 2n \tag{3.4}$$

Where L_M is the maximized likelihood for the fitted model, and n is the number of unknown parameters in the model that must be estimated (Manly *et al*. 2002).

3.5.4 DYNAMIC MODEL SCALE ADJUSTMENT

Objective 3 contained two separate spatio-temporal subdivisions that differ from the previous two analyses. This was necessary to differentiate whale selection during opposing tidal phases which dominate the dynamics of the study area (Kopach 2004). One model used ebb tide current surfaces, the other used flood tide current surfaces. The whale presence points were also partitioned based on their time of census corresponding to either ebb or flood tides. Not all years of census data had corresponding location acquisition times needed for the partition. There were an insufficient number of *July-only* occurrences to maintain the temporal stratification used for Objectives 1 and 2.

A new temporal stratification was applied to model ebb and flood current variables against whale locations occurring during these tides respectively. Segregation was performed using information from the Xtide Harmonic Tide Predictor (Flater 2006), based on the central tide prediction algorithm developed by Schureman (1924). Tidal predictions were calculated for Tofino BC, and were amended for the distance from the study area as Tofino + 1 hour. Presence occurrences were plotted on the tidal chart for the corresponding time and date. Kopach (2004) made ADCP measurements for ebb and flood tides no less than one hour from slack tide, the peak or trough between tides. This ensured the strongest respective currents were present during measurement. Accordingly, whale census times occurring within one hour before or after slack tide were omitted (Figure 3.11).

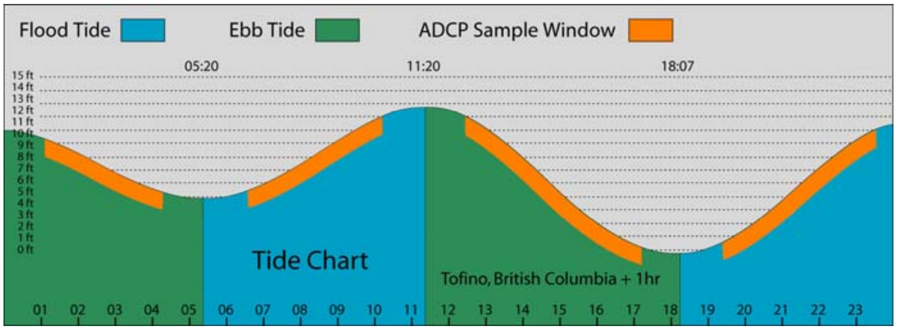

Figure 3.11 Tide chart from an unspecified date showing ebb and flood tidal phases, and ADCP sample windows occurring no less than one hour from slack tides.

Considering the reduced spatial extent of the ADCP transects, the extent of the study area had to be slightly decreased to include the dynamic surfaces used in Objective 3. As a result, the AIC results could not be compared with the models derived in Objective 2, because of the different number of presence and available points. However, Manly *et al.* (2002) states that the subset RSF should explain similar patterns of occurrence as in the original study area.

3.5.5 MODEL VALIDATION USING ROC

The RSF determines the relative likelihood of use, and can be used as a predictive surface. The accuracy of the static RSF model was tested using withheld validation data by means of a Receiver Operating Characteristic (ROC) curve in *SPSS 13*. This method iteratively cross-classifies the validation data with a dichotomous variable whose values are derived from the estimated logistic probabilities (Hosmer and Lemeshow 2000).

The area under the ROC curve provides a measure of the model's ability to discriminate between actual presence or absence points. If the model predicts too many false positives or false negatives, the area under the curve is reduced. The ROC curve plots sensitivity vs. 1- specificity across all possible probabilistic thresholds (Hosmer and Lemeshow 2000). Sensitivity is the number of correctly classified presences over the total number of true presences, and specificity is the number of predicted absences over the total number of true absences. The resulting plotted curve exhibits whether the static RSF model predicts better than a purely random model.

The final results of this investigation are presented in chronological order as each objective builds upon the previous. Figure 3.12 summarizes each process step, and the overall methodological flow.

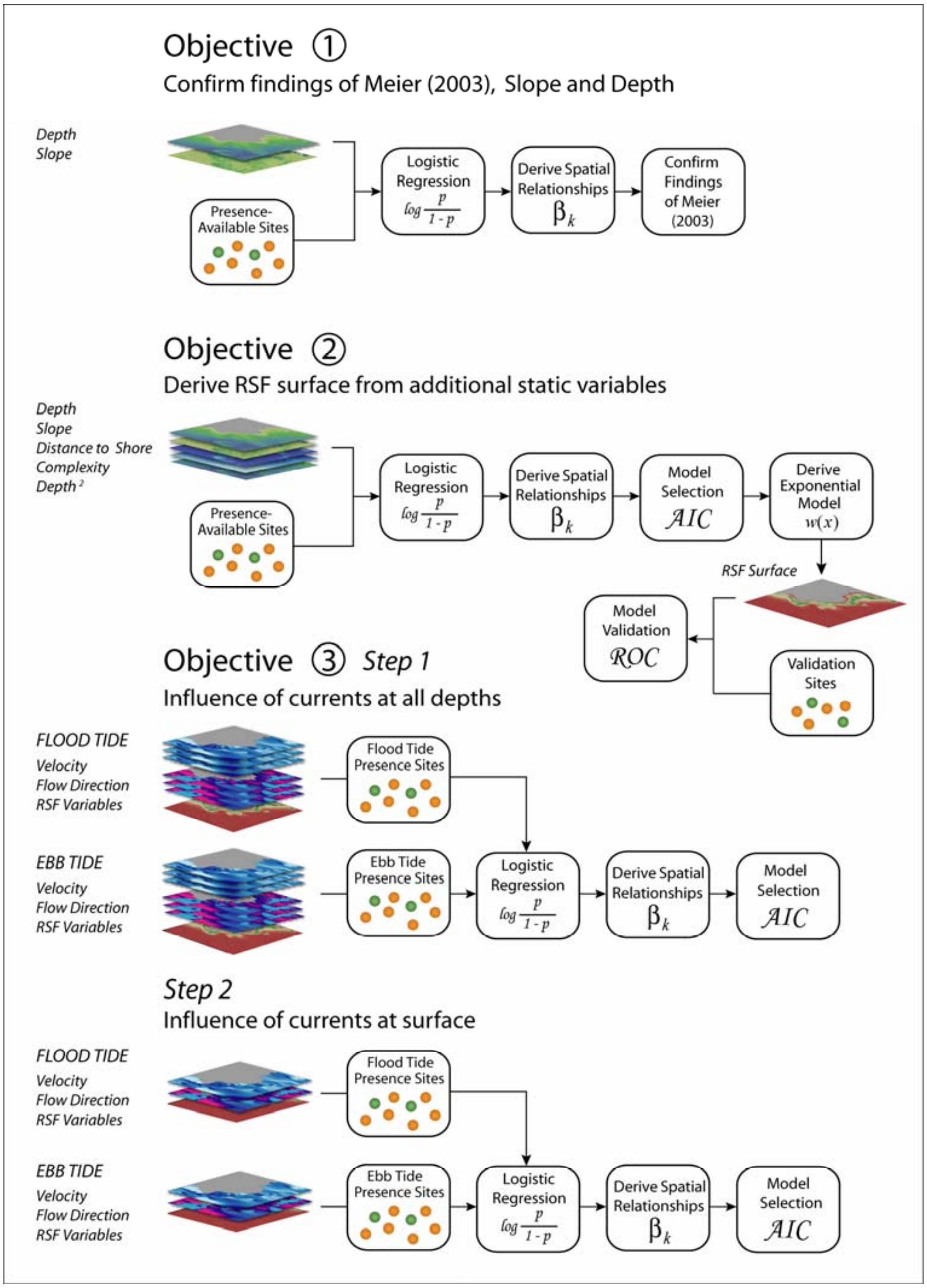

Figure 3.12 Analysis flow-chart.

4 RESULTS

4.1 CROSS-CORRELATIONS

Spearman's rank correlations for the habitat variables revealed that depth and distance from shore are too correlated and could not be used in the same model (R^2 = 0.86). Both variables were used in exclusivity during modeling. The remaining variables showed no correlation exceeding the predetermined threshold of R^2 = 0.7.

4.2 OBJECTIVE 1: CONFIRM INFLUENCE OF DEPTH AND SLOPE

Both depth and slope display a clear relationship with whale presence (Tables 4.1 & 4.2). This result substantiates the findings of Meier (2003), where there was a significant association with these same variables.

Table 4.1 Coefficients (β), standard error, Wald statistic and significance (*p*) for bathymetric depth as a predictor of gray whale habitat use.

Predictors	*β*	*SE*	*Wald*	*p*
Depth	-.079	.006	193.4	< .001*
Constant	-.162	.081	3.935	.047*

Nagelkerke R^2 = .090 * significance at <0.05

Table 4.2 Coefficients (β), standard error, Wald statistic and significance (*p*) for bathymetric slope as a predictor of gray whale habitat use.

Predictors	*β*	*SE*	*Wald*	*p*
Slope	.273	.027	105.1	< .001*
Constant	-.166	.060	753.8	< .001*

Nagelkerke R^2 = .046 * significance at <0.05

The Wald statistic reflects the relative importance of a habitat layer in the model, and the amount of explained variability in occurrence. Nagelkerke R^2 is a measure of model goodness-of-fit. Being a pseudo R^2, it implies only relative strength between models. Low values do not necessarily imply a weak explanation of variance or poor model fit (Crawley 2002).

The results show a negative coefficient value (β) for depth, expressing a tendency away from increasing depths, with more use of shallower waters. The slope coefficient is positive, implying increased presence in areas of greater benthic relief. Confidence in these relationships is strong, as highlighted by small standard errors, and being well within the 95% confidence interval.

These results parallel the findings of Meier (2003) where there was an avoidance of deeper waters, and selection for areas of increased bottom slope. Depth appears to have a more substantial influence; with a Wald score nearly double that of slope.

4.3 OBJECTIVE 2: RSF MODEL FROM ADDITIONAL STATIC VARIABLES

The AIC top ranked static model was simple and robust, containing only the quadratic function of depth and benthic complexity (Table 4.3). The rapid decrease in AIC score of the four top models is considerably less than all of the other candidates. The best performing models all included the quadratic function of depth, which improved model fit much more than its linear equivalent, depth. Benthic complexity appears in the top two models, and therefore also a strong indicator of whale presence.

Table 4.3 Candidate static RSF models of gray whale habitat use.

Rank	*Candidate Models*	*AIC*	Δ_i	w_i	*Nagelkerke*
1	**[Depth+Depth^2]^a+Cmplx^b**	**3121.0**	**0.0**	**0.67**	**0.247**
2	**[Depth+Depth^2]+Cmplx+Slope**	**3122.5**	**1.5**	**0.32**	**0.247**
3	[Depth+Depth^2]+Slope	3150.1	29.1	<0.01	0.236
4	[Depth+Depth^2]	3174.2	53.2	<0.01	0.227
5	Dstnc^c+Slope+Cmplx	3364.9	245.9	<0.01	0.155
6	Dstnc+Cmplx	3365.3	246.3	<0.01	0.154
7	Dstnc+Slope	3395.9	276.9	<0.01	0.142
8	Dstnc	3437.8	318.8	<0.01	0.124
9	Depth+Cmplx	3437.9	318.9	<0.01	0.125
10	Depth+Cmplx+Slope	3438.8	319.8	<0.01	0.125
11	Depth+Slope	3479.5	360.5	<0.01	0.108
12	Cmplx	3519.0	400.0	<0.01	0.091
13	Cmplx+Slope	3520.2	401.2	<0.01	0.091
14	Depth	3520.5	401.5	<0.01	0.090
15	Slope	3627.2	508.2	<0.01	0.046

Models are shown in decreasing order of importance based on overall AIC score, including difference in AIC (Δ_i), Akaike weights (w_i), and Nagelkerke R^2. [a]Quadratic function of depth, [b]Benthic complexity, [c]Distance from shore.

The negative coefficients for the quadratic of depth in combination with high Wald scores, and low standard error indicate a higher likelihood of whale occurrence in shallow waters (Table 4.4). Although the coefficient is negative, this trend does not imply that whales are selecting for the shallowest depths possible, it is not a linear association. Rather there is a trend for shallower water relative to all depths available in the study area. This relationship is visualized by plotting the coefficient of depth^2 against actual depth (Figure 4.1). It is apparent that the likelihood of occurrence is low near shore, but increases dramatically to a maximum use at depth ~10 m. Occurrence then subsides as depth increases further offshore into pelagic waters.

Table 4.4 Coefficients (β), standard error, Wald statistic and significance (p) for the top ranking static RSF models.

Rank	*Predictors*	β	*SE*	*Wald*	p
1st	Depth	.386	.031	155.9	< .001*
	Depth^2	-.018	.001	200.2	< .001*
	Cmplx	.014	.002	54.8	< .001*
	Constant	-2.85	.198	206.1	< .001*
2nd	Depth	.386	.031	155.5	< .001*
	Depth^2	-.018	.001	199.8	< .001*
	Cmplx	.013	.002	29.6	< .001*
	Slope	.027	.036	0.555	.456
	Constant	-2.86	.167	205.3	< .001*
3rd	Depth	.380	.031	153.2	< .001*
	Depth^2	-.018	.001	207.9	< .001*
	Slope	.146	.029	25.8	< .001*
	Constant	-2.55	.187	185.5	< .001*

* significance at <0.05

A strong positive relationship with areas of increasing benthic complexity is explicitly shown in all the top models with relatively strong Wald scores, and small standard errors. The role of benthic complexity is extraneous compared to the variability accounted for by depth^2. Slope also appears in these top models, but offers very little in terms of increasing model fit. It provides no change in Nagelkerke R^2 in the top two models, because slope is very insignificant and has no Wald score.

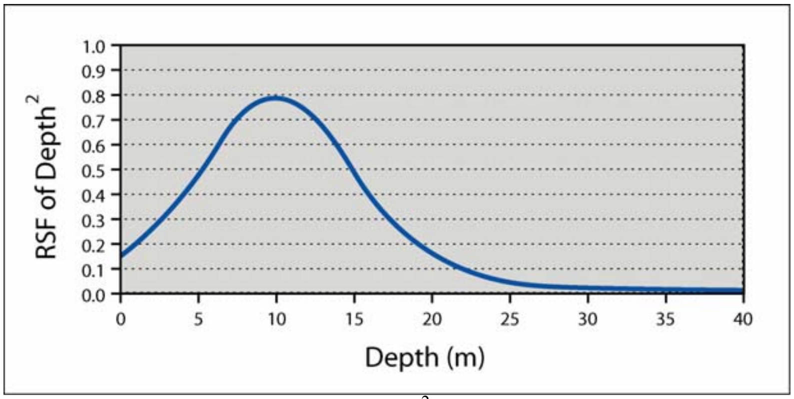

Figure 4.1 Likelihood of use at depth^2 vs. depth in meters.

The top ranked AIC model was used to create the RSF surface by entering the logistic coefficients into the exponential model using *Raster Calculator* in *ArcMap* (Eqaution 4.1).

$$w(x) = exp\{[(0.386)(depth) - (0.18)(depth^2)] + (0.014)(cmplx)\} \tag{4.1}$$

The RSF clearly exhibits increased likelihood of use in shallow, nearshore areas of high benthic complexity (Figure 4.2).

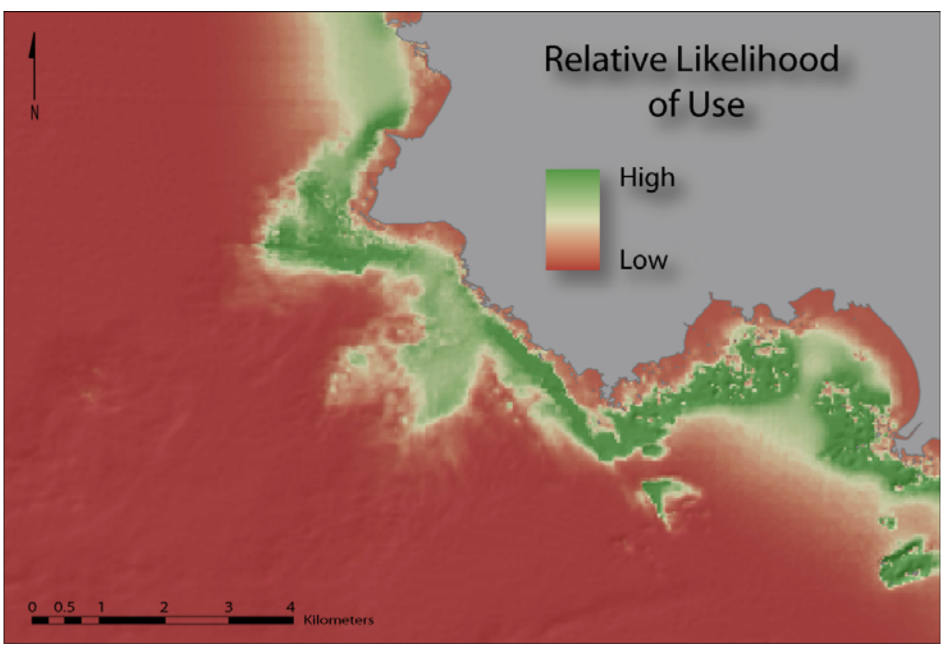

Figure 4.2 The RSF surface.

The ROC curve expresses how well the RSF model predicts the likelihood of habitat use (Figure 4.3). The area under the curve (AUC) equals 0.77 which implies that 77% of the time, a randomly selected presence point will have a higher predictive value than a randomly selected available point. An AUC of 0.77 approaches the boundary of acceptable to excellent model discrimination (Hosmer and Lemeshow 2000).

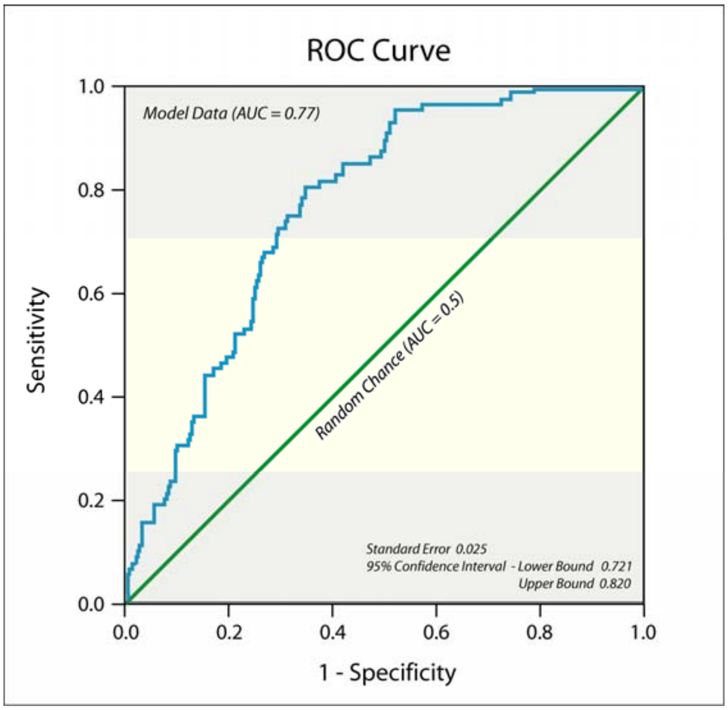

Figure 4.3 The receiver operator characteristic curve.

4.4 OBJECTIVE 3: INFLUENCE OF DYNAMIC MARINE VARIABLES

4.4.1 EBB TIDAL CURRENTS AT ALL DEPTHS

During ebb tides, north-south flowing currents occur in all of the top ranked models (Table 4.5). The negative coefficients suggest that south flowing currents have the most influence, but relatively large standard errors imply that these coefficients may be insubstantial (Table 4.6). East-west flowing currents also occur in the top model, but are surpassed by the frequency of the north-currents in the other models.

Current velocity appears to offer very little in improving model fit, only reaching the 3rd ranked model. Insignificance and high standard errors overshadow any possible effect current speed may have in this stage of the results.

Table 4.5 Candidate models for selection by whales of dynamic ebb tide current variables in combination with static RSF variables.

Rank	*Candidate Ebb Tide Models*	*AIC*	Δ_i	w_i	*Nagelkerke*
1	**[Depth+Depth^2]+Cmplx+North[a]+East[b]**	**991.7**	**0.0**	**0.57**	**0.189**
2	**[Depth+Depth^2]+Cmplx+North**	**993.3**	**1.5**	**0.26**	**0.182**
3	[Depth+Depth^2]+Cmplx+North+East +Spd[c]	995.7	4.0	0.07	0.191
4	[Depth+Depth^2]+Cmplx+North+Spd	998.7	7.0	0.02	0.182
5	[Depth+Depth^2]+Cmplx	1002.3	10.6	< .001	0.165
6	[Depth+Depth^2]+Cmplx+East	1003.9	12.2	< .001	0.169
7	[Depth+Depth^2]+Cmplx+Spd	1004.6	12.9	< .001	0.168
8	[Depth+Depth^2]+Cmplx+East+Spd	1006.4	14.7	< .001	0.176

Models are shown in decreasing order of importance based on overall AIC score, including difference in AIC (Δ_i), Akaike weights (w_i), and Nagelkerke R^2. [a]North/South trending ebb tidal currents (all depths), [b]East/West trending ebb tidal currents (all depths), [c]Current velocity (all depths).

Table 4.6 Coefficients (β), standard error, Wald statistic and significance (*p*) for the top ranking ebb tide dynamic models.

Rank	*Predictors*	*β*	*SE*	*Wald*	*p*
1st	Depth	.457	.071	41.0	< .001*
	Depth^2	-.020	.003	44.9	< .001*
	Cmplx	.021	.004	33.3	< .001*
	North Top	-.020	.011	3.27	.071
	North Mid	-.010	.014	.472	.492
	North Bot	-.010	.010	1.03	.310
	East Top	-.014	.007	4.37	.037*
	East Mid	.000	.010	.000	.995
	East Bot	.015	.010	2.38	.123
	Constant	-5.42	.473	131.1	< .001*
2nd	Depth	.466	.071	43.2	< .001*
	Depth^2	-.020	.003	48.1	< .001*
	Cmplx	.020	.004	32.4	< .001*
	North Top	-.010	.010	.995	.318
	North Mid	-.016	.014	1.17	.279
	North Bot	-.013	.010	1.90	.169
	Constant	-5.25	.456	132.8	< .001*
3rd	Depth	.452	.072	39.9	< .001*
	Depth^2	-.019	.003	41.6	< .001*
	Cmplx	.021	.004	31.9	< .001*
	North Top	-.022	.012	3.29	.070
	North Mid	-.016	.015	1.08	.298
	North Bot	-.010	.010	.880	.348
	East Top	-.017	.007	5.29	.022*
	East Mid	-.006	.012	.252	.616
	East Bot	.020	.011	3.18	.075
	Spd Top	-.020	.015	1.77	.183
	Spd Mid	.006	.016	.144	.705
	Spd Bot	-.008	.013	.402	.526
	Constant	-5.10	.550	85.7	< .001*

* significance at <0.05

4.4.2 FLOOD TIDAL CURRENTS AT ALL DEPTHS

Similar to ebb tides, north and south trending currents consistently appear in all of the top ranked flood models (Table 4.7). The sign of the coefficients dictates that they are flowing south, but any meaning is negated by high standard error, weak Wald scores, and a lack of statistical significance of the parameter (Table 4.8). Current velocity is in the top ranked model and is significant at the top and bottom of the water column.

Table 4.7 Candidate models for selection by whales of dynamic flood tide current variables in combination with static RSF variables.

Rank	*Candidate Flood Tide Models*	*AIC*	Δ_i	w_i	*Nagelkerke*
1	**[Depth+Depth^2]+Cmplx+North+East+Spd**	**890.9**	**0.0**	**.56**	**0.251**
2	[Depth+Depth^2]+Cmplx+North	894.2	3.3	.11	0.234
3	[Depth+Depth^2]+Cmplx+North+Spd	895.9	5.0	.05	0.239
4	[Depth+Depth^2]+Cmplx+North+East	896.3	5.4	.04	0.239
5	[Depth+Depth^2]+Cmplx+East+Spd	898.7	7.9	.01	0.236
6	[Depth+Depth^2]+Cmplx+Spd	905.4	14.5	< .001	0.223
7	[Depth+Depth^2]+Cmplx+East	909.0	18.1	< .001	0.219
8	[Depth+Depth^2]+Cmplx	911.8	20.9	< .001	0.209

Models are shown in decreasing order of importance based on overall AIC score, including difference in AIC (Δ_i) Akaike weights (w_i), and Nagelkerke R^2.

Table 4.8 Coefficients (β), standard error, Wald statistic and significance (*p*) for the top ranking flood tide dynamic models.

Rank	*Predictors*	*β*	*SE*	*Wald*	*p*
1st	Depth	.669	.098	46.2	< .001*
	Depth^2	-.031	.004	49.8	< .001*
	Cmplx	.016	.004	14.9	< .001*
	North Top	-.035	.011	10.3	.001*
	North Mid	.013	.018	.491	.484
	North Bot	.002	.015	.022	.882
	East Top	.000	.015	.001	.980
	East Mid	-.031	.018	2.86	.091
	East Bot	.030	.011	7.16	.007*
	Spd Top	-.052	.021	5.95	.015*
	Spd Mid	.042	.022	3.70	.054
	Spd Bot	-.035	.018	3.90	.050*
	Constant	-5.03	.637	62.3	< .001*
2nd	Depth	.699	.097	51.7	< .001*
	Depth^2	-.033	.004	58.8	< .001*
	Cmplx	.011	.004	8.75	.003*
	North Top	-.023	.007	10.3	.001*
	North Mid	-.018	.013	1.74	.187
	North Bot	.028	.012	5.52	.019*
	Constant	-5.75	.560	105.1	< .001*
3rd	Depth	.697	.098	50.6	< .001*
	Depth^2	-.033	.004	56.7	< .001*
	Cmplx	.012	.004	9.63	.002*
	North Top	-.031	.010	8.94	.003*
	North Mid	.006	.017	.108	.743
	North Bot	.012	.014	.659	.417
	Spd Top	-.031	.019	2.63	.105
	Spd Mid	.020	.019	1.15	.283
	Spd Bot	-.011	.013	.724	.395
	Constant	-5.34	.619	74.3	< .001*

* significance at <0.05

Current direction emerged in the top ranking ebb and flood models, while the influence of velocity was generally small. In both tidal phases, it is evident that these models have a better fit than the static model alone. This is likely a genuine beneficial impact of currents on occurrence, but could be a result of spurious confounding by the large number of variables presented to each model. Here, direction and velocity are a composite of three separate depths, and have been forced into the candidate models for a preliminary look at their combined role, if any.

It is apparent that modeling variables at all depths weakens the statistical significance and Wald scores, because deep areas may not reflect what is happening on the surface where whale occurrence was recorded. For this reason, there is an apparent trend of strong variables occurring at the surface of the water column. The following supplementary analysis was performed to isolate surface variables that may increase overall model fit when added individually.

4.5 EVALUATION OF SURFACE CURRENT VARIABLES

The ranking of surface models demonstrated south flowing currents having a considerable influence. The addition of north-south currents out performed the static RSF for both the ebb and flood models, (Tables 4.9 & 4.11). This is shown by an improved Naglekerke R^2, and noticeably smaller AIC score during both tidal phases, but model fit is not better than using all depths combined (Tables 4.6 & 4.9).

The AIC rankings ordered the candidate models in the same arrangement for both ebb and flood tides. No variation in order may indicate that foraging behavior of whales may be consistent across tides, and that a changing tide has no affect on habitat use. Velocity consistently provided little explanation during either tidal phase, with negligible improvement of model fit.

4.5.1 EBB TIDE SURFACE CURRENTS

Table 4.9 Ebb tide surface current candidate models.

Rank	*Ebb Tide Surface Models*	*AIC*	Δ_i	w_i	*Nagelkerke*
1	**[Depth+Depth^2]+Cmplx+Topnorth[a]**	**994.0**	**0.0**	**.58**	**.175**
2	[Depth+Depth^2]+Cmplx	1002.3	8.3	.009	.165
3	[Depth+Depth^2]+Cmplx+Topeast[b]	1002.7	8.7	.008	.166
4	[Depth+Depth^2]+Cmplx+Topspd[c]	1003.4	9.4	.005	.166

Models are shown in decreasing order of importance based on overall AIC score, including difference in AIC (Δ_i) and Akaike weights (w_i), and Nagelkerke R^2. [a]North/South trending ebb tide surface currents, [b]East/West trending ebb tide surface currents, [c]Surface current velocity.

Table 4.10 Coefficients (β), standard error, Wald statistic and significance (p) for the ebb tide surface current models.

Rank	*Predictors*	*β*	*SE*	*Wald*	*p*
1st	Depth	.468	.070	44.0	< .001*
	Depth^2	-.021	.003	49.5	< .001*
	Cmplx	.019	.003	30.6	< .001*
	Topnorth	**-.022**	**.007**	**9.9**	**.002***
	Constant	-5.15	.448	132.4	< .001*
2nd	Depth	.464	.071	43.3	< .001*
	Depth^2	-.021	.003	48.9	< .001*
	Cmplx	.019	.003	29.0	< .001*
	Constant	-4.97	.445	124.5	< .001*
3rd	Depth	.456	.071	41.3	< .001*
	Depth^2	-.020	.003	46.6	< .001*
	Cmplx	.018	.004	24.9	< .001*
	Topeast	**.010**	**.008**	**1.6**	**.214**
	Constant	-4.89	.452	117.1	< .001*
4th	Depth	.466	.071	43.6	< .001*
	Depth^2	-.021	.003	49.0	< .001*
	Cmplx	.019	.004	30.0	< .001*
	Topspd	**.011**	**.011**	**.931**	**.334**
	Constant	-5.22	.520	101.1	< .001*

* significance at <0.05

Negative north-south coefficients imply that these currents are flowing south during each tidal phase. Large Wald scores and very low standard errors during each tide suggest these currents play a significant role in the models (Tables 4.10 & 4.12). During flood tides in particular, the Wald score for south moving currents exceeds that of benthic complexity. East-west flowing currents offer nothing in explaining either the ebb or flood surface models. Exceedingly small Wald scores, large standard error, and no statistical significance confirm this.

4.5.2 FLOOD TIDE SURFACE CURRENTS

Table 4.11 Flood tide surface current candidate models.

Rank	*Flood Tide Surface Models*	*AIC*	Δ_i	w_i	*Nagelkerke*
1	**[Depth+Depth^2]+Cmplx+Topnorth**	**898.0**	**0.0**	**0.56**	**.226**
2	[Depth+Depth^2]+Cmplx	911.8	13.8	< .001	.209
3	[Depth+Depth^2]+Cmplx+Topeast	913.7	15.7	< .001	.210
4	[Depth+Depth^2]+Cmplx+Topspd	913.8	15.8	< .001	.209

Models are shown in decreasing order of importance based on overall AIC score, including difference in AIC (Δ_i) and Akaike weights (w_i), and Nagelkerke R^2

Table 4.12 Coefficients (β), standard error, Wald statistic and significance (p) for the flood tide surface current models.

Rank	*Predictors*	β	*SE*	*Wald*	*p*
1st	Depth	.699	.096	52.8	< .001*
	Depth^2	-.033	.004	59.5	< .001*
	Cmplx	.012	.004	10.0	.002*
	Topnorth	**-.028**	**.007**	**15.3**	**< .001***
	Constant	-5.84	.556	110.6	< .001*
2nd	Depth	.718	.097	54.6	< .001*
	Depth^2	-.034	.004	62.9	< .001*
	Cmplx	.011	.004	8.3	.004*
	Constant	-5.64	.558	102.0	< .001*
3rd	Depth	.717	.097	54.4	< .001*
	Depth^2	-.034	.004	62.5	< .001*
	Cmplx	.011	.004	8.4	.004*
	Topeast	**-.002**	**.007**	**.122**	**.727**
	Constant	-5.66	.562	101.4	< .001*
4th	Depth	.719	.097	54.5	< .001*
	Depth^2	-.034	.004	62.7	< .001*
	Cmplx	.011	.004	8.3	.004*
	Topspd	**-.001**	**.010**	**.009**	**.926**
	Constant	-5.62	.594	89.5	< .001*

* significance at <0.05

5 DISCUSSION

This investigation provided a case study to examine the effectiveness of logistic regression and marine GIS to derive further insight into marine habitat-use by grey whales. The results reveal a critical link with depth and benthic complexity as chief predictors of whale presence, and a possible link with south flowing currents at the top of the water column. These findings support the effectiveness of the exponential RSF model as a tool for marine ecological analysis.

The suitability of logistic regression in a marine context was assessed by confirming the findings of Meier (2003) in Objective 1. This previous investigation detected a disproportionate use of depth and benthic relief. Whales were found to occur in areas of decreasing depth and increased relief. Using logistic regression, the present analysis discovered the same tendencies; with occurrence corresponding to shallow waters and greater bathymetric slope. Although these findings are in agreement, their goodness-of-fit cannot be compared directly, as two divergent statistical techniques were used. This stage of the analysis offered confidence to proceed with the derivation of a RSF surface using additional static habitat variables.

Objective 2 introduced novel static variables of benthic topographical complexity, distance to shore, and a polynomial function of depth. The greatest model improvements were provided by the quadratic function of depth. This variable fits the likelihood of finding whales at depths they use most, approximately 10 m. The relationship between whales and depth is not linear, if this were the case they would be in waters either too shallow or too deep. Increased occurrence at 10 m may be the proximity from the highly productive intertidal zone that offers the greatest quantity of pelagic prey. This area boasts nutrient rich detritus, kelp beds, and sunlight, while providing adequate maneuverability for an adult grey whale. Meier (2003) suggested that whales may conserve more metabolic energy exploiting shallow waters than having to propel themselves to greater depths to forage.

The best model in Objective 2 was not explained by depth alone, but in combination with benthic topographical complexity. Complexity may in fact be an improved measure of slope, just like the quadratic function provided a better fit over linear depth. Slope, is present in the top models, but fails to show any significance and is very weak in association with complexity (Tables 4.3 & 4.4). The increased explanation may be due to benthic complexity providing a much better representation of seafloor variability, and prey biodiversity.

The validation of the RSF (AUC = 0.77) asserts that the exponential model provides consistent predictions of areas with an increased likelihood of whale presence. Garner (1994) and Bass (2000) state that whales select for areas of densest prey. It can be assumed that the densest prey is located in areas identified by the RSF model, with the highest concentrations tied to specific, static habitat features. This may be true, but the model still cannot discern if the whales are shifting within the areas indicated by the RSF surface. Clearly there are more patterns to identify during other periods of the season, and this model has the potential to do so. It has been previously inferred that whales express diverse selection patterns throughout a single season (Meier 2003). This shift would be undetectable using the present RSF since it was derived from temporally specific presence sightings.

Objective 3 of this investigation determined whether current velocity and direction improved the predictions of the static model. It was a biological assumption that adding currents would improve the model. The consistent appearance of south flowing currents at the surface may indicate that they are influential in gray whale habitat use. One explanation may be that during flood tides, where model fit was the highest, the prevailing direction of flow is north (Kopach 2004). Areas of south trending currents would indicate an area of turbulence as they converge with the prevailing northward flow. It has been established that turbulent areas are found to correspond with areas of increased mysid density (Kopach 2004).

During ebb tides, southward currents were also influential in the model. In this case, the orientation of Flores Island suggests that south flowing currents would be heading away, perpendicularly from the coast. These offshore flowing currents may be a result of deflection from nearshore convergence zones. There may be some type of foraging benefit in these out-flowing riptides, perhaps turbulence.

Current layers occurring at the top of the water column explained more variability in occurrence than deeper layers. This is possibly due to whale presence being recorded at the ocean surface. Therefore, it would be expected that these variables have a stronger relationship with surface presence, or at least offer more confidence in shallow waters where vertical deviation is limited. The role of currents are, in all probability, indirect and their connection to whales is perceivably convoluted. There is some indication that current dynamics do influence occurrence, but whales likely choose to move freely about the study area.

Overall, the exponential model performed well, and can be considered a suitable statistical method for marine spatial analysis. It is well adapted to this unique ecological setting because it forgives many of the assumptions and conditions required by other habitat models. Ensuing interpretations must be biologically feasible, and should be held in contrast to any shortcomings of the model.

As stated by Fielding and Bell (1997), the greatest difficulty ecological processes can create for this model is that some of the available locations may be similar, and possibly identical, to used locations. The ROC validation attempts to discriminate between true and false positives as predicted by the final RSF. The model itself is initially constructed on the belief that whales are present in areas where the independent variables are significantly different than in areas of whale absence. Considering that sampling is not exhaustive, model predictions may have been weakened by unobserved use occurring in the exact same conditions as unused available habitat.

The goodness-of-fit of the logistic regression is expressed by a pseudo R^2 that was consistently low in all of the models. Once more, it should be noted that this expresses only the relative strength of each model within this specific investigation, but is commonly low in other logistic analyses as well. This tendency of low fit in dichotomous models is a result incorporating the available data post-hoc. If availability is going to be truly representative of what is actually available in the study area, a very large sample of habitat is needed. The result is a proportionally large and unbalanced number of zeros acting as constants in the regression, weakening the fit of the model overall. If every point acts as a constant, there is no way of predicting a relationship at all. This cannot be avoided because there must be a representative number of available sites throughout the study area. If these are reduced, there will be better predictions of the *y* variable (more presence), and the fit of the model is increased, but reducing this number would no longer truly represent availability. For that reason, the models remained statistically sound at the expense of having lowered Nagelkerke values.

Lower than expected model fit may also be a result of other independent variables present in the study area that were not considered in the investigation. An effect of marine ecosystem complexity and the limited ability to collect comprehensive habitat data. If unaccounted parameters are not available to explain variability in whale presence, then model fit is further diminished. Sampling errors and GIS layer inaccuracies may have also cumulatively affected the quality of the RSF. Concerns that the sampling effort was concentrated too highly along the inner transect may have inflated the incidence of nearshore use. The spatial accuracy of the presence data is also uncertain. Whale census data have been collected for over a decade, and during this time the accuracy of GPS field units has increased significantly, reducing spatial error by up to 100 m in some cases. It is possible that early presence data may have affected the precision of the model.

The RSF model is a best attempt at representing conceptually intuitive associations between whales and their habitat. Despite ambiguities introduced by the marine environment, the model has successfully provided some guiding insight into the spatial ecology of gray whales.

5.1 FUTURE RESEARCH

Improving model results would require more thorough field measurements of available habitat, particularly for currents. The quality and resolution of the static variables are superior to the dynamic surfaces. Anchoring a buoy mounted ADCP for long-term measurements, at an increased spatial frequency, would be essential. These additional data would improve interpolation and provide more consistency across layers. Supplementary points would permit fully three dimensional interpolations, instead of using three layers at separate depths. Complete 3D interpolation was attempted initially, but the existing ADCP data were too sparse[5].

[5] See Appendix B, chapter 3.2: Marine object representation.

There was a negligible difference in the influence of ebb versus flood tidal phases in the results. An additional model should explore whale occurrence during slack tides. This is the period at the peak of flood tides and the trough of ebb tides where there is no significant current movement. This additional analysis may offer further clarification about the role of dynamics.

Considering the strong influence of benthic complexity in the RSF, it would be valuable to sample and derive a complementary benthic substrate type layer. Complex areas are indicative of swarming mysids, but gray whales also forage on benthic amphipods in muddy substrate. This soft substrate was not suitably represented in the model, and could offer further explanation of occurrence.

The quality of the presence location measurements is ample, but could be improved. More accurate instrumentation can be used during transect sampling, such as differential GPS and modified vessel-whale referencing techniques. If the same individual was observed more than once in a single census, these sightings are treated as two separate occurrences of presence. Future initiatives should include the use of photo-id of whales to derive RSF's at the individual level.

If the RSF model were applied in the study area, to inform a management plan for example, it would be necessary to perform a more rigorous validation. The ROC was suitable for this study, but a *K-fold* cross-validation would be much more comprehensive (Blum *et al.* 1999). It is an extensive iterative process that permits the use of all data for both model building, and validation (Kohavi 1995).

Breaking types of habitat use into separate behavioral categories would also be of interest. Although it is not possible to view the whales actions below the surface, refined inference can be made to discern the type of behavior at the time of observation. Behavioral observations were present only in a small portion of the data, not enough to partition a usable sample. Narrowing down certainty around behavior below the surface would require monitoring devices such as time depth recorders and digital acoustic recording tags to determine submerged spatial interactions (Woodward and Winn 2006; Malcolm *et al.* 1996).

A beneficial future analysis would occur at multiple spatial and temporal scales. A core advantage of GIS is altering layers to efficiently derive models at differing scales.

The results at the present scale are satisfactory, but only explain presence over a fixed spatio-temporal range. Multiple scales offer the most insight into species spatial organization, therefore analysis must eventually extend beyond the confines of the present study area. It is certain that habitat use changes throughout their migration, including periods where whales are eating tons of food in a single day, to eating nothing for months at time (Nerini 1984). It would be interesting to test this predictive model elsewhere in British Columbia, but more interestingly, at the contrasting reaches of their greater home range.

6 CONCLUSION

In this investigation, it has been demonstrated that there is a strong link between gray whale habitat-use and relatively shallow depths along the coast. The best performing RSF model also included increasing benthic topographical complexity as a predictor of whale occurrence. In addition, the effect of current dynamics on foraging whales was examined, where the influence of prevailing current flow direction had not been previously explored. It was found that areas with south flowing surface currents may potentially have an affect on habitat use. Current speed and subsurface current dynamics offered little improvement to the model. The adapted binary logistic regression in the form of the exponential model worked well in identifying patterns of resource use in a marine environment.

The results confirmed previous habitat-use analysis in the study area, and incorporated new and significant explanatory variables. A high quality multi-beam sonar bathymetric surface provided a robust foundation for the successful static habitat RSF surface, with predictions found to be acceptable to excellent. It is likely that ocean currents do play a role in habitat use, but only additional comprehensive field measurements will provide an improved analysis. It is commonly acknowledged that coarse input data is a regular offering to the majority of marine ecological analyses.

The marine environment is an ideal space to examine the advantages of GIS based habitat modeling. Common concepts of relatedness and object associations were frequently reassessed when dealing with disconnected objects, three dimensions, fluid dynamics and an enigmatic submerged species. Recent developments in habitat models and progress in GIS representations has made complex analysis more efficient. The combination has ultimately provided a better understanding of marine and coastal ecology.

WORKS CITED

Ardron, J. (2002) A GIS Recipe for Determining Benthic Complexity: An Indicator of Species Richness. In: *Marine Geography, GIS for Oceans and Seas*. (ed. J. Breman). ESRI Press, Redlands, California. pp. 169-175.

Baretta-Bekker, J.G., Duursma, and E.K., Kuipers, B.R. (1992) Encyclopedia of marine sciences. Springer-Verlag, Berlin. 309 pp.

Bass, J. (2000) Variations in gray whale feeding behaviour in the presence of whalewatching vessels in Clayoquot Sound, 1993-1995. Ph.D. Dissertation, Department of Geography, University of Victoria, Victoria, British Columbia.

Blum, A., Kalai, A., and Langford, J. (1999) Beating the hold-out: Bounds for the K-fold and progressive cross-validation. *Proceedings of the twelfth annual conference on computational learning theory*. University of California at Santa Cruz.

Boyce, M.S., and McDonald, L.L. (1999) Relating populations to habitats using resource selection functions. *Trends in Ecology and Evolution*. 14: 268-272

Bremen, J., Wright, D. J. and Halpin, P. N. (2002) The Inception of the ArcGIS Marine Data Model. In: *Marine Geography: GIS for the Oceans and Seas* (ed. J. Bremen) pp. 3-9. ESRI Press, Redlands, CA.

Burnham, K.P., and Anderson, D.R. (2002) Model selection and multi-modal inference: A practical information-theoretic approach. Springer-Verlag. New York. pp 488.

Calambokidis, J., Darling, J. D., Deecke, V., Gearin, P., Gosho, M., Megill, W., Tombach, C. M., Goley, D., Toropova, C. and Gisborne, B. (2002) Abundance, range and movements of a feeding aggregation of gray whales (*Eschrichtius robustus*) from California to southeastern Alaska in 1998. *Journal of Cetacean Research Management* 4: 267-276.

Crawley, M.J. (2002) Statistical computing, an introduction to data analysis using S-Plus. John Wiley & Sons Ltd. Chichester, West Sussex, England. pp 761.

Darling, J. D., Keogh, K. E. and Steeves, T. E. (1998) Gray whale (*eschrichtius robustus*) habitat utilization and prey species off Vancouver Island, B.C. *Marine Mammal Science* 14: 692-720.

Duffus, D.A. (2006) Personal Communication. *Issues of scale and selection*. Professor of Geography, Graduate Chair, Head of the University of Victoria Whale Research Laboratory. University of Victoria, British Columbia.

Duffus, D. A. (1996) The recreational use of gray whales in southern Clayoquot Sound, Canada. *Applied Geography* 16: 179-190.

Dunham, J. S. and Duffus, D. A. (2001) Foraging patterns of gray whales in central Clayoquot Sound, British Columbia, Canada. *Marine Ecology Progress Series* 223: 299-310.

Dunham, J. S. and Duffus, D. A. (2002) Diet of gray whales (*Eschrichtius robustus*) in Clayoquot Sound, British Columbia, Canada. *Marine Mammal Science* 18: 419-437

Erickson, W.P., McDonald, T.L., Skinner, R. (1998) Habitat Selection using GIS data: A Case Study. *Journal of Agricultural, Biological, and Environmental Statistics*. 3(3): 296-310.

ESRI, Environmental Systems Research Institute Inc. (2006). *ArcMap, ArcInfo, Geostatistical Analyst, ArcToolbox, and Spatial Analyst.* Redlands, California.

Fielding, A.H., and Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24(1): 38-49.

Garner, S. (1994) An exploratory investigation of gray whale (*Eschrichtius robustus*) spatial foraging strategy and the effect of vessel traffic. *Master's Thesis*, University of Victoria, Victoria, British Columbia.

Grebmeier, J.M., Overland, J.E., Moore, S.E., Farley, E.V., Carmack, E.C., Cooper, L.W., Frey, K.E., Helle, J.H., McLaughlin, and McNutt, S.L. (2006) A major ecosystem shift in the northern Bering Sea. *Science.* 311(5766): 1461-1464

Gregr, E.J., and Trites, A.W. (2001) Predictions of critical habitat for five whale species in the waters of coastal British Columbia. *Canadian Journal of Fisheries and Aquatic Science*. 58:1265-1285.

Guisan, A. and Zimmermann, N. E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.

Hooker, S.K., and Baird, R.W. (2001) Diving and ranging behaviour of odontocetes: A methodological review and critique. *Mammal Review* 31:81-105

Hosmer, D.W., and Lemeshow, S. (2000) Applied Logistic Regression 2nd Edition. John Wiley & Sons, Inc. New York. 375 pp.

Huston, M.A. (2002) Critical Issues for Improving Predictions, Introductory Essay. In: *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds. J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall and F. B. Samson). Island Press, Washington. pp 125-132.

Kerr, K.A. (2005) Nearshore Oceanography and Planktonic Prey (Family Porcellanidae) of Gray Whales, *Eschrichtius robustus*, in Clayoquot Sound, British Columbia. MSc. Thesis. Department of Geography, University of Victoria, Victoria, BC. pp 134.

Kohavi, R. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *Fourteenth international joint conference on artificial intelligence (IJCAI).* Montreal, Quebec.

Kopach, B.W. (2004) Fine-scale circulation as a component of gray whale (*Eschrichtius robustus*) habitat in Clayoquot Sound, British Columbia. M.Sc Thesis. Department of Geography, University of Victoria, Victoria, BC. 111 pp.

Levin, S.A. (1992) The problem of pattern and scale in ecology. *Ecology*, 73:1943-1967

Li, R. and Saxena, N. K. (1993) Development of an integrated marine Geographic Information System. *Marine Geodesy* 16: 293-307.

Lockwood, M. and Li, R. (1995) Marine Geographic Information systems: what sets them apart? *Marine Geodesy* 18: 157-159.

Littaye, A., Gannier, A., Laran, S. and Wilson, J. P. F. (2004) The relationship between summer aggregation of fin whales and satellite-derived environmental conditions in the northwestern Mediterranean Sea. *Remote Sensing of Environment* 90: 44-52.

Lowell, H.S., Erickson, W.P., Howlin, S., Preston, K., and Goldstein, M.I. (2003) Estimating Resource Selection Functions Using Spatially Explicit Data. *A Report*: USDA forest Service, Alaska, and Western Ecosystems Technology Inc., Wyoming. 10 pp.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., and Hines, J.E. (2006) Occupancy Estimation and Modeling, Inferring Patterns and Dynamics of Species Occurrence. Elsevier Inc. San Diego. 324 pp.

Malcolm, C. D., Duffus, D. A. and Wischniowski, S. G. (1996) Small scale behaviour of large scale subjects: diving behaviour of a gray whale (*Eschrichtius robustus*). *Western Geography* 5: 35-44.

Manly, B.F.J. (1992) Resource Selection By Animals: Statistical Design and Analysis for Field Studies. 1st Ed. Kluwer Academic Publishers. Dordrecht, Netherlands. 192 pp.

Manly, B.F.J., Lyman, L.L., Thomas, D.L., McDonald, T.L., and Erickson, W.P. (2002) Resource Selection By Animals: Statistical Design and Analysis for Field Studies. 2nd Ed. Kluwer Academic Publishers. Dordrecht, Netherlands. 221 pp.

Manly, B.F.J. (2006) Personal Communication. *Sampling probabilities and the role of the denominator in the exponential model*. Western EcoSystems Technology Inc. Cheyenne, Wyoming.

Manly, B.F.J. (2002) Estimating Resource Selection Function with line transect sampling. *Journal of Applied Mathematics and Decision Sciences*. 6(4): 213-228.

Marceau, D.J. (1999) The scale issue in social and natural sciences. *Canadain Journal of Remote Sensing*. 25(4):347-356

Marceau, D.J., Howarth, P.J., and Gratton, D.J. (1994) Remote sensing and the measurement of geographical entities in a forested environment. 1: the scale and spatial aggregation problem. *Remote Sensing of Environment*. 49(2): 93-104

Megill, W., Stelle, L. L. and Woodward, B. (2003) Surveys for gray whales, *Eschrichtius robustus*, near Cape Caution, British Columbia, summer 2003 pp. 17. Coastal Ecosystems Research Foundation.

Meier, S.K. (2003) A multi-scale analysis of habitat use by gray whales (*Eschrichtius robustus*) in Clayoquot Sound, British Columbia, 1997-99. M.Sc Thesis. Department of Geography, University of Victoria, Victoria, BC. pp 154.

Moore, S. E., Grebmeier, J. M. and Davies, J. R. (2003) gray whale distribution relative to forage habitat in the northern Bering Sea: current conditions and retrospective summary. *Canadian Journal of Zoology* 81: 734-742.

Moore, S. E., Watkins, W. A., Daher, M. A., Davies, J. R. and Dahlheim, M. E. (2002) Blue Whale habitat associations in the Northwest Pacific: analysis of remotely-sensed data using a Geographic Information System. *Oceanography* 15: 20-25.

Nerini, M. (1984) A review of gray whale feeding ecology. In: *The gray whale, Eschrichtius robustus* (eds. M. L. Jones, S. L. Swartz and S. Leatherwood) pp. 423-450. Acedemic Press Inc., Orlando, Fla.

Nielsen, S.E. (2005) Habitat ecology, conservation, and protected population viability of grizzly bears (*Ursus arctos L.*) in west-central Alberta, Canada. PhD Dissertation, University of Alberta, Canada.

Nielsen, S.E. (2006) Personal Communication. *Sampling probabilities in calculating RSF's*. Post Doctoral Fellow, Department of Biological Sciences, Univeristy of Alberta.

Nielson, R., Manly, B.F.J., McDonald, L.L. (2003) A preliminary study of the bias and variance when estimating a resource selection function with separate samples of used and available resource units. *Proceedings of the First Annual Conference on Resource Selection;* Laramie, Wyoming; January 13-15, 2003.

O'Neill, R.V., De Angelis, D.L., Waide, J.B., and Allen, T.F.H. (1986) A Hierarchical concept of ecosystems. Princeton University press, Princeton, New Jersey.

Pearce, J.L., and Boyce, M.S. (2005) Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*. A Review: British Ecological Society. 8 pp.

Patterson, H.M. (2004) Small-scale distributions and dynamics of the mysid prey of gray whales (*Eschrictius robustus*) in Clayoquot Sound, British Columbia, Canada. M.Sc thesis. Department of Geography, University of Victoria, Victoria, BC. pp 170.

RD Instruments. (1996) Acoustic Doppler Current Profilers. Principles of operation: A practical primer. RD Instruments, San Diego, CA. 36pp.

Rice, D. W. and Wolman, A. A. (1971) The life history and ecology of the gray whale (Eschrichtius robustus). *American Society of Mammologists, Special Publication no. 3*: 142.

Schureman, P. (1924) A Manual of the Harmonic Analysis and Prediction of Tides. U.S. Coast and Geodetic Survey Special Publication No. 98. Washington D.C. 416 pp

Short, C.J. (2005) A multiple trophic level approach to assess ecological connectivity and boundary function in marine protected areas: A British Columbia example. MSc. Thesis. Department of Geography, University of Victoria, Victoria, BC. 108 pp.

SPSS Inc. (2006) SPSS Predictive Analytics Software. Chigaco, Illinois.

Suring, L.H., Erickson, W.P., Howlin, S., Preston, K., and Goldstein, M.I. (2003) Estimating resource selection functions using spatially explicit data. Conference Proceedings: *First International Conference on Resource Selection.* Laramie, Wyoming, January 13-15, 2003.

Urbanski, J. A. and Szymelfenig, M. (2003) GIS-Based mapping of benthic habitats. *Estuarine, Coastal and Shelf Science* 56: 99-109.

Valavanis, V. D. (2002) *Geographic Information Systems in Oceanography and Fisheries*. Taylor & Francis, London.

Woodward, B.L., and Winn, J.P. (2006) Apparent lateralized behaviour in gray whales feeding off the central British Columbia coast. *Marine Mammal Science.* 22(1):64-73

Wright, D.J., and Scholz, A.J. (2006) Place Matters: Geospatial tools for marine science, conservation, and management in the Pacific Northwest. Oregon State University Press, Corvallis. 305 pp.

Wright, D.J. (2002) *Undersea With GIS*. (ed. D.J. Wright) ESRI Press, Redlands 253 pp.

Wright, D. J. (2000) Down to the Sea in Ships: The Emergence of Marine GIS. In: *Marine and Coastal Geographic Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 1-10. Taylor & Francis, Philadelphia.

Wright, D. J. and Goodchild, M. F. (1997) Data from the deep: implications for the GIS community. *International Journal of Geographical Information Science* 11: 523-528.

XTide Program (1998) *XTide: Harmonic Tide Clock and Tide Predictor*. David Flater <http://www.flaterco.com/xtide/index.html> accessed: March 18, 2006.

Yen, P.P.W., Sydeman, W.J., Morgan, K.H., and Whitney, F.A. (2005) Top predator distribution and abundance across the eastern Gulf of Alaska: Temporal variability and ocean habitat associations. *Deep Sea Research Part II*. 52: 799-822.

Young, P., Parkinson, S., and Lees, M. (1996) Simplicity out of complexity in environmental modeling: Occam's Razor revisited. *Journal of Applied Statistics.* 23(2&3): 165-210.

APPENDIX A

DERIVATION OF THE EXPONENTIAL MODEL

The traditional logistic regression (Equation 1.1) requires a sample of presence (1) where use was observed, and a sample of absence (0), where use was not observed. At each of these points, corresponding habitat variables are measured, and the standard statistical software provides the following logistic regression equation expressing the relationship between them:

$$log \frac{p}{1-p} = \beta_0 + \beta_1 \chi_1 + \beta_2 \chi_2 + ... + \beta_k \chi_k \tag{1.1}$$

The logistic regression equation does not provide probability directly, it only generates relationship coefficients for each covariate, but probability (*p*) is still bound by the logistic function that constrains the dichotomous, proportional data. Simple algebraic manipulation solves for the probability of use as a function of *x* (Equation 1.2) (Crawley 2002). The numerator here acts as the resource selection probability function (RSPF) (Nielson *et al*. 2003):

$$p = \frac{e^{\beta_0 + \beta_1 \chi_1 + \beta_2 \chi_2 + ... \beta_k \chi_k}}{1 + e^{\beta_0 + \beta_1 \chi_1 + \beta_2 \chi_2 + ... \beta_k \chi_k}} \tag{1.2}$$

In this form, the model holds the assumption that the study area contains samples of presence and absence in approximate proportion to their occurrence on the landscape during a single period of selection (Keating and Cherry 2004). Yet this is not the case, because the derivation of pseudo-absence occurs outside the sampling period. Therefore the RSPF (numerator) is assumed by Manly *et al*. (2002) to take an exponential function form to incorporate proportional sampling probabilities:

$$w^*(x_i) = exp(\beta_0 + \beta_1 \chi_1 + \beta_2 \chi_2 + ... + \beta_k \chi_k) \tag{1.3}$$

The RSPF equation (1.3) holds the condition that the sampling probabilities of presence and absence are known. Again, equation 1.2 assumes that the two samples are in approximate proportion to each other, however this is not the circumstance since absence was sampled separately, and after the fact. Therefore Manly *et al*. (2002) states that each presence unit must be selected with probability P_u of being independently selected during presence sampling, and that the separate sample of available units is selected with probability P_a of being independently selected form any other unit, without replacement. These sampling probabilities are linked to the constant, parameter β_0 (Erickson 1998):

$$\tau(x_i) = \frac{exp\{log_e[\frac{(1-P_a)P_u}{P_a}] + \beta_0 + \beta_1\chi_{i1} + \beta_2\chi_{i2} + ... + \beta_k\chi_{ik}\}}{1 + exp\{log_e[\frac{(1-P_a)P_u}{P_a}] + \beta_0 + \beta_1\chi_{i_1} + \beta_2\chi_{i_2} + ... + \beta_k\chi_{i_k}\}} \tag{1.4}$$

In the majority of cases, these sampling probabilities are still unknown. It is very difficult, if not impossible to identify the proportion of sampled presence units to actual species presence on the landscape. As a result, the true sampling probabilities or sample fractions that the RSPF estimation requires, are rarely provided. Given that these probabilities are included as part of the constant, but unknown, the entire constant is dropped from the equation (Manly 2002). The eliminated isolated constant is expressed:

$$\beta_0' = log_e[\frac{(1-P_a)P_u}{P_a}] + \beta_0 \tag{1.5}$$

Removal of this parameter is acceptable because there is no knowledge of the sampling probability P_u, so the constant is meaningless. It is simply an extra parameter that scales the intercept to reflect how rare or common a species is on the landscape, which is unknown in the first place (Nielsen pers. com. 2006). β_0 cannot be estimated, but the final RSF can still be calculated, which is the remaining exponential equation (Manly *et al*. 2002):

$$w(x) = exp(\beta_1\chi_1 + \beta_2\chi_2 + ... + \beta_k\chi_k) \tag{1.6}$$

The exponential model (1.6) now only effectively estimates relative likelihood of occurrence rather than calculating true probability. Firstly since the sampling probabilities were removed from the equation by eliminating the constant, and secondly because the proportion of used and available samples has been arbitrarily determined. Pearce and Boyce (2005) explain why this is so; given that presence and available locations are sampled independently, the proportion of these samples does not reflect the true prevalence of the species in the population.

They continue with an example of 100 observations of 20 presence and 80 absence locations, the probability of occurrence is 0.2 [p = 20/(20+80)]. If the absence sample is increased to 200, then the probability of occurrence changes abruptly to 0.09 [p = 20/(20+200)]. There is currently no accepted method of deriving the correct number absence units on the landscape. As long as the derivative sample is representative of what is available, then the proportion of available units to presence units is thereby set solely by the researcher (Pearce and Boyce 2005).

Agreeing that predictions of the RSF are interpreted as relative likelihoods, then there is no need to constrain the results between 1 and 0. If there are no constraints, then the denominator is removed entirely, which in practice does not really matter, if the proportion of presence units is very small. The results of the exponential model will then be approximately equal to the traditional logistic regression model, and reflect the proportional probabilities of habitat use (Manly pers. com. 2006).

WORKS CITED

Crawley, M.J. (2002) Statistical computing, an introduction to data analysis using S-Plus. John Wiley & Sons Ltd. Chichester, West Sussex, England. pp 761.

Erickson, W.P., McDonald, T.L., Skinner, R. (1998) Habitat Selection using GIS data: A Case Study. *Journal of Agricultural, Biological, and Environmental Statistics*. 3(3): 296-310.

Keating, K.A., and Cherry, S. (2004) Use and interpretation of logistic regression in habitat-selection studies. *Journal of Wildlife Management.* 68(4):774-789.

Manly, B.F.J., Lyman, L.L., Thomas, D.L., McDonald, T.L., and Erickson, W.P. (2002) Resource Selection By Animals: Statistical Design and Analysis for Field Studies. 2nd Ed. Kluwer Academic Publishers. Dordrecht, Netherlands. 221 pp.

Manly, B.F.J. (2006) Personal Communication. *Sampling probabilities and the role of the denominator in the exponential model*. Western EcoSystems Technology Inc. Cheyenne, Wyoming.

Nielsen, S.E. (2006) Personal Communication. *Sampling probabilities in calculating RSF's*. Post Doctoral Fellow, Department of Biological Sciences, Univeristy of Alberta.

Nielson, R., Manly, B.F.J., McDonald, L.L. (2003) A preliminary study of the bias and variance when estimating a resource selection function with separate samples of used and available resource units. *Proceedings of the First Annual Conference on Resource Selection;* Laramie, Wyoming; January 13-15, 2003.

Pearce, J.L., and Boyce, M.S. (2005) Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*. A Review: British Ecological Society. 8 pp.

APPENDIX B

DEVELOPMENT AND APPLICATIONS IN MARINE GIS

ABSTRACT

The application of GIS in the marine environment is growing prevalent in numerous fields of research. However, the transition of a terrestrially based geographic tool to a complex marine world is particularly challenging. Main issues in marine GIS concern adapting to the inherent complexities of the oceans, while representing disconnected objects with fuzzy boundaries within a fully dynamic three dimensional space. Existing land-based GIS systems are not entirely equipped, at this time, for georeferencing or storing this unique and specialized data. Recent advances in commercial GIS software packages are fitting the needs of many marine investigations.

TABLE OF CONTENTS

1 INTRODUCTION

The Earth is predominantly enveloped by a complex marine environment that is home to an abundance of ecological processes. This system generates key mechanisms that affect the entire planet. However, in many ways, we know more about the surfaces of our near planetary neighbours than of our coastal waters (Bartlett 2000; Wright and Goodchild 1997).

Our understanding of terrestrial habitats is extensive and has greatly increased through the use of geographic information systems (GIS). GIS is described by Goodchild (2000) as the software that is used for handling, displaying, analysing, and modeling information about the locations of phenomena and features on the Earth's surface. The conventional application of GIS has been on land, and all of the techniques, methods and software development have stemmed from this terrestrial origin.

Only until recently has GIS been implemented in understanding the marine environment (Wright 2000; Li and Saxena 1993). As a result, an entirely new branch of GIS has been developed and specialized solely for use in a coastal and marine context. This offshoot has been aptly named *Marine GIS* (Wright 2000; Lockwood and Li 1995). The transition from land to sea creates intuitive problems with regard to the intricate nature of the ocean, and its ambiguous physical boundaries (Ledoux and Gold, 2004; Wright and Goodchild, 1997; Gold and Condal, 1995; Li and Saxena, 1993). Matters grow more complex when attempting to integrate an intrinsic third dimension, or temporal dynamics to a marine habitat model. GIS software is not yet capable of displaying or evaluating a fully integrated 3D marine habitat model (Valavanis 2002; Su 2000; Li 1994). Currently however, GIS software packages do provide 3D extensions that assist in making marine models more complete (Breman *et al.* 2002).

Marine GIS does not operate far from its terrestrial roots, and still borrows from the more refined land-based techniques. Despite the complexities of adapting GIS to the ocean, the research benefits are just as beneficial underwater as they are on land.

2 MARINE GIS

2.1 INTENT OF MARINE GIS

Oceans are intricate systems that present a large number of variables and processes that can, in-turn, be combined in an infinite number of combinations. To analyze and interpret this environment, it is first necessary to partition it into manageable, conceptually based geographic categories and objects. These groupings tend to have two essential parts: location and attributes. Conveniently, these elements make oceanographic data highly suitable for spatial analysis using GIS; where locations and their attributes can be stored, georeferenced, visualized, and analyzed (Valavanis 2002; Stanbury and Starr 1999). In the literature, most marine applications that use GIS appear to have fallen under the blanket term of *Marine GIS*. This term does not merely identify these applications as non-terrestrial, but rather denotes a growing and increasingly independent branch of GIS itself.

The methods used in marine GIS result from the unique approaches and considerations that have been taken into account when faced with the complexity of the oceans. This environment is four dimensional, where variables affected by time and depth are impacted much more rapidly than variables on land (Li and Saxena 1993). This is not to say that marine GIS is autonomous from its land based counterpart, rather a different suite of questions need to be presented in marine investigations. Valvanis (2002) illustrates two such questions that seldom occur in terrestrial GIS that deal with dynamics: (*i*) how frequently should variables or processes be mapped and (*ii*) the resolution at which mapping or data gathering should be carried out. These concerns are poignant in marine GIS due to the high cost and difficulty of data collection (Wright and Goodchild 1997; Lockwood and Li 1995). Oceanographic datasets also tend to be immense, and data arrives at a much higher rate, which taxes existing commercial GIS packages (Smith 2003; Wright 2000). These questions are a result of the time, and effort encountered during an oceanographic investigation, when once completed, the final outputs are found to have grown obsolete due to the rapid change of the study area (Gold and Condal 1995).

An approach must be sought that is tailored to the marine environment, which seeks both equilibrium and efficiency when considering the frequency of field excursions. Oceanographers are further restrained by existing GIS software and analysis techniques that have been developed in static, rigid environments that bear small resemblance to the ever-changing, multi-dimensional oceans. Therefore the goal at the forefront of marine GIS is to develop this niche as its own separate entity with exclusive GIS software extensions, data models, visualization tools, and monitoring equipment (Gold *et al.* 2003; Valavanis 2002; Wright 2000; Wright and Goodchild 1997; Gold and Condal 1995; Li and Saxena 1993). From this, it will be possible to combine expert knowledge with data and spatial analysis techniques to develop a comprehensive marine model to "understand what and where things are, and how and why they are where they are" (Valavanis 2002).

2.2 TRANSITION OF TERRESTRIAL GIS TO MARINE GIS

2.2.1 MARINE GIS BEGINNINGS

The GIS archetype has evolved from its land-based origins to an aquatic space, and is constantly adapting to this new and foreign environment (Goodchild 2000; Valavanis 2002). Currently, GIS can be defined by three fundamental components: the first includes georeferenced data with descriptive attributes that represent objects in the real world. The second component is the hardware and software that is used to capture, store, update, visualize and manipulate these data. The third component is the geographically enlightened human resource that implements and operates the GIS (Valavanis 2002; Longley *et al.* 2001; Stanbury and Starr 1999; Burrough and McDonnell 1998; Lockwood and Li 1995).

There has been a recent shift in the evolution of GIS; Geographic Information Science is the suite of fundamental issues that arise in the development and application of Geographic Information Systems (Goodchild 2000). Marine GIS is inclusive to this science, a result of the many developmental and application issues spanning its progress. In fact, Geographic Information Science may play a more important role in marine GIS due to the multiplicity of problems that have been faced by implementing a terrestrially evolved science in a marine environment. This struggle explains why GIS has not yet had the success in the oceans, than it has had within more terrestrial spheres of investigation (Bartlett 1993).

The beginnings of marine GIS occurred when novel marine research using GIS prompted the appearance of abstracts, papers, and technical reports appearing in the early 1990's (Valavanis 2002; Wright 2000). Acknowledging the significant benefits of GIS, these investigations were the first attempts to apply this tool in a marine context. Goodchild (2000) stated that much can be learned from applying existing GIS software in this new environment. Applications of marine GIS share less than one third of the history of GIS as a whole (Valavanis 2002; Bartlett 2000). Much was discovered during this pioneering phase of marine GIS, due to the difficulties of application, forcing existing land-based methods to adapt and evolve. The majority of this initial development occurred within the academic sector, not through progress in commercial GIS. Many of the advances in marine GIS came from academic visionaries who were not bound by the conservatism of commercial GIS at that time. At the moment, the core of marine GIS still rests in the academic sector, but commercial vendors are now responding to the growing needs of this specialist group (Bartlett and Wright 2000).

2.2.2 EARLY APPLICATIONS AND BEYOND

Original explorations of marine GIS occurred in the coastal zone, not in the true pelagic or deepwater oceans as we know them. There was plenty of activity occurring on the coasts, and many saw GIS holding an appropriate investigative role in this region. Bartlett (2000) sees the coast as one of the most hazardous locations to live, considering that 40% of the world's population resides along it. GIS could be used to minimize the human and economic consequences of flooding, erosion and particularly ease concerns surrounding changes in the global climate (Smith 2003; Breman *et al.* 2002; Bartlett 2000). Along with environmental vulnerability, GIS plays an important economic role on the coast. The majority of the worldwide gross national product is derived from activities directly or indirectly linked with coastal zones (Zeng *et al.* 2001). The following decades have shown marine GIS to be indispensable for coastal zone management, because of its ability to handle such a diversity of geographic inputs (Smith 2003; Bartlett 2000; Lockwood and Fowler 2000; Li and Saxena 1993).

Marine GIS applications soon moved into the pelagic and deepest parts of the oceans (Wright 2000). Currently there is a multitude of research employing GIS in all facets of marine investigation: coastal and submerged vegetation mapping, coastal bathymetry mapping, wetland research, flood and natural hazard research, coastal and open ocean oceanographic process modeling, deep ocean bathymetry mapping, marine geomorphology, coastal ecological modeling, tectonics, deep environments research, habitat mapping, and coastal resource management to name a few (Smith 2003; Baxter and Shortis 2002; Valavanis 2002; Zeng *et al.* 2001; Bartlett 2000,1993; Goldfinger 2000; Hooge *et al.* 2000; McAdoo 2000). It was during this growth into deeper waters that researchers encountered problems with existing GIS technologies. Li and Saxena (1993) found that underwater research was limited due to the lack of appropriate marine specialized GIS. There has been a steady evolution towards such a specialization, but the realm of marine GIS is still restricted by the terrestrial roots of the system.

2.2.3 MARINE GIS STRUCTURAL AND SOFTWARE ISSUES

There is a good understanding of how to apply GIS to data collected from a dry-land surface. Landmasses are covered with geodetic control points that are highly accurate and only move by tectonic means. These can be used to accurately find and survey the *x,y* location of features on the surface (Longley *et al.* 2001). This is the environment that GIS was developed for: a static plane with a rigid coordinate system placed upon it. However, when GIS is applied to the oceans, few things are of a static nature that can be mapped except perhaps depth (Goodchild 2000). There is a significant lack of control points in this environment, with no buildings or street corners to measure from, usually only GPS coordinates from floating vessels (Li and Saxena 1993). The tides are constantly rising and falling, and shorelines are persistently changing profile, and there is no global vertical datum to accurately measure a '*z*' value, or depth (Bartlett and Wright 2000).

The oceans have presented new challenges that are forcing a radically different way of representing phenomena that are independent of an absolute coordinate system so characteristic of GIS (Goodchild 2000). Aquatic features have fuzzy boundaries, are dynamic, and exist in a full three dimensions. It is not completely suitable for land-based GIS to fully apply to the marine environment (Shyue and Tsai 1996). Within these aquatic applications, there is an escalation of issues concerning scale as well as accuracy, where the dynamic nature of pelagic systems and the occurrence of variability over many scales blur the linkage between processes. This spreads biotic interactions over spatial scales that exceed those prevalent in terrestrial systems (Goodchild 2000; Hyrenbach *et al.* 2000).

A variety of problems exist when the spatial data structure of a terrestrial GIS is used in a marine context. An example is that so much of the available marine information is collected as points, while the structure within a terrestrial GIS is now mostly based on vectors and rasters. The rasters that do exist represent discrete entities common on land, not the dynamic, blurred entities common to oceans. Many of the analytical operations needing to be performed on marine data are restricted by the available toolset. Problems are posed in terms of the only available, but inappropriate GIS structure, limiting the quality of the analysis products (Gold and Condal 1995). Static, terrestrial GIS is good for operations such as overlaying, buffering, reclassifying, but trying to assimilate marine data using these traditional analytical methods often proves difficult (Valavanis 2002; Stanbury and Starr 1999; Li and Saxena 1993). Common GIS packages do not have the special functions required for marine applications, like dealing with the large size of oceanographic datasets, and the speed at which they arrive. Beyond this, the ability to process spatial oceanographic data is still largely unavailable in commercial GIS (Zeng *et al.* 2001; Wright and Goodchild 1997). This commercial software been designed around cartographic metaphors that are optimized for land (Bartlett 2000). For example, many GIS packages can easily accommodate satellite images, but trying to find a package that can display underwater sidescan sonar is very challenging (Li and Saxena 1993).

The development of improved and specialized marine GIS software has been a slow process. Zeng (2001) explains that this is a direct result of marine and coastal complexity, and the difficulties creating software that can model it appropriately (ie: dynamic 3D). There is also a lack of marine data in general, that further inhibits this development process. In addition to these physical limitations, political scenarios also tend to delay the advancement of the software. There is a lack of communication between coastal experts and GIS developers. This is problematic on its own, but also results in the commercial vendors not taking risks in the investment and development of marine specialized GIS. On the whole, there have been steady developments in marine GIS software provoked by growing concerns of global change and marine resource management issues (Wright and Goodchild 1997).

Most of the specialized advancements have come from user specific needs and independent programming. Object oriented programming, UML, and Visual Basic are key additions to popular packages like *ArcGIS* that permit adjustments for unforeseeable marine applications (Bremen *et al.* 2002; Hooge *et al.* 2000). Marine investigators are often stuck with creating their own stand-alone code and extensions for individual projects. It is suggested that the most advancement in marine GIS software will come from the sharing of these extensions and adaptations amongst specialists (Hooge *et al.* 2000). Therefore, it is suggested that GIS packages should be selected based on their ease of extensibility. A commercial package should be evaluated on its capacity to incorporate existing code and the malleability of its user interface (Hatcher and Maher 1999).

2.2.4 MARINE ENTITY CLASSIFICATION AND REPRESENTATION

Within a GIS, the terrestrial nature of its functionality may impede processes for classifying or building topologic relationships for marine data. There is a definitive lack of clear boundaries between marine entities, which also tend to be topologically disconnected (Urbanski and Szymelfenig 2003; Gold and Condal 1995). Unfortunately, these entities need to be defined in some manner to proceed with analysis. Various classification methods have been adopted from complex terrestrial GIS studies, and appear to work well underwater. Conventional GIS classifiers use Boolean logic which allocates an entity to a certain class in a true or false, yes or no manner (Zeng *et al.* 2001). This approach will not work in an environment where entities are blended together along their edges. The solution lies with object oriented classifiers that permit varying degrees of allotment between classes along a continuous scale. It uses the *fuzzy set* concept of partial membership, where there is a gradual transition between classes. An example would be the shift from sand to mud on the ocean floor, where boundary areas are going to be a mix of both (Urbanski and Szymelfenig 2003). This object oriented approach also adapts to the dynamic nature of these entities, as it is more flexible for simulating changes in time (Zeng *et al.* 2001).

Baxter (2002) used artificial neural networks to classify substrate habitat types and found the result more accurate than common terrestrial methods such as traditional regression and maximum likelihood. This learning classifier works exceptionally well with blended entities, because of their ability to learn and predict patterns of a non-parametric nature. The benefits of neural networks arise through the incorporation of additional variables into the classification decisions made (ie: slope and depth).

In terrestrial investigations, most entities such as forests and pipelines are connected and tend not to move in a rapid dynamic nature. Therefore, building topology between these objects is straightforward and easily updated. This is not the case in the marine environment where objects are more mobile and usually disconnected. Again, there is a need to better represent the location and relationships between marine objects. The development of a dynamic algorithmic alternative is showing great promise within marine GIS (Ledoux and Gold 2004; Gold *et al.* 2003; Gold 2000; Gold and Condal 1995). The Voronoï tessellation specifies the spatial relationship of unconnected objects by linking them with surrounding objects using Delaunay triangulation. This web of connections is just that - a series of linear attachments that can be easily modified as required, even dramatically without rebuilding the topology. These connections automatically shift as an object moves within the GIS, where triangulations extend behind the object and shorten in front of it (Gold and Condal 1995). These dynamic adaptations are the essence of marine GIS, where clever innovations fuel the progress of marine specializations (Valavanis 2002).

2.3 MARINE DATA TYPES

One of the intrinsic setbacks of marine research is the effort required in data collection. It takes specialized equipment and a lot of preparation to do so. It is well known that the maritime environment is often harsh, and can present a fair share of climactic adversity. Additional preparation is required, since the ocean by nature is somewhat alien to us. The cost of data collection alone is often prohibitive (Baxter and Shortis 2002; Lockwood and Li 1995; Li and Saxena 1993), where a large oceanographic research vessel normally costs upwards of \$15000 - USD\$20000 per day to operate (Wright and Goodchild 1997). Financial strain is augmented by the extended period of time it takes to amass a usable amount of data (Li and Saxena 1993). This results in a general lack of data overall, where it is merely sampled and never exhaustive (Lockwood and Li 1995). Therefore, the most taxing problem is simply not knowing what is going on below the surface, introducing the pressing challenge of acquiring more reliable data. Without adequate data, even the most sophisticated analytical techniques are rendered useless (Bartlett and Wright 2000).

Some examples of common marine data types are shown in Figure 2.1 which illustrates the breadth of categories, and how they are represented in the GIS. It is extremely difficult to provide data for all points in the sample space, resulting in sparse oceanographic data on the whole. This often forces researchers to combine marine geographic data for very different objectives in a GIS, and the original intended use of these data are often overlooked or minimized (Greene *et al.* 2006; Stanbury and Starr 1999; Lockwood and Li 1995).

Technological developments are increasing the ability to collect higher quality marine data, and more of it (Bremen *et al.* 2002). Examples would be: sidescan and multibeam sonar, satellite and aerial imagery, acoustic doppler current profilers, stereo video systems, towed underwater gliders, water column salinity and temperature transponders (Baxter and Shortis 2002; Bremen *et al.* 2002). This instrumentation is providing a very realistic representation of the marine environment in all three dimensions (Figure 2.2). Time series data is introducing the needed dynamic portion of this representation. A marine GIS should be designed to store this spatiotemporal data; it is essential in understanding dynamics and seasonality (Valavanis 2002). Currently though, the quality of time series data in marine applications is weak and needs development (Bartlett and Wright 2000).

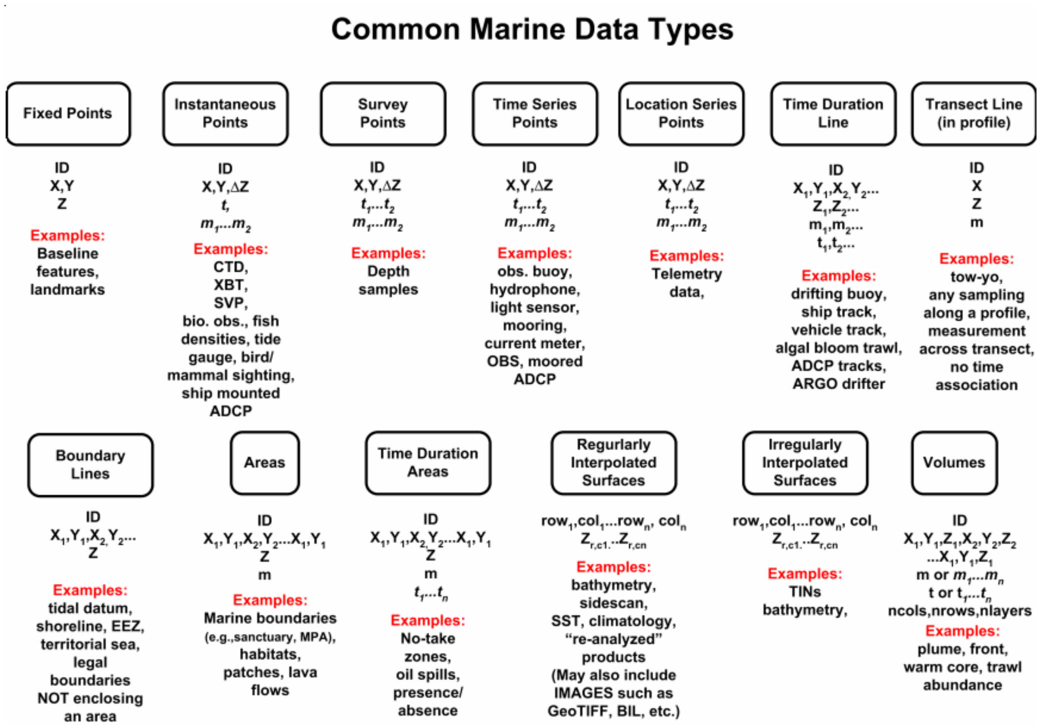

Figure 2.1 Common marine data types and examples of what they represent in a marine GIS (Bremen 2002).

Knowing the range of marine data types, it is very important to also collect quality, comprehensive metadata (Greene *et al.* 2006). Strictly defined, metadata are data about data which describe contents and define handling instructions (Longley *et al.* 2001). Often a variety of data is blended together in a GIS, and metadata keeps the user informed about the quality and particulars of the marine data at hand. It also prevents loss of data, and should be produced at the same time as data collection. In many respects it is as valuable as the data itself (Valavanis 2002; Stanbury and Starr 1999; Shyue and Tsai 1996; Lockwood and Li 1995).

Figure 2.2 A representation of the many available methods for collecting marine data (Rutgers 2005).

2.4 THE MARINE GIS DATA MODEL

Quality marine data create the foundation to perform spatial analysis. These data are intended to represent natural pelagic features within the GIS; however, the information collected is often too rich to be included in the database, it must be simplified by a data model (Li 2000). Data models lay the framework where natural entities and objects are represented in digital form (Wright and Goodchild 1997). Since any GIS database is a model of reality, it is very important to have a strong structural knowledge of the marine habitats under study (Urbanski and Szymelfenig 2003). It is subsequently vital to establish a conceptual model of the marine environment through extensive planning and designing, before implementing a GIS database (Bremen *et al.* 2002; Bartlett 2000). Considering the complexity of the marine and coastal environment, a well designed data model is crucial to the success of the entire information system (Li 2000). This data structure helps users conceive their investigative purpose, and organize databases in an intelligent manner.

The most basic spatial concept is that of location; for example, the location of a temperature buoy will give spatial meaning to that associated dataset. This relationship needs to be expressed within the database, and provide meaning to these locations. It is common geographical practice to begin any investigation by asking where, and then why. It is the purpose of the data model to best represent these relationships and then provide answers (Valavanis 2002; Duffus 1996). Again, most data models have been developed and used successfully in the terrestrial arena. These include: groundwater contamination models, climate models, hydrological models and soil loss equations. Many of the techniques involved in combining marine models with a GIS are still weakly investigated (Bartlett and Wright 2000).

Until recently, marine and coastal GIS practitioners were using the *coverage* data model; this has been practical, but coverage's have their limitations. Features are grouped together as homogeneous collections of points, lines and polygons with generic, one and two dimensional behaviour. These do not represent the dynamic complexity of marine feature behaviours (Bremen 2002). Fortunately, there has been an evolution of data models specifically tailored to better represent the marine environment. They are based on the terrestrial models that function similarly on things like vector and raster data classification, aggregation, proximity, statistical analysis, and interpolation (Valavanis 2002; Longley *et al.* 2001). These models are more capable of harnessing the power of GIS, by combining and overlaying themes, conducting spatial analysis, and performing queries among objects in two or more layers (Stanbury and Starr 1999). Since these models are attempting to simplify a complex system, they work by reducing parts of this system into themes or layers (Figure 2.3). For example, bathymetric data forms one of the basic layers in a marine GIS model (Li and Saxena 1993).

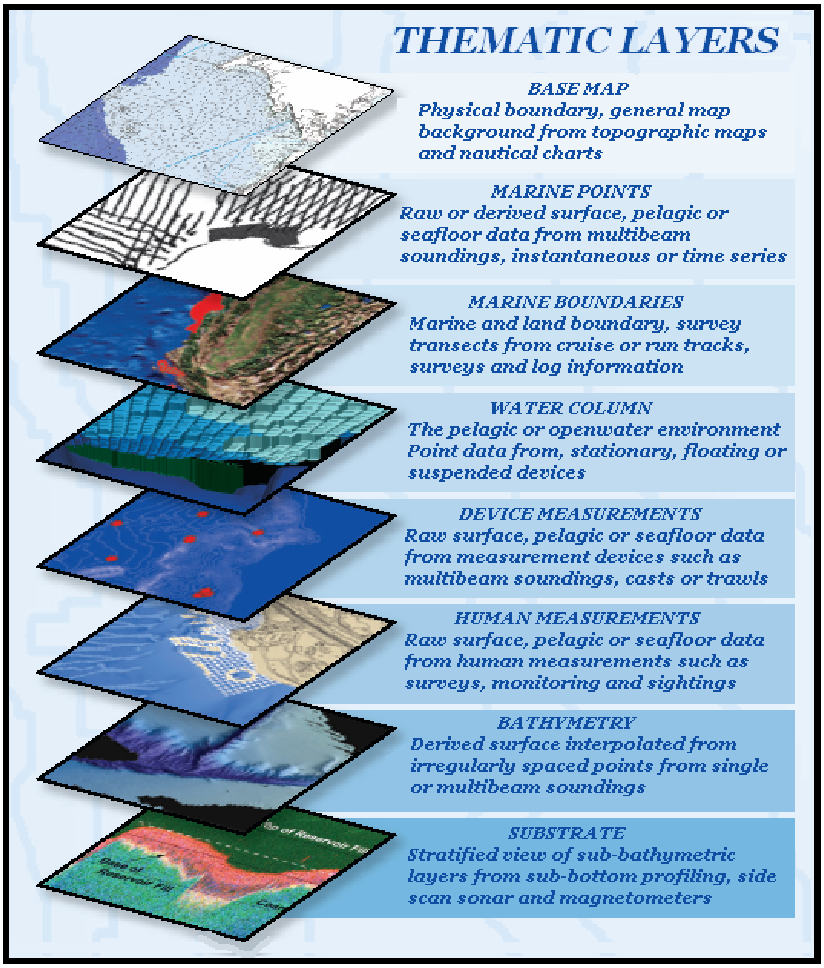

Figure 2.3 Simplified environmental thematic layers within the marine data model (ESRI 2006).

There has also been progress beyond these general abilities of the marine data model to accommodate specialized marine entities. The evolution of the model's capacity to perform various applications corresponds directly with advancements in technology. The first marine data models were housed in PC 386 computers (Li and Saxena 1993). The objective of this early analysis was to create spatial data structures for various ocean related applications and to animate some marine operations, which was quite optimistic considering the technology at hand. Many improvements have been made by Li (2000) where data models now include entities represented by surface and volumetric digital objects. This same paper describes specific models for shoreline, seafloor and time series data.

The Environmental Systems Research Institute (ESRI), the world leader in GIS software and technology, has produced the most convincing and comprehensive marine data model to date[6]. This model provides complete data structures that better integrate oceanic features. It considers how marine and coastal data can best be represented in three and four dimensions by including a volumetric model to represent multidimensionality and dynamics of these marine data and processes. It more accurately represents location and spatial extent, and has thus improved the analogous spatial analysis techniques by better characterizing the behavior of marine objects in the database (Bremen *et al.* 2002; Smith 2003). It also acknowledges historic requests for more extensibility needed to program specific marine applications (Hooge *et al.* 2000; Hatcher and Maher 1999), but the model itself provides more complexity for imputing, formatting, geoprocessing, sharing of marine data, and deals more effectively with scale dependencies (ESRI 2005). ArcGIS also includes an object oriented data model that assists in this endowment of more realistic marine behaviors. Other beneficial adaptations of the model include the shape fields in the attribute table having a *z* value for depth, which would formerly only have room for *x* and *y* in a coverage model. There are also adaptations for temporally dependent points. Tidal variance, wave activity, and even atmospheric pressure can be incorporated into the model (ESRI 2005; Smith 2003; Bremen *et al.* 2002). Marine data models should, as a minimum, incorporate the type of data, quality of data, metadata, and interpretive processes (genealogy) specific to marine data types (Greene *et al.* 2006).

The current level of marine data modeling is finally paying heed to the specialized needs of marine and coastal GIS investigators. This environment is being simplified in a more realistic manner, and retaining more real-world behaviors. However, this reduction of complexity opens the door for misrepresentation, and the introduction of error into the model.

---

[5] Accessible online at www.esri.com > support > data models > marine data model.

2.5 VARIABILITY AND UNCERTAINTY IN MARINE GIS

As GIS technology and data modeling improve, so does the intensity of analysis, and the aesthetic of the visual outputs. Amidst brilliant maps and other striking products, the problematic aspects of GIS are easily concealed or forgotten. Particular care should be devoted to checking for errors, because these outputs are capable of charming the user into a false sense of accuracy and precision, unwarranted by the data at hand (Von Meyer *et al.* 2000). After decades of GIS development there is still inadequate attention to how errors arise and propagate (Burrough and McDonnell 1998). All too often the products of a marine GIS are trusted for major decisions with minimal consideration about the impacts of error and inaccurate data (Greene *et al.* 2006). Identifying weaknesses in preliminary data, uncertainty in geographic concepts, and monitoring error propagation, all benefit the comprehension of the final product.

Advantages of the marine GIS model include fitting data within a pre-designed structure, and simplifying marine phenomena. However, the stochastic qualities of an actual ecosystem do not permit sufficient sampling to accurately reconstruct it within the model. This generates uncertainty around their existing associations in reality. Reducing complexity involves including or excluding biologically significant features; it is typically unfeasible to include them all (Horne 2002). The reduction process may become arbitrary through misconceptions of what the parts of the system are significant. The model only provides a logical best attempt at conceptualizing these phenomena, but uncertainty remains in the final model.

Uncertainty can be defined as the completeness of a digital representation, and the general measure of representation quality (Longley *et al.* 2001). Most data models assume that data attributes have been described and measured exactly, which is not the case. When uncertain data is used in quantitative analysis, the results contain these initial errors (Von Meyer *et al.* 2000; Burrough and McDonnell 1998). Ambiguity is inherent in the marine environment simply because this system is out of view and difficult to sample, presenting problems of accuracy and scale (Goodchild 2000).

There are gaps in the sampling density of most marine data, in both spatial and temporal dimensions (Von Meyer *et al.* 2000; Li and Saxena 1993). Compounding this problem, the variation in these phenomena is continuous, while the sampling points are often discrete (Lucas 2000). Also, most marine investigations are broad scale and limit analysis within finer scales, resulting in reduced map and data resolution (Shyue and Tsai 1996; Valavanis 2002). Once more, this lack of data is a consequence of adverse sampling conditions and the high cost of marine sampling in general (Baxter and Shortis 2002; Lockwood and Li 1995; Li and Saxena 1993). Since most marine data is sparse, geostatistical interpolation methods are occasionally used to fill in the gaps, creating a certain amount of distortion in the outputs (Li and Saxena 1993). Data error may also be introduced through human factors such as sampling bias, inexperience with measurement instruments, or general lack of sampling fundamentals (Burrough and McDonnell 1998). This error may be further amplified through lack of skill with GIS software and concepts. Wright and Bartlett (2000) explain that just by using GIS, does not necessarily mean it is used well. Knowledge of the amount of error can only benefit an investigation. If data collection or methodological inaccuracies propagate error, the truth in these results is questionable (Stanbury and Starr 1999). It is known that, by nature, coastal processes are fraught with uncertainties and therefore marine GIS should be able to cope with them (Zeng *et al.* 2001).

There are a variety of statistical tools that assist in quantifying error, and test model sensitivity. Examples include Monte Carlo simulations, Receiver Operating Characteristic curves, and a variety of error propagation models (Burrough and McDonnell 1998). With these methods, it is possible to better understand the effects of uncertainty, which in all actuality, is easier than finding ways of dealing with it (Longley *et al.* 2001). The best means to reduce uncertainty is to avoid it in the first place. This opportunity arises during preliminary data collection, where optimizing the sampling strategy and increasing sample density will improve the foundation for analysis (Burrough and McDonnell 1998). Also, the simultaneous collection of abundant metadata, increases the future reliability of the actual data (Bartlett 2000; Stanbury and Starr 1999).

3 THREE DIMENSIONAL GIS

3.1 LIMITATIONS OF 3D GIS IN MARINE MODELING

The traditional demands of a GIS were met years ago, where spatial information is captured, structured, manipulated, analyzed and presented. These operations have been developed for a two dimensional terrestrial plane. GIS is now being applied in a 3D environment and, not surprisingly, the existing functionality fails to operate in this setting (Zlatanova *et al*. 2002). Ideally a 3D GIS should offer the same capabilities as a 2D system, each with its own semantics, and full topology. Although many 3D GIS models have been reported, they are all dissimilar and have a multitude of both strengths and weaknesses (Rahman *et al.* 2000). Zlatanova *et al*. (2002) assert that the consensus on a 3D topological model has not yet been achieved.

3D GIS is still primitive, there is a need for sophisticated data structures to model and simulate the marine environment. Currently, these applications are either impossible with existing GIS packages, or they are highly deprived by the lack of proper 3D data structures (Li 1994; Valavanis 2002). Many marine GIS investigators desiring 3D functionality are forced to perform spatial analysis in a 2D, land-based GIS. There is an overall lack of 3D GIS software, and existing packages are unable to derive, manipulate, query and analyze 3D structures (Stoter and Zlatanova 2003; Valavanis 2002; Zeng *et al.* 2001; Shyue and Tsai 1996; Lockwood and Li 1995; Li and Saxena 1993).

In fact, true 3D GIS does not exist at this point in time. Most GIS data still consists of a location at *x,y* with a *z* value measured for depth, or height. Additionally, there is difficulty capturing an accurate *z* value because of poor georeferencing control (Raper 2000). Depth measurements contend with a continually varying tidal value, that should consider a nineteen year planetary epoch to georeference accurately (Smith 2003). This is still merely point data with an additional attribute that can only be described as 2.5 dimensional; it is not fully 3D (Baxter and Shortis 2002; Bartlett and Wright 2000). The derived surfaces appear as 3D, but their attributes only occupy a single plane, and not all points in space.

The technology does exist to efficiently collect 2.5D data, but these methods are commonly biased towards one or two dimensions. Fox and Bobbit (2000) describe a tow-yo that is suspended beneath a vessel and moved vertically through the water column collecting various data. Shyue and Tsai (1996) mention that dynamic 4D data collection is possible with an ADCP that uses sonar to capture ocean current speed and direction; it is commonly towed as well. Disproportionate dimensional bias occurs where more vertical measurements are collected by the tow-yo, and the ADCP data has limited horizontal coverage along ship tracks. For proper 3D analysis and visualization, all three dimensions must be considered equally, they are separate yet complimentary (Bartlett and Wright 2000).

Beyond this dimensional bias, there are a multitude of hindrances in the collection of 3D data and the development of 3D GIS software. Again, the cost of collecting 2.5D data is much more than that of 2D, and is difficult and time consuming to assemble. This added complexity also allows more error to be associated with the data. The structure to house this data is much more intricate, and in many ways, non-existent. There is a heavy burden put on processors, and the storage space requirement is exorbitant (Rahman *et al.* 2000). To have the necessary hardware for 3D analysis is very costly (Stoter and Zlatanova 2003; Valavanis 2002; Shyue and Tsai 1996). At the moment, despite efforts at automation, most GIS 3D modeling is done manually which is very labour intensive and time consuming (Stoter and Zlatanova 2003; Rahman *et al.* 2000).

Most GIS database management systems (DBMS) are not designed to handle 3D data, except for the addition of space for a *z* coordinate. Data is often spread between several systems for storage and visualization. This results in system inconsistency and general additional costs of time and money (Stoter and Zlatanova 2003; Zlatanova *et al.* 2002; Rahman *et al.* 2000). The visualization of 3D objects is the most developed aspect of GIS; therefore, the focus of software development should rather be on DBMS improvement (Valavanis 2002).

3.2 3D GIS SOFTWARE & MARINE OBJECT REPRESENTATION

Efforts are being made to create better 3D GIS systems, but there is currently no package that incorporates a 3D database with 3D data integration, and visualization all-in-one (Valavanis 2002). Although they have not yet been perfectly integrated, the combination of GIS data structure and 3D visualization is very powerful (Su 2000). 3D applications in most commercial GIS's are only for visualization and not other 3D functionality (Stoter and Zlatanova 2003). There have been few efforts to build analytical tools for 3D GIS (Cook *et al.* 1998). Visualizations that contend with large datasets require efficient software and hardware to accommodate them. The visualizations work to create feature layers similar to a 2D GIS; however it would be impossible to blend these layers together for accurate marine simulation and interaction (Stoter and Zlatanova 2003; Su 2000). These 3D images have been found beneficial for conceptualizing various animal habitats (Hehl-Lange 2001). Li and Saxena (1993) used 3D fly-over simulations to traverse a marine setting, which they found to be an excellent method to practice costly marine operations in a virtual space beforehand.

Advancement in GIS visualization has been rapid and sophisticated because this technology is borrowed from the superior 3D realms of Computer Assisted Drawing (CAD) and computer gaming. Virtual Reality (VR) is growing in popularity for spatial geographic visualization. One can walk through a model to examine various spatial phenomena and gain a better conceptual understanding (Rahman *et al.* 2000). Immersion into a 3D VR was used by Cook *et al.* (1998) to explore multivariate spatial data using dynamic statistics which displayed 3D scatterplots above a corresponding 3D simulation of the study area. These advancements show great promise, but in the interim 2.5D GIS visualizations require better physical characteristic properties such as texture, colour, and shading (Stoter and Zlatanova 2003).

Most traditional GIS vendors provide extensions for 3D navigation and animation, but still only in 2.5 dimensions (Zlatanova *et al.* 2002; Bartlett and Wright 2000). Three dimensional representation and integration cannot be solved by simply adding a 3D extension onto an existing two dimensional GIS package (Li 1994). There are a small number of extensions available that attempt to provide a solution for 3D representation and analysis (Zlatanova *et al.* 2002). These are popular packages that hold the largest portions of the market, they include: *3D Analyst* and *ArcScene* from ESRI's *ArcGIS*; *Imagine VirtualGIS* from ERDAS Inc.; *GeoMedia Terrain* from Intergraph Inc.; and *PAMAP GIS Topographer* from PCI Geomatics (Zlatanova *et al.* 2002; Rahman *et al.* 2000). It is possible to see that none of these are specific for marine applications, except perhaps the separate *ArcGIS* data model. The model includes a feature called *placeholders* which attempts to represent the fluidity of marine data (Bremen *et al.* 2002). These packages provide the ability to drape rasters, generate surfaces and now perform volume computations, but do not extend beyond 2.5D data (Figure 3.1).

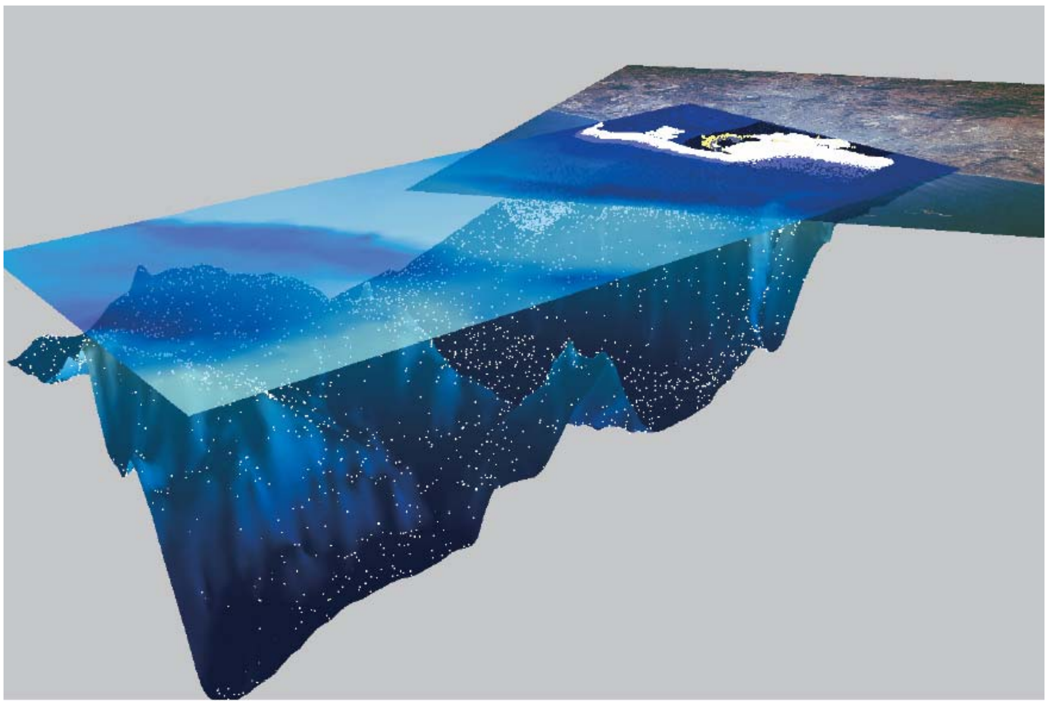

Figure 3.1 *x, y, z* coordinates interpolated to create a bathymetric surface (ESRI 2006).

Most packages still mainly work with vector data, where rasters are only incorporated to improve visualizations. They permit fly-by animations, but none allow 3D data structuring and analysis Many of these systems are lacking full 3D geometry for 3D object representation, and a fully integrated solution has yet to be offered (Zlatanova *et al.* 2002). Hooge *et al.* (2003) attempted to extend marine animal home ranges into 3D but was hampered by ArcView's 3D Analyst, which lacks volumetric capabilities. Similarly, Su (2000) explained that there are no volume 3D GIS's, only surface 3D. Meaning that a surface can be created that looks 3D, but can only holds attribute values on the surface itself. A true 3D volume would appear three dimensional, but also have values at all points in space.

Few things in the ocean can be represented by lines and polygons. There is a need for a more appropriate 3D representation (Li and Saxena 1993). The goal is to build a topological model that incorporates fully 3D marine objects. The problem with modeling marine objects is that they are represented by unconnected points that have three dimensional coordinates which tend to move over time, and the datasets have abnormal distributions (Ledoux and Gold 2004).

A variety of marine object representations have been developed over the past decade. Li (1994) described 3D representations being categorized into two groups: surface and volume[7]. Surface representations include: grids, shape models, facet models, and boundary representations. Boundary representations work by bounding objects with low dimensional elements such as lines, polygons, and polyhedrons. This is optimal for representing real-world objects because the boundaries can be measured and transferred directly into the representation. Conveniently, most rendering engines are based on boundary representations (i.e. triangles, pyramids). However, modeling rules and constraints become very complex when building topology between these objects (Stoter and Zlatanova *et al.* 2002).

Volume base representations are comprised of: 3D arrays, needle models, octrees and Constructive Solid Geometry (CSG). CSG has its origins in CAD, which produces straightforward shapes like cubes spheres and cylinders. They are simple to design, but topological relationships between them can grow very complex (Stoter and Zlatanova 2003). More contemporarily relevant representations include the Triangulated Irregular Network (TIN), which works well for 2.5D marine topological relationships, and the voxel representation. A voxel is a volume element that is similar to a 3D pixel which creates a 3D raster that is suitable for modeling continuous marine phenomena. Disadvantages of the voxel include its high resolution that demands large computer power and a large amount of storage space. Also, the surface is never regular and always appears rough. The voxel is only good for reconstruction of solid 3D objects, but does not address any spatial modeling aspects (Ledoux and Gold 2004; Stoter and Zlatanova 2003; Rahman *et al.* 2000).

---

[7] For a more detailed explanation of these representations please refer to Li (1994) and Li (2000).

3D data modeling in GIS is not only concerned with the geometric and attribute aspects of the data, but is also the topological relationships of the data. The topology of spatial data must exist so that connectedness between objects can be determined. It would be impossible to model 3D oceanographic data in a 2D GIS because the topology is not appropriately defined and the movement of objects could not be incorporated (Ledoux and Gold 2004). Bartlett in Zeng *et al*. (2001) proposes a method for an object oriented CSG data model for use in raster based 3D space using voxels. This method develops the framework for a true 3D GIS that has the potential to model marine habitats. Perhaps the best 3D marine topological model resides with the three dimensional Voronoï diagram which is an expansion of Gold and Condals (1995) 2D version. This method builds topology between unconnected 3D objects, which allows for the dynamic movement inherent to marine phenomena and perhaps the initial steps toward a fully integrated 3D GIS structure (Ledoux and Gold 2004).

A recent application of the Voronoï tesselation is the 3D Delauney tetrahedral network (TetNet), using a specialized extension from ESRI. The tool attempts to link 2.5D points through all positions in space, and interpolate a triangulated network between them. Although 2.5 dimensional, the final network works to fill the volume between the points, providing a very realistic representation of 3D space (Figure 5).

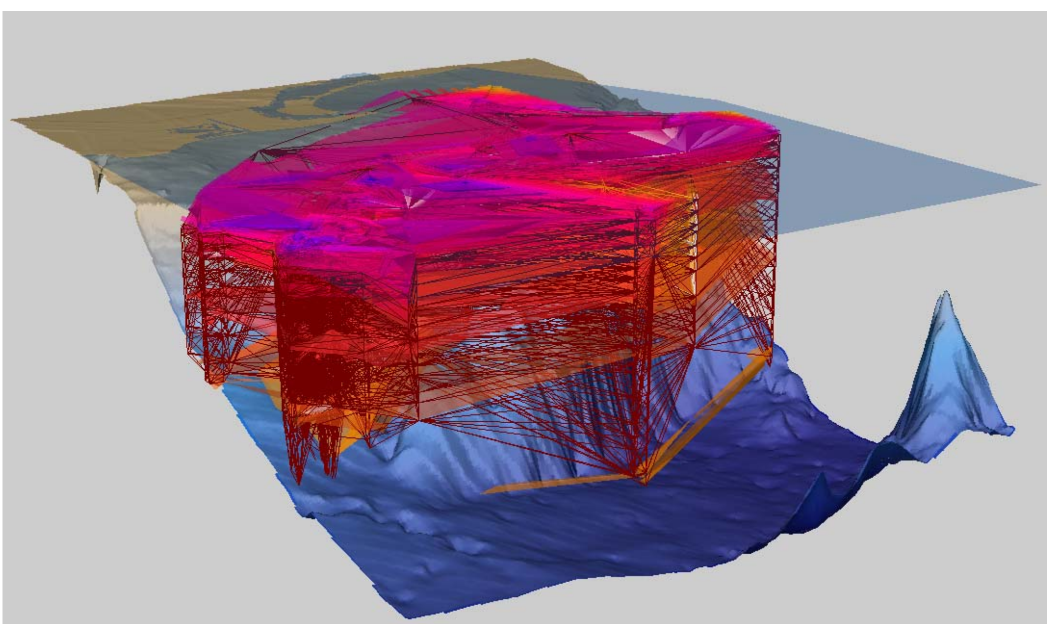

Figure 3.2 Tetrahedral Network interpolation of ocean salinity (ESRI 2006).

The next stage of 3D GIS development will focus on methods for modelling the dynamic fourth dimension (time series 3D data). The data and software requirements for 4D representations would be nearly prohibitive, demonstrating why the oceans remain a challenging environment to apply GIS (Goodchild 2000).

4 CONCLUSION

The application of GIS to the characterization and analysis of the marine environment has seen large developments in recent years. Marine GIS has progressed alongside existing GIS technologies to develop an exclusive approach to complex marine applications. The unique demands of underwater geography are obvious, and the adaptability of existing GIS software has been a limiting factor. The main obstacles include using software that has been structured upon a static, somewhat two dimensional, environment with very rigid coordinate systems. Using this configuration in a setting that is characteristic of floating disconnected objects, in a fully dynamic three dimensions has produced many challenges (Greene *et al.* 2006; Breman 2002; Wright 2000).

The requirements of marine GIS have been partially resolved through the development of specific concepts and tools for use in present GIS packages. These include ESRI's marine data model, and software extensions to support data storage and visualization. The data model assists, but does not provide comprehensive topological relationships between 3D marine objects. The major contribution to multi-dimentional GIS has come only in the form of 3D surface visualization. Currently, the Tetrahedral Network is perhaps the most representative simulation of attributes in aquatic space. To a large extent however, the majority of representations still exist only in 2.5 dimensions. Overall 3D GIS is still relatively primitive, there is a significant need for sophisticated data structures for the modeling and simulation of the marine environment (Zlatanova *et al.* 2002).

The chief focus of marine GIS still remains on establishing protocols specifically developed to accommodate unique marine data types. This includes the provision of suitable software, but also the development of a reliable method to clearly identify marine data types, quality, and interpretive processes. This will ensure quality products, but also will act to reduce error and uncertainty in an already ambiguous field of research (Greene *et al.* 2006).

WORKS CITED

Bartlett, D. J. (1993) GIS and the Coastal Zone: An Annotated Bibliography pp. 25. National Center for Geographic Information and Analysis. Report 93-9.

Bartlett, D. J. (2000) Workings on the Frontiers of Science: Applying GIS to the Coastal Zone. In: *Marine and Coastal Geographical Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 11-24. Taylor & Francis, Philadelphia.

Bartlett, D. J. and Wright, D. J. (2000) Epilogue. In: *Marine and Coastal Geographical Information Systems* (eds. D. J. Wright and D. Bartlett) pp. 309-315. Taylor & Francis, Philadelphia.

Baxter, K. and Shortis, M. (2002) Identifying Fish Habitats: the use of spatially explicit habitat modeling and prediction in marine research. In: *The 14th Annual Colloquium of the Spatial Information Research Centre* pp. 12, University of Otago, Dunedin, New Zealand.

Breman, J., Wright, D. and Halpin, P. N. (2002) The Incepetion of the ArcGIS Marine Data Model. In: *Marine Geography: GIS for the Oceans and Seas* (ed. J. Breman) pp. 3-9. ESRI, Redlands, CA.

Burrough, P. A. and McDonnell, R. A. (1998) *Principles of Geographic Information Systems: Spatial Information Systems and Geostatistics*. Oxford University Press Inc., New York.

Calambokidis, J., Darling, J. D., Deecke, V., Gearin, P., Gosho, M., Megill, W., Tombach, C. M., Goley, D., Toropova, C. and Gisborne, B. (2002) Abundance, range and movements of a feeding aggregation of gray whales (*Eschrichtius robustus*) from California to southeastern Alaska in 1998. *Journal of Cetacean Research Management* 4: 267-276.

Cook, D., Cruz-Neira, C., Kohlmeyer, B. D., Lechner, U., Lewin, N., Nelson, L., Olsen, A., Pierson, S. and Symanzik, J. (1998) Exploring environmental data in a highly immersive virtual reality environment. *Environmental Monitoring and Assessment* 51: 441-450.

Duffus, D. A. (1996) The recreational use of gray whales in southern Clayoquot Sound, Canada. *Applied Geography* 16: 179-190.

Dunham, J. S. and Duffus, D. A. (2001) Foraging patterns of gray whales in central Clayoquot Sound, British Columbia, Canada. *Marine Ecology Progress Series* 223: 299-310.

Dunham, J. S. and Duffus, D. A. (2002) Diet of gray whales (*Eschrichtius robustus*) in Clayoquot Sound, British Columbia, Canada. *Marine Mammal Science* 18: 419-437.

ESRI (2006) The Marine Data Model [Internet]. Environmental Systems Research Institute, Redlands, CA.
Available from: <http//support.esri.com/index.cmf?Fa=downloads.dataModels> [Accessed November 19, 2006]

Gold, C. (2000) An Algorithmic Approach to Marine GIS. In: *Marine and Coastal Geographic Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 37-52. Taylor & Francis, Philadelphia.

Gold, C., Chau, M., Dzieszko, M. and Goralski, R. (2003) The "Marine GIS" - Dynamic GIS in Action pp. 6. Dept. Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Hong Kong.

Gold, C. M. and Condal, A. R. (1995) A spatial data structure integrating GIS and simulation in a marine environment. *Marine Geodesy* 18: 213-228.

Goldfinger, C. (2000) Active Tectonics: Data Aquisition and Analysis with Marine GIS. In: *Marine and Coastal Geographical Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 237-254. Taylor & Francis, Philadelphia.

Goodchild, M. F. (2000) Foreword. In: *Marine and Coastal Geographical Information Systems* (eds. D. J. Wright and D. J. Bartlett). Taylor & Francis, Philadelphia.

Grebmeier, J.M., Overland, J.E., Moore, S.E., Farley, E.V., Carmack, E.C., Cooper, L.W., Frey, K.E., Helle, J.H., McLaughlin, and McNutt, S.L. (2006) A major ecosystem shift in the northern Bering Sea. *Science.* 311(5766): 1461-1464

Greene, H.G., Bizzarro, J.J., Tilden, J.E., Lopez, H.L., and Erdey, M.D. (2006) The benefits and pitfalls of Geographic Information Systems in marine benthic habitat mapping. *In*: Wright, D.J., and Scholz, A.J. Place Matters: Geospatial tools for marine science, conservation, and management in the Pacific Northwest. Oregon State University Press, Corvallis. 305 pp.

Hatcher, G. A. and Maher, N. (1999) Real-Time GIS for Marine Applications. In: *Marine and Coastal Geographical Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 137-147. Taylor & Francis, Philadelphia.

Hehl-Lange, S. (2001) Structural elements of the visual landscape and their ecological functions. *Landscape and Urban Planning* 54: 105-113.

Hooge, P. N., Eichenlaub, W. M. and Solomon, E. K. (2000) Using GIS to Analyze Animal Movements in the Marine Environment, Manuscript pp. 20, Gustavus, AK.

Horne, B. V. (2002) Approaches to Habitat Modeling: The Tensions Between Pattern and Process and Between Specificity and Generality. In: *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds. J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall and F. B. Samson) pp. 63-72. Island Press, Washington D.C.

Hyrenbach, K. D., Forney, K. A. and Dayton, P. K. (2000) Marine protected areas and ocean basin management. *Aquatic Conservation: Marine and Freshwater Ecosystems* 10: 437-458.

Ledoux, H. and Gold, C. (2004) Modelling Oceanographic Data with the three-Dimensional Voronoi Diagram. In: *Dept. Land Surveying and Geo-Informatics* pp. 6pp. Hong Kong Polytechnic University, Hong Kong.

Li, R. (1994) Data structures and application issues in 3-D Geographic Information Systems. *Geomatica* 48: 209-224.

Li, R. (2000) Data Models for Marine and Coastal Geographic Information Systems. In: *Marine and Coastal Geographic Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 25-36. Taylor & Francis, Philadelphia.

Li, R. and Saxena, N. K. (1993) Development of an integrated marine Geographic Information System. *Marine Geodesy* 16: 293-307.

Littaye, A., Gannier, A., Laran, S. and Wilson, J. P. F. (2004) The relationship between summer aggregation of fin whales and satellite-derived environmental conditions in the northwestern Mediterranean Sea. *Remote Sensing of Environment* 90: 44-52.

Lockwood, M. and Fowler, C. (2000) Significance of Coastal and Marine Data Within the Context of the United States National Spatial Data Infrastructure. In: *Marine and Coastal Geographic Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 261-278. Taylor & Francis, Philadelphia.

Lockwood, M. and Li, R. (1995) Marine Geographic Information systems: what sets them apart? *Marine Geodesy* 18: 157-159.

Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (2001) *Geographic Information Systems and Science*. John Wiley and Sons Ltd., Chicester, West Sussex, England.

Lucas, A. (2000) Representation of Variability in Marine Environmental Data. In: *Marine and Coastal Geographical Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 53-74. Taylor & Francis, Philadelphia.

Malcolm, C. D., Duffus, D. A. and Wischniowski, S. G. (1996) Small scale behaviour of large scale subjects: diving behaviour of a gray whale (*Eschrichtius robustus*). *Western Geography* 5: 35-44.

McAdoo, B. G. (2000) Mapping Submarine Slope Failures. In: *Marine and Coastal Geographic Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 189-204. Taylor & Francis, Philadelphia.

Megill, W., Stelle, L. L. and Woodward, B. (2003) Surveys for gray whales, *Eschrichtius robustus*, near Cape Caution, British Columbia, summer 2003 pp. 17. Coastal Ecosystems Research Foundation.

Nerini, M. (1984) A review of gray whale feeding ecology. In: *The gray whale, Eschrichtius robustus* (eds. M. L. Jones, S. L. Swartz and S. Leatherwood) pp. 423-450. Acedemic Press Inc., Orlando, Fla.

Rahman, A. A., Zlatanova, S. and Pilouk, M. (2000) The 3D GIS Software Development: Global Efforts From Researchers and Vendors: Manuscript. In: *ESRI* pp. 13, Redlands, CA.

Raper, J. (2000) 2.5 and 3D GIS for Coastal Geomorphology. In: *Coastal and Marine Geographic Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 129-136. Taylor & Francis, Philadelphia.

Rutgers (2005) Rutgers University Coastal Ocean Observation Lab [Internet] Available from: <http://marine.rutgers.edu/cool> [Accessed March 22, 2005]

Shyue, S.W. and Tsai, P.Y. (1996) A study on the dimensional aspect of the marine Geographic Information Systems. *Geomatica* 48: 674-679.

Smith, S. (2003) Marine GIS - Where Multidimensionality Presents Special Challenges. In: *GIS Weekly, October 6-10* pp. 12.

Stanbury, K. B. and Starr, R. M. (1999) Applications of Geographic Information Systems (GIS) to habitat assessment and marine resource management. *Oceanologica Acta* 22: 699-703.

Stoter, J. and Zlatanova, S. (2003) 3D GIS, Where Are We Standing?: Manuscript pp. 6, Delft University of Technology, Delft, The Netherlands.

Su, Y. (2000) A User-Friendly Marine GIS for Multi-Dimensional Visulaization. In: *Marine and Coastal Geographic Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 227-236. Taylor & Francis, Philadelphia.

Urbanski, J. A. and Szymelfenig, M. (2003) GIS-Based mapping of benthic habitats. *Estuarine, Coastal and Shelf Science* 56: 99-109.

Valavanis, V. D. (2002) *Geographic Information Systems in Oceanography and Fisheries*. Taylor & Francis, London.

Von Meyer, N., Foote, K. E. and Huebner, D. J. (2000) Information Quality Considerations for Coastal Data. In: *Marine and Coastal Geographic Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 295-308. Taylor & Francis, Philadelphia.

Wright, D. J. (2000) Down to the Sea in Ships: The Emergence of Marine GIS. In: *Marine and Coastal Geographic Information Systems* (eds. D. J. Wright and D. J. Bartlett) pp. 1-10. Taylor & Francis, Philadelphia.

Wright, D. J. and Goodchild, M. F. (1997) Data from the deep: implications for the GIS community. *International Journal of Geographical Information Science* 11: 523-528.

Zeng, T. Q., Zhou, Q., Cowell, P. and Huang, H. (2001) Coastal GIS: Functionality versus applications. *International Archives for Photogrammetry and Remote Sensing* 34, part 2W2: 343-356.

Zlatanova, S., Rahman, A. A. and Pilouk, M. (2002) 3D GIS - Current Status and Perspectives. In: *Proceedings of the Joint Commision on Geo-Spatial Theory, Processing and Application* pp. 6, Ottawa.

APPENDIX C

MARINE GIS RELEVANCE FOR GRAY WHALES

1 PREDICTIVE MODELING FOR WHALE CONSERVATION

Marine GIS is growing as an indispensable analysis tool for many disciplines. Marine biologists and ecologists are finding GIS useful for investigating whale habitat use (Grebmeier *et al*. 2006; Christensen 2002; Moore *et al.* 2002). A species of particular interest is the gray whale (*Eschrichtius robustus)*, by acting as a flagship species, it can be used to initiate conservation projects and develop marine protected sites (Zacharias and Roff 2001). This species also serves as a bio-indicator of environmental variability, by occurring in areas of high secondary productivity, and responding to changes in its production (Moore *et al.* 2003). Gray whales also work as a keystone species by circulating nutrients in the benthic sediment into the water column (COSEWIC 2004). Facing a multitude of post-whaling threats, GIS predictive models help to better understand the habitat selection of this enigmatic cetacean (Moore and Clarke 2002; Moses and Finn 1997).

Predictive models are used to determine where in a particular landscape a species may be found. Biologists, or habitat and species managers, may then make decisions based on the results of these models (Guisan and Zimmermann 2000). These decisions are usually made in regard to conservation issues in the context of habitat disruption and population sustainability. Compiling all of the ecological information necessary to make such assessments is a difficult task. This is especially true for natural systems, where much uncertainty and complexity is present. It is the role of a predictive model to simplify these systems so that biologists and managers can better base their decisions (Maurer 2002). Using a GIS to house the predictive model provides benefit through easy visualization of the model results. In its simplest form, the model should be able to, predict the distribution of biotic entities on the basis of ecological parameters that are believed to be the driving forces of their distribution (Guisan and Zimmermann 2000). Once these environmental factors are determined, areas of special concern can be delineated (Short 2005).

In the case of predictive modeling for whales, the dynamic marine environment poses a particular challenge due to the added complexity introduced by movement, fuzzy boundaries, and three dimensions (Breman *et al.* 2002). Nonetheless, the ecological variables that coerce whale presence and distribution can be identified with some additional effort. Using gray whale habitat for example, the environmental features such as water depth, sea floor bathymetry, substrate type, prey patchiness, sea surface temperature (SST) and salinity can all be included as layers in a GIS predictive model (Moore *et al.* 2003; Darling *et al.* 1998). GIS can quantify the extent and scale of these associations between cetaceans and their habitat, supporting prediction of whale locations beyond the study areas using GIS. This predictive knowledge is particularly useful for outlining vital habitat for whale species of dire concern. The northern right whale has the smallest population of all the worlds baleen cetaceans (Clapham *et al.* 1999). Conservation efforts are intensive to prevent the extinction of this once heavily hunted species. A GIS predictive model was developed by Moses and Finn (1997) using SST and bathymetry as predictors of right whale habitat use. The results of their analysis was applied to a recovery plan for the species. The necessity for geographical management is mounting for some whale populations due to increasing anthropogenic stress on these animals and their habitat.

2 GRAY WHALE CONSERVATION STATUS

The U.S National Marine Fisheries Service (NMFS) under direction of the Marine Mammal Protection Act (16 U.S.C 1361 *et seq*.; the MMPA) reviewed the status of the eastern pacific gray whale stock in November 1984, and changed their status from endangered to threatened. Again in 1994 the status was reviewed and changed from threatened to non-threatened (Rugh *et al.* 1999). This displays a remarkable recovery of the eastern population from severe consumptive whaling. By 1900 some whalers considered this population to be in effect, extinct (Clapham *et al.* 1999). Although this population has convalesced, the NMFS review continues to suggest that, "As this stock reaches carrying capacity, research should persist on human impacts to critical habitat"

(Rugh *et al.* 1999). The Committee on the Status of Endangered Wildlife in Canada (COSEWIC) gives this same population a status of *special concern*. It also states that as this group reaches its carrying capacity, it will be limited by available feeding habitat (COSEWIC 2004). This focus on an inadequate food supply is particularly alarming in regard to a study by Moore *et al.* (2003) which suggests the discovery of 354 dead, emaciated gray whales in 2000 is a result of reduced food availability. The western Pacific population remains critically endangered with perhaps only a couple hundred individuals remaining (Clapham *et al.* 1999). The reduction in benthic secondary productivity exists in association with a variety of other environmental pressures on the eastern population.

3 POST-WHALING THREATS

Gray whales received international protection from commercial whaling in 1938 (Clapham *et al.* 1999). However, a multiplicity of contemporary natural and anthropogenic threats may possibly be affecting the species. The predominant naturally occurring danger is an abrupt reduction of food in the Alaskan summer feeding grounds (Moore *et al.* 2003). This resulted in the death of over one third of individuals which summered along the BC coast over four years from 1998 to 2002 (COSEWIC 2004). The cause of this shortage or quality of prey is believed to be a result of changing hydrographical conditions (Grebmeier and Dutton 2000). Of the human generated pressures facing this species, offshore oil and gas exploration is the most profound, with noise being the largest concern (COSEWIC 2004; Moore and Clarke 2002; Rugh *et al.* 1999; Weller *et al.* 2002). Gray whales experience changes in surface-dive and respiration behaviour, spatial distribution, and have abandoned feeding areas in relation to increased human activities and associated noise (Weller *et al.* 2002). Dahlheim (1987) found significant changes in gray whale calling rates and call structure when exposed to human produced noise. Sound sources connected with oil and gas exploration include: aircraft overflights, boat traffic, drilling equipment, production platforms, and airgun blasts used in seismic surveys. These activities have shown acute behavioural avoidance in migrating gray whales (Moore and Clarke 2002; Würsig *et al.* 1999). Secondary impacts of the oil and gas industry include oil spills which can damage prey, their skin and baleen (Rugh *et al.* 1999). Twenty six oil affected gray whale carcasses were found after the *Exxon Valdez* spill in Alaska (Clapham *et al.* 1999).

Gray whales provide a large economic contribution in the form of whale watching. However, with substantial growth of this industry in recent years, increased public and private boat traffic is affecting their behaviour and distribution (Duffus 1996; Rugh *et al.* 1999). Bursk (1989) found that gray whales regularly changed their speed and deviated from their course in the presence of whale watching vessels. Rugh *et al*. (1999) describe similar behaviours where the whales alter their speed and respiration when followed by whale watching boats. Duffus (1996) discovered that a population of foraging gray whales were being displaced from their feeding areas as a direct result of increasing whale watching traffic. Disruption of their feeding patterns poses some ecological concern, knowing that these animals will not feed to any great extent after leaving these areas to migrate south (Malcolm *et al.* 1996; Rice and Wolman 1971). Other human generated concerns are those of fatalities as a result of impact with boats, and entanglement with fishing gear (COSEWIC 2004; Moore and Clarke 2002; Clapham *et al.* 1999; Rugh *et al.* 1999).

WORKS CITED

Bremen, J., Wright, D. J. and Halpin, P. N. (2002) The Inception of the ArcGIS Marine Data Model. In: *Marine Geography: GIS for the Oceans and Seas* (ed. J. Bremen) pp. 3-9. ESRI, Redlands, CA.

Bursk, M. (1989) Response of whales to whale watching in southern California. p.11. *In*: Proceedings of the Workshop to Review and Evaluate Whale Watching Programs and Management Needs. November 1988, Monterey California. 53 pp.

Christensen, B. (2002) Marine Mammal and Human Patterns of Use. In: *Marine Geography: GIS for Oceans and Seas* (ed. J. Bremen) pp. 161-167. ESRI, Redlands, CA.

Clapham P. J., Young S. B. and Robert L. Brownell J. (1999) Baleen whales: conservation issues and the status of the most endangered populations. *Mammal Review* 29: 35-60.

COSEWIC (2004) COSEWIC assessment and update status report on the gray whale (Eastern North Pacific population) Eschrichtius robustus in Canada. Commitee on the Status of Endangered Wildlife in Canada., Ottawa.

Dalheim, M.E. (1987) Bio-acoustics of the gray whale (*Eschrichtius robustus*) Ph.D Thesis, University of British Columbia, Vancouver, BC 315 pp.

Darling, J. D., Keogh, K. E. and Steeves, T. E. (1998) Gray whale (*eschrichtius robustus*) habitat utilization and prey species off Vancouver Island, B.C. *Marine Mammal Science* 14: 692-720.

Duffus, D. A. (1996) The recreational use of gray whales in southern Clayoquot Sound, Canada. *Applied Geography* 16: 179-190.

Grebmeier, J. M. and Dutton, K. H. (2000) *Benthic processes in the northern Bering/Chukchi seas: status and global change. Impacts of changes in sea ice and other environmental parameters in the Arctic. Report to the Marine Mammal Commision,* Bethesda, Md.

Grebmeier, J.M., Overland, J.E., Moore, S.E., Farley, E.V., Carmack, E.C., Cooper, L.W., Frey, K.E., Helle, J.H., McLaughlin, and McNutt, S.L. (2006) A major ecosystem shift in the northern Bering Sea. *Science.* 311(5766): 1461-1464

Guisan, A. and Zimmermann, N. E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.

Malcolm, C. D., Duffus, D. A. and Wischniowski, S. G. (1996) Small scale behaviour of large scale subjects: diving behaviour of a gray whale (*Eschrichtius robustus*). *Western Geography* 5: 35-44.

Maurer, B. A. (2002) Predicting Distribution and Abundance: Thinking Within and Between Scales. In: *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds. J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall and F. B. Samson) pp. 125-132. Island Press, Washington.

Moore, S. E., Grebmeier, J. M. and Davies, J. R. (2003) gray whale distribution relative to forage habitat in the northern Bering Sea: current conditions and retrospective summary. *Canadian Journal of Zoology* 81: 734-742.

Moore, S. E. and Clarke, J. T. (2002) Potential impacts of offshore human activities on gray whales (*Eschrichtius robustus*). *Journal of Cetacean Research Management* 4: 19-25.

Moore, S. E., Watkins, W. A., Daher, M. A., Davies, J. R. and Dahlheim, M. E. (2002) Blue Whale habitat associations in the Northwest Pacific: analysis of remotely-sensed data using a Geographic Information System. *Oceanography* 15: 20-25.

Moses, E. and Finn, J. T. (1997) Using Geographic Information Systems to predict North Atlantic Right Whale (*Eubalaena glacialis*) habitat. *Journal of Northwest Atlantic Fisheries Science* 22: 37- 46.

Rice, D. W. and Wolman, A. A. (1971) The life history and ecology of the gray whale (Eschrichtius robustus). *American Society of Mammologists, Special Publication no. 3*: 142.

Rugh, D. J., Moto, M. M., Moore, S. E. and DeMaster, D. P. (1999) *Status Review of the Eastern North Pacific Stock of Gray Whales*. U.S. Dep. Commer., NOAA Technical Memorandum NMFS-AFSC-103.

Short, C.J. (2005) A multiple trophic level approach to assess ecological connectivity and boundary function in marine protected areas: A British Columbia example. MSc. Thesis. Department of Geography, University of Victoria, Victoria, BC. 108 pp.

Weller, D. W., Ivashchenko, Y. V., Tsidulko, G. A., Burdin, A. M. and Brownell Jr., R.L. (2002) *Influence of seismic surveys on western gray whales off Sakhalin Island, Russia in 2001*. A Report: Kamchatka Institute of Ecology and Nature Management, Russian Academy of Sciences, Kamchatka, Russia.

Würsig, B., Weller, D. W., Burdin, A. M., Blokhin, S. A., Reeve, S. H., Bradford, A. L. and Brownell Jr., R. L. (1999) Gray whales summering off Sakhalin Island, Far East Russia: July-October 1997. In: *Unpublished contract report submitted by Texas A&M University and the Kamchatka Institute of Ecology and Nature management, February 1999* pp. 101.

Zacharias, M. A. and Roff, J. C. (2001) Use of focal species in marine conservation and mangement: a review and critique. *Aquatic Conservation: Marine and Freshwater Ecosystems* 11: 59-76.