Spring 2009

A Predictive Habitat Model for Endangered White Abalone Restoration Planning in Southern California

Shinobu Okano
California State University, Monterey Bay

Follow this and additional works at: https://digitalcommons.csumb.edu/sns_theses

Recommended Citation
https://digitalcommons.csumb.edu/sns_theses/22

This Master's Thesis (Open Access) is brought to you for free and open access by the School of Natural Sciences at Digital Commons @ CSUMB. It has been accepted for inclusion in SNS Master’s Theses by an authorized administrator of Digital Commons @ CSUMB. For more information, please contact digitalcommons@csumb.edu.
A PREDICTIVE HABITAT MODEL FOR ENDANGERED WHITE
ABALONE RESTORATION PLANNING IN SOUTHERN CALIFORNIA

A Thesis
Presented to the
Faculty of the
Division of Science and Environmental Policy
California State University Monterey Bay

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
in
Coastal and Watershed Science and Policy

by
Shinobu Okano
Spring 2009
The Undersigned Faculty Committee Approves the

Thesis of Shinobu Okano

In Partial Fulfillment of the

Requirements for the Degree

Master of Science in

Coastal and Watershed Science and Policy

A PREDICTIVE HABITAT MODEL FOR ENDANGERED WHITE

ABALONE RESTORATION PLANNING IN SOUTHERN CALIFORNIA

Rikk Kvitek, Chair
Division of Science and Environmental Policy

Corey Garza
Division of Science and Environmental Policy

John Butler
Southwest Fisheries Science Center, La Jolla,

Marsha Moroh, Dean
College of Science, Media Arts, and Technology

Spring 2009
ABSTRACT

A Predictive Habitat Model for Endangered White Abalone Restoration Planning in Southern California
by Shinobu Okano
Master of Science in Coastal and Watershed Science and Policy California State University Monterey Bay, 2009

A rapid, reliable and cost effective means for identifying species-habitat relationships is urgently needed to support management planning to preserve and restore for the federally listed endangered white abalone (*Haliotis sorenseni*). Despite an ongoing recovery effort for this depleted species, little is known about the distribution of the white abalone. The aim of this study was to develop a predictive white abalone habitat model from high-resolution multibeam bathymetry data by analyzing relationships between occurrence patterns and geomorphology of the seafloor at Tanner Bank in California, where the presence white abalone has been well documented using ROV video transect surveys. We hypothesized that there are predictable relationships between the occurrence of white abalone and measurable seafloor characteristics including depth, slope, rugosity, Topographic Positioning Index (TPI), and substrate types that can be derived from bathymetric digital elevation models (DEM). Analyses were based on a Generalized Linear Model (GLM) and Ecological Niche Factor Analysis (ENFA). ENFA was used to generate pseudo-absences since reliable absences were not available in the dataset. The GLM with ENFA-weighted pseudo-absence was used to derive a predictive habitat map in a geographic information system (GIS). Evaluation by a Receiver Operating Characteristic (ROC) curve indicated a high accuracy of model performance. The initial results from the application of the Tanner Bank derived model to bathymetry data from Carrington Point, Santa Rosa Island, Ca (a different site where white abalone were once abundant but are now absent) supports the broad utility of this model as a tool for identifying potential outplanting sites for white abalone recovery efforts. This modeling approach also has potential utility in the conservation and management planning for the heavily depleted green, pink and pinto abalone, as well as other scarce benthic species.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>SECTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER 1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. METHODS</td>
<td>6</td>
</tr>
<tr>
<td>2.1. Study site</td>
<td>6</td>
</tr>
<tr>
<td>2.2. Data description</td>
<td>6</td>
</tr>
<tr>
<td>2.2.1. Response variable</td>
<td>9</td>
</tr>
<tr>
<td>2.3. Modeling approach</td>
<td>11</td>
</tr>
<tr>
<td>2.3.1. Generalized Linear Model (GLM)</td>
<td>11</td>
</tr>
<tr>
<td>2.3.2. Variable selection</td>
<td>11</td>
</tr>
<tr>
<td>2.3.3. Polynomial function</td>
<td>12</td>
</tr>
<tr>
<td>2.3.4. Multivariate analysis</td>
<td>12</td>
</tr>
<tr>
<td>2.3.5. Spatial Autocorrelation</td>
<td>13</td>
</tr>
<tr>
<td>2.3.6. Model evaluation</td>
<td>14</td>
</tr>
<tr>
<td>3. RESULTS</td>
<td>16</td>
</tr>
<tr>
<td>3.1. Variable selection</td>
<td>16</td>
</tr>
<tr>
<td>3.2. Polynomial test</td>
<td>18</td>
</tr>
<tr>
<td>3.3. Multivariate analysis</td>
<td>20</td>
</tr>
<tr>
<td>3.4. Spatial Autocorrelation</td>
<td>20</td>
</tr>
<tr>
<td>3.5. Model evaluation</td>
<td>22</td>
</tr>
<tr>
<td>3.6. Mapping the habitat model</td>
<td>24</td>
</tr>
<tr>
<td>4. DISCUSSION</td>
<td>27</td>
</tr>
<tr>
<td>Interpretation of the habitat model</td>
<td>27</td>
</tr>
<tr>
<td>Model limitation</td>
<td>27</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>A list of predictor variable descriptions</td>
<td>8</td>
</tr>
<tr>
<td>Table 2</td>
<td>Spearman rank correlation matrix table</td>
<td>17</td>
</tr>
<tr>
<td>Table 3</td>
<td>Variable descriptions and comparison between presence points and random points</td>
<td>18</td>
</tr>
<tr>
<td>Table 4</td>
<td>Results of polynomial analysis</td>
<td>19</td>
</tr>
<tr>
<td>Table 4</td>
<td>AIC and accuracy from AUC for the models</td>
<td>20</td>
</tr>
<tr>
<td>Table 5</td>
<td>Summaries of GLM and ALR</td>
<td>23</td>
</tr>
<tr>
<td>Table 6</td>
<td>Contingency tables of GLM and ALR</td>
<td>24</td>
</tr>
<tr>
<td>Table 7</td>
<td>Summery of accuracy evaluation for GLM and ALR</td>
<td>24</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. Schematic diagram of the research strategy........................................5
Figure 2. Multibeam bathymetry image in shaded relief of Tanner Bank .............7
Figure 4. The density histogram of data in relation to each continuous environmental variable ......................................................................................16
Figure 4. Relationships between probability of white abalone occurrence and each selected predictor variables .................................................................19
Figure 5. Correlograms of spatial autocorrelation for four variables used in the final model ........................................................................................................21
Figure 6. Correlograms of residual spatial autocorrelation for the models ..........22
Figure 7. ROC of the models ..............................................................................23
Figure 5. Predicted white abalone habitat at Tanner Bank created from multivariate GLM ........................................................................................................25
Figure 6. Map of northern Channel Islands, California .....................................25
Figure 7. The GLM model was extrapolated to Carrington Point, Santa Rosa Island ..............................................................................................................26
ACKNOWLEDGEMENTS

I want to thank the Department of Science and Environmental Policy for giving me permission to commence this thesis in the first instance. I have furthermore to thank my advisor Dr. Rikk Kvitek for giving me his full support in guiding me with suggestions and encouragement to go ahead in all the time of research for and writing of this thesis. I am indebted to Dr. Corey Garza of the help and advice. I am also grateful to Dr. John Butler donating the valuable data and giving me helpful information. I would like to thank Melissa Nueman in National Oceanic Atmospheric Administration for her interest in this project and funding. I am obliged to Earl H. Myers and Ethel M. Myers Grant for Oceanographic and Marine Biology for also funding this project. I would like to thank Dr. Marc LosHuetos for his support during proposal writing. Thanks for Dr. Fred Watson for his help with R and model building. My friends in Seafloor mapping lab had supported me in various ways and cheering me up. I want to thank them for all their help, support and interest. I would like to thank most especially to my family for their great help in difficult times.
1. INTRODUCTION

The white abalone (*Haliotis sorenseni*) is the most heavily depleted among the seven abalone species found in California (Hobday et al. 2001) and has been listed as an endangered species under the Endangered Species Act since 2001. As a consequence of the listing, NOAA released its White Abalone Recovery Plan in 2008 which outlines recommended actions required for the species’ recovery. The primary goal of the recovery plan is to establish self-sustaining populations throughout its historic range. As rare and endangered species recovery plans require knowledge of required environmental factors for the species (Carroll et al, 2003), previous white abalone studies focused on identifying habitat parameters that white abalone need to survive (Hobday and Tegner 2000; Hobday et al. 2001; Lafferty et al. 2004; Butler et al. 2006).

The white abalone range extends from Point Conception, California to Baja California, Mexico, in depths between 30 and 65 meters (Davis et al., 1998; Hobday et al., 2001; NMFS, 2008). Young abalone seek cover in rock crevices, under rocks, and under adult sea urchin spines (Tegner 1989, NMFS 2006). The occurrence of adults is higher in open low relief areas amongst relatively large rocks, usually near the rock-sand interface and is associated with brown alga of *Laminaria forlowii* but not associated with other algae (*Pelagophycus porra* or *Eisenia arborea*) (Hobday and Tegner 2000, Hobday et al. 2001). The presence of sand channels may be important for the movement and concentration of algal drift macroalgae on which white abalone feed (Hobday and Tegner 2000).

Because most organisms are associated with specific habitat types, the spatial distribution of many species can be predicted by linking appropriate environmental parameters with the occurrence pattern of the species (Guisan and Zimmermann 2000). In order to quantify white abalone population and distribution of the habitat, most of the studies have been focused on the habitat preferences of the species. Since white abalone live below safe SCUBA depth (Lafferty et al. 2004), most studies have been conducted using manned and unmanned submersibles. Prior to ESA listing in 1999, the first major white abalone habitat study was conducted using the manned submersible Delta in waters
off the coast of southern California, including the Channel Islands and at the Osborn, Farnsworth, Tanner and Cortes offshore banks, where white abalone used to be abundant (Hobday and Tegner 2000; Hobday et al. 2001; Lafferty et al. 2004). Physical and biological habitat preferences of white abalone were identified from these submersible observations (Lafferty et al. 2004). Although the results of the habitat association analyses were informative, determination of the distribution and quantity of the habitat could not be obtained without detailed seafloor maps.

The requisite mapping of underwater habitats can be achieved with remote sensing technology, such as high-resolution multibeam sonar. In 2002, the Seafloor Mapping Lab at California State University Monterey Bay conducted multibeam surveys at Tanner Bank to produce high-resolution bathymetry maps that were used to design and guide subsequent ROV video surveys of white abalone habitat. ROV surveys were conducted from 2002-2006 by NOAA's Southwest Fisheries Science Center. The study revealed that high-resolution digital elevation models (DEMs) provided by multibeam sonar and used to characterize seafloor morphology, are effective tools for estimating the distribution of habitat suitable to white abalone (Butler et al. 2006). Although white abalone habitat preference has been studied and the technology to map the seafloor habitat has been available, landscape ecology models that objectively and quantitatively link these habitat preferences to seafloor morphology data sets have not been developed for white abalone.

Species-habitat relationships can be modeled as a method to predict habitat suitability and subsequently display the spatial distribution of habitat using Geographic Information System (GIS). In recent years, statistical habitat models have become a fundamental tool in the area of species conservation, reserve design, and population assessment. Habitat models derived from a combination of GIS technology and multivariate statistical analyses can represent spatial configuration of species. Carroll et al. (1999) created a multiple logistic regression model to predict distribution of the fisher (Martes pennanti) using data from presence/absence of the species and vegetation layer from satellite imagery. Predictive models have also been developed for benthic marine habitats using data from remote sensing technologies such as multibeam and sidescan sonar and Remotely Operated vehicle (ROV). Young et al. (in prep) developed predictive habitat models for three rockfish species in Cordell Bank National Marina
Sanctuary, California. These studies and others have used a variety of modeling techniques for relating species presence/absence data to habitat characteristics including, Logistic Regression (Cabeza et al., 2004), Classification and Regression Trees (CART) (Bourg et al. 2005), Ecological Niche Factor Analysis (ENFA) (Hirzel et al. 2002, Bryan and Metaxas 2007), Generalized Linear Models (GLMs) (Johnson et al., 2004; Iampietro et al. 2008), and Generalized Additive Models (GAMs) (Lehmann, 1998; Pearce et al., 2001). GLMs are popular and often used for modeling species distributions because they have been proven to be robust and can be easily applied in a geographic information system (GIS) (Guisan and Zimmermann 2000).

The ability to apply the prediction of a habitat models developed for a single species at one sight to other locations where it may persists is also of interest in most conservation and management strategies. Vanreusel et al. (2007) tested the transferability of habitat-based predictive distribution models for two regionally threatened butterflies within and among three nature reserves in northeastern Belgium. They found all models were transferable among the independent areas. Habitat model transferability was tested also for marine benthic species with positive results. Iampietro et al. (2008) used predictive habitat models for two of three rockfish species developed at Cordell Bank, California to successfully predict rockfish species presence at Del Monte Shalebeds of Monterey Bay, California.

Habitat models are especially useful for benthic organisms such as abalone and rockfish due to their close association with seafloor features. For bottom dwelling organisms, physical attributes of the seafloor such as depth, habitat complexity, slope, and substrate type are important in structuring their habitat and provide useful parameters for predicting patterns of spatial distribution (Freeman and Rogers 2001). Multivariate statistics which are often used for modeling predictive species distribution in GIS require both presence and absence data as response variables (Guisan and Zimmermann, 2000). In building a reliable model, absence data should indicate the entire area deemed unsuitable for the species. Habitat modeling is a particularly useful and effective tool in detecting areas with a high potential for re-colonization for rare and endangered species conservation and restoration (Hirzel et al. 2002; Engler et al., 2004; Lütolf et al., 2005). However, these models often have unreliable absence data, because sampling efforts for
rare species often target only those areas judged a priori most likely to have the species, with the identification of absence not included as an objective of the study.

One approach in creating a habitat model using a dataset with unreliable absence points is Ecological Niche Factor Analysis (ENFA) (Hirzel et al, 2002). Similar to Principal Component Analysis, ENFA summarizes habitat variables into a few uncorrelated and standardized ecological factors that explain species' ecological distribution (Hirzel et al., 2002). Hirzel et al. (2002) incorporated a suite of GIS and statistical tools into the BioMapper software application which can visualize species' habitat preferences as habitat suitability maps. Since ENFA does not require absence data, this method is widely used for various cryptic, rare and endangered species habitat modeling procedures (Sattler et al. 2007; Braunisch et al. 2008). The weakness of this approach is the tendency of ENFA models to over-predict species habitat because they lack absence data to restrict the predictions in environmentally inappropriate locations (Engler et al., 2004).

Another approach for dealing with unreliable species absence data is to randomly generate pseudo-absences over the study area based on presence-only data and modeling these results in logistic regression models such as GLM and GAM (Zaniewski et al. 2002; Engler et al. 2004). Compared with an ENFA-only model, this approach provides slightly better predictions, but at the risk of randomly sampled pseudo-absence points falling into areas suitable for the species (Engler et al, 2004). To address this weakness, Engler et al. (2004) developed a two-step approach to first create an ENFA model from presence-only data, followed by randomly selecting pseudo-absences from an area which the ENFA model identified as unsuitable for the species. These combined GLM-ENFA models enhance the quality of GLM-based potential distribution maps so as to provide better accuracy than either model alone.

The purpose of this study was to develop a predictive habitat model for endangered white abalone to guide in determining potential outplanting sites in southern California. General approach of the study was to build a binomial GLM model linking high-resolution multibeam derived habitat raster data with georeferenced ROV white abalone observation data (Figure 1). A sub-set of the ROV observation data that was not use in the development of the model, and was retained for subsequent validation and
quantification of model accuracy. Because the original dataset was not considered to have reliable measures of absence due to the nature of the ROV survey design, pseudo-absence data were generated for use in the modeling. The resultant model was then used to predict potential habitat for white abalone at other sites where white abalone are believed to have occurred in the past.

Figure 1. Schematic diagram of the research strategy. Habitat models based on distributions of white abalone were built using binomial GLM. Since the original dataset was not considered to have reliable measures of absence, and pseudo-absence data were therefore used in the modeling. The final model were then transferred the model to predict potential habitat distributions of the species in locations with different geographical settings.
2. METHODS

2.1. STUDY SITE

This study focused on the development of a predictive white abalone habitat model for Tanner Bank, California. Tanner Bank is a 17-kilometer-wide shallow rise beneath the Pacific Ocean, located 180 kilometers west of San Diego, California, at latitude 32°41' N and longitude 119°08' W (Figure 2). The Bank is characterized by strong currents and rough weather (Lewbel et al., 1981).

Tanner Bank consists of rocky ridges surrounded by sediment. The shallowest point of the bank lies at 27 meters. The ridges are mostly volcanic in origin with numerous incised channels and little sedimentary cover (Earkins et al., 2002). Remotely Operated Vehicle (ROV) transects ran mainly over the high areas of the bank at a depth range of 30m to 60m. The seafloor in these areas consists of sand with scattered rocks and occasional patches of larger rocks and boulders. Although multibeam bathymetry data covers the whole bank (80.0km²), the habitat model was created using the only data from the south-east portion of the bank (20km²) where white abalone were observed (Figure 2).

2.2. DATA DESCRIPTION

This habitat modeling study was based on two primary datasets; (1) habitat parameters collected by multibeam survey and (2) response variables which contain white abalone presence/absence location data collected by ROV survey. The multibeam sonar data of Tanner Bank were collected by California State University Monterey Bay Seafloor Mapping Lab to produce high-resolution bathymetry maps used to design and guide ROV video surveys of white abalone habitat. The ROV survey was conducted during the period of 2002-2006 by NOAA Southwest Fisheries Science Center.

The habitat parameter grids derived from bathymetric Digital Elevation Models (DEM) were used to quantify the relationship between white abalone presence/absence and each of the habitat variables in the GLMs.
Figure 2. Multibeam bathymetry image in shaded relief of Tanner Bank, California. Multibeam survey was completed in summer 2002. Black lines represent ROV transect tracks in 2002, 2004, and 2006. Red dots represent white abalone sightings. The black box indicates the area for which the GLM model was created.

Predictor variables

Selection of initial habitat variables were based on the known white abalone habitat preferences which include depth, moderately complex habitat structure, and sand-rock interface (Butler et al., 2006). Digital Elevation Models (DEMs) derived from multibeam sonar data with 2m resolution was used to create nine other habitat variables: slope, aspect, substrate, rugosity, vector ruggedness measure (VRM), Euclidean distance from sand-rock interface, and three different radii of topographic position index (TPIs) (Table 1).

Predictors were generated in the ArcGIS 9.2 extensions. Slope is a measure of the steepness of the seafloor relief; slope value is larger when an area has a greater degree of slope. Aspect shows direction of slope in terms of 360 degrees of the compass. Slope
and aspect and aspect were created in Spatial Analyst. VRM measures habitat complexity. Smaller values represent low complexity and high value represent high complexity of habitat. VRM was created in Terrain Tools in ArcGIS 9.2. Substrate was classified into two substrate types, rocky seafloor (1) and soft sediment seafloor (0), on the basis of the interpretation of the VRM, using a threshold for the break value between the two categories. Rugosity also measures habitat complexity.

Table 1. A list of predictor variable descriptions.

<table>
<thead>
<tr>
<th>Environmental Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>Depth in meter</td>
</tr>
<tr>
<td>Slope</td>
<td>Slope in degree</td>
</tr>
<tr>
<td>Rugosity</td>
<td>Ratio of terrain area to planar area</td>
</tr>
<tr>
<td>VRM</td>
<td>Ruggedness value, 0 (flat) to 1 (rugged)</td>
</tr>
<tr>
<td>Aspect</td>
<td>Degrees clockwise from north, 0 to 360</td>
</tr>
<tr>
<td>TPI 30</td>
<td>30m neighborhood radius TPI</td>
</tr>
<tr>
<td>TPI 60</td>
<td>60m neighborhood radius TPI</td>
</tr>
<tr>
<td>TPI 90</td>
<td>90m neighborhood radius TPI</td>
</tr>
<tr>
<td>TPI 240</td>
<td>240m neighborhood radius TPI</td>
</tr>
<tr>
<td>Substrate</td>
<td>Classified substrate (rock=1, sand=0)</td>
</tr>
<tr>
<td>Distance</td>
<td>Distance in meter to the rock-sand interface</td>
</tr>
</tbody>
</table>

TPIs were calculated with the extension Benthic Terrain Model (BTM). I also calculated Euclidean distance to sand-rock interface. TPI compares the elevation of each cell to the mean elevation of an area surrounding that cell and divides the landscape into classes based on how different or similar a point is to what is around it (Jenness, 2006). TPI values represent a point that is higher than the surrounding area, and negative TPI values represent a point that is lower than the surrounding area. Flat areas or constant slope shows TPI values close to zero. The TPI values were classified into six slope positions (Peak/Ridges, Upper Slope, Middle Slope, Flat/Plain, Lower Slope, and Valley Crevice) based on an index introduced by Weiss (2001). In cases where TPI values fell between -0.5 and 0.5, flat/Plain and middle slope areas were distinguished by using a threshold slope value of 4°. I used 30m, 60m and 240 m neighborhood radius TPIs.

Distance is Euclidean distance to sand-rock interface. To identify sand-rock interface, I used a fine scale TPI (30m neighborhood radius). ROV video was analyzed to verify that white abalone were found at the location categorized as Middle Slope in
TPI. The timestamp from each ROV video was linked to the timestamp in each transect shapefile in GIS so that TPI classifications can be precisely linked to the spot where the footage was collected. Comparing abalone location in the footage and TPI classification on the GIS map, I verified Middle Slope can represent rock-sand interface. Detailed descriptions of the variables are listed in Appendix I.

2.2.1. Response variable

The response variable is a binary variable, white abalone presence (1) or absence (0), from the ROV observation which shows binomial distribution. The ROV video surveys were conducted aboard the NOAA Research Vessel David Starr Jordan in 2002, 2004 and 2006 to document white abalone presence and absence as well as microhabitat conditions associated with the habitat. Each transect was approximately two-hours long and covered a length of 1 km (Butler et al., 2006). There were 45 video transects between 2002 and 2006 at Tanner Bank, in which a total of 246 white abalone were observed. The ROV’s positioning and tracking were provided by an acoustic positioning system that sends an acoustic signal from a transmitter located on the ROV to boat mounted receiver. The receiver calculates the position of the ROV relative to the vessel position. Accuracy of the ROV position is $\pm 10$ m (Butler, personal communication).

Ideally the biological survey would employ stratified random sampling to avoid adding bias to data (Green, 1979). In this case, due to rarity of the species and because the primary objective of the ROV surveys was to locate as many white abalone individuals as possible, the transect locations were restricted to those areas and habitat types thought to be most preferred by white abalone: i.e. between the 30 and 60 meter isobaths over rocky reef and the sand-and-rock reef interface (Butler et al., 2006).

Although the surveys successfully found a sufficient number of white abalone to run statistical analysis, this dataset lacked reliable absence observations needed to use GLM. For this reason, I used Ecological Niche Factor Analysis (ENFA) to create ENFA-weighted pseudo-absences due to lack of reliable absence data.

In order to assess the model predictive accuracy half of the presence/absence data were reserved from model creation. The reserved data were point locations that were
subsequently compared with the model results raster to evaluate how well the model could discriminate suitable and unsuitable habitat.

GLMs require both presence and absence as response variables. Presence locations were obtained directly from ROV video observation. Presence position data from 2002, 2004, and 2006 were pooled for the analyses.

The white abalone dataset lacks reliable absence observation due to (1) low density of the species population and (2) a sampling design not intended to sample absence data. Sampling efforts were focused in areas that were most likely to have white abalone present, thus the species’ absence may not be irrelevant to model habitat suitability. As false absences can decrease the reliability of prediction models (Chefaoui & Lobo, 2007) and true absences were not available in the original dataset, pseudo-absences were generated for later GLM analysis. Pseudo-absences should be chosen from area where the species is unlikely to be present. I followed the methods that Engler et al. (2006) developed to generate pseudo-absences using ENFA. Relying only on presence data, ENFA compares the environmental values where the species was observed to a reference set describing the whole study area. ENFA extracts two types of uncorrelated factors: (1) marginality factor (i.e. how different the species optimum habitat is from the mean habitat in the reference area) and (2) specialization factor (i.e. the ratio between the range of values for the species habitat and the reference area) (Hirzel et al., 2002). Using those factors ENFA computes habitat suitability maps (HS maps). Absence points were sampled randomly from the area where ENFA prediction was lower than 0.3, and were generated at twice the number of “presence” locations (n= 492). The ENFA model was performed in Biomapper 4.0 (Hirzel et al., 2007), a GIS and statistical tool designed to build habitat suitability models and maps. All environmental parameter grids were created in ArcGIS 9.2 (ESRI, 2008). The ESRI grid format data were then converted into IDRISI (Eastman, 1990) format in order to produce a habitat suitability map in Biomapper (Hirzel et al., 2007).
2.3. Modeling approach

2.3.1. Generalized Linear Model (GLM)

The objective of this analysis is how much variation in the presence/absence data your habitat variables account for. The model was fitted by logistic regression through a binomial generalized linear model (GLM) to investigate how seafloor morphology affects the occurrence of white abalone and to predict the species’ most suitable habitat. GLM is a generalization of multiple regression analysis used to model the relation between a response variable (i.e. presence/absence of white abalone) and a set of explanatory variables (i.e. habitat parameters). All GLM analyses were performed using the function in Marine Geospatial Ecology Tools (MGET) (Version 0.7a, Roberts et al., in review) which is an open-source software program that utilizes the R statistical package and creates multivariate habitat models in ArcGIS. A binomial GLM with logit link function was used. Because the ROV search for white abalone were concentrated on and around the bank top the modeling site was necessarily limited to that area (Figure 3).

The steps involved in constructing and analyzing a habitat model were as follows (1) selection or rejection of predictor variables based on results of correlation and regression tests and descriptive statistics (mean and standard deviation) with non-parametric Mann-Whitney U-test, (2) testing the derived polynomial function, (3) multivariate analysis of the selected predictors, and (4) testing presence of spatial autocorrelation in the response variables using Moran’s I.

2.3.2. Variable selection

The traditional approach to statistical model building involves minimizing the number of variables in the model while still explaining the patterns observed in the data (Hosmer & Lemeshow 2000). Variables were examined individually before running multivariate analysis to exclude unnecessary predictors and to evaluate the response curve of the predictor variables.

Descriptive statistics were used to analyze white abalone habitat selection of the study sites. Statistical associations among abalone and its predictors were examined using, Mann-Whitney U-test, categorical variables are from fisher’s exact test. Variable descriptions and comparison between presence points (n=246) and randomly sampled
points (n=246) created within the modeling site. The predictor variable correlations were examined using Spearman rank correlation. If a strong correlation between variables was found, only one of them was used to avoid multicollinearity.

2.3.3. POLYNOMIAL FUNCTION

In general, a linear relationship between environmental variables and response variables is very rare in ecology (Lek et al. 1996). The shapes of species response curves to environmental gradients are critical for better prediction of species/habitat distribution (Austin et al., 1994). A GLM is more flexible and better suited for analyzing relationships of spatial data because they do not force data into unnatural scales and allow for non-linearity (Guisan et al., 2002). To explore the response of the species to environmental factors, initial analyses were performed for each continuous variable. Assuming depth, slope, VRM and rugosity had polynomial relationships, the degree of the polynomial function was evaluated using stepwise backwards elimination. Stepwise backwards elimination is a technique to eliminate the variables one by one until all the variables remaining in the model are significant (p ≤ 0.05). The second degree polynomial (y ~ x + x^2) was used to test all continuous variables, and if a second degree polynomial was not significant, it was eliminated and the model was re-run in the non-polynomial form (y ~ x). If the model was still not significant, the variable was considered nonsignificant and excluded from later multivariate analysis. Because TPIs and substrate type are categorical variables they were excluded from the polynomial test.

2.3.4. MULTIVARIATE ANALYSIS

Once the significant variables had been identified the next step was to assess all possible combinations of variables in multivariable analysis to select the best performance model. Interactions among variables were also examined. In this study, an optimal model was identified using model selection procedure based on the Akaike Information Criterion (AIC), a measure of goodness of fit. Smaller AIC values indicate the better fit model.
2.3.5. Spatial Autocorrelation

Statistical testing to predict spatial pattern is often complicated by spatial autocorrelation. Spatial autocorrelation occurs when the values of variables at near by locations are not statistically independent from each other (Dormann et al., 2007). This phenomenon is common in habitat study as species' spatial patterns are often controlled by habitat types, such as depth, habitat complexity, and substrate types and therefore natural systems almost always have autocorrelation in the form of patchiness or gradient (Legendre, 1993; Dormann et al., 2007). However, the occurrence of spatial among data points violates the basic assumption of non-independence of data in most regression models. Legendre (1993) suggests that positive spatial autocorrelation could underestimate the standard errors of regression model thus inflates type I error (i.e. increases the rate of over-prediction). Moreover, Lennon (2000) argues that spatial autocorrelation can also affect estimates of the relative importance of environmental predictors, a phenomenon referred to as a "red shift".

The most common method used to measure spatial autocorrelation is Moran's I and Geary's c, which measure how similar a data point is to its neighboring points (Legendre, 1991). Moran's I was used to test for the presence of spatial autocorrelation in the variables used for the GLM model (presence/absence) and to generate correlograms for both response and environmental variables. Moran's I usually varies between 1.0 and -1.0, where 1.0 indicates positive correlation (clustered) and -1.0 indicates negative correlation (dispersed) (Diniz-Filho et al., 2002). Zero indicates no spatial autocorrelation (random). Moran's I coefficients then were then plotted against the geographical distance class. Known as a correlogram, this graphical representation allows for a spatial display of the similarity between neighboring data points.

The preliminary analysis indicated that spatial autocorrelation was present in the original dataset with significant spatial autocorrelation in Moran's I ($p < 0.001$) in all four variables. One way to account for spatial autocorrelation in subsequent analyses is to remove it by resampling individual data points at distances larger than the minimum distance at which spatial autocorrelation occurs (Legendre, 1993; Guisan and Zimmermann, 2000). However, the disadvantage of this approach is that it results in a net loss of valuable information and decreases the sample size of a survey.
The alternative method is to incorporate spatial autocorrelation into the model by adding distance-based weight matrix based on spatial neighborhood structures (Legendre, 1993). Augustin et al. (1996) introduced autologistic regression (ALR) models which incorporate an autocorrelated error in the regression model as an approach to correct for these biases. In autocovariate was calculated for the neighborhood sizes at which spatial autocorrelation of the response variable become zero. While A simplified formula of GLM is:

\[ \text{logit}(p) = b_1x_1 + b_2x_2 + b_3x_3 \]

ALR model is:

\[ \text{logit}(p) = b_1x_1 + b_2x_2 + b_3x_3 + \text{autocovariance} \]

The ALR model allow and investigator to: (1) examine the influences of spatial autocorrelation in the GLM model and determine the distance at where Moran’s I become near zero; and (2) estimate autocovariance using the distance and add to the final GLM model. Analysis of Moran’s I correlograms and development of the ALR model were then performed in the open-source R-CRAN software (version 2.6.1) (R Development Core Team, 2007) using the ‘spatial’, ‘ncf’ and ‘spdep’ packages.

Although both the non-spatial model (GLM) and spatial model (ALR) were constructed and examined in order to examine their predictive capability, the objective of this study was to test the hypothesis that there is a measurable relationship between white abalone occurrence and the habitat thus it was out of my scope to determine whether spatial or non-spatial model should be used to model white abalone habitat. This analysis was performed only to obtain an estimate of spatial autocorrelation in the data and account for any influences on the predictive capability of the predictive GLM habitat model.

### 2.3.6. Model evaluation

The prediction accuracy of the final GLM and ALR models, the ability of the model to separate presence and absence of the species in MGET. The method used to evaluate prediction accuracy was a measure of area under Receiver Operating Characteristic (ROC), which is called Area Under Curve (AUC) (Fieldings and Bell, 1997). ROC plots all sensitivity values (true presence) on the y axis versus 1-specificity
values (false presence) on the x axis. Sensitivity describes a probability that a prediction result provides “presence” (predicted presence) when the species is actually present (true presence). Specificity describes a probability that a prediction result provides “absence” (predicted absence) when the species is not absent (true absence). Each point on the ROC plot represents a sensitivity/specificity pair corresponding to a particular decision threshold. Accuracy depends on the cutoff value. A stricter cutoff will increase specificity (the power to predict true negative) but decrease sensitivity (the power to predict true positive) (Fieldings and Bell, 1997). MGET automatically choose the optimal cutoff value for the ROC, thus the optimum cutoff values were used to discriminate presence against absence. AUC quantifies the ability of the model to discriminate between presence and absence, and does not depend on the cut-off threshold thus avoiding the supported subjectivity in the threshold selection process (Lobo et al., 2008). The value of the AUC falls between unitless measurements of 0.5 and 1.0. An AUC of 0.5 suggests the model predictions are random. According to Hosmer and Lemeshow (2002), an AUC value of above 0.7 describes an acceptable level of performance, between 0.8 and 0.9 is excellent, and above 0.9 is outstanding. The coefficients of the variables for the GLM and ALR models were ranked based on their relative importance as indicated by z-values to examine the change in ranks.

The final model was mapped to display white abalone habitat at Tanner Bank. Multibeam bathymetry data are available at several locations in northern Channel Islands (Figure 6). Carrington Point at Santa Rosa Island was selected to map the predictive habitat model.
Figure 4. The density histogram of data in relation to each continuous environmental variable. Note that Depth, Log Slope, Log Rugosity, and Log VRM showed clear polynomial response while Distance from Middle Slope shows linear relationship and Aspect shows no pattern.

3. RESULTS

3.1. Variable selection

Correlation between predictor variables

Correlations among the predictor variables are summarized in Table 2. Correlation among predictor variables was generally high. Most variables are significantly correlated with a Spearman rank correlation of $r > 0.5$. Rugosity was significantly correlated with slope, VRM and substrate types ($r > 0.8$, $p < 0.001$). Depth and habitat complexity variables (rugosity and VRM) have positive correlations, indicating that the shallower part of the study site has complex habitat structure. Another clear correlation is between substrate type and VRM ($r = 0.80$, $p < 0.001$). This strong correlation is because the substrate grid was reclassified from the VRM grid. All variables except aspect showed significant correlation.
Table 2. Spearman rank correlation matrix table. Correlation coefficients in italic are insignificant.

<table>
<thead>
<tr>
<th></th>
<th>Depth</th>
<th>Aspect</th>
<th>Log Slope</th>
<th>Log Rug</th>
<th>Log VRM</th>
<th>Distance</th>
<th>Substrate</th>
<th>TPI 30</th>
<th>TPI 60</th>
<th>TPI 90</th>
<th>TPI 240</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td></td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspect</td>
<td>0.45</td>
<td></td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log</td>
<td>0.48</td>
<td>0.13</td>
<td></td>
<td>0.59</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>0.47</td>
<td>0.04</td>
<td>0.68</td>
<td>0.68</td>
<td>0.61</td>
<td>0.60</td>
<td>-0.23</td>
<td>0.52</td>
<td>0.59</td>
<td>0.59</td>
<td>0.68</td>
</tr>
<tr>
<td>Log</td>
<td>0.45</td>
<td>0.11</td>
<td>-0.31</td>
<td>-0.34</td>
<td>-0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rug</td>
<td>0.40</td>
<td>0.24</td>
<td>0.59</td>
<td>0.68</td>
<td>0.61</td>
<td>0.60</td>
<td>-0.23</td>
<td>0.52</td>
<td>0.59</td>
<td>0.59</td>
<td>0.68</td>
</tr>
<tr>
<td>VRM</td>
<td>0.40</td>
<td>0.24</td>
<td>0.59</td>
<td>0.68</td>
<td>0.61</td>
<td>0.60</td>
<td>-0.23</td>
<td>0.52</td>
<td>0.59</td>
<td>0.59</td>
<td>0.68</td>
</tr>
<tr>
<td>Distance</td>
<td>0.40</td>
<td>0.24</td>
<td>0.59</td>
<td>0.68</td>
<td>0.61</td>
<td>0.60</td>
<td>-0.23</td>
<td>0.52</td>
<td>0.59</td>
<td>0.59</td>
<td>0.68</td>
</tr>
</tbody>
</table>

**Descriptive statistics**

Descriptive statistics (mean and standard deviation) were used to initially assess predictor variables. Variable were compared between presence points (n=246) and random points (n=246) created within the modeling site. P-values for continuous variables are from non-parametric Mann-Whitney U-test, categorical variables are from fisher’s exact test (Table 3). I found that five of six habitat parameters of white abalone presence points were significantly different from average values in the modeling site when tested separately. According to the result, white abalone presence was significantly associated with depth; white abalone at Tanner Bank occurred on average at a depth of 44.7 m. This depth range is very similar to the results of existing white abalone studies at different locations (Davis et al., 1998; Haaker et al., 2000; Hobday et al., 2001; NMFS, 2008). Distance to sand-rock interface is also a strong predictor, and the mean distance of 3.5m ±3.22 is within ROV positioning error. The average slope is 4.54 degree, which is the cutoff value of TPI category between “Flat/Plain” and “Middle Slope”, which indicates that the boundary of these categories are very close to rock-sand interface where white abalone are likely to occur. Although both VRM and rugosity explain habitat complexity in term of seafloor complexity, Log-transformed VRM is more significant variable than log-transformed rugosity. Aspect has a nonsignificant relationship to white abalone occurrence (U-test, p=0.164).
Table 3. Variable descriptions and comparison between presence points (n=246) and random points (n=246) created within the modeling site. P-values for continuous variables are from non-parametric Mann-Whitney U-test, categorical variables are from fisher’s exact test.

<table>
<thead>
<tr>
<th>Environmental Parameters</th>
<th>Presence</th>
<th>Random</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Continuous variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth</td>
<td>-44.49</td>
<td>4.44</td>
<td>-55.94</td>
</tr>
<tr>
<td>Slope</td>
<td>4.54</td>
<td>3.43</td>
<td>4.06</td>
</tr>
<tr>
<td>Rugosity</td>
<td>1.0068</td>
<td>0.0126</td>
<td>1.1272</td>
</tr>
<tr>
<td>VRM</td>
<td>0.0015</td>
<td>0.0015</td>
<td>0.0002</td>
</tr>
<tr>
<td>Aspect</td>
<td>151.58</td>
<td>112.81</td>
<td>159.68</td>
</tr>
<tr>
<td>Distance</td>
<td>3.50</td>
<td>3.22</td>
<td>11.37</td>
</tr>
<tr>
<td><strong>Categorical variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPI 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPI 60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPI 240</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substrate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Variable selection**

The first variables eliminated were rugosity and slope because both variables are strongly correlated with VRM (table 2) and weaker explanatory variables than VRM (table 3). Substrate was eliminated because of its strong correlation with VRM ($r > 0.8$). All TPI categories were strongly correlated with VRM ($r > 0.50$), and were thus eliminated from multivariate analysis. The aspect variable was eliminated because it showed no significant effect on white abalone occurrence. As a result, the following variables were selected from variable selection analysis; depth, VRM, distance to rock-sand interface.

**3.2. Polynomial test**

The results of backwards elimination show that depth, slope, rugosity, and VRM have a second-order polynomial response, and Distance shows a linear response (Table 3). Depth is the strongest predictor of potential white abalone habitat when used as a second-degree polynomial term (deviance = 118.12, AIC = 124. The distance to rock-sand interface demonstrated a negative linear relationship to the occurrence of abalone. This suggests that the further the observation was from the rock-sand interface the lower the probability of being able to predict the occurrence of white abalone. Aspect did not
show any apparent pattern. The response curves of selected predictor variables are shown in figure 4. The curves show clear probability response to the predictors.

Table 4. Results of polynomial analysis. Starting from higher degree of polynomial terms, insignificant variables were eliminated from the GLM models until all the variables become significant. The forms chosen for each variable are in bold. Depth and VRM show second degree polynomial while distance show linear term.

| Variables     | Estimate coefficient | Standard Error | z-value | Pr(>|z|) | Signif. | Residual Deviance | AIC  |
|---------------|----------------------|----------------|---------|---------|---------|------------------|------|
| Null model    | -0.6931              | 0.1913         | -3.624  | 0.000   | ***     | 158.58           | 158.58|
| **Second-order** |                     |                |         |         |         |                  |      |
| Depth         | -1.23                | 0.40           | -3.08   | 0.002   | **      | 118.12           | 124.12|
| Depth²        | -0.01                | 0.00           | -3.21   | 0.001   | **      |                  |      |
| VRM           | -4.98                | 1.73           | -2.88   | 0.004   | **      |                  |      |
| VRM²          | -0.36                | 0.12           | -3.01   | 0.002   | **      | 136.34           | 142.00|
| Distance      | -0.11                | 0.16           | -0.71   | 0.475   |         | 132.05           | 138.05|
| Distance²     | -0.01                | 0.01           | -0.45   | 0.655   |         | 132.34           | 136.34|
| **Linear**    |                      |                |         |         |         |                  |      |
| Distance      | -0.19                | 0.06           | -3.23   | 0.001   | **      | 132.34           | 136.34|

Note: Significance codes: \( p < 0.0001 \) * ***; \( p < 0.001 \) **; \( p < 0.01 \) *; \( p < 0.05 \) *; \( p < 0.1 \) *.

Figure 4. Relationships between probability of white abalone occurrence and each selected predictor variables (a) Depth, (b) Log VRM, (c) Distance from rock-sand interface according to univariate models.
3.3. **Multivariate analysis**

The presence of white abalone was significantly associated with depth, seafloor ruggedness and distance to rock-sand interface. Among the three variables I used, depth is the strongest explanatory factor. The selected variables suggest habitat parameters that are of major habitat requirements for the species. All possible combinations of selected variables are shown in table 4. According to AIC, the model which includes all three variables is the model with greatest explanatory power (AIC = 246.56).

### Table 4. AIC and accuracy from AUC for Models are listed in ascending ΔAIC.

<table>
<thead>
<tr>
<th>Model Structure</th>
<th>AIC</th>
<th>ΔAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null model</td>
<td>471.75</td>
<td></td>
</tr>
<tr>
<td>Depth (<em><strong>) + Depth² (</strong></em>)+ LogVRM (<em><strong>)+ LogVRM² (</strong></em>))</td>
<td>254.45</td>
<td>7.89</td>
</tr>
<tr>
<td>Depth (<em><strong>) + Depth² (</strong></em>)+ Distance (**)</td>
<td>248.62</td>
<td>2.06</td>
</tr>
<tr>
<td>Depth (<em><strong>) + Depth² (</strong></em>)+ LogVRM (<em>)+ LogVRM² (</em>)+ Distance (**)</td>
<td>246.56</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Note: Significance codes: p<0.001 "***"; p<0.01 "**"; p<0.05 "*"

3.4. **Spatial Autocorrelation**

*Correlograms for variables*

The correlograms of the residuals for the variables are shown in Figure 5. The range of geographical distance was divided approximately 70m in each distance class bin. A correlogram of abalone presence/absence (Figure 5a) suggests that the presence of white abalone occurrence has strong positive spatial autocorrelation (clustered) that occurs over small distance with a continuous decrease in autocorrelation up to around 500m. Although there were moderate negative and positive autocorrelation at medium-distance range within the variables, predictor variables (Figure 5b, c, d) showed approximately the same spatial autocorrelation pattern. This pattern indicates that spatial variation in the habitat parameters is structured in patches at larger scales while relatively homogenous across smaller spatial scales. The distance up to where spatial autocorrelation is observed, approximately 500m in this case, can be interpreted as the average patch size in the variables (Diniz-Filho et al., 2002; Diniz-Filho et al., 2003).
Correlograms of spatial autocorrelation for four variables used in the final model: (a) response (presence/absence), (b) depth, (c) LogVRM, (d) distance from sand-rock interface.

**Correlograms for the models**

The correlogram of the residual for the GLM model displays a strong short-distance positive spatial autocorrelation (Figure 6a). Moran’s I values were binned each 50m. The Moran’s I becomes near zero at 500m, thus the radius for autocovariance for ALR model was set at 500m to eliminate short-distance spatial autocorrelation. Although including autocovariance did not completely remove the spatial structure in the GLM, significant reduction of short-distance spatial autocorrelation was observed especially the first two distance classes (200m radius) (Figure 6b).
Figure 6. Correlograms of residual spatial autocorrelation for (a) GLM (non-spatial model) and (b) Autologistic regression (spatial model). The neighborhood radius for autocovariance is 500m. The structure of spatial dimension (autocovariance) was incorporated in ALR model. Note that short-distance spatial autocorrelation is reduced in the ALR model. Moran’s I values were binned each 100m.

Model outcomes

Summary outcomes for the GLM and ALR are shown in Table 5. Depth was the most important variable after adding autocovariance in the GLM, but a change in the relative magnitude of estimated coefficients was observed. In the ALR model, distance to sand-rock interface is the weakest variable and it is barely not significant (p = 0.087). Although the relative importance changed, the coefficient estimates did not change.

3.5. Model evaluation

ROC curves for the GLM and ALR are shown in Figure 7. Although evaluation of the GLM model prediction is outstanding (AUC = 0.935), by accounting for spatial autocorrelation, the ALR model showed better predictive performance (AUC = 0.966). The contingency tables are shown in Table 6. The optimum threshold (cutoff) probability values to distinguish presence and absence for GLM and ALR are 0.476 and 0.486, respectively. Overall prediction accuracy for the GLM was 0.886 (Table 7). This means GLM has an 88.6% probability that it can correctly discriminate presence and absence with 90.2% probability of presence (sensitivity) and 87.8% probability of absence (specificity) (Table 6). After incorporating spatial autocorrelation into the model, the accuracy was improved to 91.1%, up 4.5% over the GLM. Also Type I error was
decreased from 0.122 to 0.089, and Type II error was decreased from 0.098 to 0.089 (Table 6). The ALR was also a better fitting model (AIC = 195.53) than the GLM (AIC = 246.56).

Table 5. Summaries of (a) GLM and (b) ALR. The tables report the estimated coefficients, standard error, z-value, significance, and the rank of importance of variables. In the ALR model, spatial autocorrelation was incorporated into the model by adding autocovariance as a variable.

| GLM      | Estimate coefficients | Std. Error | z value | Pr(>|z|) | Rank |
|----------|-----------------------|------------|---------|----------|------|
| Depth    | -1.958                | 0.302      | -6.475  | 0.000    | ***  |
| Depth²    | -0.023                | 0.003      | -6.672  | 0.000    | ***  |
| LogVRM   | -2.631                | 1.135      | -2.317  | 0.020    | *    |
| LogVRM²   | -0.188                | 0.080      | -2.355  | 0.019    | *    |
| Distance  | -0.106                | 0.037      | -2.896  | 0.004    | **   |

| ALR      | Estimate coefficients | Std. Error | z value | Pr(>|z|) | Rank |
|----------|-----------------------|------------|---------|----------|------|
| Depth    | -1.507                | 0.358      | -4.213  | 0.000    | ***  |
| Depth²    | -0.017                | 0.004      | -4.266  | 0.000    | ***  |
| LogVRM   | -2.727                | 1.413      | -1.930  | 0.054    |    |
| LogVRM²   | -0.204                | 0.098      | -2.075  | 0.038    | *    |
| Distance  | -0.079                | 0.046      | -1.711  | 0.087    |    |
| Autocovariance | 13.104          | 2.317      | 5.655   | 0.000    | ***  |

Note: Significance codes: *p<0.05, **p<0.01, ***p<0.001

Figure 7. ROC of (a) GLM model and (b) ALR model in which spatial autocovariate is incorporated. The curves represent true presence fraction against false presence fraction computed for all possible cutoff points between 1 and 0. AUC value for GLM model is 0.886 and ALR model is 0.966. The optimum cutoff values for the GLM and ALR are 0.476 and 0.486, respectively.
Table 6. Contingency tables of (a) GLM and (b) ALR. Model predictions against actual observations. Optimum cutoff values for GLM and ALR are 0.476 and 0.486, respectively.

<table>
<thead>
<tr>
<th>(a) GLM</th>
<th>Actual Presence</th>
<th>Actual Absence</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Presence</td>
<td>107</td>
<td>43</td>
<td>150</td>
</tr>
<tr>
<td>Predicted Absence</td>
<td>16</td>
<td>203</td>
<td>219</td>
</tr>
<tr>
<td>Total</td>
<td>123</td>
<td>246</td>
<td>369</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) ALR</th>
<th>Actual Presence</th>
<th>Actual Absence</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Presence</td>
<td>111</td>
<td>25</td>
<td>136</td>
</tr>
<tr>
<td>Predicted Absence</td>
<td>12</td>
<td>221</td>
<td>233</td>
</tr>
<tr>
<td>Total</td>
<td>123</td>
<td>246</td>
<td>369</td>
</tr>
</tbody>
</table>

Table 7. Summary of accuracy evaluation for GLM and ALR.

<table>
<thead>
<tr>
<th></th>
<th>GLM</th>
<th>ALR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.886</td>
<td>0.911</td>
</tr>
<tr>
<td>Positive prediction value (precision)</td>
<td>0.787</td>
<td>0.836</td>
</tr>
<tr>
<td>Negative prediction value (specificity)</td>
<td>0.947</td>
<td>0.953</td>
</tr>
<tr>
<td>Prediction-conditioned fallout (Type I error)</td>
<td>0.213</td>
<td>0.164</td>
</tr>
<tr>
<td>Prediction-conditioned miss (Type II error)</td>
<td>0.053</td>
<td>0.047</td>
</tr>
</tbody>
</table>

3.6. Mapping the habitat model

The aim of this study is to create habitat models which are transferable to different locations. ALR models are not recommended to extrapolate to a new area since spatial structure is different in each location (Dormann et al, 2007), the GLM model was applied to map white abalone habitat at tanner Bank. The resulting map of the GLM model presents a good prediction on the overall distribution of white abalone (Figure 5). The GLM model was also mapped potential white abalone habitat in Carrington Point, Santa Rosa Island (Figure 7).
Figure 5. Predicted white abalone habitat at Tanner Bank created from multivariate GLM. Warm colors indicate higher probability of white abalone occurrence and cold colors indicate lower probability of occurrence. Red dots indicate white abalone presence data which were reserved for accuracy assessment.

Figure 6. Map of northern Channel Islands, California. The areas in green were multibeam surveyed, which can be mapped white abalone habitat using the GLM model.
Figure 7. The GLM model was extrapolated to Carrington Point, Santa Rosa Island. Warm colors indicate higher probability of white abalone occurrence and cold colors indicate lower probability of occurrence.
4. DISCUSSION

Interpretation of the habitat model

The objective of this study was to create a habitat model to predict suitable white abalone habitat. The key assumption of this study was that white abalone were consistently associated with certain seafloor feature so that suitable habitat could be modeled using physical benthic parameters. Although the final GLM habitat model is very simple and includes only three physical, DEM derived variables, the results suggest that the white abalone habitat can be explained with high accuracy using multibeam bathymetry data alone.

The modeling results are consistent with existing knowledge of white abalone habitat preferences. Depth was the strongest predictor in the model (Table 4), and the predicted depth range is very similar to the ranges previously reported from direct observation (Hobday and Tegner 2000; Hobday et al. 2001; Rogers-Benett et al. 2002; Lafferty et al. 2004; Butler et al. 2006). The distance to rock-sand interface was also a strong predictor in the GLM model, and consistent with the observation that white abalone utilize sand channels to catch drifting algae (Lafferty et al. 2004). The GLM model identified moderate ruggedness as an important variable, which is consistent with white abalone's reported preference for low seafloor complexity (Butler et al. 2006).

Model limitation

White abalone habitat modeling has been limited by the availability of adequate habitat data. Although the result of the accuracy assessment for the final predictive model indicates an excellent fit with observational data, the model is based only on physical attributes of the habitat derived from bathymetric data. To improve the model’s predictability and transferability to locations with different habitat settings, adding other important habitat variables should be considered. Such considerations may include algae coverage models and more accurate rock-sand interface classifications derived from sidescan sonar data. Santos (1993) found that subtidal algal cover could be modeled using physical environmental variables such as depth (availability of light), bottom type, slope,
wave exposure and the amount of bottom sediment. Here, I was not able to build an algal model due to a lack of sufficient observational data on algal coverage at Tanner Bank. However, the expectation is that a *Laminaria* and *Agarum* coverage model may have significantly improved the white abalone habitat model accuracy because of the well documented association between white abalone theses algal species on Tanner Bank (Butler et al. 2006). There may also be important interactions between algal distribution and current flow relevant to white abalone occurrence because of their dependence on drift algae for food.

Another limitation of the habitat model in this study was the resolution of the bathymetry data (2m). Observation of the sand channels in the video data indicated that they could be as narrow as 1-2m and rocks occupied by abalone could be as small as 1m in diameter, both below the minimum detection level of the available bathymetry data. Also, sand-covered rocks, which abalone do not choose for their habitat, are categorized as rock in the multibeam data, causing misclassification of the rock-sand interface. One possible solution would be to use classified backscatter intensity images (described in Cochrane and Lafferty, 2002). Backscatter data can discriminate different substrate types (e.g. rock versus sediment) based on acoustic reflection.

Unfortunately the backscatter data available for Tanner Bank could not be used in this modeling effort because they came from a towed sidescan sonar system. Data from most towed sidescan sensors lack the positional accuracy of hull-mounted multibeam systems making it difficult or impossible to adequately align the two data sets. Now, with most modern multibeam bathymetry systems capable of simultaneously collecting sidescan quality backscatter data, it is possible to obtain and properly georeference both types of data needed to further refine predictive benthic habitat models. The performance of the habitat model created in this study was very good; however, the data sets used to create this static model provide only a single snapshot in time and space (Guisan and Zimmermann, 2000). Because the data were based on adult white abalone observation (Butler et al., 2006), the model may not be a good predictor of habitat for juveniles, which may or may not have the same habitat requirements. Another possible source of error when using static models is that species distribution patterns may be altered at the modeling location and thus not fully representative of the species’ actual preferences.
The model suggests the white abalone’s suitable habitat is between 40 and 50m, yet the species was historically found at shallower depths prior to the time that exploitation occurred (Tutschulte, 1976; Hobday et al., 2001). The depth range of present distribution could be affected by human disturbance. This gives some indication that white abalone can survive in shallower water, which would make outplanting much more practical, safer and cost effective than at greater depths.

**Spatial autocorrelation**

Spatial autocorrelation can occur at all spatial scales (Dormann, 2007), and some studies suggest that it does not necessarily cause a problem in the analysis of spatially explicit ecological data (Diniz-Filho et al. 2003; Hawkins et al. 2003; Hawkins et al. 2007). However, this does not mean that one should ignore the spatial artifacts (Legendre 1993) that spatial autocorrelation can introduce in statistical analyses of spatial data. The results of the accuracy assessment showed the ALR models helped correct for spatial autocorrelation and slightly improve predictive power (AUC = 0.966, accuracy = 0.911) though the predictive accuracy of the GLM model was very high (AUC = 0.935, accuracy = 0.886). The GLM model has the advantage that it could be applied more easily because it does not require calculation of autocovariance. Although adding spatial autocorrelation into the model changed the relative importance of environmental variables, the strongest variable (depth) remained the most important variable in the ALR model, and VRM and distance to sand-rock interface were still strong predictors. These results indicate that autocorrelation in the non-spatial GLM did not create a significant bias in terms of predictive accuracy.

**Management implications**

Species specific benthic habitat models created from multibeam bathymetry data have great potential for species conservation and management; particularly for rare species where there are often insufficient observational data on distribution and abundance available to make a management plan. The predictive maps developed here delineating suitable habitat areas can serve as the basis for a white abalone outplanning
strategy that prioritizes areas for possible restoration efforts and as well as helping to identify locations where white abalone may still exist.

The modeling methods developed in this study have the potential to be applied to other subtidal abalone species, including, green (Halioitis fulgens), pink (Halioitis corrugata) and pinto (Halioitis kamtschatkana) abalone which are heavily depleted and listed as Species of Concern (Rogers-Bennett et al. 2002). Estimating current abundance of these species is difficult because there is no baseline information available (Rogers-Bennett et al. 2002). In cases such as these, predictive habitat modeling could be used to generate estimates of population size as well as the distribution and abundance of suitable habitat: information required for a species to be considered for higher levels of protection under endangered status in the ESA (Federal Register, 2006).
REFERENCES


Young M., Iampietro P., Garza C., Robers D., Kvitek R.. in prep. Modeling Rockfish Abundance and Distribution on Cordell Bank National Marine Sanctuary, California
APPENDICES

APPENDIX A

DESCRIPTIONS OF PREDICTOR VARIABLES

Depth
As many studies have previously confirmed, depth is a strong predictor of white abalone habitat. White abalone occur in between at a depth range of about 30m to 60m (Davis et al. 1998, Hobday et al. 2001, Lafferty et al. 2004, Butler et al. 2006). Depth may possibly be a predictor of algae types which white abalone eat. Brown algae Laminaria farlowii and Agarum fimbriatum, white abalone’s primary food, are dominant at the depths of 20-50m (Tutschulte 1976, Hobday et al. 2001).

Slope
Slope is a measure of the steepness of the seafloor relief; slope value is larger when an area has a greater degree of slope. The relevant slope grid was calculated from the DEM in ArcGIS using the Spatial Analyst tool. There was a right-skewed distribution of abalone presence on the observed slope, thus natural logarithms were used to transform skewed distribution data prior to analyses.

Rugosity
Rugosity, or roughness of the seafloor, is calculated from DEM using ArcGIS extension Benthic Terrain Modeler (BTM) by comparing the ration of the surface area to the planar area of the same footprint. Directly it describes topographic roughness with a surface area to planar area ratio. Low values indicate flat smooth seafloor with high values indicating complex high-relief seafloor. As white abalone prefer complex habitat (Butler et al. 2006) rugosity can be a good predictor of presence, however, it is highly correlated with slope because it directly measures the variability in topographic gradient. There is no clear distinction between steep, flat seafloor (high slope and low rugosity) and steep, complex seafloor (high slope and high rugosity) (Sappington et al., 2007),
which may lower prediction accuracy. Natural logarithms were used to transform skewed distribution.

**Vector Ruggedness Measure (VRM)**

VRM also measures seafloor roughness. VRM is less correlated with slope since it quantifies complexity of seafloor more independently of slope and aspect than rugosity (Sappington et al., 2007), thus it can predict habitat complexity without an influence of gradient of seafloor. Natural logarithms were used to transform skewed distribution.

**Substrate**

Seafloor types were also classified into two substrate types, rocky seafloor (1) and soft sediment seafloor (0), on the basis of the interpretation of the VRM, using a threshold for the break value between the two categories. For Tanner Bank, the threshold value was 0.00025, and the areas which have a VRM value above this value were categorized as “rough”. The issue on this procedure is that artifacts are often classified as “rough” because they tend to have high VRM value. To avoid the misclassification, a hand-drawn mask is created. These artifacts are masked out from the VRM grid during the reclassification process and the areas are classified as “smooth”.

**Topographic Position Index (TPI)**

TPI was derived from the bathymetry grid. TPI compares the elevation of each cell to the mean elevation of an area surrounding that cell and divides the landscape into classes based on how different or similar a point is to what is around it (Jenness 2006). Using BTM, three TPIs were calculated with neighborhood sizes (i.e. radii) of 30m, 60m and 240m from DEM and slope grid. Neighborhood size is a critical component in analyzing TPIs. Small neighborhoods capture small features like small reefs or peaks while large neighborhoods capture larger scale features (Tagil and Jenness, 2008).

TPI values represent a point that is higher than the surrounding area, and negative TPI values represent a point that is lower than the surrounding area. Flat areas or constant slope shows TPI values close to zero. The TPI values were classified into six slope positions (Peak/Ridges, Upper Slope, Middle Slope, Flat/Plain, Lower Slope, and Valley Crevice) based on an index introduced by Weiss (2001). In cases where TPI values fell between -0.5 and 0.5, flat/Plain and middle slope areas were distinguished by
using a threshold slope value of 4°. TPI provides a useful parameter for seafloor classification and identifies abrupt changes in relief such as reef edges next to flat sand channels where the white abalone is typically found. By examining ROV video data, I determined that the boundary between the “Middle Slope” and “Flat/Plain” was most likely to be rock-sand interface where white abalone are normally found.

**Distance to rock-sand interface**

Rock-sand interface could be a strong variable for white abalone habitat models as TPI slope position categories could identify these interface locations. Although video analysis results showed white abalone always occur in or near rock-sand interface, the presence data points would not always fall where the interface is because of low positioning accuracy of ROV data (< ±10m, Butler, personal communication). In order to interpolate the positioning error in the models, I extracted the rock-sand interface (i.e. the edges between “Middle Slope” and “Flat/Plain” of TPI slope position) and calculated shortest Euclidean distance to the interface from each cell using a Euclidean Distance tool in ArcGIS Spatial Analyst Tools. This process created a new grid which showed that the closer the cells are to the nearest interface the lower the distance value.

**Algae coverage**

The white abalone feeds on macroalgae, such as *Laminaria farlowii* and *Agarum fimbriatu* (Tutschulte, 1976; Hobday et al., 2001; Lafferty et al., 2004), and Butler et al. (2006) found a strong correlation between these algae species and white abalone occurrence in Tanner Bank and also verified that presence of these algae species could be a significant parameter for white abalone habitat. Algae coverage along the ROV transect survey was analyzed by NMFS. Video footage was reviewed and algae species and the level of abundance were recorded in spreadsheets. Point shapefiles for algae coverage along the transect lines were created from these algae analysis data.

Although it is clear that including algae coverage in the model would improve model performance, I did not treat theses algae coverage data as a predictor variable for the model because the data unfortunately do not cover the entire modeling site but only where ROV transects ran in 2002. Since only the environmental parameters covering the entire modeling site in the GIS (i.e. ESRI grid format) can be used in this modeling study,
I used algae analysis after making the model to explain high/low prediction in relation to white abalone presence but did not include algae analysis in the model itself.

APPENDIX B

R STATISTICAL COMPUTING CODE

Compute spatial correlograms

# require R package "spatial"

Presence_A <- read.csv(file.choose())

par(mfrow=c(2,2), mar=c(5,5,0.5,0.5))

#Compute Spatial Correlogram

# Correlogram for Presence
topo.kr <- surf.ls(2, Presence_A)
correlogram (topo.kr, 50, pch=16,xlim=c(0,3500), ylab="Moran's I",
xlab="Geographic Distance (m)")
text(3500,0.9,"(a) Presence",cex=1.1,pos = 2)

# Correlogram for Depth
topo.kr <- surf.ls(2, Bathy_A)
correlogram (topo.kr, 50, pch = 16,xlim = c(0,3500), ylab="Moran's I",
xlab = "Geographic Distance (m)")
text(3500,0.9,"(b) Depth", cex=1.1, pos = 2)

# Correlogram for VRM
topo.kr <- surf.ls(2, Vrm_A)
correlogram (topo.kr, 50, pch = 16,xlim = c(0,3500), ylab="Moran's I",
xlab = "Geographic Distance (m)")
text(3500,0.9,"(c) LogVRM",cex=1.1, pos = 2)

# Correlogram for Distance
topo.kr <- surf.ls(2, Distance_A)
correlogram (topo.kr, 50, pch = 16, xlim = c(0, 3500), ylab = "Moran's I", xlab = "Geographic Distance (m)")
text(3500, 0.9, "(d) Distance", cex = 1.1, pos = 2)

**Autologistic regression model**

# package "ncf" is required
ALR_A <- read.csv(file.choose())
attach(ALR_A)
model <- glm(Presence ~ tb2mbat + I(tb2mbat^2) + Ltb2mvrm + I(Ltb2mvrm^2) + dist_t6, family = binomial(link = "logit"))
correlog1.1 <- correlog(x, y, residuals(model), na.rm = T, increment = 50, resamp = 10)

# plot the first 20 distance classes:
par(mfrow = c(2, 1))
plot(correrlog1.1$correlation[1:20], pch = 16, cex = 1, lwd = 1.5, xlab = "distance", ylab = "Moran's I", cex.lab = 2, cex.axis = 1.5); abline(h = 0)

# calculate Moran's I values explicitly for a certain distance,
# and to test for its significance:

# package "spdep" is required
ALR_A.nb <- dnearneigh(as.matrix(ALR_A[1:2]), 0, 20)

# Give lower and upper distance class here
ALR_A.listw <- nb2listw(ALR_A.nb)

# turns neighbourhood object into a weighted list
GlobMT1.1 <- moran.test(residuals(model), listw = ALR_A.listw)

# Autocovariate regression

# prepare neighbour lists for spatial autocorrelation analysis
nb.list <- dnearneigh(as.matrix(ALR_A[, c("x", "y")]), 0, 5, zero.policy = T)
nb.weights <- nb2listw(nb.list)
coords <- as.matrix(cbind(x, y))
ac500 <- autocov_dist(Presence, coords, nbs = 500, type = "inverse",
zero.policy = T, style = "B")
model_alr <- glm(Presence ~ tb2mbat + I(tb2mbat^2) + Ltb2mvrm +
I(Ltb2mvrm^2) + dist_t6 + ac500, family = binomial(link = "logit"))

**Comparison of non-spatial GLM and spatial ALR models**

# building GLM model
model_glm <- glm(formula = Presence ~ tb2mbat + I(tb2mbat^2) + Ltb2mvrm +
I(Ltb2mvrm^2) + dist_t6, family = binomial(link = "logit"))

summary(model_glm)

Call:
glm(formula = Presence ~ tb2mbat + I(tb2mbat^2) + Ltb2mvrm +
I(Ltb2mvrm^2) + dist_t6, family = binomial(link = "logit"))

Deviance Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2.0158821</td>
<td>-0.3581795</td>
<td>-0.004854</td>
<td>0.5994577</td>
<td>2.4337838</td>
</tr>
</tbody>
</table>

Coefficients:

|                | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | -48.998006 | 7.577240   | -6.466  | 1.00e-10 *** |
| tb2mbat        | -1.958420  | 0.302451   | -6.475  | 9.47e-11 *** |
| I(tb2mbat^2)   | -0.022895  | 0.003432   | -6.767  | 2.53e-11 *** |
| Ltb2mvrm       | -2.630898  | 1.135296   | -2.317  | 0.02048 *    |
| I(Ltb2mvrm^2)  | -0.188106  | 0.079890   | -2.355  | 0.01854 *    |
| dist_t6        | -0.106343  | 0.036724   | -2.896  | 0.00378 **   |

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 469.75  on 368  degrees of freedom
Residual deviance: 234.56  on 363  degrees of freedom
AIC: 246.56
# summary of residuals
anova(model_glm)

Analysis of Deviance Table

Model: binomial, link: logit

Response: Presence

Terms added sequentially (first to last)

<table>
<thead>
<tr>
<th>Df</th>
<th>Deviance</th>
<th>Resid. Df</th>
<th>Resid. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>368</td>
<td>469.75</td>
<td></td>
</tr>
<tr>
<td>tb2mbat</td>
<td>1</td>
<td>106.78</td>
<td>367</td>
</tr>
<tr>
<td>I(tb2mbat^2)</td>
<td>1</td>
<td>102.83</td>
<td>366</td>
</tr>
<tr>
<td>Ltb2mvrm</td>
<td>1</td>
<td>0.49</td>
<td>365</td>
</tr>
<tr>
<td>I(Ltb2mvrm^2)</td>
<td>1</td>
<td>15.20</td>
<td>364</td>
</tr>
<tr>
<td>dist_t6</td>
<td>1</td>
<td>9.89</td>
<td>363</td>
</tr>
</tbody>
</table>

# building ALR model
model_alr <- glm(formula = Presence ~ tb2mbat + I(tb2mbat^2) + Ltb2mvrm + I(Ltb2mvrm^2) + dist_t6 + ac500, family = binomial(link = "logit"))

summary(model_alr)

Call:
  glm(formula = Presence ~ tb2mbat + I(tb2mbat^2) + Ltb2mvrm + I(Ltb2mvrm^2) + dist_t6 + ac500, family = binomial(link = "logit"))

Deviance Residuals:
  Min        1Q  Median        3Q       Max
 -2.844852  -0.273203  -0.002601  0.320599  2.563181

Coefficients:
  Estimate  Std. Error    z value Pr(>|z|)
(Intercept)  -42.758306   9.190066  -4.653   3.28e-06  ***
tb2mbat      -1.506568   0.357599  -4.213    2.52e-05  ***
I(tb2mbat^2)  -0.016952   0.003974  -4.266    1.99e-05  ***
Ltb2mvrm     -2.726956  1.412652  -1.930     0.0536    .
I(Ltb2mvrm^2) -0.204057   0.098320  -2.075     0.0379    *
dist_t6      -0.079321   0.046370  -1.711     0.0872    .
ac500         13.104039  2.317140   5.655    1.56e-08  ***
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 469.75  on 368 degrees of freedom
Residual deviance: 181.53 on 362 degrees of freedom
AIC: 195.53

Number of Fisher Scoring iterations: 8

# summary of residuals
anova(model_alr)
Analysis of Deviance Table

Model: binomial, link: logit
Response: Presence

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev
NULL     1  469.75          368        469.75
  tb2mbat 1   106.78          367        362.97
 I(tb2mbat^2) 1    102.83          366       260.14
     Ltb2mvrm 1     0.49          365       259.65
 I(Ltb2mvrm^2) 1    15.20          364      244.45
     dist_t6 1     9.89          363      234.56
      ac500 1    53.03          362      181.53