

A BATHYMETRIC-BASED HABITAT MODEL FOR YELLOWEYE ROCKFISH  
(*SEBASTES RUBERRIMUS*) ON ALASKA'S OUTER KENAI PENINSULA

A Thesis

Presented to the Faculty of  
Alaska Pacific University  
In Partial Fulfillment of the Requirements  
For the Degree of  
Master of Science in Environmental Science

By

Joshua D. Mumm

December 2015

I grant Alaska Pacific University the non-exclusive right to use this work for the purpose of making single copies available to the public on a not-for-profit basis if the University's circulating copy is lost or destroyed.

---

Joshua D. Mumm

Date \_\_\_\_\_

December 2015

A BATHYMETRIC-BASED HABITAT MODEL FOR YELLOWEYE ROCKFISH  
(*SEBASTES RUBERRIMUS*) ON ALASKA'S OUTER KENAI PENINSULA

by

Joshua D. Mumm

THESIS

APPROVED: Thesis Committee

\_\_\_\_\_  
Chairperson  
Bradley Harris, Ph.D.  
Assistant Professor  
Alaska Pacific University

\_\_\_\_\_  
Date

\_\_\_\_\_  
Thesis Committee  
Roman Dial, Ph.D.  
Professor  
Alaska Pacific University

\_\_\_\_\_  
Date

\_\_\_\_\_  
Thesis Committee  
William Bechtol, Ph.D.  
Fisheries Consultant  
Bechtol Research

\_\_\_\_\_  
Date

APPROVED: \_\_\_\_\_  
Tracy M. Stewart, Ph.D.  
Academic Dean

\_\_\_\_\_  
Date

## ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Brad Harris for his guidance and encouragement throughout this project. His enthusiasm, motivating work ethic, and knack for seeing past myriads of distractions to focus on the issues that matter were especially appreciated when I was discouraged and lost in inconsequential minutiae. Committee member Dr. Roman Dial opened my eyes to the power of R and provided a wealth of invaluable insight into the modeling process. Dr. William Bechtol's expertise on the biology and management of rockfish in the area contributed much to the committee. I am deeply indebted to Mike Byerly for selflessly sharing the ROV and multibeam data he and others at the ADF&G office in Homer meticulously acquired and organized. Dr. Franz Mueter provided advice on a preliminary investigation. I owe Dr. Alexandre Hirzel a postcard for using his 'postcard-ware' to complete the Ecological Niche Factor Analysis. Lastly, although I only briefly corresponded with them directly, clearly this project was largely based on the work of Pat Iampietro, Drs. Mary Young, Rikk Kvitec, and others at the California State University, Seafloor Mapping Lab. I thank them for inspiration.

## ABSTRACT

Motivated primarily as part of a habitat-based stock assessment, we explored the feasibility of modeling yelloweye rockfish (*Sebastes ruberrimus*) habitat in Southcentral Alaska using high-resolution multibeam bathymetry. A generalized linear model was developed with bathymetrically derived terrain metrics (rugosity, slope, bathymetric position index, and distance-to-rock) as predictor variables. The model was parameterized and validated using remotely operated vehicle observations. When evaluated for the Chiswell Island training area, the model correctly classified 96.0% (n = 100) of a reserved set of presence/absence validation points (Cohen's Kappa = 0.92; AUC = 0.98). When evaluated for the independent Nuka Island testing area, the overall accuracy was 82.5% (n=332; Kappa = 0.65; AUC = 0.95). This study demonstrates that suitable yelloweye habitat can be modeled with reasonable accuracy using high-resolution multibeam bathymetry, and such a model is fairly portable among sites along the Kenai Peninsula's outer coast.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	i
ABSTRACT.....	ii
TABLE OF CONTENTS.....	iii
LIST OF TABLES.....	iv
LIST OF FIGURES.....	v
GENERAL INTRODUCTION.....	1
CHAPTER 1: BACKGROUND: AN OVERVIEW OF HABITAT MODELING AND THE BIOLOGY & MANAGEMENT OF DEMERSAL ROCKFISH IN ALASKA	
1.1 Habitat Models.....	4
1.2 Rockfish.....	9
CHAPTER 2: A BATHYMETRIC-BASED HABITAT MODEL FOR YELLOWEYE ROCKFISH ON ALASKA’S OUTER KENAI PENINSULA	
2.1 Introduction.....	16
2.2 Methods.....	18
2.3 Results.....	28
2.4 Discussion.....	32
GENERAL DISCUSSION.....	45
REFERENCES.....	48
TABLES.....	57
FIGURES.....	65
APENDIX.....	94

**LIST OF TABLES**

<b>Table 1.</b> Predictor variables and scales considered for the ENFA and GLMs.....	57
<b>Table 2.</b> Ranked performance of univariate models .....	58
<b>Table 3.</b> GLM parameters .....	59
<b>Table 4.</b> Error matrix for the Chiswell GLM in the Chiswell area .....	60
<b>Table 5.</b> Error matrix for the Chiswell GLM in the Nuka area.....	61
<b>Table 6.</b> Error matrix for the Nuka GLM in the Nuka area .....	62
<b>Table 7.</b> Summary of the Chiswell ENFA .....	63
<b>Table 8.</b> Summary of the Nuka ENFA.....	64

## LIST OF FIGURES

<b>Figure 1.</b> North Gulf District.....	66
<b>Figure 2.</b> Chiswell Island study area.....	67
<b>Figure 3.</b> Rugosity (VRM) in the Chiswell area.....	68
<b>Figure 4.</b> Distance-to-rock (DTR) in the Chiswell area.....	69
<b>Figure 5.</b> Coarse bathymetric position index (BPI) in the Chiswell area.....	70
<b>Figure 6.</b> Fine BPI in the Chiswell area.....	71
<b>Figure 7.</b> Slope in the Chiswell area.....	72
<b>Figure 8.</b> Habitat suitability from the ENFA in the Chiswell area.....	73
<b>Figure 9.</b> Probability of habitat in the Chiswell area predicted from the Chiswell GLM.....	74
<b>Figure 10.</b> Nuka Island study area.....	75
<b>Figure 11.</b> VRM in the Nuka area.....	76
<b>Figure 12.</b> DTR in the Nuka area.....	77
<b>Figure 13.</b> Coarse BPI in the Nuka area.....	78
<b>Figure 14.</b> Fine BPI in the Nuka area.....	79
<b>Figure 15.</b> Slope in the Nuka area.....	80
<b>Figure 16.</b> Habitat suitability from the ENFA in the Nuka area.....	81
<b>Figure 17.</b> Probability of habitat in the Nuka area predicted from the Chiswell GLM.....	82
<b>Figure 18.</b> Probability of habitat in the Nuka area predicted from the Nuka GLM.....	83
<b>Figure 19.</b> Correlograms for various scales of BPI.....	84
<b>Figure 20.</b> Correlograms for various scales of VRM.....	85
<b>Figure 21.</b> Correlograms for various scales of DTR.....	86



<b>Figure 22.</b> Correlogram of variables considered in final selection for Chiswell area .....	87
<b>Figure 23.</b> Correlogram of variables considered in final selection for Nuka area .....	88
<b>Figure 24.</b> Univariate models in the Chiswell area.....	89
<b>Figure 25.</b> Univariate models in the Nuka area .....	90
<b>Figure 26.</b> ROC plots for the GLMs .....	91
<b>Figure 27.</b> Distribution of yelloweye across the first two ecological factors in Chiswell area ....	92
<b>Figure 28.</b> Distribution of yelloweye across the first two ecological factors in Nuka area.....	93

## GENERAL INTRODUCTION

Habitat models and species distribution models predict the potential or realized distribution of a species based on environmental variables. They have been applied to a variety of fishery management and conservation issues such as: identifying potential marine protected areas (Ardron et al. 2002; Ardron and Wallace 2005; Embling et al. 2010); and delineating essential fish habitat, as mandated by the Magnuson-Stevens “Sustainable Fisheries Act” (DOC 1997; Valavanis et al. 2004, 2008; Rooper et al. 2014; Miller et al. 2015). The habitat model presented here was motivated primarily for use as part of a habitat-based abundance estimate for yelloweye rockfish (*Sebastes ruberrimus*) in southcentral Alaska.

Manned submersible, or more cost effective Remotely Operated Vehicle (ROV), surveys have become the standard method for estimating the density of demersal rockfish, largely because the rocky habitat where these species occur precludes traditional trawl surveys, and closed swim bladders embolize when brought to the surface (O’Connell and Carlile 1993, 1994; Nasby-Lucas et al. 2002; Johnson et al. 2003; Yoklavich 2007; Byerly et al. 2007, 2015; Green et al. 2014). Traditionally the habitat delineations used in habitat-based rockfish assessments have been derived from visually interpreting bathymetric survey data, usually by a trained expert familiar with the local geology and characteristics of the specific type of sonar product (Greene et al. 1999, 2007; Nasby-Lucas 2002; Yoklavich et al. 2007). The disadvantage of this method is that it is time consuming, reliant on the availability of a particular expert, and because it is subjective, is prone to bias and non standardization among areas. In contrast, the purely analytic algorithmic approach presented here should yield cost-effective, reproducible, standardized results among areas.

In Chapter 1 I provide an overview of the field of habitat modeling with particular focus on the problems inherent to unreliable absences and the methods developed in response. I also provide background information on the biology and management of yelloweye in Alaska.

Techniques have been previously developed for modeling the distribution of three species of rockfish — rosy (*S. rosaceus*), yellowtail (*S. flavidus*), and greenstriped (*S. elongatus*) — off the coast of California using submersible observations and high-resolution multibeam bathymetry (Iampietro et al. 2005, 2008; Young et al. 2010). In Chapter 2, I capitalize on previously acquired analogous data (Byerly et al. 2007, 2015) to explore the feasibility of using a similar approach to model the distribution of yelloweye rockfish habitat in southcentral Alaska.

## **CHAPTER 1**

### **BACKGROUND:**

#### **AN OVERVIEW OF HABITAT MODELING, AND THE BIOLOGY & MANAGEMENT OF DEMERSAL ROCKFISH IN ALASKA**

## 1.1 HABITAT MODELS

Ecologists have long recognized that the ecological requirements of species, loosely described as their niches, and their distributions are related. Hutchinson (1957) defined an environmental niche as the n-dimensional hypervolume in the multidimensional space of environmental factors that affect the welfare of a species. Realizing the range of conditions under which a species could potentially exist is greater than the range of conditions under which a species actually does exist, especially after the effects of predation and competition, he distinguished fundamental from realized niches. Habitat models in essence, project this multidimensional hypervolume of a niche onto the three dimensions of physical space, and, less commonly, time. More concretely, habitat models predict the distribution of a species' habitat — either fundamental if predicted from theoretical physiological constraints, or realized if derived from field observations — based on environmental variables, with habitat defined as the place where an organism is ordinarily found (Begon et al. 2006; Araju and Guisan 2006).

Habitat modeling has become commonplace over the past few decades with the proliferation of geographic information systems (GIS) facilitating the integration of increasingly available spatial environmental and species occurrence datasets over large areas. Early habitat models focused on the terrestrial realm, largely because collecting data — both species presence and environmental — was easier on land than underwater, particularly over large spatial extents because most satellite and airborne remote sensing technologies do not penetrate the sea surface, and also because of the temporal variability and dynamics of water bodies (Valavanis et al. 2008). Plants, because of their permanence, and highly terrain dependent species of animals, such as mountain goats, were among the earliest species to be successfully modeled with GIS-based habitat models (Fischer 1990; Fitzgerald and Lees 1993; Holmes 1993; Gross 2002).

Despite the challenges associated with modeling marine habitats, interest has exploded over the past two decades, largely because of the growing availability of spatially explicit marine environmental datasets, especially high-resolution remotely-sensed bathymetry collected using multibeam echo sounders (MBES) (Iampietro et al. 2005; Valavanis et al. 2008; Brown and Blondel 2009). Although MBES have widely emerged as the tool of choice for seafloor habitat modeling, primarily because of their ability to collect both bathymetry and backscatter information simultaneously over a full coverage swath of the seafloor, the older single beam acoustic ground discrimination systems and side scan sonars both have their respective merits and have also both been employed with some degree of success in marine benthic habitat models (Parnum et al. 2009).

The popularity of habitat modeling has generated a diverse array of habitat models. While an exhaustive review of the various model types is beyond the scope of this study, a brief overview of the field is warranted. Beginning with the commonalities, in virtually all habitat models: (1) the study area is depicted as a raster map, that is, a full coverage geospatial grid of equally sized adjacent cells; (2) each cell is assigned values for a range of environmental variables to form the set of independent variables; (3) the dependent variable is the species occurrence data observed for a subset of the cells; and (4) the habitat model itself is the function which classifies the cells of the study area as accurately as possible as either suitable or unsuitable habitat based on the environmental variables (Hirzel et al. 2002). A major distinction amongst the huge array of published habitat models relates to the type of species occurrence data used, specifically whether a model requires: (1) presence and absence data; or (2) presence data only.

## **Presence/Absence Models**

Presence/absence models, which use group discriminative analyses, were developed first and are more classically intuitive. Here the species distribution data contain both presence and absence points. Regression based generalized linear models (GLMs), and their even more generalized extension generalized additive models (GAMs), are the most popular types of presence/absence habitat models currently in use. Much of their popularity over ordinary multiple regression is due to the ease at which they accommodate non-normally distributed and heteroscedastic predictor variables including ordinal and even categorical data. Another major advantage of GLMs is that they can constrain the response variable to a meaningful range of values through the use of a link function which relates the linear predictor — the linear combination of environmental variables and their parameters — to the response variable. For example, in species habitat modeling the response variable is commonly desired to be binomially distributed within the range of 0 (unsuitable) to 1 (suitable). In this case, the logit function is used to linearize the binomially distributed response variable and relate it to the linear predictor. Other popular presence/absence habitat models include canonical correspondence analysis (e.g. for rockfish, Stein et al. 1992), ensembles of regressions or classification trees (e.g., Moore et al. 2009), and neural networks (e.g., Fitzgerald and Lees 1993). One of the main advantages of presence/absence models over presence-only models is that their accuracy is easily evaluated by comparing the predictions output from the model to observed presence and absence points. The main disadvantage of presence/absence models is that they require absence points.

A common problem in habitat distribution modeling is the unreliability of absence points. While the observation of a species guarantees both presence and suitability of habitat at that location, the inverse is not true; lack of detection at a point does not necessarily indicate that that

point is unsuitable habitat. Habitat modelers have termed these locations *false absences* (Hirzel 2002). False absences arise from either of two situations: (1) the species was in fact at that location but was not detected; or (2) the species really was not at that location, even though the habitat was suitable. The first situation is especially common when dealing with small, inconspicuous, or otherwise hard to detect organisms, and can be accounted for with occupancy estimation techniques, although these require repeat sampling, where occupancy is modeled as the product of probability of detection and the probability of occurrence (Mackenzie et al. 2006; Kery 2010). The second situation is typical of rare and especially heavily exploited populations where the realized niche is small relative to the fundamental niche. Here exploitation, predation, competition, or other factors keep a species confined to a small area of occupied habitat relative to the greater amount of available suitable habitat (Soberón 2007). The second situation also occurs when an ecosystem has not reached static equilibrium, as in the case with recently introduced invaders or colonizers, or never will, in the dynamic case of metapopulations.

### **Presence-Only Models**

Largely as a way to circumvent the aforementioned difficulties associated with unreliable absence points, methods have been developed to model species habitat distributions using only presence points. Also known as profile methods, these approaches do not require absence points for fitting. Instead of comparing the environmental characteristics of the set of presence points to the absence points, profile methods compare the environmental characteristics of the presence points to the background environmental characteristics, where background is defined as the greater study area. They compare the realized niche to the totality of available environmental conditions. The first widely known profile habitat modeling method was the climatic envelope approach developed by Australian botanists in the 1980s and implemented in the BIOCLIM



package (Busby 1991). More recently ecological niche factor analysis (ENFA) was developed and packaged in the Biomapper software (Hirzel et al. 2002), which has been at least partially superseded by the maximum entropy method implemented in the MaxEnt program (Phillips et al. 2006). Although attractive because they avoid the problems caused by unreliable absences, presence-only models have certain disadvantages. Chief amongst these disadvantages is the tendency of presence-only models to be overly inclusive in the amount of area they classify as habitat (Hirzel et al. 2002). This tendency has been attributed directly to the lack of absence points to restrict the predicted habitat output from the model, thus a 'perfect' habitat model could classify the entire study area as suitable habitat, at least when evaluated using the classic percentage accuracy metrics. Essentially, presence-only models are not penalized for errors of commission (false positives). A related challenge with presence only models is evaluation of their accuracy given the lack of absence data available for validation, although several performance metrics have been proposed (Hirzel et al. 2006; Monk et al. 2010).

Given the advantages and classic familiarity of regression based presence/absence models such as GLMs, yet their often times difficult to satisfy requirement for absence points, techniques have been developed to create pseudo-absences with GLMs. The simplest method is to pick points at random across the entire study area and use these pseudo-absences as actual absences in a presence/absence model such as a GLM (e.g., Hirzel et al. 2001). However, the random selection of this method runs the risk of treating suitable habitat as absence, thereby reducing model performance. To reduce the probability of selecting good habitat as a pseudo-absence, Engler et al. (2004) developed a two-step method wherein the pool of cells from which pseudo-absences are randomly selected is restricted to that subset of the study area identified by a preliminary ENFA as poor habitat. This decreases the probability that a pseudo-absence point

is selected from an area that is, in fact, good habitat. Arguably, this two-step ENFA-GLM method combines the respective strengths of the classic regression-based GLM with the presence-only ENFA.

## 1.2 ROCKFISH

### General Biology

Rockfish, (*Sebastes spp.* and *Sebastolobus spp.*; order Scorpaeniformes; family Scorpaenidae) are extremely diverse, with ~102 species worldwide, the majority of which (~96 species), are distributed across the North Pacific (Love et al. 2002). At least 24 species occur along the outer coast of the Kenai Peninsula (Russ et al. 2013). Rockfish have several unique life history characteristics. Eggs are fertilized internally, months after mating, are primitively matrotrophically viviparous, and are parturated as fully formed larvae (Love et al 2002). Extreme longevity — 205 y for rougheye (*S. aleutianus*) and 121 y for yelloweye in Alaska — likely evolved as an adaptation to profound episodic recruitment, wherein decades often separate oceanic conditions supportive of successful recruitments (Munk 2001; O’Connell and Brylinsky 2003). Together these dramatically k-selected life history characteristics (Pianka 1970), along with late maturation (22 y for female yelloweye), low natural mortality, limited dispersal, and closed swim bladders, predispose rockfish, especially demersal species such as yelloweye, to classic vulnerability to overfishing (O’Connell and Funk 1986; Bechtol 1998).

Yelloweye occur at depths of 15–549 m, but are more typically found on hard rocky bottom from 91 to 180 m, with size and age generally increasing with depth (Love et al. 2002; Johnson et al. 2003). They are found from Baja to Umnak I. (Kramer and O’Connell 2004).

## **In Habitat Models**

Rockfish, particularly the demersal species, are excellent candidates for habitat models. Unlike pelagic species which migrate to follow shifting water bodies, demersal rockfish exhibit high site fidelity and are closely associated with permanent rock outcroppings (Carlson and Haight 1972; Johnson et al. 2003; Iampietro et al. 2008; Rooper et al. 2010; Hannah and Rankin 2011; Yoklavich et al. 2000). Yelloweye distributions in Alaska are strongly related to the three-dimensional geomorphology of the seafloor with the highest densities found over areas of broken rock and boulders (Stein et al. 1992; O'Connell and Carlile 1993). Some demersal rockfish may spend their entire lives on the same rock pile (Carlson and Haight 1972; Hannah and Rankin 2011). These rock outcroppings are relatively easy to detect with high-resolution bathymetry (Iampietro et al. 2005; Ardron and Wallace 2005).

Young et al. (2010) successfully modeled the distribution of three species of rockfish — rosy (*S. rosaceus*), yellowtail (*S. flavidus*) and greenstriped (*S. elongatus*) — off California using a combination of submersible observations and multibeam bathymetry. For each species they incorporated several depth derived terrain variables into a binomial logistic GLM to predict the probability of presence for that species (Iampietro et al. 2005, 2008; Young et al. 2010).

## **North Gulf Fishery**

Rockfish of the Kenai Peninsula's outer coast play important ecological roles and have been long pursued by both the commercial and recreational fisheries. For ADF&G (Alaska Department of Fish and Game) commercial groundfish management purposes, the coast is defined as the North Gulf District, part of the Cook Inlet Management Area and bounded on the east by Cape Fairfield and Point Adam to the west (5 AAC 28.305; Figure 1). Commercial harvests in the North Gulf District peaked at 502,000 lb in 1995, but were capped at 150,000 lb

in 1998 because of rapid harvest increases, sustainability concerns, and limited stock assessment information (Trowbridge et al. 2008). This cap was based on historical catch averages (Bechtol 1998), similar to a Tier 6 approach applied by the North Pacific Fishery Management Council (NPFMC) for groundfish assessment in which only catch data are available (NPFMC 2014). Harvests have subsequently fluctuated largely in concert with market prices and competing economic opportunities afforded by alternative local fisheries, primarily salmon and halibut, with rockfish reaching a near record low of 25,000 lb in 2007 before climbing to 60,500 lb in 2014 (Trowbridge et al. 2008).

While at least 24 species of rockfish occur in the North Gulf District, catches are dominated by the pelagic black rockfish *S. melanops* and the demersal yelloweye rockfish *S. ruberrimus*, comprising approximately 50% and 30% of the catch respectively (Trowbridge et al. 2008). Pelagic species are harvested mostly using jig gear, while demersal species are harvested mostly with longline. In 2005, the directed fishery for demersal rockfish was eliminated allowing harvest of demersal rockfish only as bycatch, mostly to the halibut and Pacific cod longline fisheries but also incidental to the directed jig fisheries for pelagic rockfish and lingcod (Russ et al. 2013). The current study focused on yelloweye because demersal species are expected to be more conducive to terrain-based habitat models than pelagic species and because yelloweye are more vulnerable to overfishing, with slower growth and maturation, and more limited dispersion than black rockfish (Bechtol 1998; Johnson et al. 2003; Hannah and Rankin 2011). As evidence of their extremely episodic recruitment, in recent years (2001 to 2004) the commercial catch of yelloweye in Lower Cook Inlet has been dominated by fish of a single year class, those recruited in 1969 (Trowbridge et al. 2008).

## **West Coast Status**

Yelloweye abundances off California, Washington, and Oregon are estimated to be at 10% of pre-exploitation levels, far below the customary 25% threshold used to define overfishing (Taylor and Wetzel 2011). Puget Sound yelloweye and two other demersal rockfish species have been listed as threatened under the Endangered Species Act (Drake et al. 2010). The particular life history traits for demersal rockfish in general, and especially yelloweye, suggest that these species will be extremely slow to recover from overfishing, with estimates ranging from 50 to 500 y for Washington yelloweye stocks (Taylor 2011).

## **Management in Alaska**

Management authority for yelloweye in Alaska varies by location and fishery. Recreational fisheries are managed by ADF&G, both within state waters (0-3 nautical miles [nmi] from shore) and within the Exclusive Economic Zone (EEZ; 3-200 nmi). Commercial fisheries in state waters are managed exclusively by the State (5 AAC 28.010). In southeast Alaska, the State also manages yelloweye as part of the demersal shelf rockfish (DSR) fishery in the EEZ through an extended jurisdiction program with oversight by the NPFMC. Commercial DSR fisheries in federal waters outside of southeast Alaska are managed exclusively by the NPFMC.

Despite their ecologic and economic importance and susceptibility to overfishing, neither a comprehensive stock assessment nor a coastwide abundance estimate has been completed for rockfish of the North Gulf District. The current harvest cap is somewhat arbitrary, being based on historic catches, rather than biologically significant reference points (Trowbridge et al. 2008). Information about distribution and abundance is limited, primarily to several index sites along the coast where ADF&G has completed both high-resolution MBES bathymetric surveys and

video surveys of rockfish using a Remotely Operated Vehicle (ROV). Habitat-based stock assessments are well suited to species with patchy heterogeneous distributions highly dependent on habitat, such as demersal rockfish (Nasby-Lucas et al. 2002; Yoklavich et al. 2007; Tissot et al. 2007; Rooper et al. 2010).

The management plan for yelloweye in the Southeast Region of Alaska provides a feasible example of how a habitat-based stock assessment might be conducted in the Central Region. Prior to 1992, catch limits in the Southeast Region were based on historic catch averages. Beginning in 1992, catch limits for Southeast yelloweye have been set relative to the fishery-independent biomass estimates based on manned submersible or ROV line transect surveys (O'Connell et al. 1991; O'Connell and Carlisle 1993). These biomass estimates are simply the mean densities observed in the submersible or ROV line transects, expanded by an estimate of the total habitat in the district. The habitat delineations have a variety of sources including expert interpreted sidescan and/or multibeam sonar (Greene et al. 1999, 2007), high catch per unit effort (CPUE) as recorded in commercial logbook data buffered by 0.5 km, and rocky features on NOAA nautical charts buffered by 0.5 mi (Green et al. 2014). Of note, O'Connell and Carlisle (1993) only intended these estimates of habitat area for *interim* use until they accomplished their *ultimate goal* of developing a *quantitative predictive model to estimate density of yelloweye rockfish and other DSR species based on one or more parameters reflective of structural habitat complexity*.

The management strategy for yelloweye in the North Gulf District is not yet as developed as in the Southeast Region. The ADF&G is gaining a reasonable estimate of yelloweye densities within habitat strata based on ROV video surveys conducted at several index sites along the outer coast, however no coastwide abundance estimate can be calculated because the total area of

habitat in the district is not yet known. The current project aimed at this knowledge gap by investigating the feasibility of modeling the distribution of yelloweye habitat using remotely sensed MBES bathymetry. Unlike the Southeast Region, habitat in the Central Region cannot be estimated from logbook data because logbooks are not required of Central Region commercial harvesters. Full coverage multibeam surveys have been completed for the entirety of the index sites mentioned previously in addition to a course grid (~4 km survey line spacing) over most of the remaining coast (Figure 1). Conceptually, the habitat model produced by this project will serve as a bridge to convert multibeam bathymetry to predicted rockfish habitat, and at least partially realizes the objective proposed by O'Connell and Carlisle (1993) 22 years ago.

**CHAPTER 2**

**A BATHYMETRIC-BASED HABITAT MODEL FOR  
YELLOWEYE ROCKFISH ON ALASKA'S OUTER KENAI PENINSULA**



## 2.1 INTRODUCTION

Several life history characteristics predispose demersal rockfish such as yelloweye rockfish (*Sebastes ruberimus*), to being classically vulnerable to overexploitation. Chief amongst these characteristics are their profoundly k-selected traits of low productivity, episodic recruitment, late maturation, and low dispersion. Furthermore, demersal rockfish are difficult to survey using traditional methods because their rocky habitat precludes trawl surveys, and closed swim bladders embolize when brought to the surface, thereby inhibiting extractive mark-recapture surveys (Gotshall 1964). Yelloweye stocks off the west coast of the United States are severely overfished with abundances currently estimated at ~10% of pre-exploitation levels and the Puget Sound population segment listed as threatened under the Endangered Species Act in 2010 (Drake et al. 2010; Taylor and Wetzel 2011).

Despite their susceptibility to overfishing, ecological importance, and economic value in both recreational and commercial fisheries, neither a comprehensive stock assessment nor a districtwide abundance estimate has been completed for yelloweye in Southcentral Alaska. Instead catch limits are static and based on historic catch averages (5 AAC 28.365). Preferably, catch limits are set relative to biologically significant reference points such as an estimate of abundance or biomass. Perhaps the most effective method of estimating the abundance of a heterogeneously distributed species closely associated with specific habitats such as yelloweye is with a habitat-based abundance estimate where the densities observed within habitat strata are expanded by the total areal extent of habitat in the management unit (e.g., Nasby-Lucas et al. 2002; Yoklavich et al. 2007 ). In contrast to Southcentral Alaska, yelloweye catch limits in Southeast Alaska are tied to this type of habitat-based abundance estimate. The Alaska Department of Fish and Game (ADF&G) is using a Remotely Operated Vehicle (ROV) to

estimate the density of yelloweye within habitat strata at several index sides along the outer coast of the Kenai Peninsula (Byerly et al. 2007, 2015). However, before a districtwide abundance estimate can be calculated, an estimate of the total area of suitable habitat in the district is required. The current project aimed to bridge this knowledge gap by producing a model for predicting potential yelloweye habitat from high-resolution bathymetry in Southcentral Alaska.

Young et al. (2010) successfully modeled the distribution of three species of rockfish — rosy (*S. rosaceus*), yellowtail (*S. flavidus*) and greenstriped (*S. elongatus*) — off California using a combination of submersible observations and multibeam echosounder (MBES) bathymetry. For each species they incorporated several depth derived terrain variables into a binomial logistic generalized linear model (GLM) to predict the probability of presence for that species. The current study used previously acquired MBES and ROV data to explore the feasibility of modeling yelloweye rockfish habitat in the Chiswell Island and Nuka Island study areas using an approach similar to that developed in California (Figure 1; Iampietro et al. 2005, 2008; Young et al. 2010).

### **Research Question**

Can yelloweye habitat on the outer coast of the Kenai Peninsula be accurately modeled using high-resolution bathymetry? Primary objectives were:

- 1) Determine if yelloweye within the Chiswell and Nuka study areas are preferentially distributed across bathymetrically derived terrain variables.
- 2) Determine the most parsimonious combination of terrain variables for predicting the distribution of yelloweye habitat within the Chiswell study area.
- 3) Evaluate the accuracy of a GLM in predicting the distribution of yelloweye habitat within the Chiswell training area.

- 4) As a further test of portability and robustness, evaluate the performance of the Chiswell habitat model in the independent Nuka evaluation area.

## 2.2 METHODS

### Study Areas

The Kenai Peninsula's outer coast is the ~200 km long span from Prince William Sound to Kachemak Bay (Figure 1). The coast is characterized by rugged, steeply incised glacial fjords and direct exposure to the Gulf of Alaska. For ADF&G commercial groundfish management purposes, the coast is defined as the North Gulf District, part of the Cook Inlet Management Area and bounded on the east by Cape Fairfield and Point Adam to the west (5 AAC 28.305).

Two different study areas along the outer coast were used in this analysis: the Chiswell Island study area and the Nuka Island study area (Figures 2 and 10). These were selected from amongst the four areas in the district where both MBES bathymetry and ROV rockfish surveys have been conducted. The Chiswell area was used for most of the variable selection and parametrization, while the Nuka area was reserved as a mostly independent testing area.

More precisely, the roles of the areas in the study were slightly more nuanced. Both areas were considered when selecting which scale of each type of terrain variable to include in the scope of the final variable selection process. However, to maintain the independence of the Nuka area for use as a test of the portability of the Chiswell model, the final variable selection from amongst the best scale of each type of terrain variable, was done using only the Chiswell area. Additionally, after testing the portability of the Chiswell model in the Nuka area, the GLM was reparametrized to the Nuka area for comparative purposes.

### Chiswell Island

The ~17 by 28 km (161 km<sup>2</sup>) Chiswell Island study area is located ~50 km southwest of Seward (Figures 1 and 2). The Chiswell Islands are granitic and, typical of the outer coast, contain numerous steep rocky walls and submerged rock piles (Wilson and Hults 2012). The depth within the study area ranged from 0 to 303 m, but only depths between 15 and 150 m were used in the analysis because only this depth range was surveyed with the ROV.

### Nuka Island

The Nuka Island area is ~10 by 20 km (96 km<sup>2</sup>), ~50 km west of the Chiswell area and ~60 km ESE of Homer (Figures 1 and 10). In contrast to the granitic bedrock of the Chiswell area, the bedrock in the Nuka area is metasedimentary. The study area contains areas of relatively shallow rocky relief separated by deep roughly parallel mud and sand filled troughs. The depth ranged from 0 to 250 m, but similar to the Chiswell area, areas shallower than 15 m and deeper than 150 m were not surveyed by the ROV and were excluded from the analysis.

### **Data Acquisition**

All field sampling was completed prior to and independent from the current study.

### ROV Video Surveys

ADF&G surveyed the Chiswell Islands for rockfish using a Deep Ocean Engineering, Phantom HD 2+2, ROV in 2004 and 2005 (Byerly et al. 2015). ROV position was determined using a Tracklink 1500MA Ultra Short Baseline (USBL) acoustic tracking system coupled to dGPS enabled Trimble AG132 receiver, Furuno SC-60 GPS compass and Applied Geomechanics, MD900-TW pitch/roll sensor. Approximately 69 transects, each 500 m long (~29 km in total), were surveyed with video. Yelloweye were observed at 164 points along these transects.

The Nuka ROV survey was completed in 2009 using the same equipment as was used in the Chiswell area except that a Kongsberg MRUD replaced the pitch/roll sensor. Approximately 82 transects, each 300 m long (~25 km in total), were surveyed. Yelloweye were observed at 169 points.

### Bathymetric Surveys

The Chiswell bathymetry was compiled from two separate hydrographic surveys. The area north of Lone Rock (59° 34.18'N) was surveyed by NOAA in 2000 using a Reson Seabat 8101 (240 kHz) MBES integrated with an Applanix POS/MV pitch/roll sensor and CSI MBX-3 dGPS receiver (data available://www.ngdc.noaa.gov). The area south of Lone rock was surveyed by Golder Associates, Inc. under ADF&G contract in 2006 using a Reson 8124 (600 kHz) MBES integrated with a pitch/roll sensor and Trimble Ag 132 dGPS receiver (Byerly et al. 2007). A digital elevation model (DEM) with 3 m horizontal resolution was mosaicked from these multibeam data.

The Nuka area was surveyed by ADF&G and Terrasond, Inc. in 2008, using a Reson Seabat 7125 (400 kHz) MBES, Applanix POS/MV pitch/roll sensor, and base station corrected GPS using Trimble 5700 receivers. A 3 m DEM was created from these survey data.

## **Analysis**

### Terrain Variables

Four types of terrain variables were derived from the depth rasters: Bathymetric position index (BPI), rugosity, slope, and distance-to-rock (DTR). All terrain variables related to the surface morphology of the seafloor; backscatter information from the MBES surveys was not included. A variety of scales were considered for each type of terrain variable, because fish associate with the seafloor at a variety of scales (Wilson et al. 2007; Anderson and Yoklavich

2007; Monk et al. 2011). BPI, rugosity, and DTR were each calculated at four different scales, corresponding to different sizes of neighborhoods used to calculate each of the metrics for a given cell (Table 1; Figures 3–7,11–15). Since depth and slope can be calculated without consideration of surrounding cells, only one scale was considered. The BPI and rugosity variables were calculated from the DEM using the Benthic Terrain Modeler (Wright et al. 2005).

Rugosity was calculated as the vector rugosity measure (VRM) and is a metric of the variance in three-dimensional orientation of vectors orthogonal to the surface of the cells (Sappington et al. 2007).

BPI is the difference in depth between a given cell and the mean depth of the cells in the surrounding neighborhood. It is used to distinguish ridgetops which have positive BPI values from valley bottoms which have negative BPI (Weiss 2001). A BPI value near 0 may be either nearly flat or midslope.

DTR was calculated as the Euclidean distance (m) to the nearest cell with a VRM value greater than a threshold value of 0.001 for DTR 7, 5 and 3, and 0.020 for DTR 21. Young et al. (2010) used a VRM threshold of 0.001 to distinguish soft sediment from rock. The 0.001 value, while somewhat arbitrary, appeared reasonable within both of the current study areas for the three finer scales of DTR. For DTR 21 however, the VRM threshold was increased to 0.020 because this better distinguished rocky areas from soft sediment.

#### *Presence and Absence Points*

A split sample method was used with 70% of the presence points applied to fit the model and the remaining 30% reserved to evaluate the accuracy of the model, except in the case of the Chiswell model being applied to the Nuka area. The latter case allowed all of the Nuka presence

and absence points to be used in the accuracy assessment, because none of these points were used to fit the Chiswell model.

### Ecological Niche Factor Analysis (ENFA)

Selecting absence points was more complex than presence points owing to imperfect detectability and false absences. To reduce the number of false absences, absence points were only selected from areas along transects that were identified by a preliminary ENFA as poor habitat, following Young et al.'s (2010) adaptation of the Engler et al. (2004) method of selecting pseudo-absences.

Ecological niche factor analysis (ENFA) is a method of modeling habitat distributions that does not require absence points. The method compares the n-dimensional space occupied by the species along n-environmental gradients to the multidimensional characteristics of the background or greater study area. This approach is similar to other multidimensional variable reduction techniques, such as principle component analysis (PCA), in combining multiple collinear predictor variables into a few 'super' variables or factors that account for the majority of the variation in the environmental data based on eigenvectors of predictor variable covariance matrices. However, unlike PCA where the factors are oriented orthogonal to one another, in ENFA the factors are constructed such that they are given easily interpreted ecological meaning. In ENFA, the first factor is termed the *marginality factor* which captures how different the occupied niche is from the totality of available environmental conditions. Subsequent factors are *specialization factors* which describe the breadth of the occupied niche.

ENFA's chief advantage over more traditional presence/absence habitat modeling techniques such as GLMs and GAMs is in avoiding the problem of false absences by relying

only on presence points. The main disadvantage of ENFA is a tendency to overestimate the amount of suitable habitat (Engler et al. 2004).

As part of the ENFA, in addition to the marginality and specialization *factors*, overall marginality and specialization *values* were calculated. Marginality is a measure of how different the mean of the species frequency distribution ( $\mu_S$ ) across an environmental gradient is from the global or greater study area mean ( $\mu_G$ ), standardized by the standard deviation of the global distribution ( $\sigma_G$ ) (Hirzel et al. 2002):

$$M = \frac{|\mu_G - \mu_S|}{1.96\sigma_G} \quad (1)$$

Specialization is defined as the ratio of the standard deviation of the global distribution ( $\sigma_G$ ) to the standard deviation ( $\sigma_S$ ) of the focal species:

$$S = \frac{\sigma_G}{\sigma_S} \quad (2)$$

In practice, both marginality and specialization are calculated over multiple dimensions; the univariate definitions presented above are for conceptual explanatory purposes. For most species, marginality ranges from 0 to 1, with large values indicating a large difference in conditions between where the species is found and the average in the study area. The raw specialization value is somewhat difficult to interpret since it ranges from 1 to  $\infty$ , so is often expressed as its inverse, tolerance. Tolerance ranges from 0 to 1, with 0 indicating a very specialized or stenocious species, and 1 indicating a species tolerant to a wide variety of environments.

The ENFA in this study used the Biomapper version 4.0 software (Hirzel et al. 2007). All environmental rasters as well as the training set of presence points were converted to the IDRISI RST format required by Biomapper, while taking care to properly mask and co-register the layers so that they covered the exact same extent and all cells were perfectly aligned.



All environmental layers were normalized with the Box-Cox transformation (Box and Cox 1964). The Box-Cox transformation was used for all variables except where the Box-Cox transformed version caused terminal errors in the ENFA algorithm (due to discontinuous or very large values), in which case the raw, non Box-Cox transformed raster was used, following Hirzel et al. (2002). Although normality is theoretically desirable for extracting factors based on eigenvectors, empirically the ENFA algorithm is fairly robust to non-normality (Hirzel et al. 2002). For several layers, both the Box-Cox transformation and the raw variable caused terminal errors. These variables were excluded from the analysis. The broken stick method, with extreme optima, the harmonic mean algorithm, and 10 cross validations were used for both areas.

To convert the continuous habitat suitability raster from the ENFA (range 0 to 100, with higher values being more suitable) to a binomial suitable/unsuitable map, a habitat suitability score threshold of 3 was used for Chiswell and 12 for Nuka. These values were chosen such that 95% of training presence points were classified as suitable habitat.

Although Engler et al. (2004) created pseudo-absences by selecting points from among *all areas* where the organism was not detected, the surveyed transects within the current study areas were extensive enough that pseudo-absences were selected *only from areas that were surveyed* with the ROV and where yelloweye were not detected, following the methods of Young et al. (2010). Along this subset of the transects, the absence points were selected randomly. As with the presence points, 70% of the absence points in the Chiswell area were used to fit the model while the remaining 30% were reserved for an accuracy assessment.

Three performance indices were calculated to evaluate the habitat suitability score output from the ENFA (Hirzel et al. 2006). The absolute validation index ( $0 < AVI < 1$ ) is the proportion of presence points with a suitability score  $> 50$  and indicates how well the model

discriminates highly suitable from unsuitable areas. To account for chance agreement, the contrast validation index ( $0 < CVI < AVI$ ) was calculated by subtracting the AVI expected from a null model that would predict suitability at random. The Boyce Index is less dependent on a particular threshold than AVI and CVI, and can range from -1 to 1, with 0 expected from a chance model and 1 a perfect model.

### Variable Selection

The significance of each type of terrain variable at each scale for each study area in predicting yelloweye presence or absence was determined using simple logistic regression and the Wald test of significance. Because a local optimum was suspected for depth and BPI, a quadratic transformation of each of these variables was included.

The predictive power of the different scales of terrain variables were ranked for each type of terrain variable within each study area based on differences in AIC scores (Burnham and Anderson 2004). Collinearity among scales of a given type of terrain variable were examined with correlograms based on Pearson's correlation coefficient. Because most scales of a given type of variable were correlated, and because including multiple collinear 'independent' variables in the same model can cause overfitting, exaggerate significance, and even reverse the sign of a coefficient, only the most predictive scale(s) of each type of variable was included in the scope of the final variable selection process.

Although it is common practice to include the linear term with a quadratic response term, so as not to overly constrain the shape of the response curve, exploratory plots of BPI30 vs. yelloweye presence/absence were relatively symmetric about the y-axis, so the linear term was excluded.

Collinearities among the best scales of each type of predictor variable were also examined with correlograms.

The final variable selection involved a forward stepping AIC analysis, in which the variables are added one at a time and the resulting AIC scores are compared, using the MASS package in R version 3.1.0 (Venables and Ripley 2010; R Core Team 2014). Although both study areas informed the choice of the best scale of each type of terrain variable, in an effort to maintain the independence of the Nuka area as a test of the portability of the Chiswell model, the final variable selection — from among the best scale of each type of variable — used only the Chiswell data.

#### Generalized Linear Model (GLM)

The final habitat suitability model took the form of GLM using a binomial logistic link function:

$$\text{Logit}(P) = X_1B_1 + X_2B_2 \dots X_iB_i + \alpha \quad (3)$$

where  $P$  is the probability of suitable yelloweye habitat in a given cell,  $X_i$  is the value of terrain variable  $i$  in that cell,  $B_i$  is the coefficient of that terrain variable, and  $\alpha$  is an intercept. *Logit* is the logistic link function. It both linearizes binomial logistic data and constrains the probability between 0 and 1:

$$\text{Logit}(P) = \ln \left( \frac{P}{1-P} \right) \quad (4)$$

The GLM was fit using maximum likelihood estimation, and was parametrized twice, to create two versions of the GLM. First the GLM was fit to the Chiswell area, which was the focus of this study. After the accuracy of the Chiswell GLM was evaluated — both within the Chiswell area, and then in the Nuka area — the model was refit to the Nuka area and reevaluated solely in the Nuka area to compare changes in performance and parameter weighting.

### Accuracy Assessment

The accuracy of each of the models was evaluated by producing confusion matrices and calculating the percentage of suitable points correctly classified as suitable and percentage of unsuitable points correctly classified as unsuitable. Overall accuracy was calculated as the percentage of all the ground truth points correctly classified. The ground truth points were the set of suitable and unsuitable habitat points based on the ROV observations. To more closely examine the accuracies of the individual classes (suitable and unsuitable), producer and user accuracies were calculated. These widely used measures of remotely sensed classification accuracy differ in their denominators. Producer's accuracy is the percentage of all the pixels assigned to a particular class that were classified correctly, while user's accuracy is the percentage of all the ground truth points of a particular class that were classified correctly. Sensitivity is the producer's accuracy for presence points, while specificity is the producer's accuracy for absence points (Fielding and Bell 1997).

Additionally, Cohen's Kappa was calculated for each model using the irr package (Cohen 1960; Gamer et al. 2012). Like percent agreement, Cohen's Kappa is based on a confusion matrix created using a fixed threshold, but is a more stringent test of the performance of a classification model because it accounts for chance agreement.

Finally, a receiver operator characteristics (ROC) analysis was completed for each of the various model and study area combinations using the ROCR package (Fielding and Bell 1997; Sing et al. 2007). ROC curves were plotted and the area-under-the-curve (AUC) scores were calculated for each model and area. ROC plots are created by plotting the true positive fraction (sensitivity) against the false positive fraction (1 - specificity) at various thresholds. Possible AUC scores range from 0 to 1, with 0.5 being expected from a completely random classification

and 1.0 indicating a perfect classification with no false positives. The AUC is the probability that a randomly chosen suitable point would have a higher probability of being suitable than a randomly chosen unsuitable point. ROC curves are useful for evaluating the performance of classification models that output continuous responses, because unlike the confusion matrix based measures, they do not require that the response first be binomially reduced (Pearce and Ferrier 2000). Thus, they evaluate the performance of a classification model independent of any specific threshold. The ROC plots were also used to select appropriate thresholds to distinguish suitable from unsuitable habitat.

All three of these performance metrics (percent agreement, Kappa, and AUC) were calculated for: (1) the Chiswell model against the reserved validation set of the Chiswell presence and absence points; (2) the Chiswell model applied to the Nuka area using all of the Nuka points; and (3) the Nuka model applied to the Nuka area using the reserved validation points.

## 2.3 RESULTS

### Distributions Across Univariate Gradients

Yelloweye presence was significantly related to each scale of each of the investigated terrain variables (simple logistic regression; Wald test of significance;  $p < 0.001$ ) (Table 2).

Although the linear versions of BPI30 for both areas and BPI60 for the Chiswell area were not significant predictors of yelloweye presence, the quadratic versions of both of these variables were significant, indicating dome-shaped responses, or local rather than extreme optima.

Specifically, yelloweye were observed more: in (VRM) and near (DTR) rugose areas; steep areas, shallow areas; areas with positive large scale BPI; and areas of either positive or negative, but not neutral, small scale BPI (Figures 24 and 25).

## Variable Selection

For each type of terrain variable, all scales were strongly correlated (Pearson's correlation coefficient  $> 0.5$ ), except the smallest scale of BPI, BPI30, was only weakly correlated with the largest scale, BPI240 ( $r < 0.45$ ) (Figures 19–21).

A scale factor of 7 was chosen as the best scale of VRM for inclusion in the scope of the final variable selection process, based on  $\Delta AIC$  values of the single variable models (VRM7; Table 2). For simplicity and to avoid relying on rugosity calculated at two different scales, DTR7 was included in the final scope because this was the DTR scale corresponding to the best VRM scale (VRM7), even though DTR3 and 5 were stronger predictors (Table 2). The two most predictive scales of BPI were the linear version of BPI240 and the quadratic version of BPI30. The quadratic and linear version of the depth term had similar predictive power, so both were considered in the scope of the final stepwise AIC variable selection.

To summarize, the scope of variables considered in the final stepwise AIC variable selection process were VRM7, BPI240, BPI30<sup>2</sup>, DTR7, Depth, Depth<sup>2</sup>, and Slope (Figures 22–25). This was determined by first examining collinearity among various scales within each type of terrain variable, then selecting the best weakly-correlated scale(s) of variable(s) from each type, based on AIC scores for the univariate models.

The stepwise AIC process found the most parsimonious model for suitable yelloweye habitat in the Chiswell area included VRM7, DTR7, Slope and BPI240 (Table 3).

The ROC curve for the Chiswell area suggested a probability value of 0.5 as the best threshold for distinguishing unsuitable from suitable habitat (Figure 26). A threshold of 0.5 to 0.7 is often used for these types of GLM habitat models (Hirzel and Guisan 2002).

## Accuracy Assessment

### Chiswell GLM in Chiswell Area

Using a probability threshold of 0.5, the GLM fit to the Chiswell training points evaluated against the Chiswell validation points ( $n = 100$ ) yielded an overall accuracy of 96% (Table 4; Figure 9). Both the producer's and user's accuracies for both presence and absence were also all 96%, yielding a significant ( $p < 0.001$ ) Cohen's Kappa of 0.92. A Kappa value  $> 0.75$  indicates "excellent agreement" (Landis and Koch 1977). The AUC was 0.997, with AUC  $> 0.9$  indicating "outstanding" discrimination (Hosmer and Lemeshow 2004).

### Chiswell GLM in Nuka Area

When the GLM as parametrized in the Chiswell area was applied to the Nuka area and evaluated against the entire set of presence and absence points in the Nuka area ( $n = 332$ ), the overall accuracy dropped to 82.5%, comprised of a producer's accuracy for presence points of 95.7% and 69.3% for absence (Table 5; Figure 17). Cohen's Kappa for this classification was significant ( $p < 0.001$ ) at 0.65. A Kappa value between 0.40 and 0.75 has been interpreted as "good agreement" (Landis and Koch 1977). The AUC was 0.952.

### Nuka GLM in Nuka Area

When the GLM was reparametrized to the Nuka area, the accuracy of the model evaluated against the validation set of Nuka points, while retaining the same set of four predictor variables, increased as compared to the GLM fit to the Chiswell area (Tables 5 and 6; Figure 18). The overall accuracy increased to 89.0% comprised of a producer's accuracy of 88.0% for presence and 90.0% for absence. Cohen's Kappa increased to 0.78. In contrast to the accuracy metrics based on the fixed threshold, the threshold-independent AUC of the reparametrized Nuka

GLM evaluated against the validation Nuka points was very similar (AUC = 0.953) to the AUC of the Chiswell GLM tested against the entire set of Nuka presence and absence points..

## **ENFA**

In both the Chiswell and Nuka areas, the marginality values (2.25 and 1.81 respectively) indicated that the terrain where yelloweye were observed was much different than the average terrain in each study areas (Tables 7 and 8). The tolerance values (0.52 and 0.58) indicated moderate specialization. The coefficients of the marginality factor in both areas indicated that yelloweye were observed in more rugose, near rugose, steeper and shallower areas than the average in each area. The marginality and specialization can be seen graphically by comparing the means and dispersion of frequency distributions of yelloweye occurrence relative to the greater study areas along the marginality and 1<sup>st</sup> specialization factors (Figures 27 and 28).

The composition of the marginality factor was nearly identical between areas, comprised mostly of VRM21 and with the ranked relative contribution of the individual variables identical except that the ranking of BPI60 and BPI240 was reversed (Tables 7 and 8). The specialization factors were comprised substantially of DTR in both areas. However the remaining composition of the specialization factors differed between areas, primarily in that BPI was more important in the Nuka area. The marginality factor combined with the first 3 specialization factors accounted for > 75% of the information in the training set of presence points in each area. Finally, the percentage of information accounted for by each factor was similar between areas.

As measures of performance, in the Chiswell area, the AVI (0.50), CVI (0.43) and Boyce index (0.64) all indicate the habitat suitability score from the ENFA was a good discriminator of suitable from unsuitable habitat. In Nuka area the AVI (0.30), CVI (0.30) and Boyce Index (0.44) indicate fair discrimination.



## 2.4 DISCUSSION

This study demonstrated that the distribution of yelloweye habitat on the outer coast of Alaska's Kenai Peninsula can be modeled with reasonable accuracy using several terrain variables derived from high-resolution MBES bathymetry.

Specifically, to revisit the primary objectives: (1) The distribution of yelloweye habitat was significantly related to each of the investigated terrain variables. (2) The most parsimonious combination of predictors for yelloweye habitat in the Chiswell area included a moderate scale of (i) rugosity and (ii) distance-to-rock, (iii) a broad scale BPI, and (iv) slope. (3) A GLM combining these variables was an excellent predictor of yelloweye habitat in the Chiswell training area. (4) This habitat model was fairly robust across study areas.

### **Comparison to Previously Published Models**

A previous study to predict presence/absence of rosy, yellowtail, and greenstriped rockfish in the Cordel Bank Marine Sanctuary (CBMS) off California reported overall accuracies of 96, 92 and 92% with Kappas of 0.89, 0.71, 0.62, respectively (Iampietro et al. 2008; Young et al. 2010). These accuracies are similar to the 96% overall accuracy and Kappa of 0.92 for the Chiswell model when evaluated in the area it was fit.

The accuracy of the Chiswell model tested in the independent Nuka area was also similar to the previously published accuracy of one of the two CBMS models when evaluated in an independent study area. Iampietro et al. (2008) found the CBMS model for greenstriped to be 71% accurate (Kappa = 0.42) when evaluated at the independent Del Monte shalebeds (DMSB) off California. In contrast however, their CBMS model for rosy was unsuccessful at predicting habitat in the DMSB, classifying the entire DMSB study area as habitat. The authors attributed the poor performance of the rosy model in the DMSB to the different depth ranges of the study

areas and the influence of depth in the rosy model. For comparison, when the Chiswell model was applied to the Nuka area the accuracy was 82.5% (Kappa = 0.65). To summarize, the Chiswell model performed slightly better than the best of the Californian models, both when evaluated within the area it was fit and when fit to the independent Nuka area.

One main difference among studies was the inclusion of depth as an explanatory variable in the California models; depth was considered for the Chiswell and Nuka models but excluded by the stepwise AIC process. This could explain why the Chiswell model was more portable across areas than the California models. While demersal rockfishes do exhibit preferred depth ranges (Richards 1986; Johnson et al. 2003; Rooper 2008), it could be that the rosy habitat preference observed by Young et al. (2010) was more directly linked to another variable closely correlated with depth, such as the depth distribution of rugose rock outcrops. Perhaps, since depth was only related to the distribution of rosy in CBMS tangentially, the apparent depth preference observed there, did not hold true in the DMSB.

Monk et al. (2012) reported AUCs ranging from 0.54 to 0.96 for GLMs predicating the distribution of nine reef fishes off Australia based on several MBES terrain variables (including rugosity, BPI, depth, and distance-to-reef). The AUC's of their best performing models compare similarly to the AUC's of both the Nuka and Chiswell GLMs.

### **ROC Analysis**

The ROC analysis indicated excellent performance of the GLM habitat models developed for the North Gulf District, both within the areas they were fit and when applied to different study areas. As expected, the Chiswell model performed better in the Chiswell area where it was fit, than in the independent Nuka testing area (AUC = 0.987 vs 0.952).

Unexpectedly, refitting the GLM to the Nuka area did not substantially improve the AUC over that from the Chiswell model (0.952 for Chiswell model in Nuka area vs. 0.953 for Nuka in Nuka). However, this is confounded by the different sets of validation points used. All Nuka presence/absence points were used to evaluate the Chiswell model applied to the Nuka area since none of these were used in fitting the model, while the training points were excluded from the ROC evaluation of the GLM refit to the Nuka area. Although it is somewhat circular to evaluate the performance of a model using the same set of points used to fit it, the ROC plot for the Nuka model in the Nuka area using the training points is included as ancillary support of the earlier hypothesis that the unexpected lack of increase in AUC accompanying the refit of the GLM in the NUKA area likely reflects the different points used in the evaluation, and the relatively small sample size for the evaluation points. As expected, when the model was refit to the Nuka Area and evaluated using the validation points, the AUC increased (0.960), similar to that of the Chiswell model when evaluated using the Chiswell validation points (AUC = 0.987).

### **Direct vs. Indirect Environmental Variables in Habitat Models**

It is important to acknowledge that none of the investigated terrain variables are ‘directly’ driving the distribution of yelloweye. Guisan and Zimmermann (2000) discussed the types of variables used in habitat models. Ideally, habitat models rely on what they denoted as *resource* or *direct* variables which both have physiological significance to the focal species (e.g., prey density, water temperature, and salinity). In contrast, the less desirable *indirect* variables (e.g. BPI, VRM, and slope) do not present the same physiological significance. Unfortunately, resource and direct variables are the most expensive and difficult to obtain over the large areas required by habitat models, often requiring in situ measurements. Indirect variables in contrast can often be remotely sensed. Consequently many habitat models, including the current study,

are often obliged to rely on the less desirable indirect variables. The main disadvantage of relying on indirect variables as opposed to direct or resources variables is that because they are, by definition, not fundamentally directly driving the distribution of the species, a model relying on indirect variables will likely have limited portability and robustness.

All of the terrain (indirect) variables used in the present habitat model likely gain their significant relationship with yelloweye distribution only indirectly through their collinearity with either the resource variable availability of prey, or the direct variable availability of refuge. Food resources are concentrated near rock outcroppings by currents. The complex three-dimensional complexity of rugose rocky areas increases the availability of crevices for use as refuge from predation and enhanced ambush feeding opportunities (Yoklavich et al. 1999; Greene et al. 2011). The void to clast ratio of the substrate appears to be especially important to yelloweye distribution (O'Connell and Carlile 1993). BPI probably gains most of its importance as an indicator of positive topography capable of deflecting currents and concentrating prey. The rugosity based VRM and DTR variables on the other hand probably gain most of their importance from their relationship to the availability of refuge. Slope may be important because it is related to both prey and refuge availability.

### **Critiques**

The most substantial critique of the current study is its accuracy assessment. Since the absence points used in the assessment were only selected from those areas classified by the ENFA as unsuitable habitat, to some degree, the accuracy assessment is comparing the results of the ENFA to the GLM. This critique applies to the published models of Young et al. (2010) as well. The problem is rooted in the unreliable absence issue, common to all presence/absence habitat models. Selecting the absence points based on the ENFA was done in good faith, in an

effort to reduce false absences. Alternatively, if the absence points were not filtered by the ENFA, false absences likely would be selected which would induce their own biases in the accuracy assessment. The problem of unreliable absences will plague any presence/absence habitat model.

Kéry (2010) describes specific problems caused by imperfect detectability including underestimation of habitat, and confusing covariates of probability of detection with covariates of probability of occurrence. A strong argument is provided for modeling species distribution with site occupancy models, where occupancy is modeled as the product of detection and occurrence. A site occupancy model was not used here, because of the need for repeat sampling. Also, site occupancy models do not ameliorate the more operative type of false absences in this study, those absences related to the relative rarity of yelloweye and the fine scale of the observation unit. It would be unreasonable to expect a relatively rare species such as yelloweye to occupy every 3 m cell of suitable habitat. The most promising approach to dealing with this type of false absence is through a presence-only type of model, such as the ENFA used in the first step of the current analysis.

### **Limitations**

The biggest limitation for the real world application of the habitat model presented here to management is the model's reliance on MBES data. The most useful rockfish habitat model would be one capable of distinguishing habitats using only single beam bathymetry. Even though the yelloweye model developed here displays at least some degree of portability between study areas, its real world application is limited because multibeam data are only available for some, as yet relatively small, portions of the coast, whereas single beam data are available for the entire coast of Alaska. Realizing that complete coastwide full coverage multibeam data will not

be achieved for a long time (despite an annual budget of \$12 million), Elvenes et al. (2014) investigated the feasibility of modeling the distribution of surficial sediment and benthic biotopes in Norway using single beam data and compared the resultant maps to those obtained from multibeam sonar. Results were encouraging, with the single beam data yielding habitat maps very similar to maps from multibeam data. However, the data density of the single beam bathymetry appears to be greater than what is available for the North Gulf District, and mapping was also at a coarser scale. The best available single beam bathymetry readily available for the central Gulf of Alaska is Zimmermann and Prescott's (2015) layer (100 m horizontal resolution) based on the digitized and corrected smooth sheet soundings used to create NOAA navigational charts. Elvenes et al. (2014) on the other hand used Olex-derived single beam bathymetry in a heavily trafficked area. Olex is a crowdsourced system for compiling and sharing bathymetry collected by ordinary (non survey) vessels during everyday operations. Although some vessels do use the Olex system in Alaska, the participation rate is likely greater in Norway where the system was developed. However, Olex bathymetry in the North Gulf District should be investigated more thoroughly, specifically as a potential data source for a habitat model. The Kenai Fjord Tours fleet of commercial tour boats use the system, so perhaps Olex would provide a valuable bathymetric data source, at least for some of the more heavily trafficked areas.

When relying on rugosity to predict probability of occurrence, one must be attentive to survey data quality and the smoothing algorithms applied to it during post processing. Of all the investigated terrain variables, rugosity is likely the most negatively affected by poor bathymetric survey data quality. The bathymetry in the Nuka area had more survey artifacts such as roll artifacts, which appear as lines perpendicular to the transect orientation, and in general was noisier with many points of either null or erroneous data. These survey artifacts inflated the

calculated rugosity, which in turn inflated the predicted probability of suitable habitat in those areas. Not surprisingly, when the GLM was refit to the Nuka area, the relative weight of VRM and DTR decreased while the other variables, especially BPI, increased. Apparently in areas with poor quality bathymetry, BPI gains importance as a predictor of habitat as compared to the rugosity-based VRM and DTR.

### **Applications**

Despite the limitations inherent to its reliance on MBES data, the model presented here does have real world value to improving the management of DSR in Alaska. More specifically and foremost, the GLM could be used to classify both existing and anticipated MBES bathymetry into suitable and unsuitable DSR habitat.

While most of the existing MBES data in the North Gulf District have already been classified, large areas of unclassified MBES data exist in the adjacent Cook Inlet and Prince William Sound (PWS) districts. The one existing MBES dataset in the North Gulf District that has not yet been classified is the course grid spanning the western portion of the district (Figure 1). An algorithmic method as presented here could be used to classify these data, which will be key to interpreting the surrounding area where only single beam data are available.

Undoubtedly in the future, more of the North Gulf District will be surveyed with MBES. Collecting coastwide multibeam data has been identified as a priority of many nations (Elvenes et al. 2014). More locally, as evidence of the increasing availability of multibeam, the United States Geological Survey (USGS) recently signed a five-year cooperative agreement with ADF&G to jointly collect multibeam bathymetry along the Northern Gulf of Alaska. As part of this effort, 600 km<sup>2</sup> were surveyed during 2014 in the PWS district near Chenega and Cape Cleare, and 900 km<sup>2</sup> were surveyed in 2015 in the Southeast Region near Cape Spencer. None

of these areas have yet been classified as suitable or unsuitable DSR habitat. The model presented here could be used to stratify these anticipated MBES data as they become available, thereby improving the habitat area estimates used in the Southeast Region based on the best available data, and towards providing an initial estimate of habitat in Central Region.

In addition to the MBES data that have not yet been classified, existing habitat delineations in multibeam areas could be revised using the algorithmic method presented here. Previous classifications involved either manual expert interpretation in the Southeast Region (Greene et al. 1999); or semianalytic methods, based on a combination of slope, rugosity, depth and manual interpretation in the Central Region. A purely objective, algorithmic classification would reduce potential bias and improve standardization across management areas.

## **ENFA**

Although the ENFA was conducted primarily as a means of reducing false absences for use by the GLM and was not intended to be the focus of the current study, some comments about it are warranted. The habitat suitability model derived from the ENFA could conceivably be used as a stand-alone habitat model (e.g., Hirzel et al. 2001; Leverette 2005; Galparsoro et al. 2009; Monk et al. 2010, 2011). Indeed, the habitat suitability maps from the ENFA appear to do a reasonable job of classifying rock outcrops as suitable habitat and the flat areas as unsuitable (Figures 8 and 16), and the performance measures (AVI, CVI, and Boyce Index) obtained from it provide some measure of validity.

The main reasons the ENFA was not used as the primary habitat model in this study were: (1) I wanted to first test the methods developed for the CBMS and the DMSB (Iampietro et al. 2008; Young et al. 2010) as a baseline, before delving into the multitude of different modeling methods; (2) previous work warned that ENFA habitat models tend to be overly



inclusive (Engler et al 2004); (3) evaluating the accuracy of presence-only models like ENFA is more complex than classic regression-based presence/absence models such as GLMs.

The ENFA for both the Chiswell and Nuka areas indicated that yelloweye are a highly marginal species, with the occupied niche dramatically different from the average conditions available in an area, and also a stenoeious species in regards to the investigated terrain variables. The strikingly similar composition between areas of the marginality factor, and also the specialization factors, indicate that the observed relationships between the distribution of yelloweye and each of the investigated terrain metrics is consistent between areas, thereby suggesting a general ecological relationship rather than statistical coincidence.

### **The Importance of Scale**

The choice of scale is important to any habitat model. While the importance of pixel size (grain) used in the analysis is also well established, the focus here is on the importance of the size of the neighborhood used to calculate each of the individual terrain metrics (Wilson et al. 2007). Although all the single variables except the linear version of BPI30 were significant when considered independently, the relative predictive power of variables depended greatly on size of neighborhood used in the calculation, as evidenced by the large difference in AIC values among model configurations. Furthermore, the functional response varied depending on the scale of analysis. Particularly interesting is how BPI is monotonically related to yelloweye presence when calculated using a large neighborhood, yet when a small neighborhood is used the response curve becomes dome-shaped, with a local minimum centered about neutral BPI values of 0 (Figures 24 and 25). This indicates that when considered on a larger scale, yelloweye prefer to be up near the tops of broad swales and mounds, yet when considered at a finer scale, yelloweye are found both in the bottoms of small localized depression and on the tops of small

localized ridges, but not as commonly on flat or midslope locations. The large scale BPI likely is effective at distinguishing large scale hard bottom areas of positive relief from soft bottom depressions, and the small scale BPI is probably more effective at pinpointing those smaller scale structurally complex areas, which manifest as both positive and negative topography. As a related alternate explanation, this could be also due to differences in how yelloweye settle (large scale choice) and in how they choose to forage (small scale choice).

For rugosity, the larger scales of VRM performed better than the smaller scales of VRM except that in the NUKA area, VRM21 (the largest scale considered) performed poorly. The converse is true for DTR, with the smaller scales generally performing better than the larger scales. This could be attributed to either ecology or survey positional inaccuracies. As an ecological explanation, perhaps yelloweye need not actually be in rugose areas so long as they are near rugose areas, with being near rugose areas captured best by both the large scale VRM, with its large effective search radius, and the small scale DTR which inherently buffers rugose areas. Alternatively, perhaps this theoretical disjunction between smaller scale rugose features and yelloweye presence was more related to positional imprecision associated with the ROV survey. The 3 m cell size certainly pushes the limit of, and likely often exceeds, the precision of the acoustic USBL tracking system used to locate the ROV.

For BPI, both the quadratic transformations of the small scale versions and the linear versions of the large scale BPIs performed well.

These findings that the value of terrain variables in predicting habitat depends on the scale of analysis (neighborhood size) are consistent with other marine benthic habitat modeling studies (Wilson et al. 2007; Galparsoro et al. 2009; Monk et al. 2011).

## **Distributions Across the Univariate Gradients**

The relationship between yelloweye presence and each of the six best scales of the individual terrain variables were largely as expected and similar between study areas (Figures 24 and 25). The preference for shallow areas was not entirely anticipated and may be caused by collinearity between depth and some other more important driver of yelloweye distribution, e.g. the depth distribution of rugose rock outcroppings, as discussed earlier. For example, the relationship between depth and yelloweye distribution seen here may be indicative of the rugose rocky outcroppings tending to be in shallow areas, rather than depth itself being the operative driver of yelloweye distribution.

## **Future Research**

Although the most important avenue for future research is investigating the feasibility of distinguishing rockfish habitat using only single beam data, if one were to remain reliant on multibeam data, the most promising way to improve the current model would be to incorporate the backscatter intensity as one of the independent variables. Acoustic characteristics, chiefly the intensity of the return sonar signal, have been widely demonstrated to be closely linked to the physical composition (via acoustic reflectivity or hardness) of the seafloor, which in turn is significantly related to yelloweye densities (O'Connell and Carlile 1994; Ferrini and Flood 2006; Brown and Blondel 2009; Parnum et al. 2009; Brown et al. 2011).

Other potential explanatory variables are modeled or measured bottom current velocities (especially in conjunction with aspect of the seafloor) — although these are likely to be of much coarser resolution than the terrain variables used here — an ocean exposure index, or additional terrain metrics such as curvature, complexity and aspect (Ardron 2002; Burrows et al. 2008; Rooper 2014; Zimmermann *In press*). Most recently, Zimmermann (*In press*) demonstrated

how the substrate information interpolated from National Ocean Service (NOS) smooth sheets could be used in simplistic conditional habitat suitability maps for juvenile flatfish, including two study areas in the North Gulf District (Port Dick and Aialik Bay). Substrate point data on smooth sheets, while ~17 times more dense than what is found on final published navigational charts, is still probably too coarse to sufficiently delineate habitat on its own. However, perhaps an interpolated substrate layer would enhance a purely morphologic model, such as the one presented here, as an additional predictor variable.

In addition to the independent variables, the dependent variable (probability of suitable habitat) could also be improved by using acoustic tags to get more information on the relative amounts of time fish spend in various habitats. This type of resource use habitat modeling is increasingly being used for terrestrial animals (e.g., Beus 2010).

Finally, apart from the variables, alternatives to the framework of the model itself could be explored. Although GLMs are extremely popular in habitat suitability modeling, use of GAMs has increased mostly because of their ability to accommodate a more diverse range of response surfaces (Guisan et al. 2002; Rooper et al. 2014; Sigler et al. 2015). Monk et al. (2012) found the choice of specific modeling technique (GAM, GLM, or MAXENT) did not considerably influence the distributions of nine demersal fishes predicted from multibeam variables. Instead they found the type of particular occurrence dataset used to be more influential, and echoed Kery's (2010) urge for care in interpreting the output of species distribution models that do not account for probability of detection. When Monk et al. (2010) compared a variety of presence-only modeling techniques (ENFA, MaxEnt, BIOCLIM, and DOMAIN) for five demersal fishes using MBES terrain variables, they found MaxEnt generally performed slightly better than ENFA based on Kappa and AUC.

## Conclusion

This study demonstrated that yelloweye habitat can be modeled with reasonable accuracy using several terrain variables derived from high-resolution multibeam bathymetry. Furthermore, such a model displayed fair portability between areas within the North Gulf District. However, the application of such a model is limited simply because the model is entirely reliant on multibeam bathymetry and multibeam bathymetry is currently only available for certain portions of the coast. Therefore, the most imperative avenue for future research is investigating the feasibility of modeling yelloweye habitat using the less desirable but more widely available bathymetry from either NOS smooth sheets or Olex. Unfortunately, given the data density of these alternate bathymetry sources, the feasibility of modeling yelloweye habitat based purely on the surface morphology of these data is uncertain. The most promising approach to modeling yelloweye habitat in these areas with only crude depth information is by incorporating an interpolated substrate composition layer as an additional explanatory variable. The output from such a model, supplemented perhaps with buffered rocky features on smooth sheets, will undoubtedly be rough, but offers the best chance of providing an interim estimate of the areal extent of yelloweye habitat in the district, thereby bridging the knowledge gap introduced earlier necessary for achieving the ultimate goal of a districtwide abundance estimate of yelloweye. Such an interim estimate of habitat could be revised as the single beam data is supplanted with additional multibeam data anticipated in the future.

## GENERAL DISCUSSION

The primary motivation for this project was the need for an areal expansion factor to use in a districtwide abundance estimate of yelloweye rockfish. Specifically, I explored whether yelloweye habitat can be accurately modeled using high-resolution bathymetry acquired using multibeam sonar, with validation through ROV observations. I concluded that yelloweye habitat can be accurately modeled using high-resolution bathymetry, and furthermore, such a model is fairly robust and portable between sites within the North Gulf District. However, given the current limited availability of multibeam bathymetry, a habitat model entirely reliant on these high-resolution data is not sufficient for deriving a districtwide estimate of habitat. Therefore, the most imperative avenue for future research is investigating the feasibility of modeling yelloweye habitat using the poorer quality but more widely available data from either the Olex crowdsourced compilation of single beam bathymetry, or the lead line and single beam bathymetry archived on NOS smooth sheets. In all likelihood, given the data density of these alternate bathymetric sources, a successful habitat model in these areas will likely require additional explanatory variables, most promisingly, a substrate composition layer interpolated from substrate point data on smooth sheets and other sources (e.g., Zimmermann *In press*).

In light of the uncertain feasibility and anticipated poor precision of habitat delineations based on this course alternative bathymetry, a deliberate discussion of how best to proceed with the overall management strategy for DSR in the Southcentral Region of Alaska is warranted. Rather than expand densities observed by the ROV up to an absolute estimate of abundance, maintaining the observed densities as a strictly relative measure of abundance may be best. After all, the ROV surveys will provide the same essential information signal, regardless of whatever particular areal expansion factor is used. Instead of scaling the densities up to an absolute

abundance estimate based on a grossly imprecise estimate of the total areal extent of habitat, it may be best to simply maintain the ROV observations as density.

In situations with many false absences, such as the current study owing to the relative rarity of the species and small scale of observational unit, a presence-only modeling approach may be more appropriate than the presence/absence approach used here. Several methods of reducing false absences have been developed to accommodate presence/absence models. Here, I used Young et al.'s (2010) adaptation of the Engler (2004) method of reducing false absences based on a preliminary presence-only ENFA. The problem with using this two-step approach in situations where false absences are prevalent is that by filtering the absences heavily with the ENFA, then using those absences to validate the GLM, a potential circularity of logic is risked, wherein the output of the ENFA is used as truth to validate the GLM. The absences did need to be filtered somehow however. If they were not, false absences in both the training and validation set of absences would induce bias and reduce model performance. In situations like this, where heavy filtering is required, it may be best to adhere to a presence-only model. That way the problematic requirement for reliable absences is alleviated altogether. In contrast, in situations with more reliable absences owing to either a larger observational unit (e.g., groundfish trawl survey tows, Rooper et al. 2014; Sigler et al. 2015) or more ubiquitous species such as many plants, presence-absence modeling techniques such as GLMs and GAMs are appropriate and perform well.

For the presence-only step of my analysis I used ENFA, which is attractive because its output is so easily interpreted in ecological terms. The ENFA appeared to do a reasonable job of distinguishing rugose rocky areas from flat soft sediment areas, and its performance could certainly be fine-tuned if it was intended as the focus of a study. However, MaxEnt is the more

popular profile technique used in recent literature, and has shown a slight performance advantage over ENFA when used to model the distribution of demersal fish based on terrain variables derived from MBES bathymetry (Monk et al. 2010). The problem with any of these presence-only techniques is that without absences, a comprehensive satisfactory evaluation of their performance remains elusive.



## REFERENCES

- Anderson, T. J., and M. M. Yoklavich. 2007. Multiscale habitat associations of deepwater demersal fishes off central California. *Fishery Bulletin* 105(2):168–179.
- Araujo M. and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33:1677–1688.
- Ardron, J. 2002. A GIS recipe for determining benthic complexity. Pages 169–175 *in* Breman, J., (Ed.), *Marine Geography, GIS for the Oceans and Seas*.
- Ardron, J., J. Lash, and D. Haggarty. 2002. Modeling a network of marine protected areas for the central coast of BC. Living Oceans Society, Sointula, BC, Canada.
- Ardron, J.A., and S. Wallace. 2005. Modelling inshore rockfish habitat in British Columbia: A pilot study. Pages 47–68 *in* D.Wright and A.Scholtz (Eds.), *Place Matters: Geospatial Tools for Marine Science*. University of Oregon Press, Eugene, OR.
- Bechtol, W.R. 1998. A synopsis of life history and assessment of Cook Inlet rockfish. Alaska Department of Fish and Game, Regional Information Report No. 2A98–40.
- Begon, M., C. Townsend, and J. Harper. 2006. *Ecology: From individuals to ecosystems*. Blackwell Publishing Malden. MA.
- Beus, T. 2010. Habitat modeling using path analysis: Delineating mountain goat habitat in the Washington Cascades. Master's thesis. Western Washington University.
- Box, G. E. & D. R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26:211-252.
- Brown, C. J., and P. Blondel. 2009. Developments in the application of multibeam sonar backscatter for seafloor habitat mapping. *Applied Acoustics* 70(10):1242–1247.
- Brown, C. J., S. J. Smith, P. Lawton, and J. T. Anderson. 2011. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine, Coastal and Shelf Science* 92(3):502-520.
- Burnham, K.P., and D.R. Anderson. 2004. Multimodal inference: understanding AIC and BIC in model selection. *Sociological Methods & Research* 33:261–304.
- Burrows, M. T., R. Harvey, and L. Robb. 2008. Wave exposure indices from digital coastlines and the prediction of rocky shore community structure. *Marine Ecology-Progress Series*-353:1.

- Busby, J.R., 1991. BIOCLIM – A bioclimate analysis and prediction system. Chapter 10 in Margules, C.R., and M.P. Austin (Eds.), Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne.
- Byerly, M., M. Spahn, and K. J. Goldman. 2015. Chiswell Ridge lingcod ROV survey with ancillary population estimates of demersal shelf rockfish, 2005. Alaska Department of Fish and Game, Fishery Data Series No. 15–26, Anchorage.
- Byerly, M., R. Sylwester, D. Aldrich, and M. Spahn. 2007. Chiswell ridge habitat mapping and groundfish assessment. North Pacific Research Board Final Report 616:2.
- Carlson, H. R., and R. Haight. 1972. Evidence for a home site and homing of adult yellowtail rockfish, *Sebastes flavidus*. Journal of the Fisheries Research Board of Canada 29(7):1011–1014.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20(1):37–46.
- DOC (Department of Commerce), 1997. Magnuson–Stevens Act Provisions: Essential Fish Habitat (EFH). Federal Register, vol. 62, issue 244, pp. 66531–66559.
- Drake, J., E. Berntson, R. Gustafson, E. Holmes, P. Levin, N. Tolimieri, R. Waples, S. M. Sogard, G.D. Williams, and J.M. Cope. 2010. Status review of five rockfish species in Puget Sound, Washington: Bocaccio (*Sebastes paucispinis*), canary rockfish (*S. pinniger*), yelloweye rockfish (*S. ruberrimus*), greenstriped rockfish (*S. elongatus*), and redstripe rockfish (*S. proriger*). U.S. Dept. Commerce., NOAA Tech. Memo. NMFS-NWFSC-108, 234 p.
- Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17(1):43–57.
- Elvenes, S., M. F. Dolan, P. Buhl-Mortensen, and V. K. Bellec. 2014. An evaluation of compiled single-beam bathymetry data as a basis for regional sediment and biotope mapping. ICES Journal of Marine Science: Journal Du Conseil 71(4):867–881.
- Embling, C. B., P. A. Gillibrand, J. Gordon, J. Shrimpton, P. T. Stevick, and P. S. Hammond. 2010. Using habitat models to identify suitable sites for marine protected areas for harbour porpoises (*Phocoena phocoena*). Biological Conservation 143(2):267–279.
- Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology 41(2):263–274.
- Ferrini, V. L., and R. D. Flood. 2006. The effects of fine-scale surface roughness and grain size on 300 kHz multibeam backscatter intensity in sandy marine sedimentary environments. Marine Geology 228(1):153–172.

- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24(01):38–49.
- Fischer, H.S., 1990. Simulating the distribution of plant communities in an alpine landscape. *Coenoses* 5:37–43.
- Fitzgerald, R. W., & Lees, B. G. 1993. The application of neural networks to the floristic classification of remote sensing and GIS data in complex terrain. *International Archives of Photogrammetry and Remote Sensing* 29:570-570.
- Galparsoro, I., A. Borja, J. Bald, P. Liria, and G. Chust. 2009. Predicting suitable habitat for the European Lobster (*Homarus gammarus*), on the Basque continental shelf (Bay of Biscay), using Ecological-niche Factor Analysis. *Ecological Modelling* 220:556-567.
- Gamer, M., J. Lemon, I. Fellows, and P. Singh. 2012. Irr: Various coefficients of interrater reliability and agreement. R Package Version 0.84.
- Gotshall, D.W. 1964. Increasing tagged rockfish (genus *Sebastes*) survival by deflating the swim bladder. *Calif. Fish and Game* 50:253–260.
- Green, K., K. Van Kirk, J. Stahl, M. Jaenicke, and S. Meyer. 2014. 14: Assessment of the demersal shelf rockfish stock complex in the Southeast Outside District of the Gulf of Alaska *in*: Stock assessment and fishery evaluation report for the groundfish resources for the Gulf of Alaska, North Pacific Fisheries Management Council, Anchorage, Alaska pp. 751–838.
- Greene, H. G., V. M. O’Connell, and C. K. Brylinsky. 2011. Tectonic and glacial related seafloor geomorphology as possible demersal shelf rockfish habitat surrogates—Examples along the Alaskan convergent transform plate boundary. *Continental Shelf Research* 31(2):S39–S53.
- Greene, H. G., V. M. O’Connell, W. W. Wakefield, and C. K. Brylinsky. 2007. The offshore Edgecumbe lava field, southeast Alaska: Geologic and habitat characterization of a commercial fishing ground. *Mapping the Seafloor for Habitat Characterization*. *Geol.Assoc.can.Spec.Pap* 47:277–296.
- Greene, H. G., M. M. Yoklavich, R. M. Starr, V. M. O’Connell, W. W. Wakefield, D. E. Sullivan, J. E. McRea, and G. M. Cailliet. 1999. A classification scheme for deep seafloor habitats. *Oceanologica Acta* 22(6):663–678.
- Gross, J., M. Kneeland, D. Reed, and R. Reich. 2002. GIS-based habitat models for mountain goats. *Journal of Mammalogy* 83(1):218–228.
- Guisan, A., T. C. Edwards, and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling* 157(2):89–100.

- Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135(2):147–186.
- Hannah, R. W., and P. S. Rankin. 2011. Site fidelity and movement of eight species of Pacific rockfish at a high-relief rocky reef on the Oregon coast. *North American Journal of Fisheries Management* 31(3):483–494.
- Hirzel, A., and A. Guisan. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling* 157(2):331–341.
- Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* 83(7):2027–2036.
- Hirzel, A., J. Hausser, and N. Perrin. 2007. Biomapper 4.0. Laboratory for Conservation Biology, Department of Ecology and Evolution, University of Lausanne, Switzerland. URL: [Http://www2.unil.ch/biomapper](http://www2.unil.ch/biomapper).
- Hirzel, A., V. Helfer, and F. Metral. 2001. Assessing habitat-suitability models with a virtual species. *Ecological Modelling* 145(2):111–121.
- Hirzel, A. H., G. Le Lay, V. Helfer, C. Randin, and A. Guisan. 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199(2):142–152.
- Holmes, R. E. 1993. A GIS-based habitat model for mountain goats (*Oreamnos americanus*) in the North Cascades National Park Service Complex, Washington. Master's thesis. Western Washington University.
- Hosmer D. W. and S. Lemeshow. 2004. Applied logistic regression, 2<sup>nd</sup> edn. John Wiley & Sons, Hoboken, NJ.
- Hutchinson, G. E. 1957. Cold Spring Harbor Symposium on Quantitative Biology. Concluding Remarks 22:415–427.
- Iampietro, P. J., R. G. Kvitek, and E. Morris. 2005. Recent advances in automated genus-specific marine habitat mapping enabled by high-resolution multibeam bathymetry. *Marine Technology Society Journal* 39(3):83–93.
- Iampietro, P. J., M. A. Young, and R. G. Kvitek. 2008. Multivariate prediction of rockfish habitat suitability in Cordell Bank National Marine Sanctuary and Del Monte Shalebeds, California, USA. *Marine Geodesy* 31(4):359–371.
- Johnson, S. W., M. L. Murphy, and D. J. Csepp. 2003. Distribution, habitat, and behavior of rockfishes, *Sebastes spp.*, in nearshore waters of southeastern Alaska: Observations from a remotely operated vehicle. *Environmental Biology of Fishes* 66(3):259–270.

- Kéry, M. 2010. Introduction to WinBUGS for Ecologists. Academic Press, Burlington, MA.
- Kramer, D.E. and V.M. O'Connell. 2004. Guide to the northeast Pacific rockfishes, genera *Sebastes* and *Sebastolobus*, 2004 edition. University of Alaska Sea Grant, Marine Advisory Bulletin 25, 78 p.
- Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174.
- Leverette, T. L., and A. Metaxas. 2005. Predicting habitat for two species of deep-water coral on the Canadian Atlantic continental shelf and slope. Pages 467–479. *Cold-water corals and ecosystems*. Springer Berlin Heidelberg.
- Love, M. S., M. Yoklavich, and L. K. Thorsteinson. 2002. The rockfishes of the northeast pacific. Univ of California Press, Berkley, CA.
- MacKenzie, D.I., J.D. Nichols, J.A. Royle, K.H. Pollock, L.L. Baily and J.E. Hines. 2006. *Occupancy Estimation and Modeling*. Academic Press, Burlington, MA.
- Miller, R. J., C. Juska, and J. Hocevar. 2015. Submarine canyons as coral and sponge habitat on the eastern Bering Sea slope. *Global Ecology and Conservation* 4:85–94.
- Monk, J., D. Ierodiaconou, V. L. Versace, A. Bellgrove, E. Harvey, A. Rattray, L. Laurenson, and G. P. Quinn. 2010. Habitat suitability for marine fishes using presence-only modelling and multibeam sonar. *Marine Ecology Progress Series* 420:157-174.
- Monk, J., D. Ierodiaconou, A. Bellgrove, E. Harvey, and L. Laurenson. 2011. Remotely sensed hydroacoustics and observation data for predicting fish habitat suitability. *Continental Shelf Research* 31(2):S17–S27.
- Monk, J., D. Ierodiaconou, E. Harvey, A. Rattray, and V. L. Versace. 2012. Are we predicting the actual or apparent distribution of temperate marine fishes? *PloS One* 7(4):e34558.
- Moore, C. H., E. S. Harvey, and K. P. Van Niel. 2009. Spatial prediction of demersal fish distributions: Enhancing our understanding of species–environment relationships. *ICES Journal of Marine Science: Journal Du Conseil* 66(9):2068-2075.
- Munk, K. M. 2001. Maximum ages of groundfishes in waters off Alaska and British Columbia and considerations of age determination. *Alaska Fishery Research Bulletin* 8(1):12–21.
- Nasby-Lucas, N. M., B. W. Embley, M. A. Hixon, S. G. Merle, B. N. Tissot, and D. J. Wright. 2002. Integration of submersible transect data and high-resolution multibeam sonar imagery for a habitat-based groundfish assessment of Heceta Bank, Oregon. *Fisheries Bulletin* 100(4):739–751.

- North Pacific Fishery Management Council (NPFMC). 2014. Stock assessment and fishery evaluation report for groundfish resources of the Gulf of Alaska. North Pacific Fishery Management Council, Anchorage, AK. 1040 p.
- O'Connell, V. M., B. E. Bracken, and D. W. Carlile. 1991. Demersal shelf rockfish. *in* Stock assessment and fishery evaluation report for the 1992 Gulf of Alaska groundfish fishery. N. Pac. Fish. Manage. Council., Anchorage AK.
- O'Connell V. and C. Brylinsky. 2003. The Southeast Alaska Demersal Shelf Rockfish Fishery with 2004 Outlook. Alaska Department of Fish and Game, Regional Information Report No. IJ03-43.
- O'Connell, V.M, and D.W. Carlile. 1993. Habitat-specific density of adult yelloweye rockfish *Sebastes ruberrimus* in the eastern Gulf of Alaska. *Fishery Bulletin* 91(2):304–309.
- O'Connell, V. M., and D. W. Carlile. 1994. Comparison of a remotely operated vehicle and a submersible for estimating abundance of demersal shelf rockfishes in the eastern gulf of alaska. *North American Journal of Fisheries Management* 14(1):196-201.
- O'Connell, V.M. and F.C. Funk 1986. Age and growth of yelloweye rockfish (*Sebastes ruberrimus*) landed in southeastern Alaska. Pages 171–185 *in* Melteff, B,R. [ed.]. *Proceedings of the International Rockfish Symposium*. University of Alaska Sea Grant, Juneau, AK.
- Parnum, I., Siwabessy, J., Gavrilov, A., and M. Parsons. 2009. A comparison of single beam and multibeam sonar systems in seafloor habitat mapping. *Proceedings of the 3rd International Conference and Exhibition on Underwater Acoustic Measurements: Technologies and Results* :21–26.
- Pearce, J. and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*. 133:225–245.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190(3):231–259.
- Pianka, E.R. 1970. On r- and k-selection. *The American Naturalist* 104:592–597.
- Richards, L. 1986. Depth and habitat distributions of three species of rockfish in British Columbia: observations from the submersible PISCES IV. *Environmental Biology of Fishes* 17(1):13–21.
- Rooper, C. N. 2008. An ecological analysis of rockfish (*Sebastes spp.*) assemblages in the north Pacific Ocean along broad-scale environmental gradients. *Fishery Bulletin* 106(2):1–11.

- Rooper, C. N., G. R. Hoff, and A. De Robertis. 2010. Assessing habitat utilization and rockfish (*Sebastes spp.*) biomass on an isolated rocky ridge using acoustics and stereo image analysis. *Canadian Journal of Fisheries and Aquatic Sciences* 67(10):1658–1670.
- Rooper, C. N., M. Zimmermann, M. M. Prescott, and A. J. Hermann. 2014. Predictive models of coral and sponge distribution, abundance and diversity in bottom trawl surveys of the Aleutian Islands, Alaska. *Marine Ecological Progress Series* 503:157–176.
- Russ, E., C. E. Trowbridge, and C. Russ. 2013. Cook Inlet Area groundfish management report, 2005–2011. Alaska Department of Fish and Game, Fishery Management Report No. 13-04, Anchorage AK.
- Sappington, J., K. M. Longshore, and D. B. Thompson. 2007. Quantifying landscape ruggedness for animal habitat analysis: A case study using bighorn sheep in the Mojave Desert. *The Journal of Wildlife Management* 71(5):1419–1426.
- Sigler, M. F., C. N. Rooper, G. R. Hoff, R. P. Stone, R. A. McConnaughey, and T. K. Wilderbuer. 2015. Faunal features of submarine canyons on the eastern Bering Sea slope. *Marine Ecological Progress Series* 526:21–40.
- Sing, T., O. Sander, N. Beerenwinkel, and T. Lengauer. 2007. ROCR: Visualizing the Performance of Scoring Classifiers. R Package Version 1.0–2.
- Soberón, J. 2007. Grinnellian and eltonian niches and geographic distributions of species. *Ecology Letters* 10(12):1115–1123.
- Stein, D. L., B. N. Tissot, M. A. Hixon, and W. Barss. 1992. Fish-habitat associations on a deep reef at the edge of the Oregon continental shelf. *Fishery Bulletin* 90(3):540-551.
- Taylor, I. G. 2011. Rebuilding analysis for yelloweye rockfish based on the 2011 update stock assessment. National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle, WA .
- Taylor, I., and C. Wetzel. 2011. Status of the US yelloweye rockfish resource in 2011 (update of 2009 assessment model). Pacific Fishery Management Council, Portland, OR.
- Team, R. C. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2012.
- Tissot, B. N., M. A. Hixon, and D. L. Stein. 2007. Habitat-based submersible assessment of macro-invertebrate and groundfish assemblages at Heceta Bank, Oregon, from 1988 to 1990. *Journal of Experimental Marine Biology and Ecology* 352(1):50–64.
- Trowbridge, C., W. Dunne, M. Lambdin, M. Byerly, and K. Goldman. 2008. Cook Inlet Area Groundfish Management Report 1996–2004. Alaska Department of Fish and Game.

- Valavanis, V. D., S. Georgakarakos, A. Kapantagakis, A. Palialexis, and I. Katara. 2004. A GIS environmental modelling approach to essential fish habitat designation. *Ecological Modelling* 178(3):417–427.
- Valavanis, V. D., G. J. Pierce, A. F. Zuur, A. Palialexis, A. Saveliev, I. Katara, and J. Wang. 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. *Hydrobiologia* 612(1):5–20.
- Venables, W., and B. Ripley. 2010. stepAIC: MASS. R package version 7.3–9.
- Weiss, A. 2001. Topographic position and landforms analysis. Proceedings of the Poster Presentation, ESRI User Conference, San Diego, CA :200–200.
- Wilson, F.H., and Hults, C.P., comps., 2012, Geology of the Prince William Sound and Kenai Peninsula region, Alaska: U.S. Geological Survey Scientific Investigations Map 3110, pamphlet 38 p., scale 1:350,000. Available: <http://pubs.usgs.gov/sim/3110/>.
- Wilson, M. F., B. O’Connell, C. Brown, J. C. Guinan, and A. J. Grehan. 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy* 30(1–2):3–35.
- Wright, D., E. Lundblad, E. Larkin, R. Rinehart, J. Murphy, L. Cary-Kothera, and K. Draganov. 2005. ArcGIS benthic terrain modeler, corvallis, oregon, oregon state university, davey jones locker seafloor Mapping/Marine GIS laboratory and NOAA coastal services center. Available: <http://www.Csc.Noaa.gov/products/btm> .
- Yoklavich, M. M., H. G. Greene, G. M. Cailliet, D. E. Sullivan, R. N. Lea, and M. S. Love. 2000. Habitat associations of deep-water rockfishes in a submarine canyon: An example of a natural refuge. *Fishery Bulletin-National Oceanic and Atmospheric Administration* 98(3):625–641.
- Yoklavich, M. M., M. S. Love, and K. A. Forney. 2007. A fishery-independent assessment of an overfished rockfish stock, cowcod (*Sebastes levis*), using direct observations from an occupied submersible. *Canadian Journal of Fisheries and Aquatic Sciences* 64(12):1795–1804.
- Young, M. A., P. J. Iampietro, R. G. Kvitek, and C. D. Garza. 2010. Multivariate bathymetry-derived generalized linear model accurately predicts rockfish distribution on Cordell Bank, California, USA. *Marine Ecology Progress Series* 415:247–261.
- Zimmermann, M., and M. M. Prescott. 2014. Smooth sheet bathymetry of Cook Inlet, Alaska. U.S. Dep. Commer., NOAA Tech. Memo. NMFS AFSC–275, 32 p.
- Zimmermann, M., and M. M. Prescott. 2015. Smooth sheet bathymetry of the central Gulf of Alaska. U.S. Dep. Commer., NOAA Tech. Memo. NMFS–AFSC–287, 54 p.



Zimmermann, M., J. A. Reid, and N. Golden. (*In press*). Using smooth sheets to describe groundfish habitat in Alaskan waters, with specific application to two flatfishes. Deep Sea Research Part II: Topical Studies in Oceanography.

## TABLES

**Table 1.** Predictor variables and scales considered for the ENFAs and GLMs. In addition to these linear terms, quadratic terms for depth and all of the bathymetric position index (BPI) variables were considered, because local optima were suspected. No neighborhood size is provided for depth, slope or distance-to-rock (DTR), because these were calculated without consideration of surrounding cells. No inner radius is provided for vector ruggedness measure (VRM), because this is calculated using a square rather than annulus shaped neighborhood. Numbers in variable names correspond to radii in meters for BPI and diameters in cells for VRM. The numbers in the DTR names indicate which scale of VRM that DTR variable is based on. Dashes indicate not applicable.

<b>Variable</b>	<b>Inner radius (cells)</b>	<b>Outer radius (cells)</b>	<b>Inner radius (m)</b>	<b>Outer radius (m)</b>
Depth	-	-	-	-
Slope	-	-	-	-
BPI30	5	10	15	30
BPI60	15	20	45	60
BPI120	35	40	105	120
BPI240	75	80	225	240
VRM3	-	1	-	3
VRM5	-	2	-	6
VRM7	-	3	-	9
VRM21	-	10	-	30
DTR3	-	-	-	-
DTR5	-	-	-	-
DTR7	-	-	-	-
DTR21	-	-	-	-

**Table 2.** Performance of various scales of simple logistic regression models at predicting suitable yelloweye habitat ranked by AIC score within each type of terrain variable for both study areas. Significance of each univariate model is also included (Wald test ; \* indicates  $p < 0.001$ ). Highlighted variables were considered in the scope of the final stepwise AIC variable selection process for the GLM.

	CHISWELL			NUKA		
	Scale	$\Delta$ AIC	p	Scale	$\Delta$ AIC	p
VRM	21	0.0	*	7	0	*
	7	30.3	*	5	8.3	*
	5	49.4	*	3	28.3	*
	3	82.8	*	21	56.9	*
BPI	60 <sup>2</sup>	0.0	*	240	0	*
	240	6.5	*	30 <sup>2</sup>	30.0	*
	30 <sup>2</sup>	20.6	*	120	36.3	*
	120 <sup>2</sup>	25.7	*	60 <sup>2</sup>	36.8	*
	240 <sup>2</sup>	46.3	*	240 <sup>2</sup>	40.2	*
	120	77.5	*	120 <sup>2</sup>	55.4	*
	60	102.7	0.009	60	89.0	*
30	106.1	0.045	30	108.4	0.446	
DTR	3	0.0	*	5	0	*
	5	6.3	*	3	1.2	*
	7	12.7	*	7	7.8	*
	21	35.7	*	21	61.2	*
Depth	3 <sup>2</sup>	0.0	*	3	0.0	*
	3	0.9	*	3 <sup>2</sup>	4.6	*
Slope	3	0.0	*	3	0.0	*

**Table 3.** Summary of the GLM fit to each of two study areas. Values in parenthesis are p values based on the Wald test.

<b>Parameter</b>	<b>Chiswell Estimate</b>	<b>Nuka Estimate</b>
Intercept	-0.0842 (0.940)	-2.4380 (<0.001)***
DTR7	-0.0774 (0.004)**	-0.0271 (0.166)
Slope	0.1970 (0.005)**	0.1083 (0.041)*
VRM7	319.0483 (0.303)	295.6500 (0.004)**
BPI240	0.0649 (0.274)	0.1136 (0.003)**
AIC	73.4	125.2
Null deviance	316.1	321.6
Residual deviance	63.4	115.2
n (observations)	228	232

\*\*\* indicates  $p < 0.001$

\*\* indicates  $p < 0.01$

\* indicates  $p < 0.05$

**Table 4.** Error matrix for the Chiswell GLM applied to the Chiswell area. The presence and absence points used in this accuracy assessment were independent from those used to fit the model.

		<b>OBSERVED</b>		User's Accuracy
		Absent	Present	
<b>PREDICTED</b>	Absent	<b>48</b>	2	96.0%
	Present	2	<b>48</b>	96.0%
Producer's Accuracy		96.0%	96.0%	<b>96.0%</b>

Kappa = 0.92

n = 100 observations

**Table 5.** Error matrix for the Chiswell GLM applied to the Nuka area. All presence and absence points from the Nuka area were used in the evaluation because none of these points were used to fit the model.

		<b>OBSERVED</b>		User's Accuracy
		Absent	Present	
<b>PREDICTED</b>	Absent	<b>115</b>	7	94.2%
	Present	51	<b>159</b>	75.7%
Producer's Accuracy		69.3%	95.7%	<b>82.5%</b>

Kappa = 0.65

n = 332 observations

**Table 6.** Error matrix for the Nuka GLM applied to the Nuka area. The presence and absence points used in this accuracy assessment were independent from those used to fit the model.

		<b>OBSERVED</b>		User's Accuracy
		Absent	Present	
<b>PREDICTED</b>	Absent	<b>45</b>	6	88.2%
	Present	5	<b>44</b>	89.8%
Producer's Accuracy		90.0%	88.0%	<b>89.0%</b>

Kappa = 0.78

n = 100 observations

**Table 7.** Summary of the Chiswell ecological niche factor analysis (ENFA). The top row contains the percentage of information contained in the multivariate dataset accounted for by each of the four most important ecological factors. The cells below contain the coefficients for each individual terrain variables for each ecological factor. The variables are sorted by the absolute values of the marginality factor weighting with positive coefficients for the marginality factor indicating that yelloweye prefer locations with higher values than the mean value in the Chiswell study area. All variables were normalized with the Box-Cox transformation except for VRM21 which was not because extreme values caused a terminal error in the ENFA algorithm. VRM7, VRM5 and VRM3, and BPI30 also caused terminal errors in the algorithm because they were not continuous enough and were not included in the ENFA.

<b>Terrain Variable</b>	<b>Marginality Factor (40%)</b>	<b>Specialization Factor 1 (19%)</b>	<b>Specialization Factor 2 (11%)</b>	<b>Specialization Factor 3 (8%)</b>
VRM21	0.449	-0.042	-0.013	0.002
DTR3-box	-0.373	-0.035	-0.817	-0.164
DTR21-box	-0.368	-0.120	-0.113	-0.253
DTR5-box	-0.364	-0.630	0.449	0.861
DTR7-box	-0.360	0.693	0.179	-0.401
Slope-box	0.259	-0.264	-0.114	0.034
Depth-box	0.198	0.176	-0.199	0.053
BPI60-box	0.179	0.000	-0.013	0.016
BPI240-box	-0.059	0.047	-0.113	-0.048
BPI120-box	0.055	0.025	0.021	-0.004

Marginality: 2.256

Specialization: 1.921

Tolerance: 0.520



**Table 8.** Summary of the Nuka ecological niche factor analysis (ENFA). The top row contains the percentage of information contained in the multivariate dataset accounted for by each of the four most important ecological factors. The cells below contain the coefficients for each of individual terrain variables for each ecological factor. The variables are sorted by the absolute values of the marginality factor weighting with positive coefficients for the marginality factor indicating that yelloweye prefer locations with higher values than the mean value in the Nuka study area. All variables were normalized with the Box-Cox transformation except for VRM21 which was not because extreme values caused a terminal error in the ENFA algorithm. VRM7, VRM5 and VRM3, and BPI30 also caused terminal errors in the algorithm because they were not continuous enough and were not included in the ENFA.

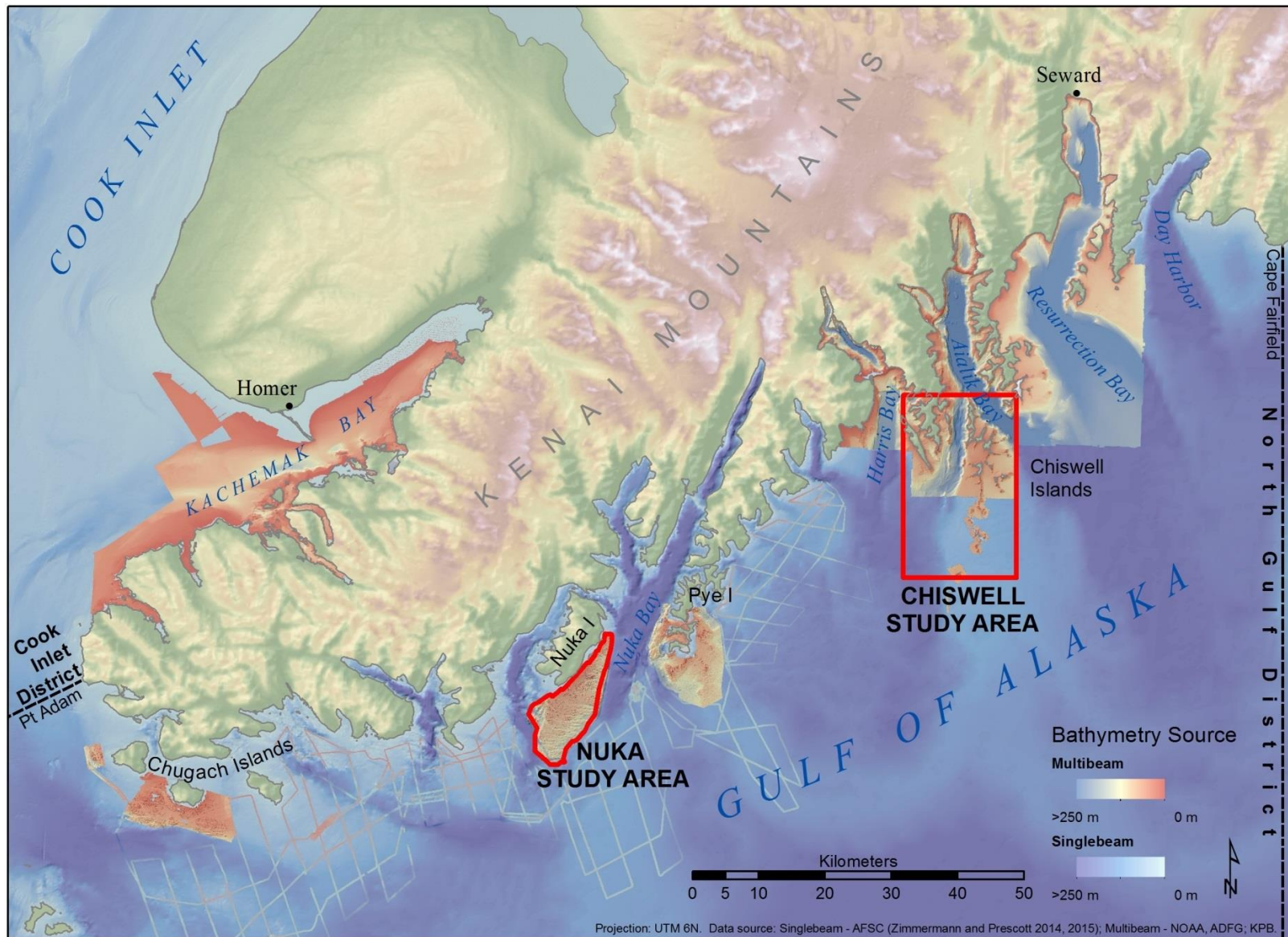
<b>Terrain Variable</b>	<b>Marginality Factor (33%)</b>	<b>Specialization Factor 1 (25%)</b>	<b>Specialization Factor 2 (10%)</b>	<b>Specialization Factor 3 (8%)</b>
VRM21	0.544	0.006	-0.101	0.055
DTR3-box	-0.334	0.256	-0.107	-0.202
DTR21-box	-0.332	-0.281	0.439	0.706
DTR5-box	-0.332	-0.011	0.258	0.086
DTR7-box	-0.324	0.470	-0.535	-0.458
Slope-box	0.279	-0.035	-0.183	-0.111
Depth-box	0.216	0.236	0.421	-0.080
BPI240-box	-0.181	-0.715	0.304	-0.362
BPI60-box	0.123	-0.050	0.134	-0.163
BPI120-box	-0.071	-0.209	-0.273	0.225

Marginality: 1.816

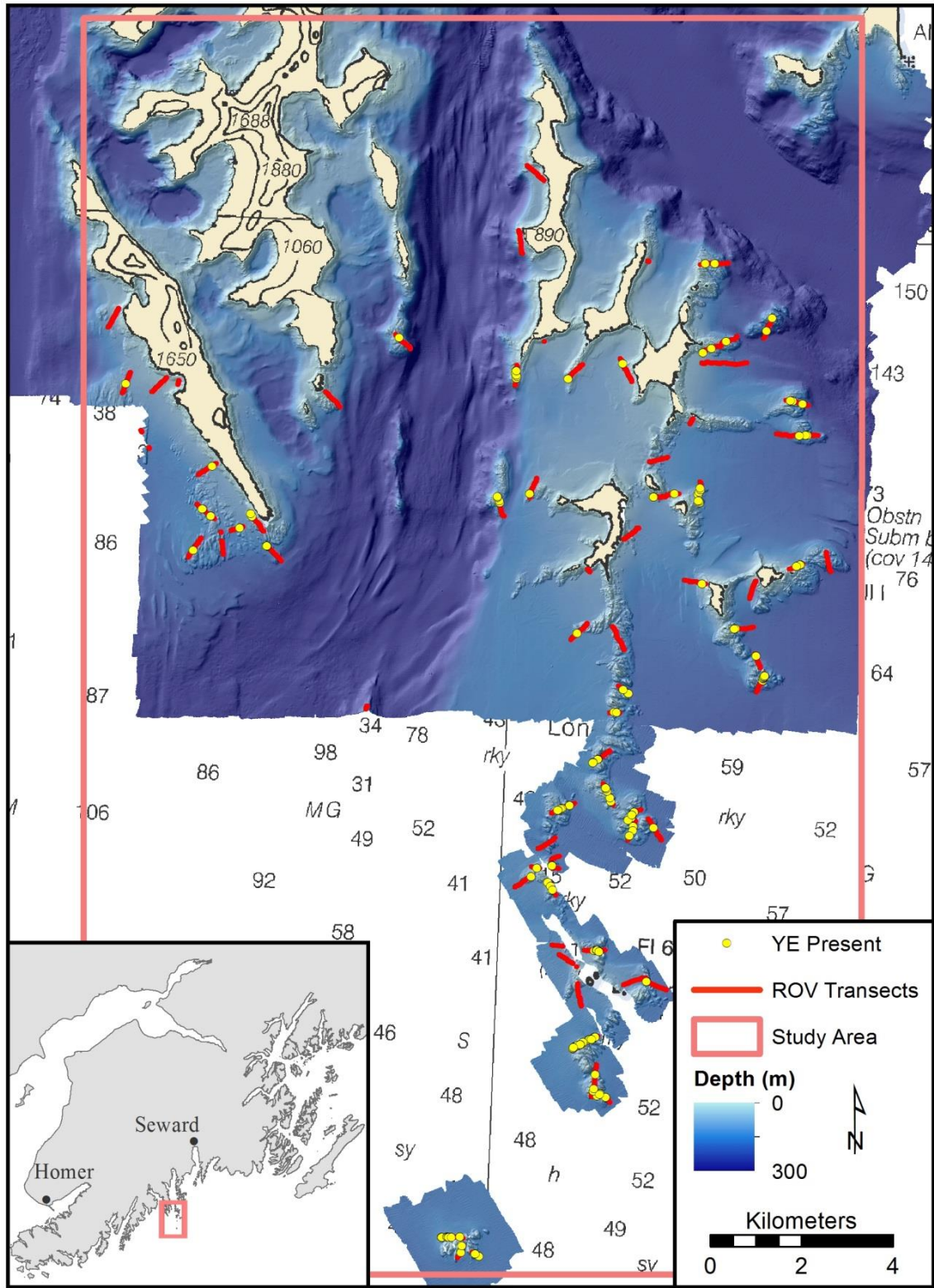
Specialization: 1.725

Tolerance: 0.579

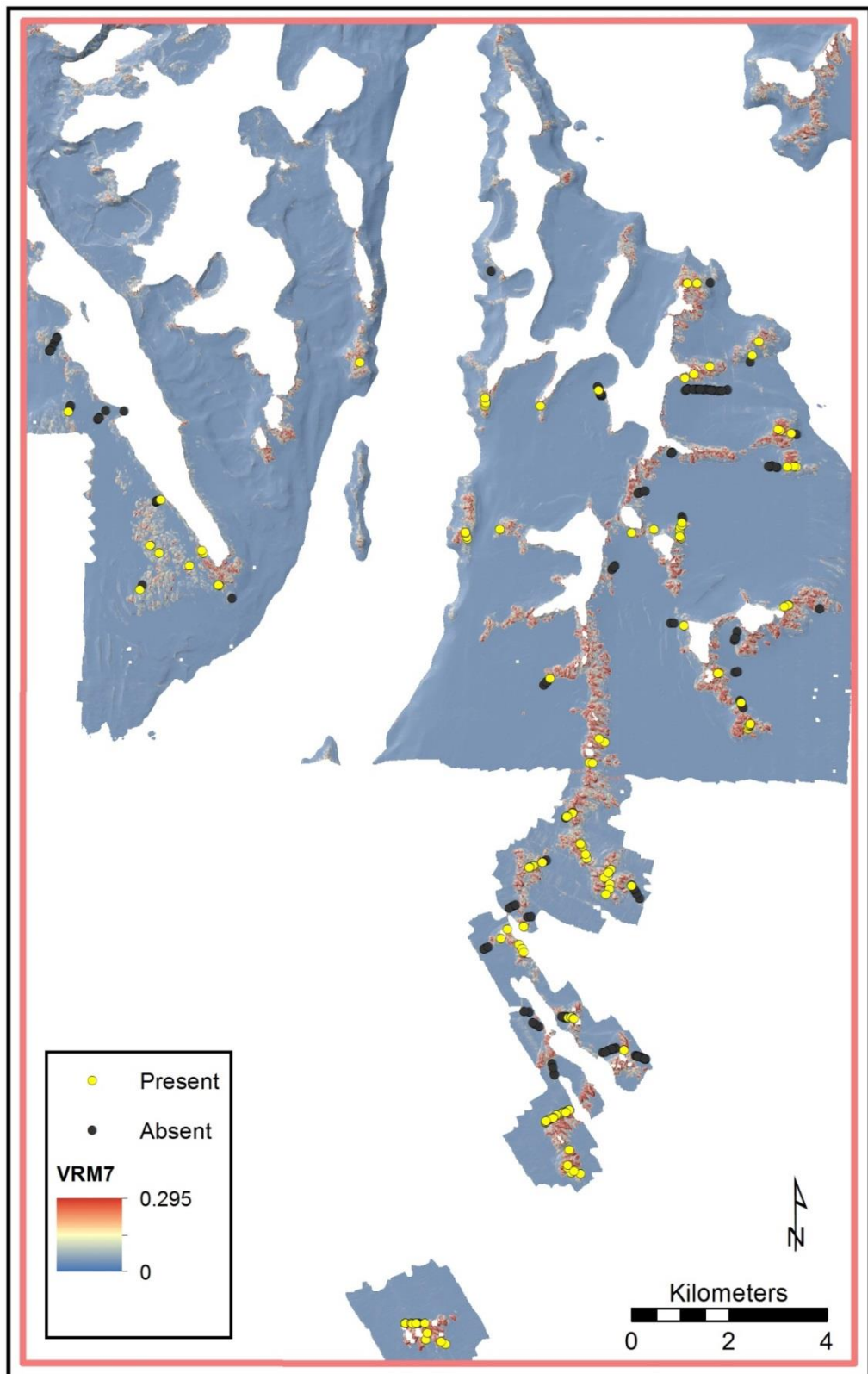
**FIGURES**



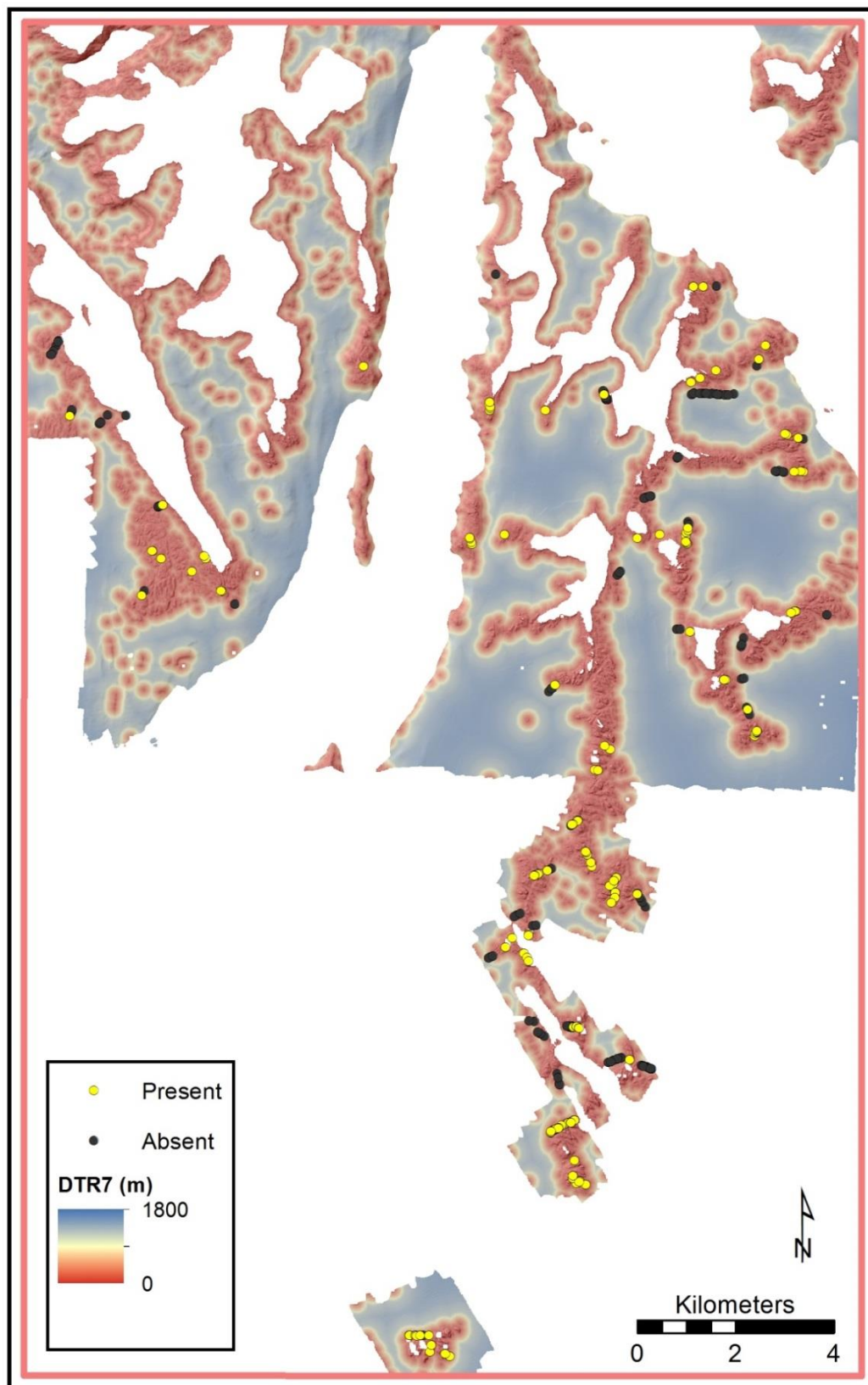
**Figure 1.** Location of the two study areas within the North Gulf ADF&G commercial groundfish management district and the best available bathymetry. Reddish bathymetry was surveyed with multibeam sonar while the blue data are from smooth sheets.



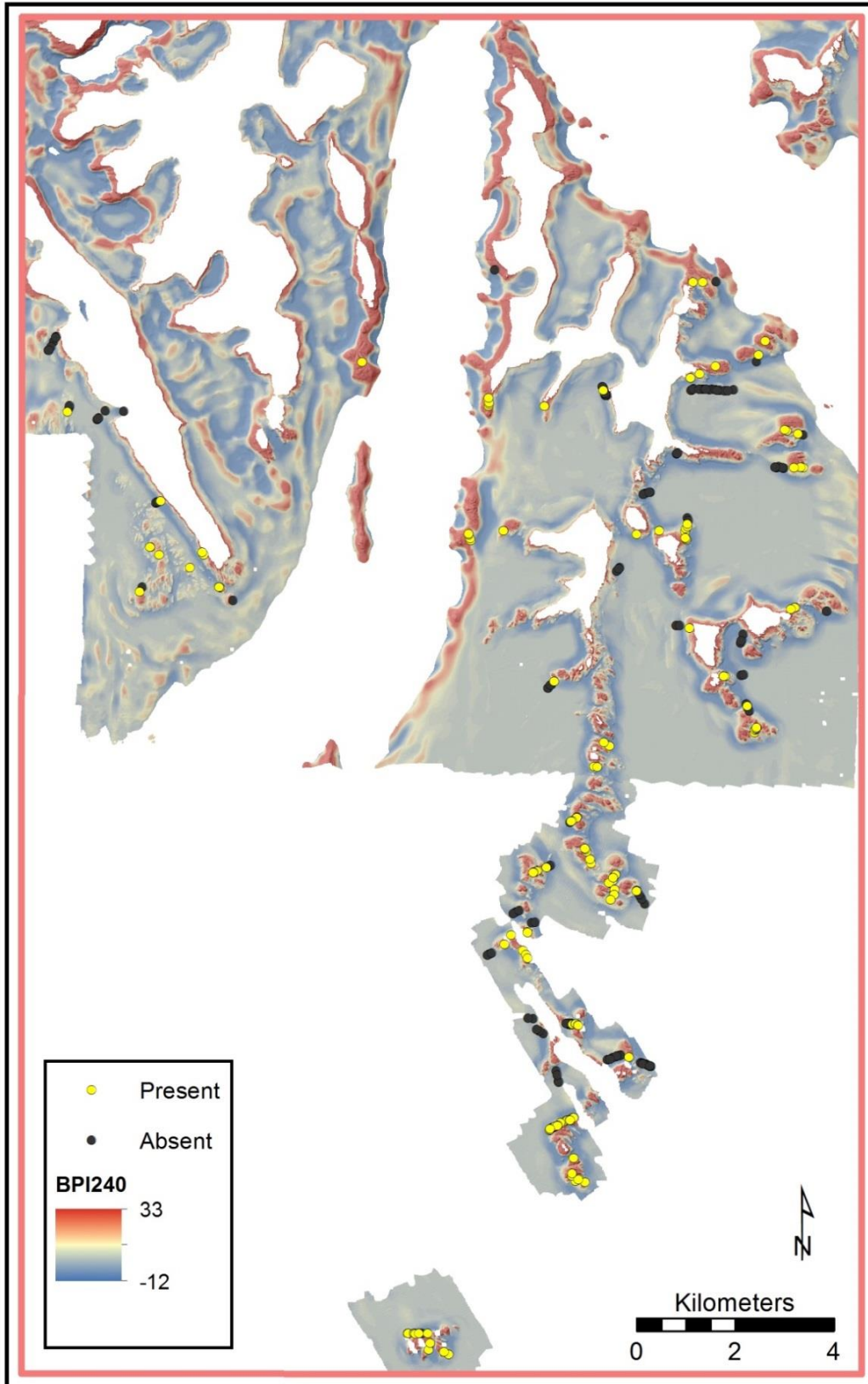
**Figure 2.** Chiswell Island Study area. Red lines are ROV transects surveyed in 2004 and 2005. Blue bathymetry is from multibeam surveys conducted between 2000 and 2006. Yellow points are the combined training and validation sets of yelloweye rockfish observations (n = 164). Soundings are in fathoms.



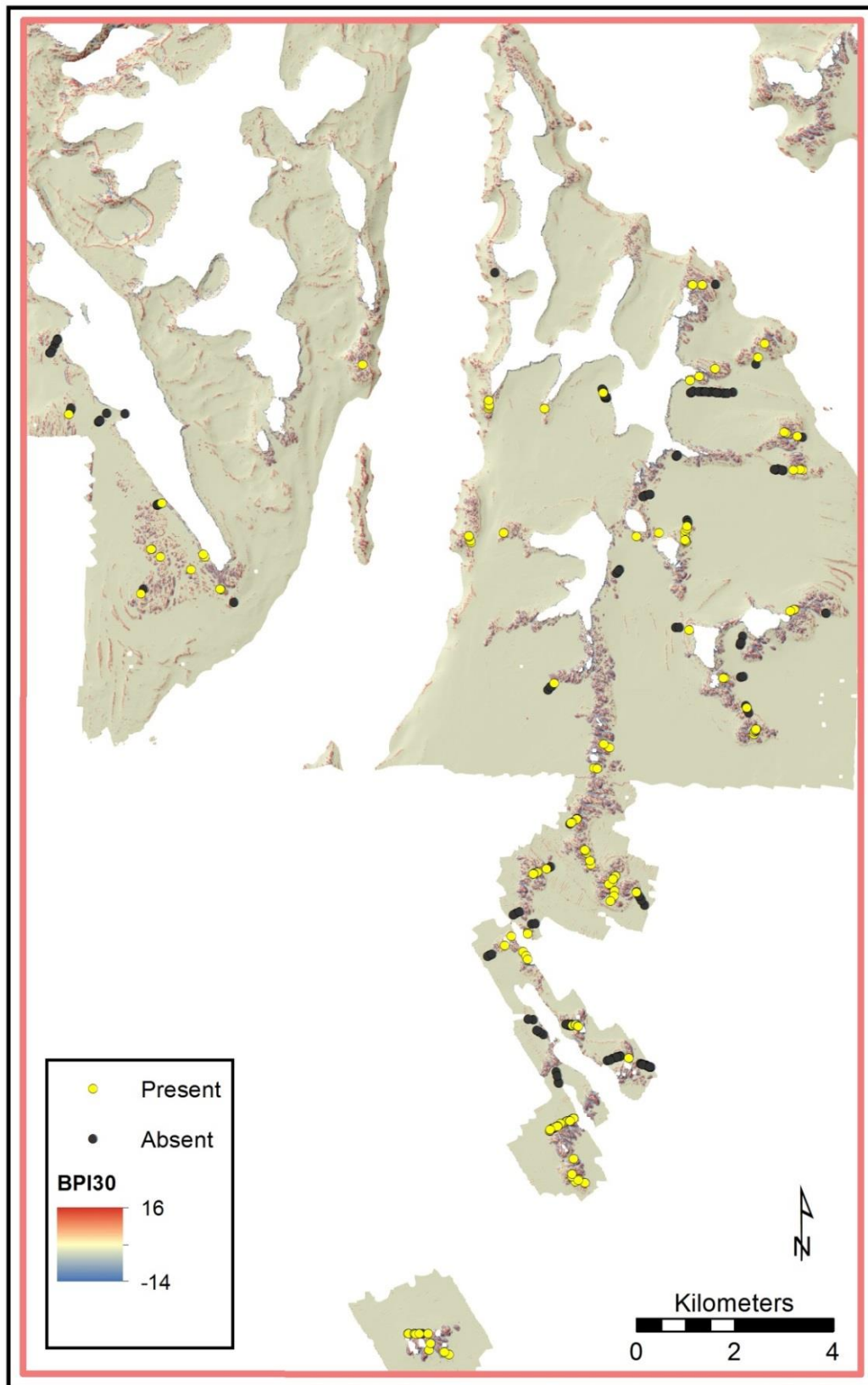
**Figure 3.** Vector rugosity measure (VRM) calculated with a scale factor of 7 in the Chiswell area. Yellow points are the combined training and validation sets of yelloweye rockfish observations. Black points are the combined training and validation set of absence points.



**Figure 4.** Distance-to-rock (DTR) calculated as distance to VRM7 peaks in the Chiswell area.



**Figure 5.** Bathymetric position index (BPI) calculated using a scale factor of 240 in the Chiswell area.



**Figure 6.** Bathymetric position index (BPI) calculated using a scale factor of 30 in the Chiswell area.



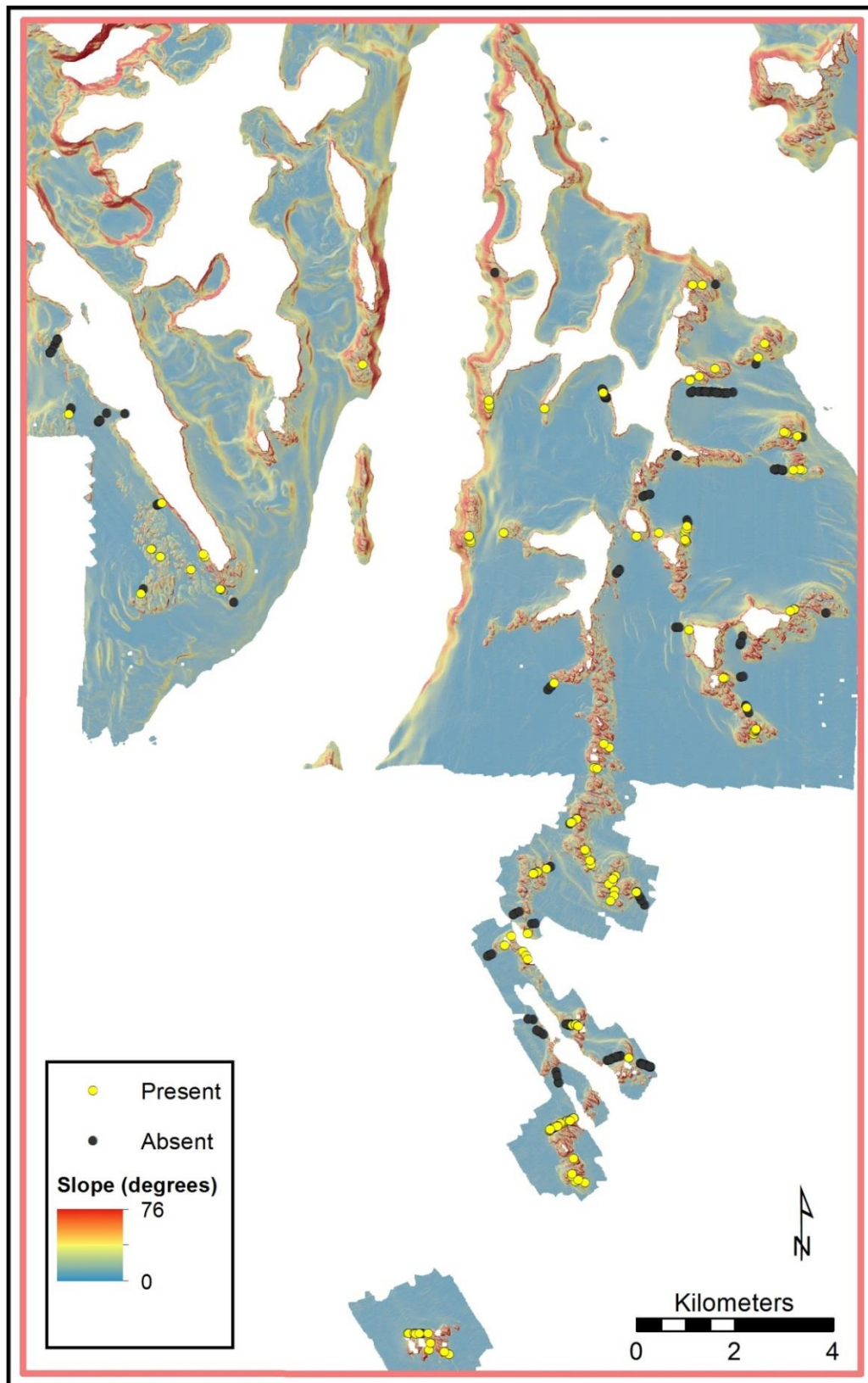
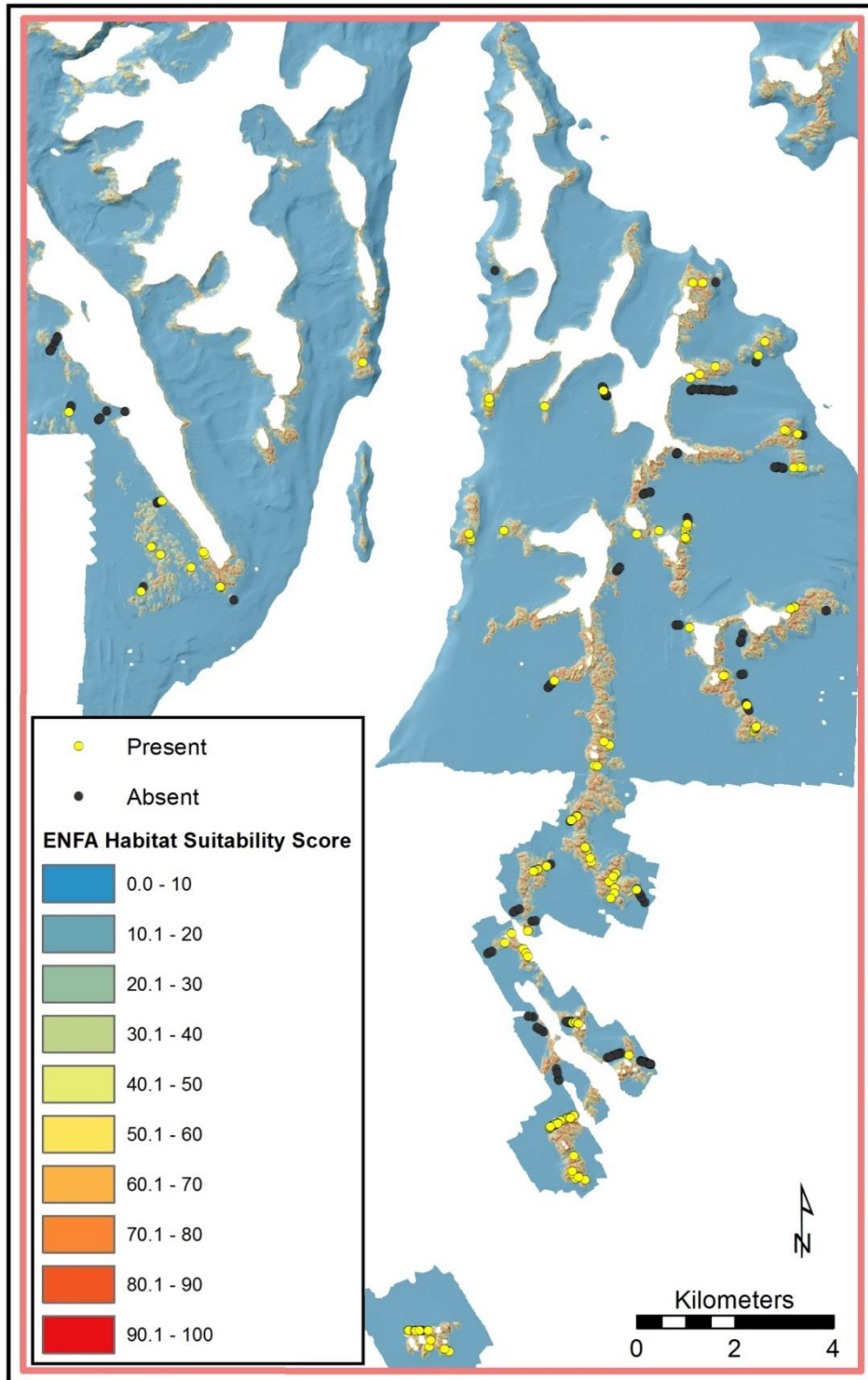
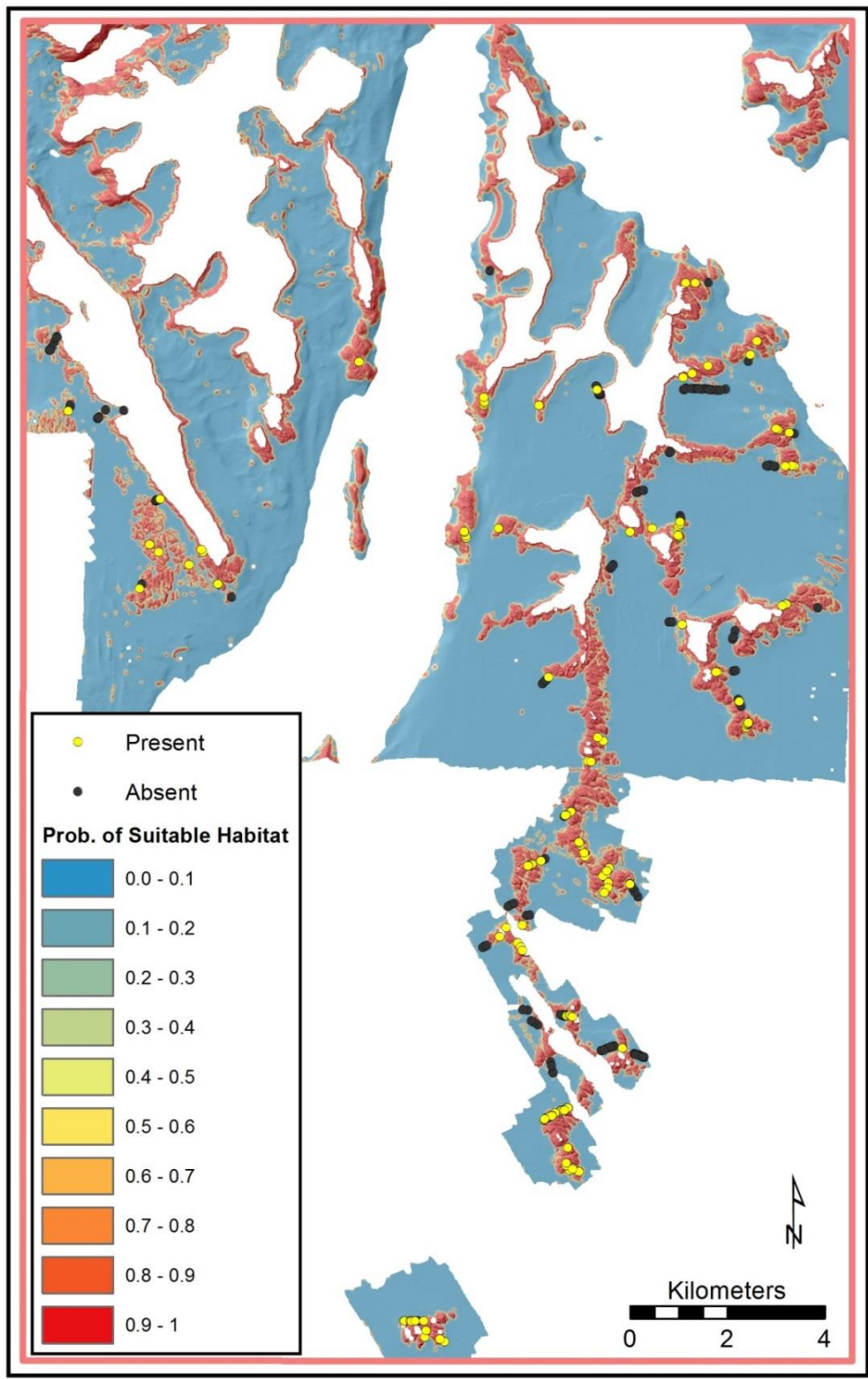


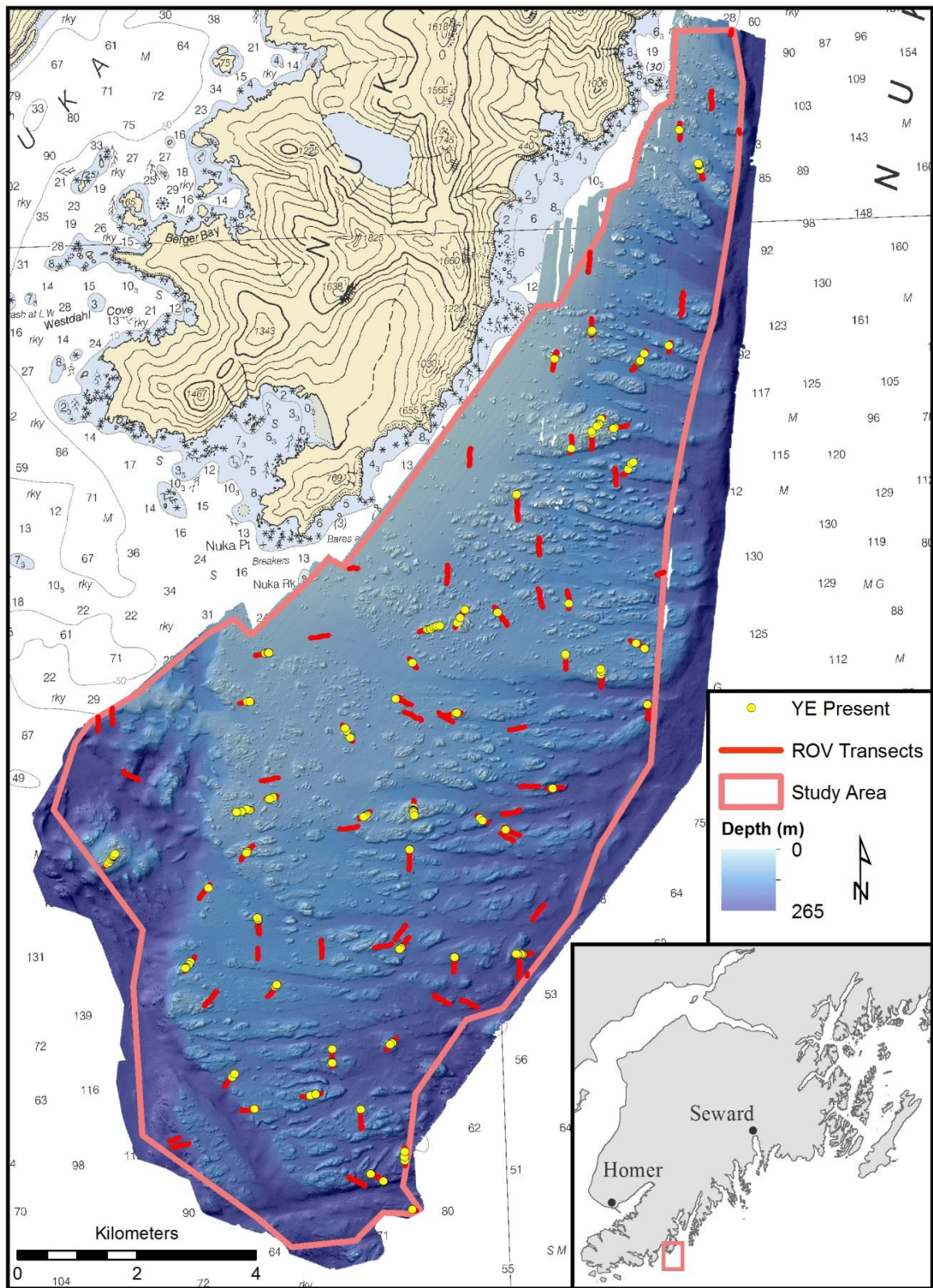
Figure 7. Slope in the Chiswell area.



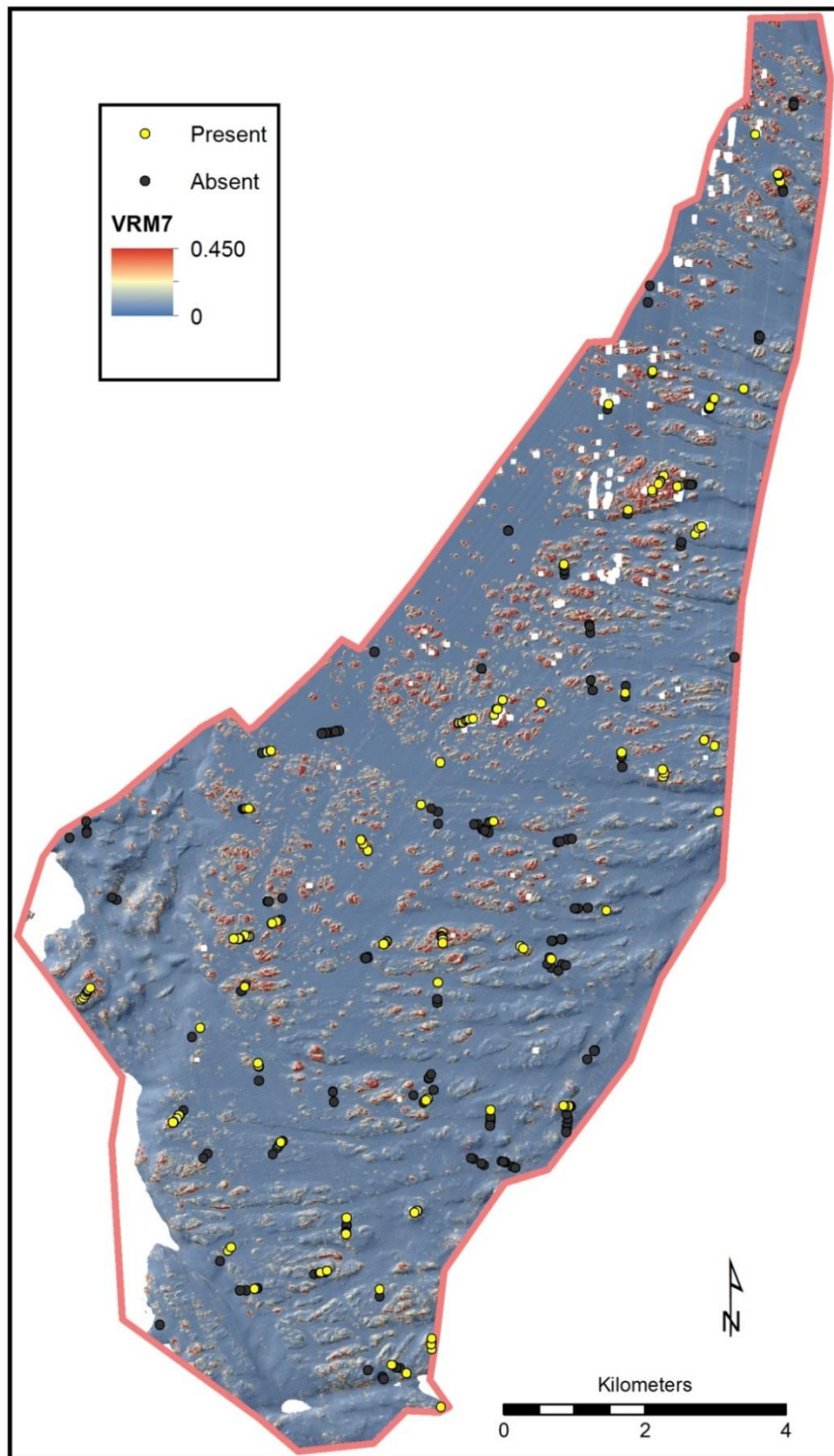
**Figure 8.** Habitat suitability score for yelloweye rockfish in the Chiswell area, based on the ecological niche factor analysis. Higher value indicates more suitable habitat. The ENFA used only presence points. Yellow points are the combined training and validation sets of yelloweye rockfish observations. Black points are the combined training and validation set of absence points.



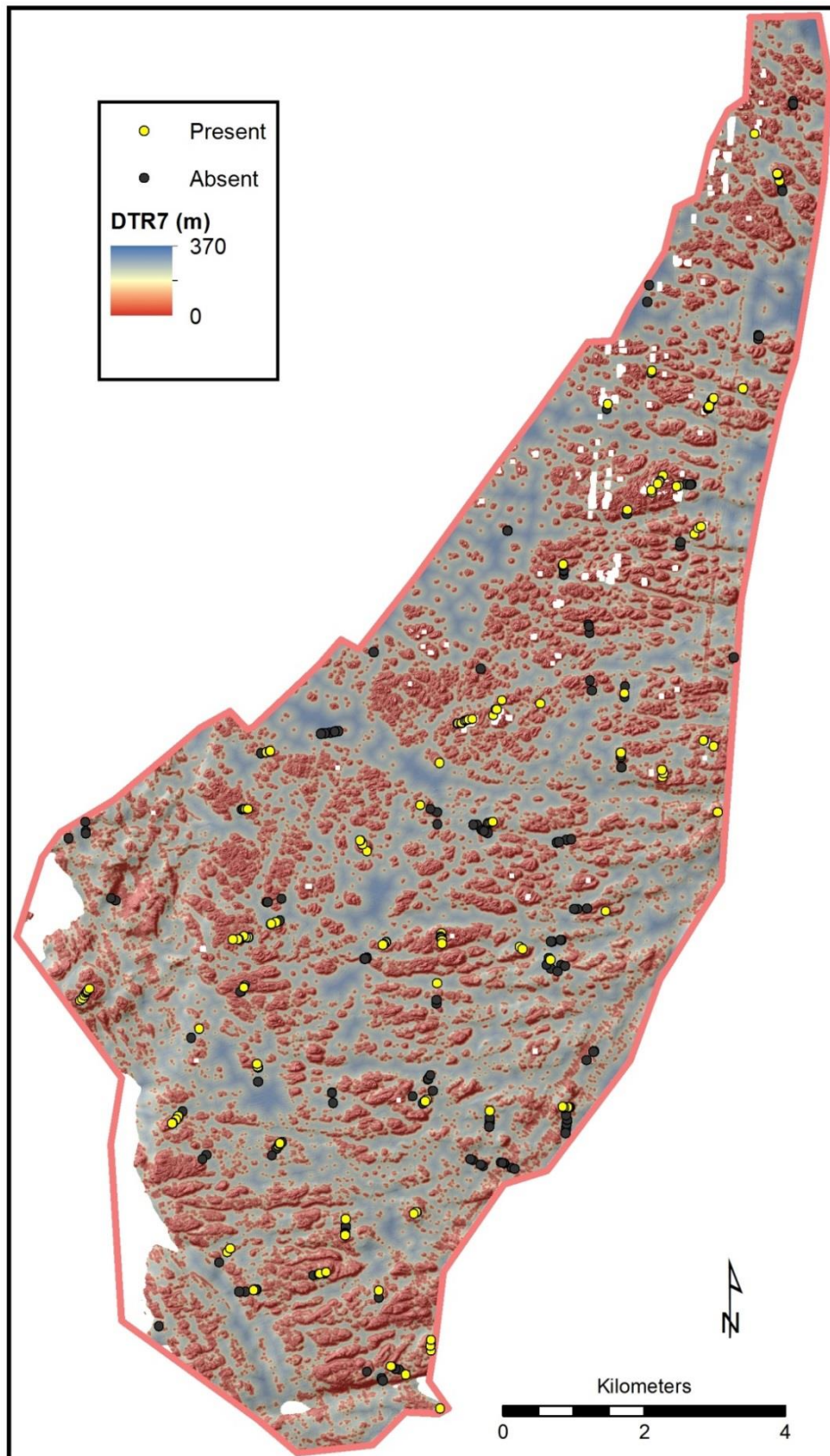
**Figure 9.** Probability of suitable yelloweye rockfish habitat in the Chiswell area based on the GLM fit to the Chiswell area using DTR7, VRM7, Slope, and BPI240. Yellow points are the combined training and validation sets of yelloweye rockfish observations. Black points are the combined training and validation set of absence points.



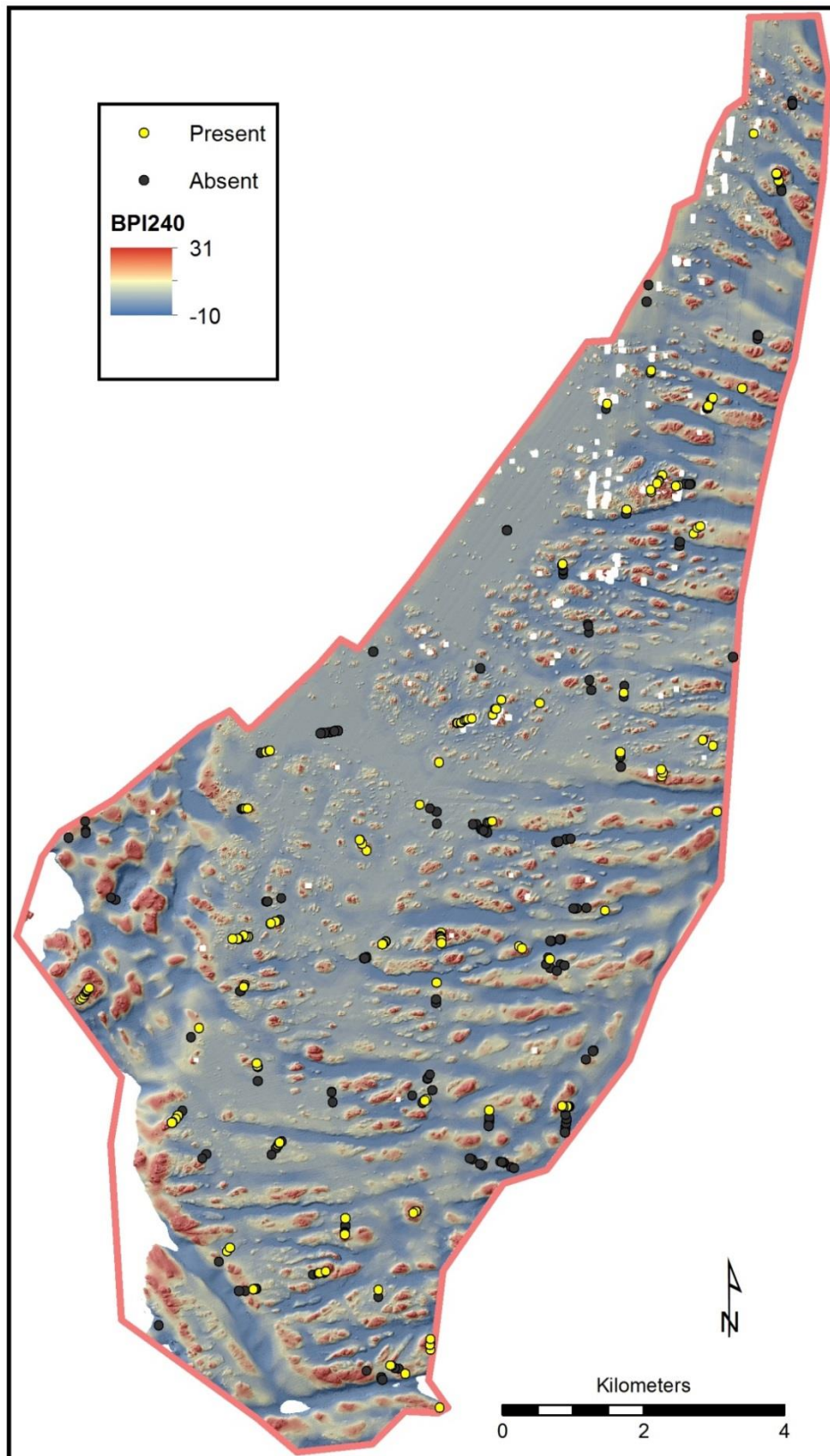
**Figure 10.** Nuka Island study area. Red lines are ROV transects surveyed in 2009. Blue bathymetry was surveyed in 2008. Soundings are in fathoms. Yellow points are the combined training and validation sets of yelloweye rockfish observations (n = 169).



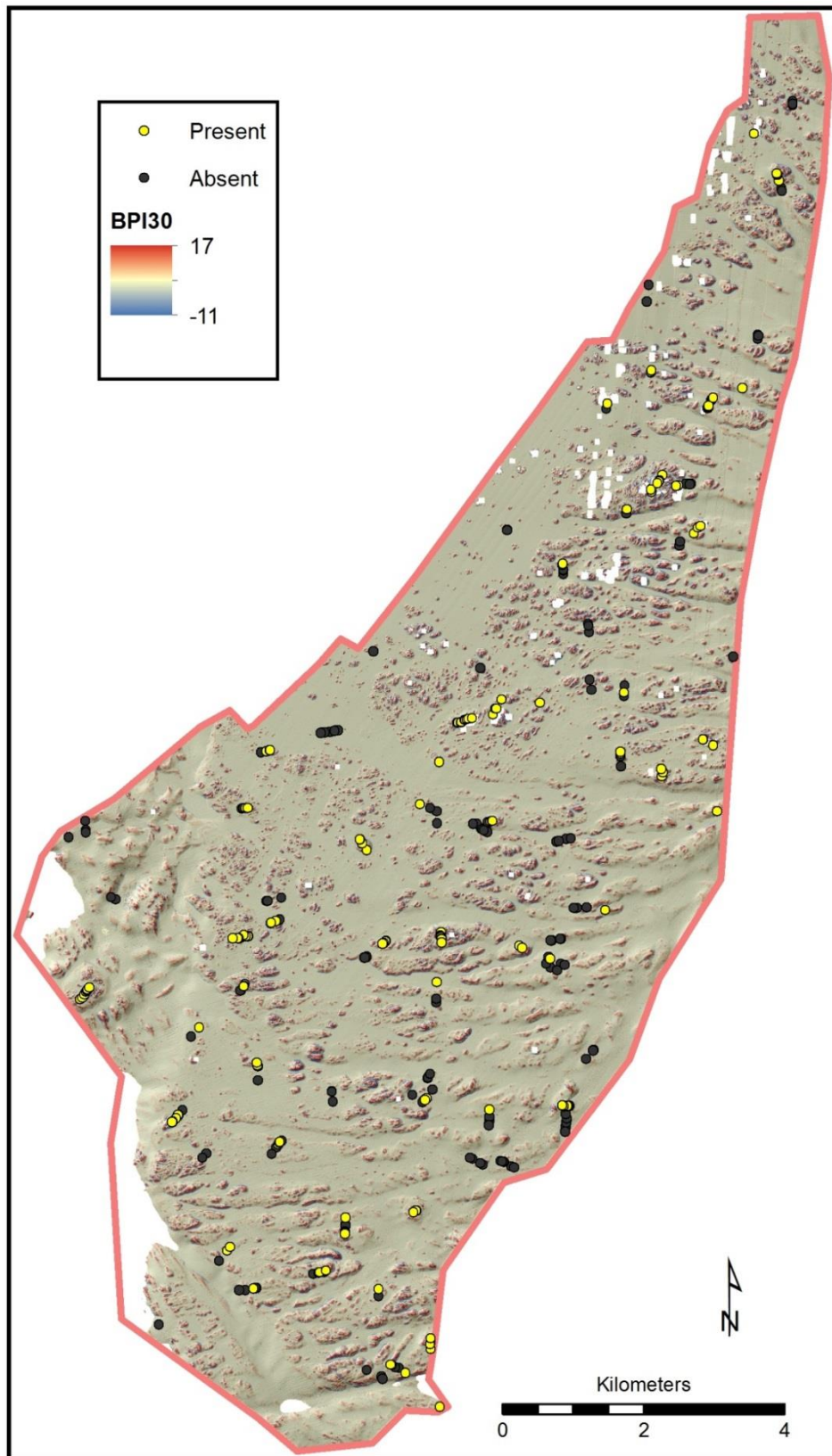
**Figure 11.** Vector rugosity measure calculated with a scale factor of 7 in the Nuka area. Yellow points are the combined training and validation sets of yelloweye rockfish observations. Black points are the combined training and validation set of absence points.



**Figure 12.** Distance-to-rock (DTR) calculated as distance to VRM7 peaks in the Nuka area.

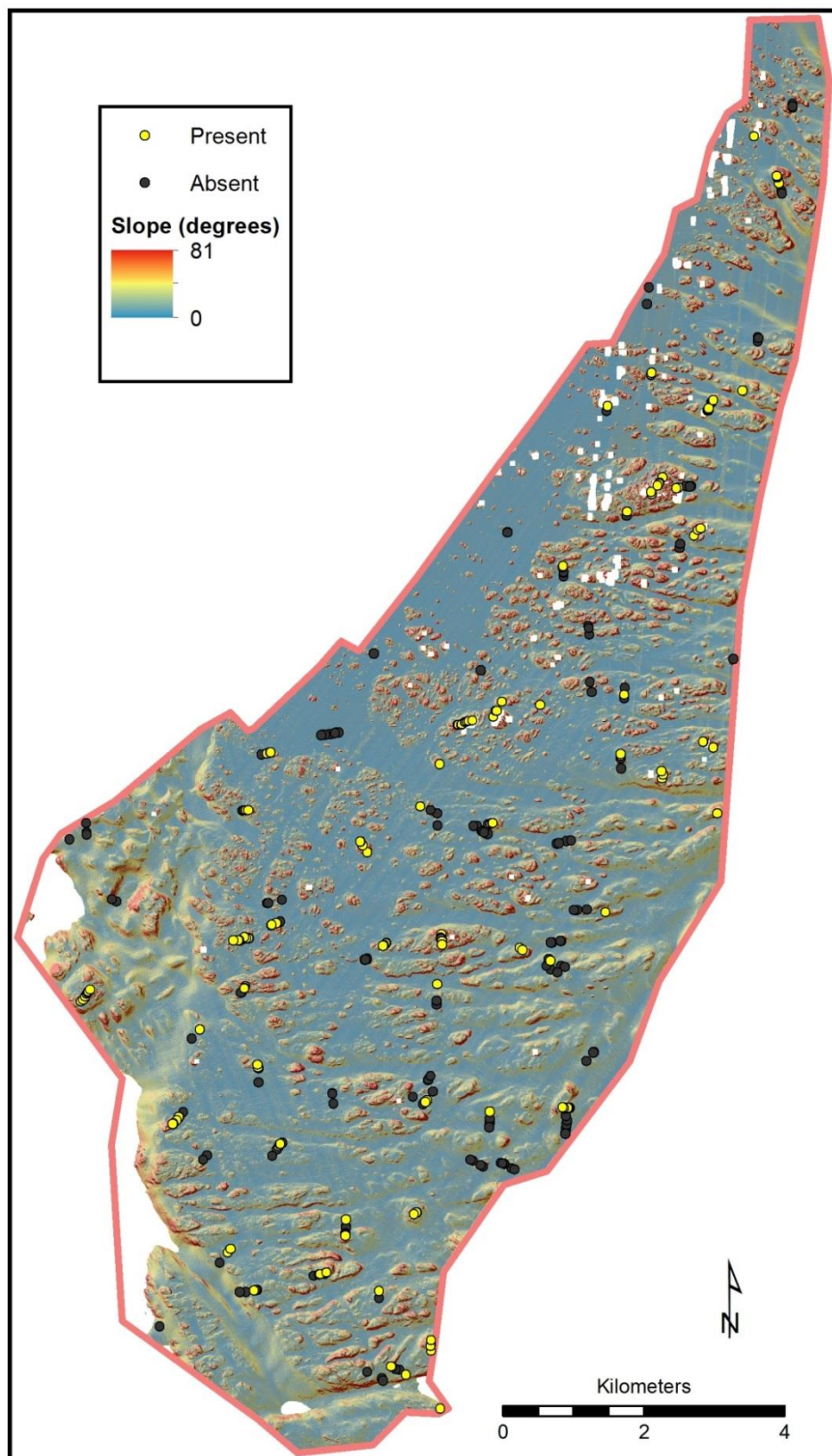


**Figure 13.** Bathymetric position index (BPI) calculated using a scale factor of 240 in the Nuka area.

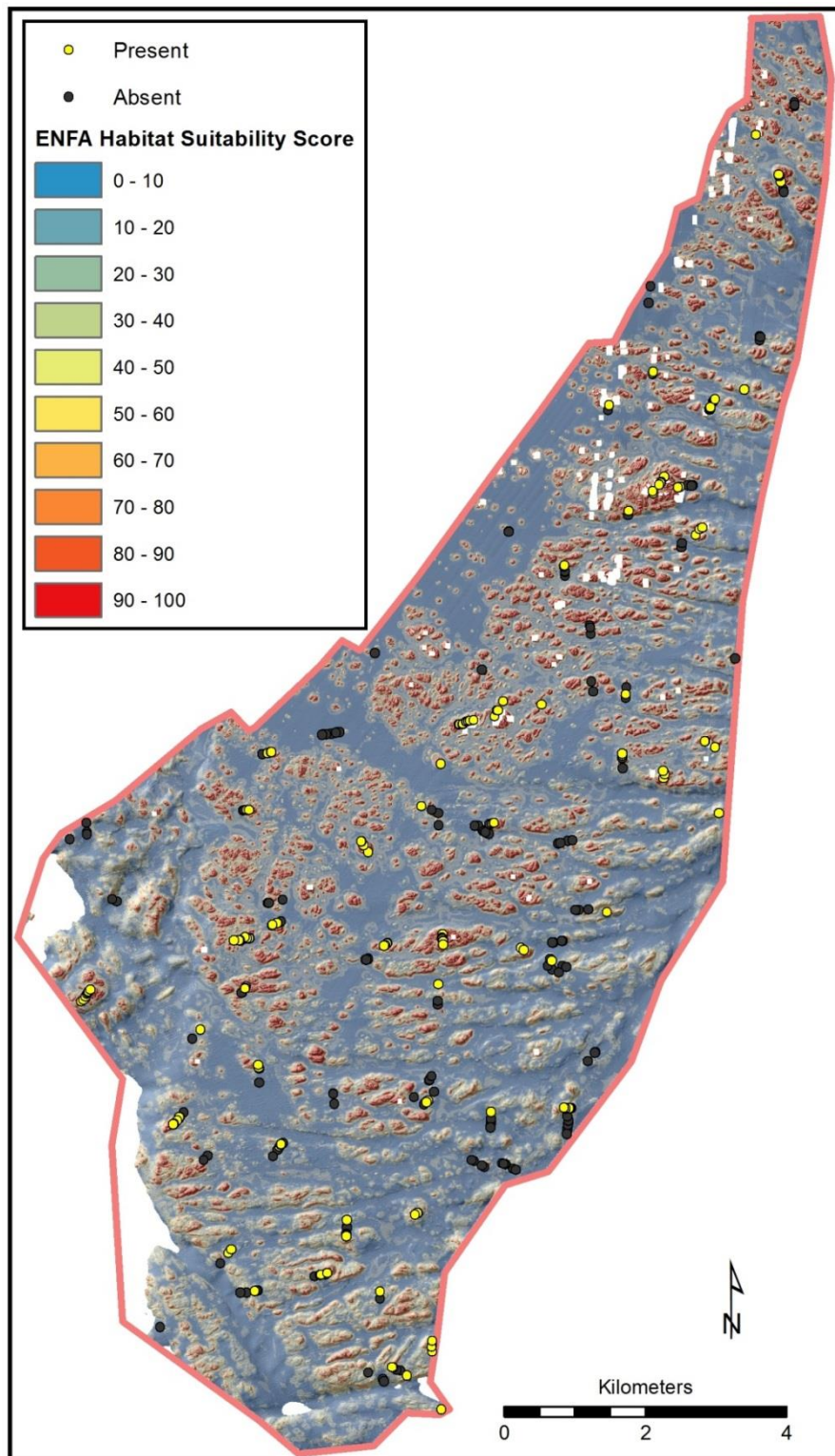


**Figure 14.** Bathymetric position index (BPI) calculated using scale factor of 30 in the Nuka area.

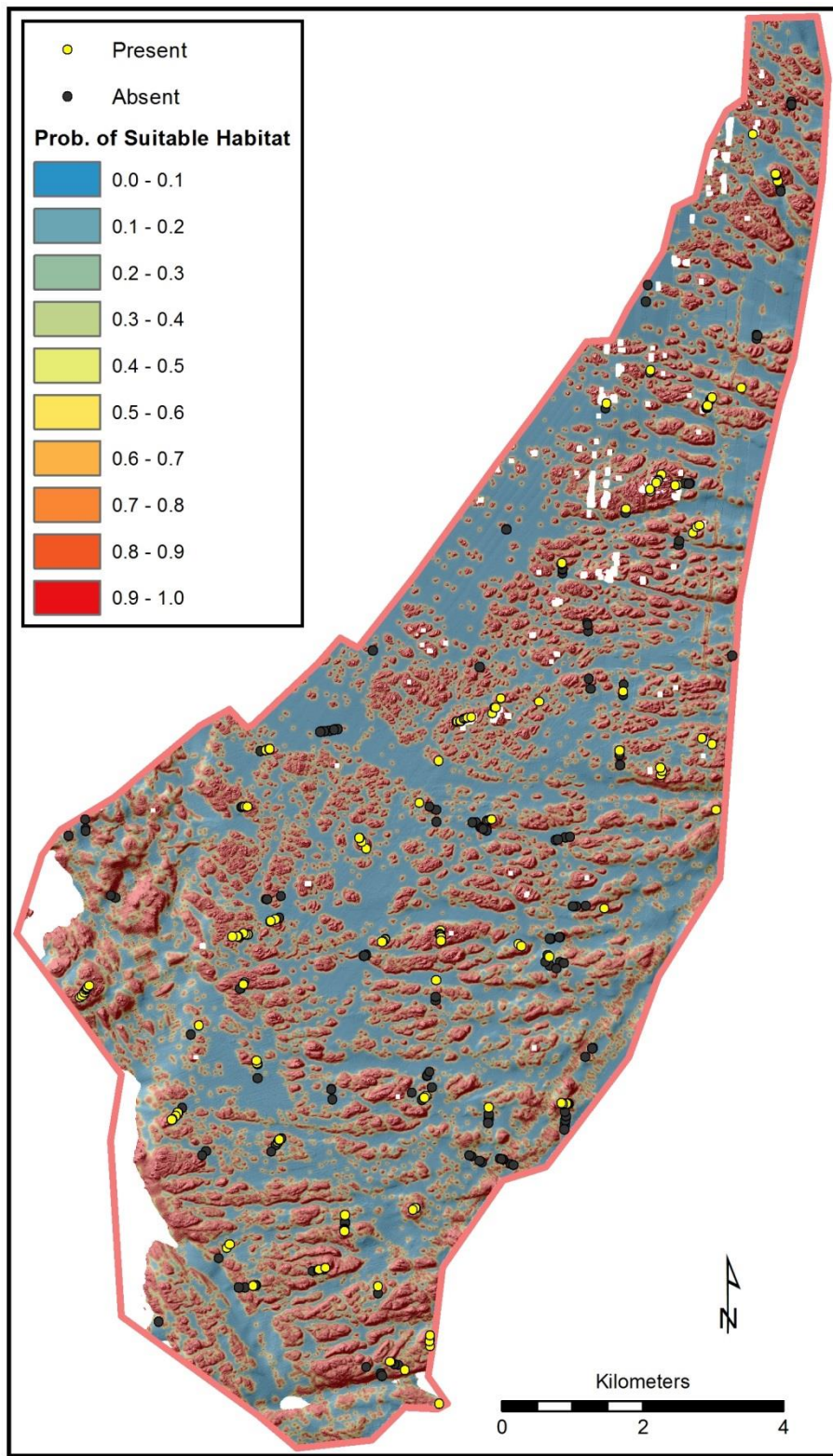




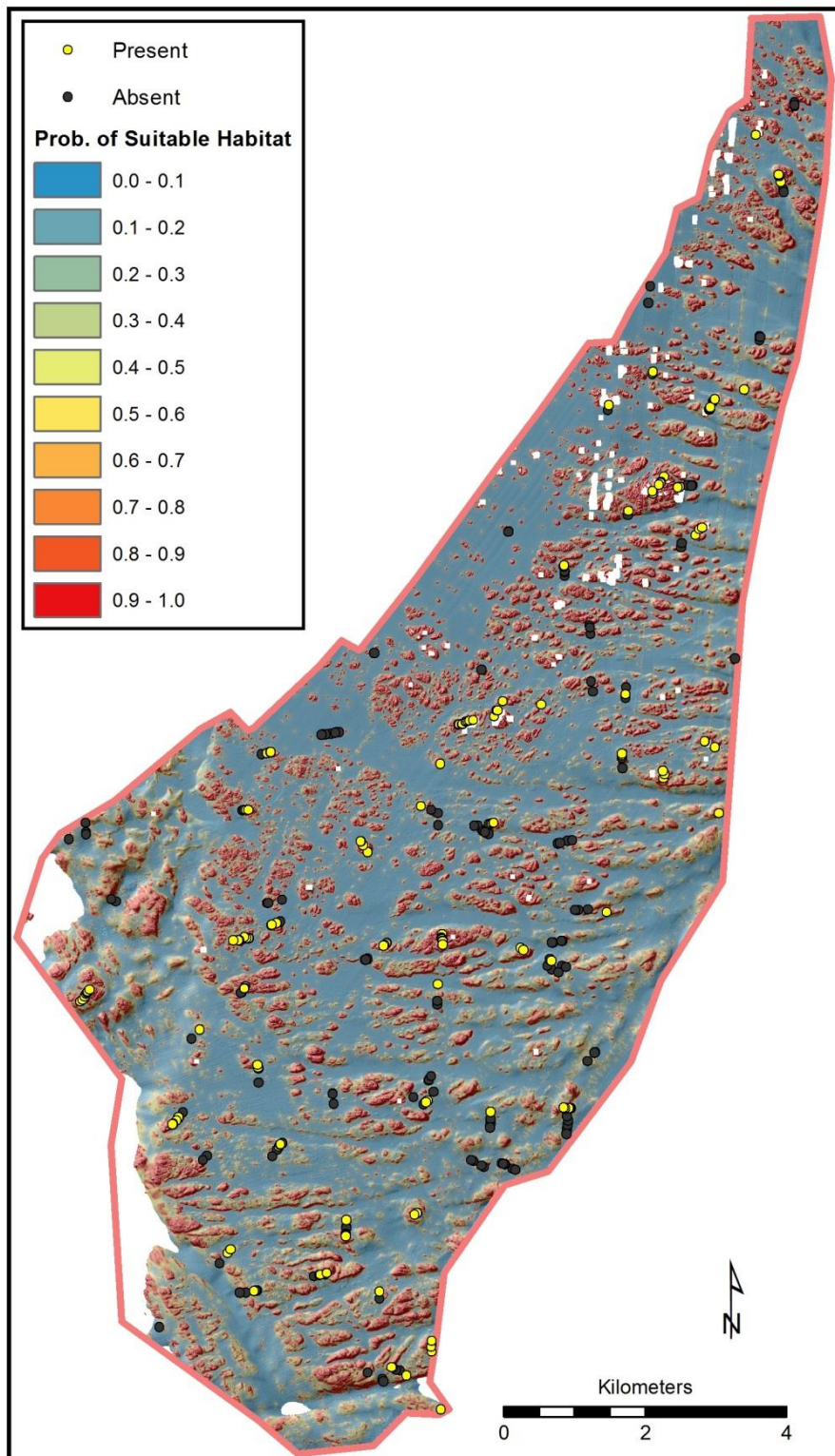
**Figure 15.** Slope in the Nuka area.



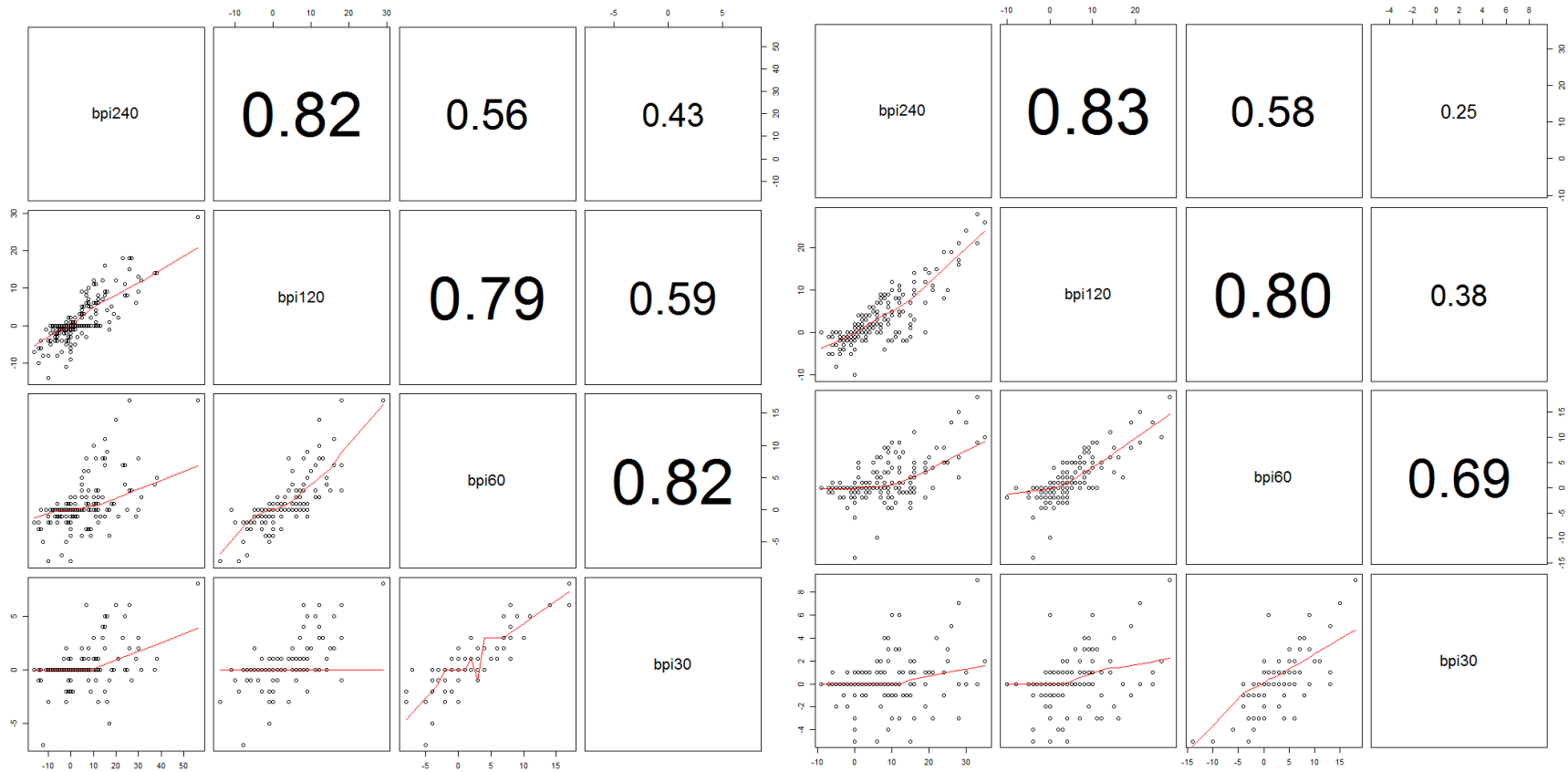
**Figure 16.** Habitat suitability score for yelloweye rockfish in the Nuka area, based on the ecological niche factor analysis. Higher value indicates more suitable habitat. The ENFA used only presence points. Yellow points are the combined training and validation sets of yelloweye rockfish observations. Black points are the combined training and validation set of absence points.



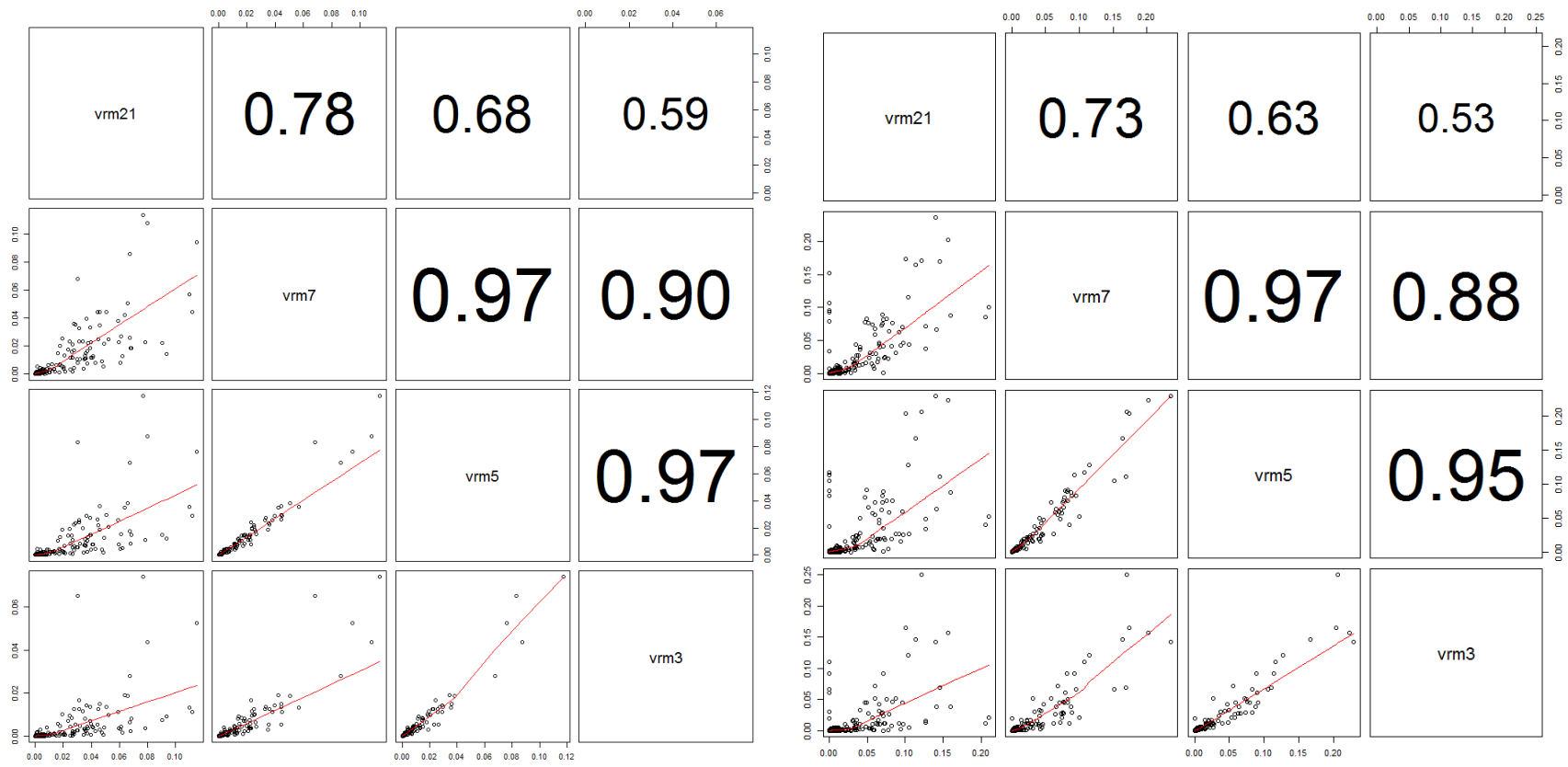
**Figure 17.** Probability of suitable yelloweye rockfish habitat in the Nuka area based on the GLM fit to the Chiswell area using DTR7, VRM7, slope, and BPI240. Yellow points are the combined training and validation sets of yelloweye rockfish observations. Black points are the combined training and validation set of absence points.



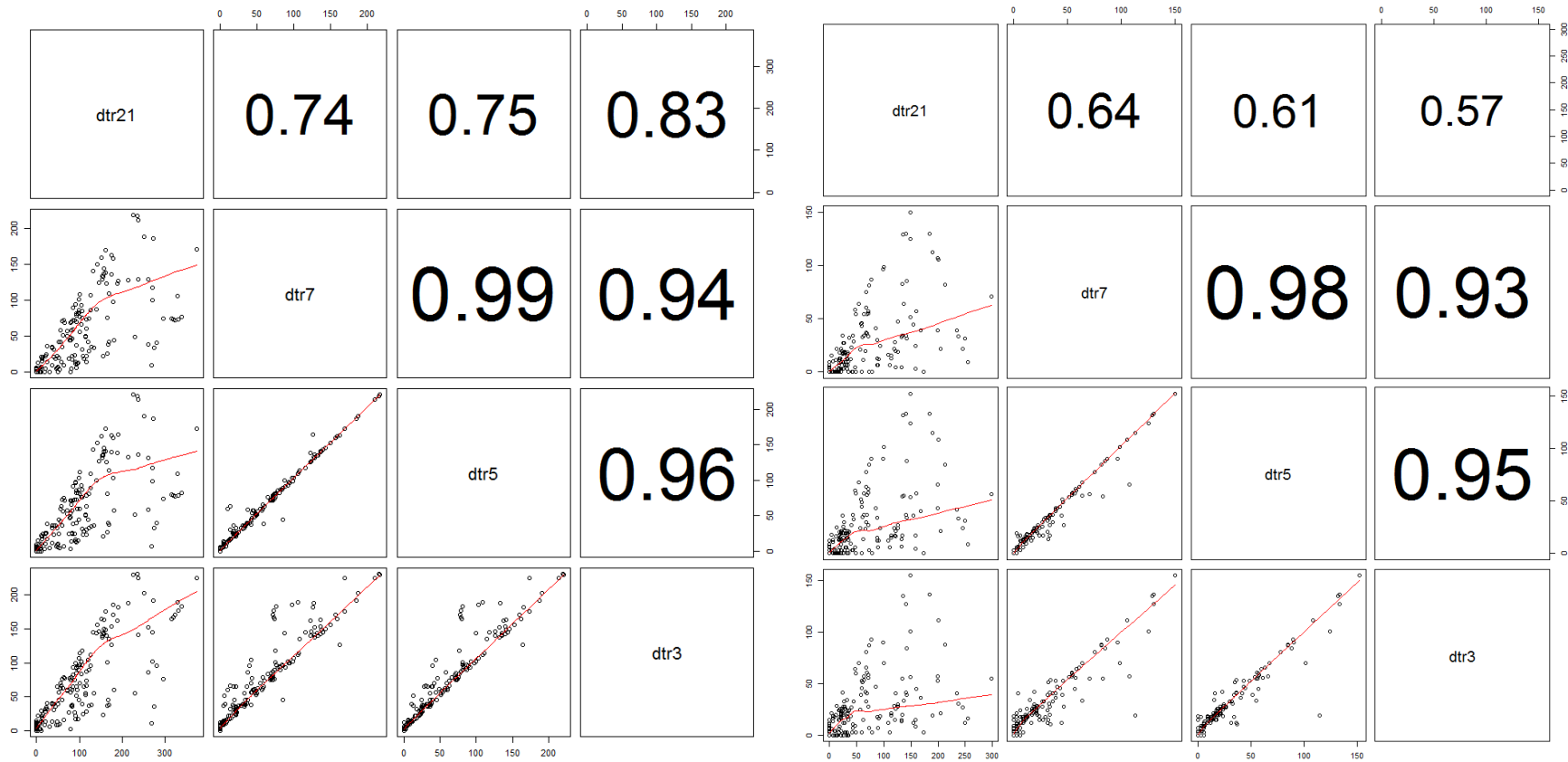
**Figure 18.** Probability of suitable yelloweye rockfish habitat in the Nuka area based on a GLM using the same variables selected for the Chiswell area (DTR7, VRM7, slope, and BPI240) reparametrized to the Nuka area. Yellow points are the combined training and validation sets of yelloweye rockfish observations. Black points are the combined training and validation set of absence points.



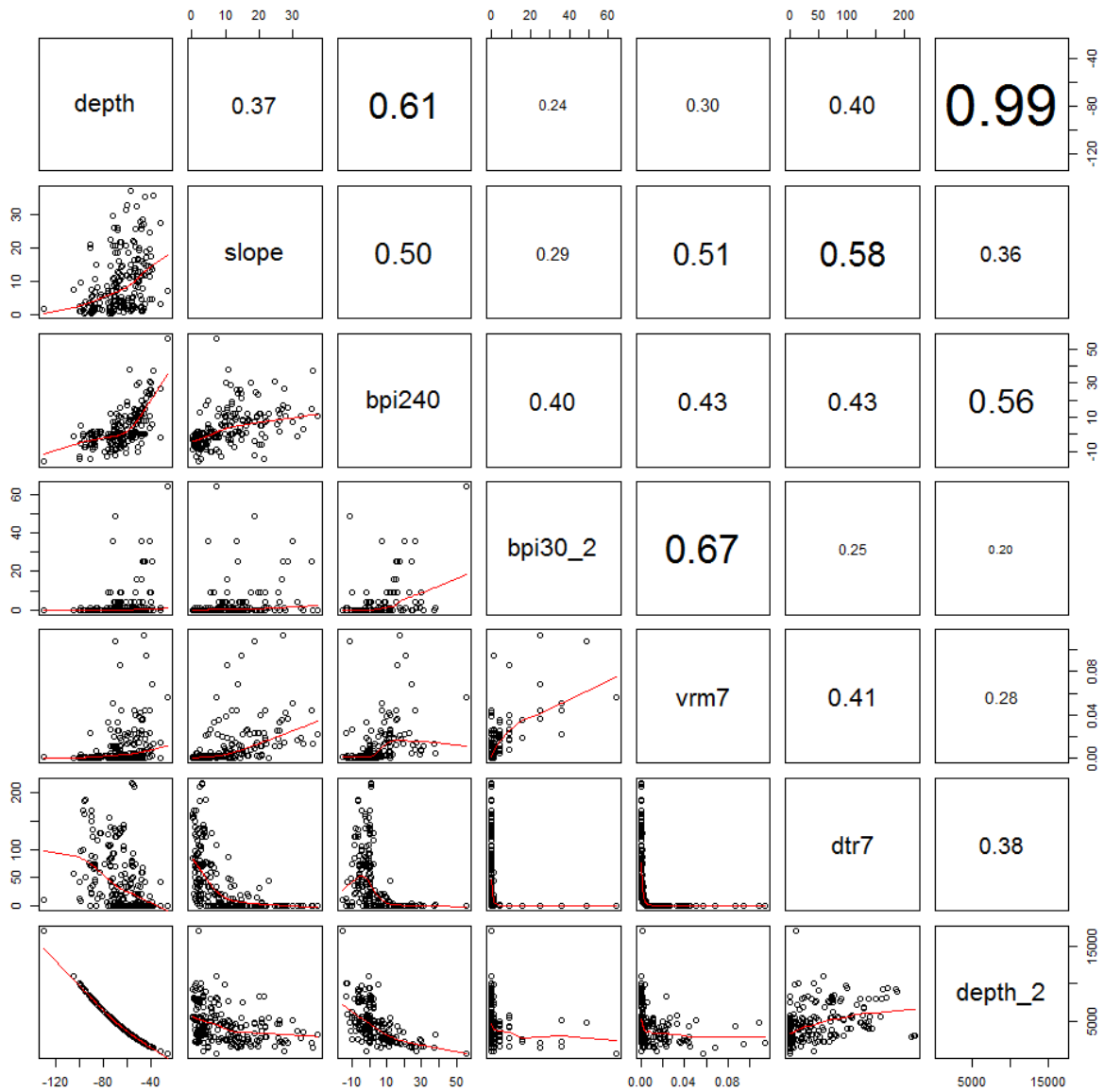
**Figure 19.** Correlogram of the four scales of BPI in the Chiswell area (left) and Nuka area (right). Values are Pearson correlation coefficients.



**Figure 20.** Correlogram of the four scales of VRM in the Chiswell area (left) and Nuka area (right).

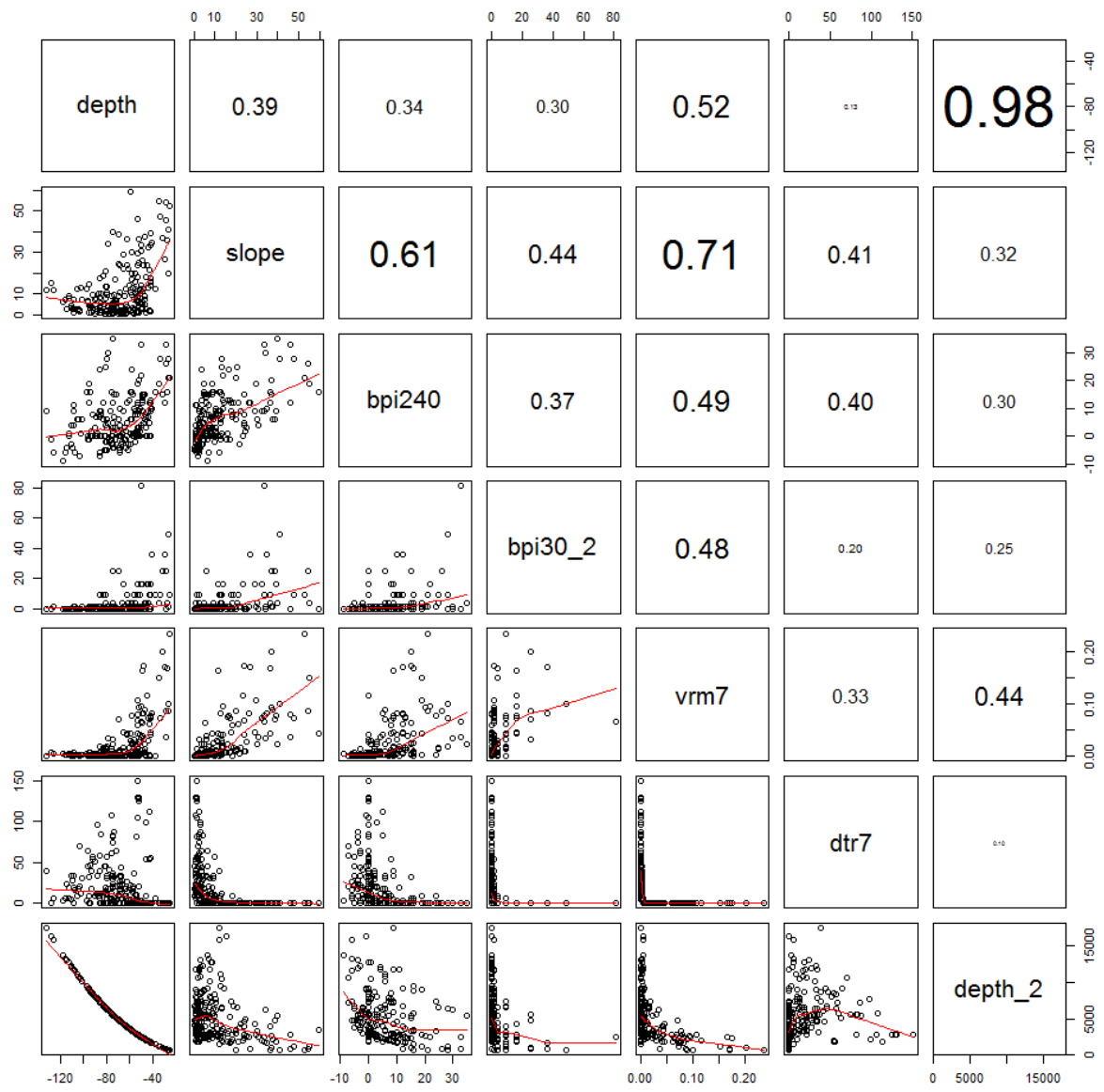


**Figure 21.** Correlogram of the four scales of DTR in the Chiswell area (left) and Nuka area (right).

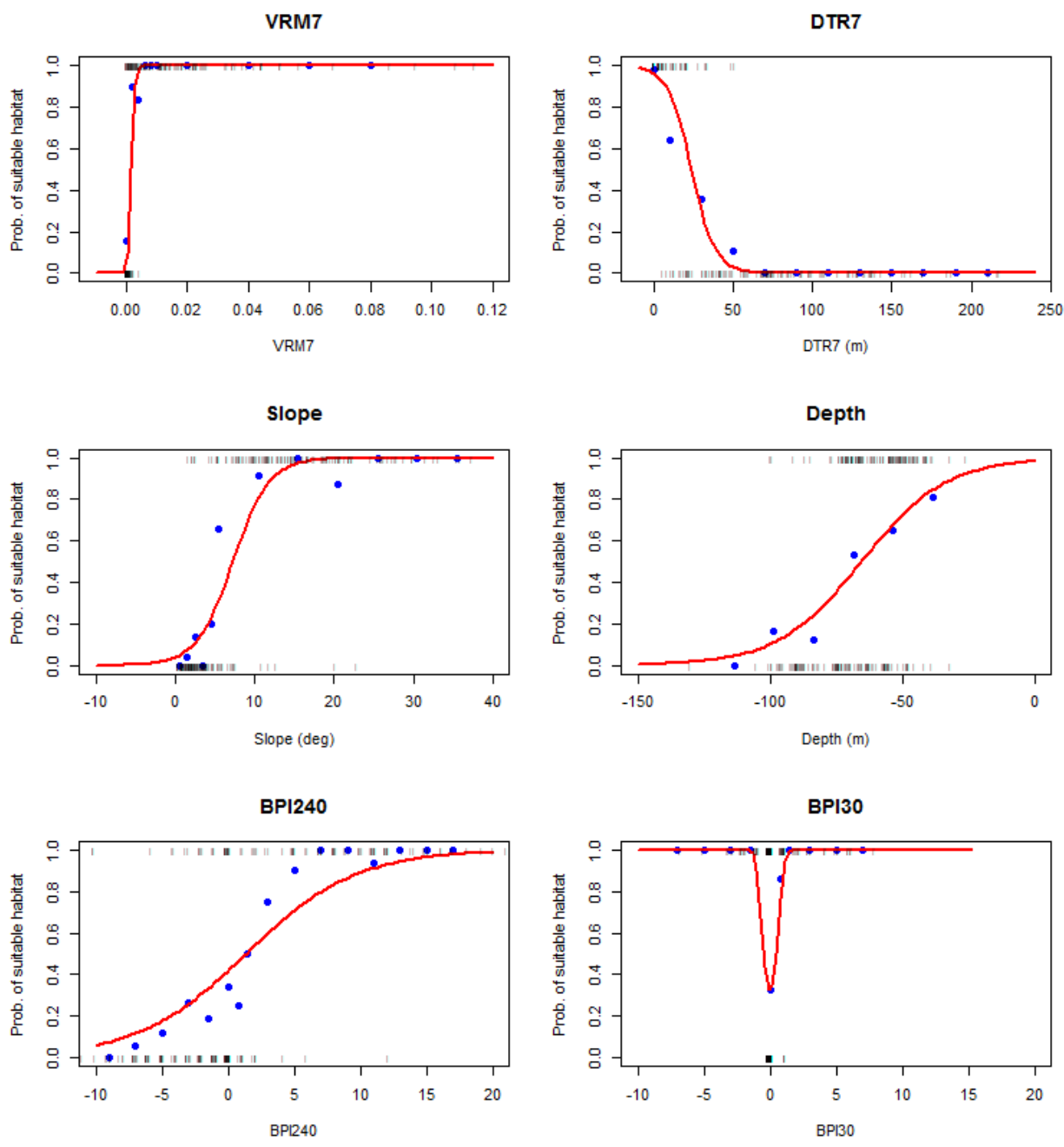


**Figure 22.** Correlogram of the six variables included in the scope of the final stepwise AIC variable selection process for the GLMs using data from the Chiswell area. Subscripts indicate quadratic transformations.

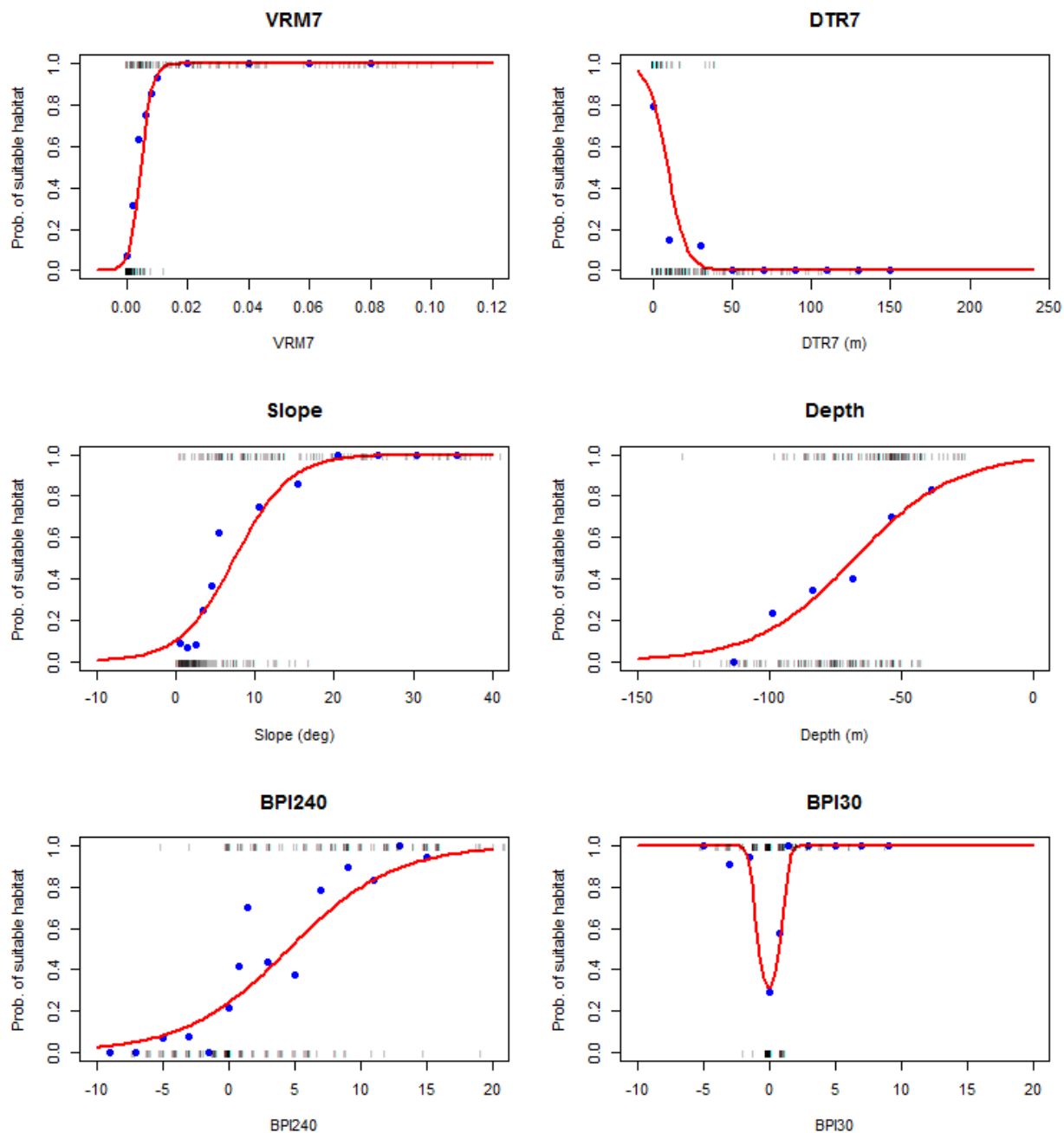




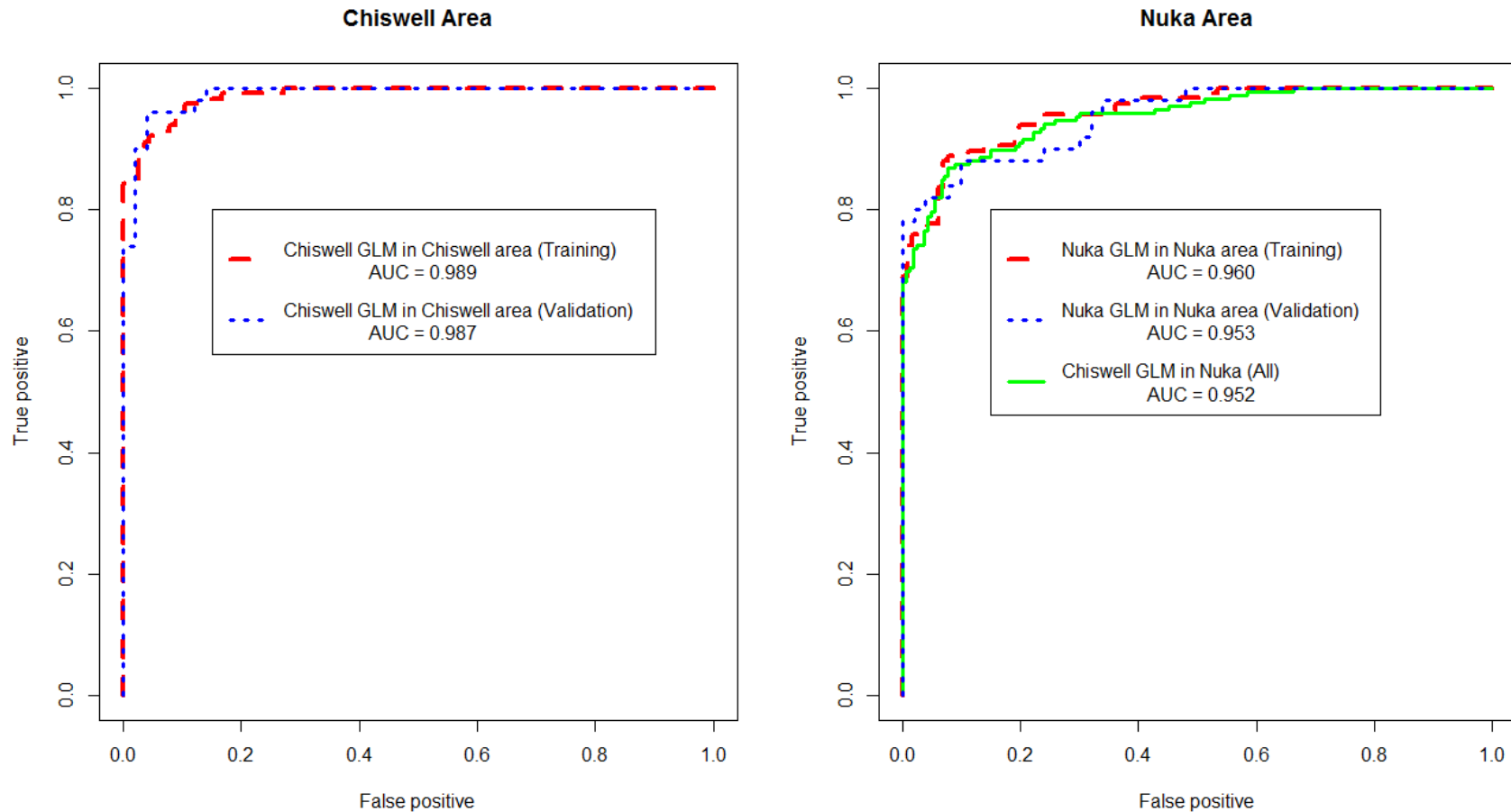
**Figure 23.** Correlogram of the six variables included in the scope of the final stepwise AIC variable selection process for the GLMs using data from the Nuka area. Subscripts indicate quadratic transformations.



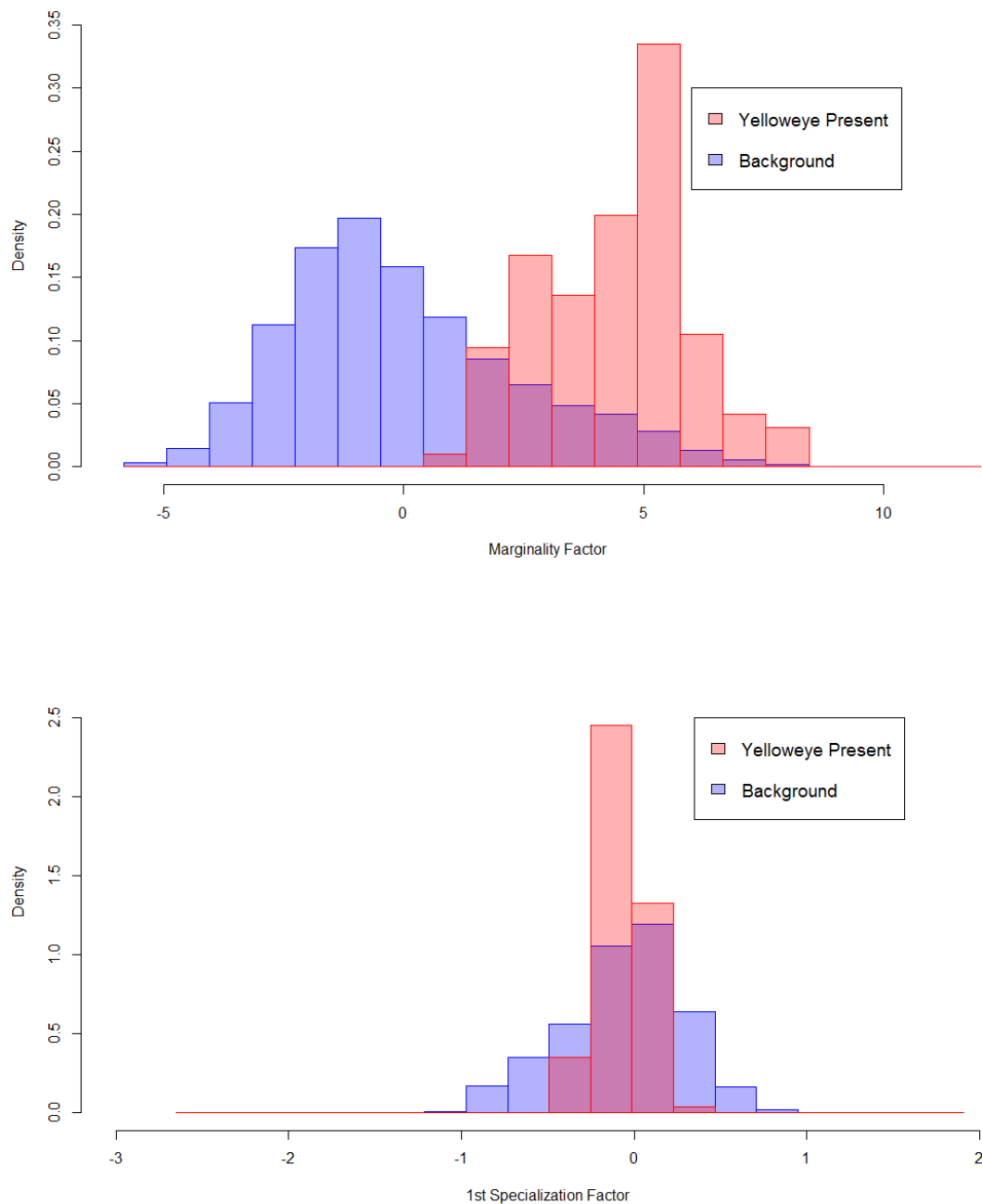
**Figure 24.** Simple logistic regression curves for the Chiswell area relating probability of suitable yelloweye habitat to the six variables selected for inclusion in the final stepwise AIC variable selection. Blue dots are the proportion of presence out of all observations for a given bin. Black ticks are individual point observations.



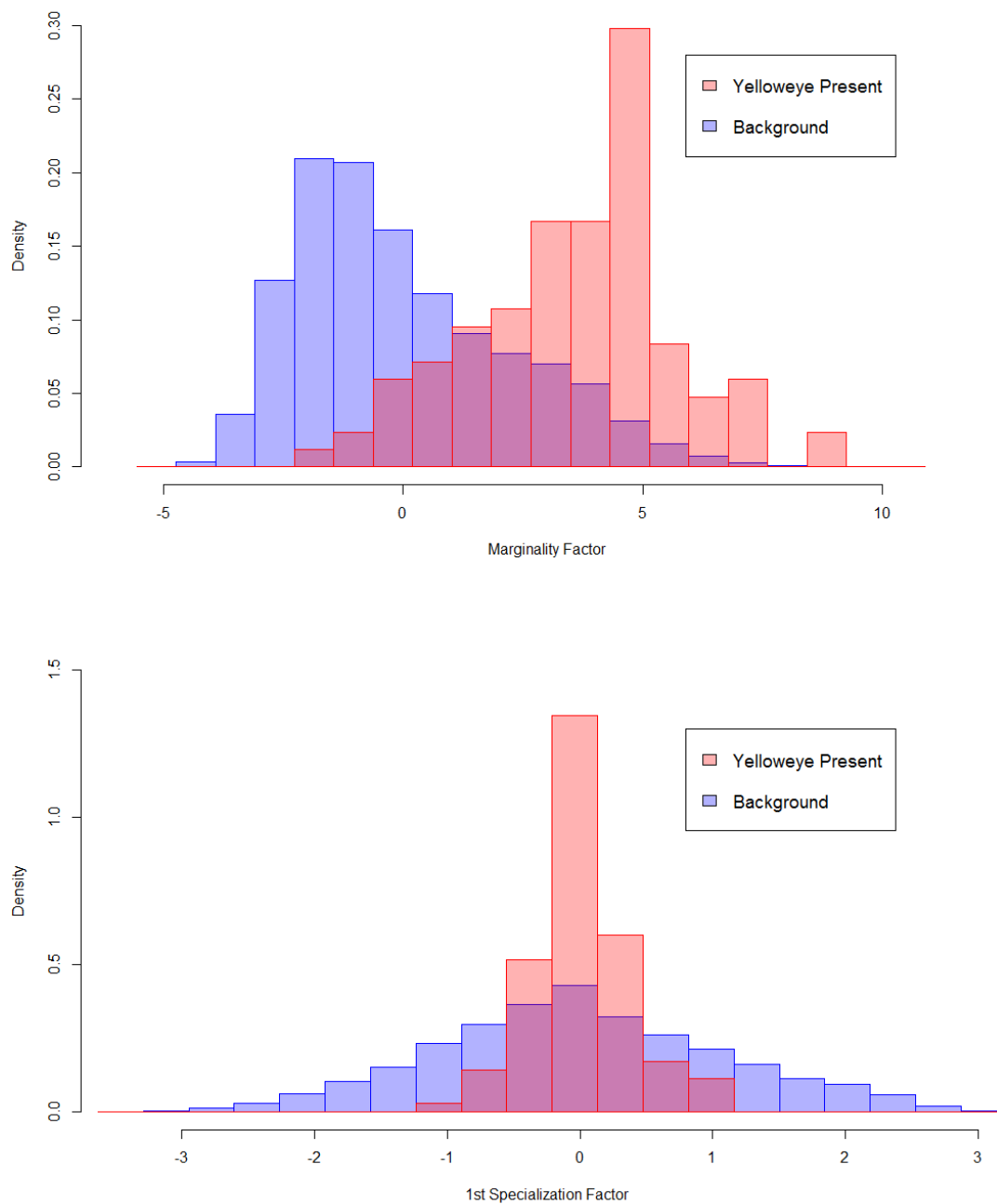
**Figure 25.** Simple logistic regression curves for the Nuka area relating probability of suitable yelloweye habitat to the six variables selected for inclusion in the final stepwise AIC variable selection. Blue dots are the proportion of presence out of all observations for a given bin. Black ticks are individual point observations.



**Figure 26.** Receiver Operator Characteristic (ROC) plots for the GLMs applied to the Chiswell area (left) and Nuka area (right). The blue lines evaluate the performance of the models using the training datasets, while red use the reserved validation set of points. The green line represents the Chiswell model validated in the Nuka area and uses the combined training and evaluation datasets because none of these data were used to fit the model. For reference, a completely random classification would appear as a diagonal line through the origin with slope = 1 and would produce an AUC of 0.5, while a perfect classification would yield an AUC of 1.0



**Figure 27.** Distribution of yelloweye across the first two ecological factors of the ENFA in the Chiswell area. Red is the distribution of yelloweye presence. Blue is the global distribution of the greater study area. Top panel is the marginality factor. Bottom panel is the first specialization factor. The second and third specialization factors were omitted to save space, but showed the same general pattern as the first specialization factor in which the distribution of the presence points was centered about a mean similar to the mean of the background distribution and more narrowly dispersed.



**Figure 28.** Distribution of yelloweye across the first two ecological factors of the ENFA in the Nuka area. Red is the distribution of yelloweye presence. Blue is the global distribution of the entire study area. Top panel is the marginality factor. Bottom panel is the first specialization factor. The second and third specialization factors were omitted to save space, but showed the same general pattern as the first specialization factor in which the distribution of the presence points was centered about a mean similar to the mean of the background and more narrowly dispersed.













349573.42	6614092.28	1	-43.9805984	10.7641001	29	8	-2	-2	16.9706001	0	0	0	3	0.0337098	0.0197911	0.0100861	0.0029589
355792.72	6609938.52	1	-66.1872025	18.2383995	-6	-7	-5	-2	21.6333008	0	0	0	0	0.0260114	0.0107038	0.0089888	0.0050938
354967.21	6607400.91	1	-65.1634979	35.0905991	8	5	5	5	0	0	0	0	0	0.0855916	0.0601711	0.0387505	0.013152
354992.37	6607482.52	1	-66.4433975	5.2332501	6	0	0	-1	9.4868298	0	0	0	0	0.0536592	0.031032	0.0219516	0.0080366
351906.46	6607203.18	1	-58.1380997	40.3535995	12	14	7	3	0	0	0	0	4.24264	0.0763509	0.0156668	0.0072534	0.0017323
350901.41	6608438.34	1	-70.7012024	9.7949104	-1	0	0	0	19.2094002	16.1555004	6.7082	8.48528	13.4164	0.0123956	0.0022684	0.0016404	0.0013169
348754.48	6594793.71	1	-82.0781021	4.84024	-2	-6	-2	0	27.6585999	3	6	9	12	0.0187779	0.0007119	0.0004381	0.000194
348663.36	6594842.77	1	-40.1402016	47.9328995	38	30	10	3	0	0	0	0	0	0.0339036	0.0379821	0.0410109	0.0308605
352115.04	6604113.88	1	-76.1732025	20.2747002	-8	-3	1	0	6.7082	0	0	0	0	0.036662	0.0255658	0.0166957	0.0078643
351557.38	6604982.35	1	-66.0145035	20.6208	0	-9	-9	-2	42.9534988	0	0	0	0	0.0409909	0.0165899	0.0122845	0.0054398
353917.75	6616537.56	1	-58.3723984	14.7188997	10	1	-1	-1	15	0	0	0	0	0.0254387	0.0175362	0.0120947	0.005089
353710.68	6616533.69	1	-51.3241005	6.68717	0	-10	-4	0	21.2131996	0	0	3	6.7082	0.0404681	0.0080596	0.0028593	0.0008391
351363.06	6605683.40	1	-49.5354004	40.2162018	17	9	9	2	3	0	0	0	3	0.0678744	0.0196865	0.0116725	0.0049427
351359.94	6605680.86	1	-47.1629982	33.6217003	20	11	11	3	0	0	0	0	3	0.0813375	0.0185788	0.0070134	0.0029919
344103.84	6610354.29	1	-66.3350983	24.4067001	9	11	9	2	3	0	0	0	4.24264	0.0378608	0.0123925	0.0059302	0.0011417
344103.34	6610354.76	1	-66.3350983	24.4067001	9	11	9	2	3	0	0	0	4.24264	0.0378608	0.0123925	0.0059302	0.0011417
343784.84	6611001.89	1	-45.7733002	16.1966	6	-4	-1	0	52.3927002	61.8465996	20.1245995	22.8472996	28.4605007	0.007037	0.002906	0.0028278	0.0015876
342913.52	6612098.07	1	-57.4389992	17.3871002	-9	-6	0	0	17.4929008	47.4342003	0	0	10.8167	0.0096024	0.0072387	0.0054361	0.0026448
342696.63	6611162.86	1	-70.5700989	4.5036702	0	-2	-2	-1	48.4664993	46.6689987	0	6.7082	48.3735008	0.0062947	0.0055261	0.0042409	0.002106













635513.85	6577314.1	1	-59.456	48.138199	15	10	3	-2	4.24264	0	0	0	0	0.121434	0.0301936	0.0195699	0.0131705
635527.26	6577334.4	1	-67.549	5.54421	7	1	-4	-3	18.9737	0	0	0	3	0.0793757	0.0142519	0.0087065	0.0021119
635594.5	6577460.6	1	-70.109	4.98803	2	5	3	1	19.2094	85.802101	6	9	12.3693	0.0151145	0.0027041	0.0010755	0.0003864
634873.74	6576363.7	1	-40.309	28.9361	12	2	4	1	3	0	0	0	0	0.124038	0.11992	0.0601372	0.0134109
635073.91	6576209.7	1	-51.151	10.1531	6	3	4	1	12.7279	12	0	0	6	0	0.0115265	0.0052638	0.0029738
635378.61	6575617.5	1	-74.182	19.5877	8	0	-5	-3	13.4164	0	0	0	0	0.0676643	0.0269931	0.0234936	0.0155721
633458.26	6575102.8	1	-36.912	28.077801	15	17	9	3	0	0	0	0	0	0.0875525	0.0511332	0.0334159	0.010914
635650.51	6571590	1	-82.12	10.7042	18	15	4	0	12.3693	0	0	0	0	0.0226551	0.0143344	0.0137124	0.0075443
634865.07	6572087.7	1	-49.97	17.258499	25	14	5	4	0	0	0	0	0	0.053775	0.0846021	0.076279	0.0492499
634878.92	6572149.9	1	-54.916	8.5229597	15	8	4	2	8.48528	0	0	0	3	0.0250575	0.0107932	0.0069896	0.002465
634275.13	6572435.6	1	-51.159	31.393101	14	13	9	0	3	21	0	0	0	0	0.0819229	0.0723823	0.0507803
636505.28	6580627.3	1	-24.192	38.368198	37	26	12	3	0	0	0	0	0	0.131005	0.0474368	0.045993	0.025541
636498.3	6580641.2	1	-17.161	18.355801	44	34	18	5	0	0	0	0	0	0.122354	0.0384623	0.0344409	0.0234789
632516.1	6573044.9	1	-52.895	3.3408699	1	-2	-1	0	21.8403	4.24264	8.48528	10.8167	13.4164	0.0162097	0.0011853	0.0012464	0.0007374
632590.87	6573175.9	1	-46.936	25.2794	8	9	3	2	3	0	0	0	0	0.0382845	0.037226	0.0205271	0.0108466
631956.75	6572846.3	1	-37.136	44.069401	18	11	3	4	3	0	0	0	0	0.111268	0.110371	0.0793285	0.0357547
632021.04	6572856	1	-40.243	17.4195	15	7	2	3	0	0	0	0	0	0.0702612	0.0707946	0.0662672	0.0430057
632098.71	6572894.1	1	-30.725	5.6899199	24	18	10	4	0	6	0	0	0	0	0.114753	0.0856579	0.0766664
630607.85	6571143.7	1	-42.201	20.6968	16	11	0	-5	15	0	0	0	0	0.131972	0.0831922	0.0868334	0.096781
629310.2	6570006.8	1	-57.081	1.0169801	0	-1	0	0	50.2892	32.311001	24.1868	25.806999	26.8328	0.0035237	0.0010213	0.0006329	0.000235