Modeling Habitat Use of Young-of-the-Year Pacific Sand Lance (*Ammodytes hexapterus*) in the Nearshore Region of Barkley Sound, British Columbia.

by

Trevor Bruce Haynes
BSc. Simon Fraser University, 2001

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In the Department of Geography

University of Victoria

Modeling Habitat Use of Young-of-the-Year Pacific Sand Lance (*Ammodytes hexapterus*) in the Nearshore Region of Barkley Sound, British Columbia.

by

Trevor Bruce Haynes
BSc. Simon Fraser University, 2001

SUPERVISORY COMMITTEE

Dr. P. Dearden, (Department of Geography)

Supervisor

Dr. C. Robinson, (Department of Geography)

Co-Supervisor

Dr. D. Duffus (Department of Geography)

Departmental Member

Dr. A. Burger (Department of Biology)

Outside Member

Supervisory Committee

Supervisor

Co-Supervisor

Departmental Member

Outside Member

ABSTRACT

Successful management of coastal ecosystems requires an understanding of the distribution of key food web species through space and time relative to environmental predictors. Here, I examined the habitat use of an important forage species, the Pacific sand lance (*Ammodytes hexapterus*), using an inductive habitat modeling approach. I examined the presence/absence of Young-of-the-Year Pacific sand lance in the intertidal/shallow subtidal habitat of Barkley Sound, British Columbia. I determined sand lance occurrence using a beach seine at low tide, which was preferred to visual and intertidal digging detection methods due to its high detection frequency, ease of use, and ability to physically capture sand lance. I constructed models using environmental data measured at two different scales: 1) empirically measured environmental data (site-specific level) and 2) GIS derived environmental data extracted with a 200m buffer (landscape level). For each scale, I employed both logistic regression and classification tree modeling procedures to construct habitat models of sand lance occurrence at 55 study sites sampled during the summer of 2003. At the site-specific level, both logistic regression and classification tree models performed similar, however, classification trees were easier to construct and interpret as well as revealing interactions among variables undetected by logistic regression. Based on a deviance pruned classification tree, *Grain Size Mean*, *Intertidal Eelgrass Presence/Absence*, *Major Substrate Low Intertidal*, and *Grain Size Sorting* influenced sand lance occurrence at this scale, with importance values of 100, 79, 75, and 61 respectively. Standardized importance was based on the overall change in node impurity in the classification tree for each variable. At the landscape level, only *Coastline Density* was significantly related to sand lance occurrence, however, it was difficult to suggest this variables direct relation to sand lance habitat use. Overall, the habitat modeling approach identified important environmental variables influencing sand lance habitat selection at two different scales and stressed the utility of field data to construct and confirm these models.

TABLE OF CONTENTS

SUPERVISORY COMMITTEE……………………………………………………………ii

ABSTRACT………………………………………………………………………………iii

TABLE OF CONTENTS……………………………………………………………………v

LIST OF TABLES…………………………………………………………………………vii

LIST OF FIGURES………………………………………………………………………ix

LIST OF APPENDICIES…………………………………………………………………xii

ACKNOWLEDGEMENTS………………………………………………………………xiii

CHAPTER 1: General Ecology of the Pacific Sand Lance and Study Objective…… 1

CHAPTER 2: General Sampling Methods for Sand Lance and Environmental Data……………………………………………………………………………………6

- 2.1 STUDY SITE…………………………………………………………………6
- 2.2 SAMPLING SUMMARY……………………………………………………9
- 2.3 SAND LANCE SAMPLING…………………………………………………10
- 2.4 YEARLY SAMPLING………………………………………………………12
- 2.5 ENVIRONMENTAL SAMPLING……………………………………………15

CHAPTER 3: Comparison of Survey Methods for Sampling Sand Lance in the Nearshore……………………………………………………………………………24

- 3.1 INTRODUCTION……………………………………………………………24
- 3.2 METHODS……………………………………………………………………24
- 3.3 RESULTS……………………………………………………………………28
- 3.4 DISCUSSION…………………………………………………………………32
- 3.5 CONCLUSION………………………………………………………………36

CHAPTER 4: Temporal Variation in Sand Lance Distribution and Abundance....37

4.1 INTRODUCTION.........37

4.2 METHODS.........37

4.3 RESULTS.........38

4.4 DISCUSSION.........43

4.5 CONCLUSION.........47

CHAPTER 5: Modeling Nearshore Habitat for Young-of-the-Year Sand Lance: An Empirical Inductive Approach.........49

5.1 INTRODUCTION.........49

5.2 METHODS.........54

5.3 RESULTS.........67

5.4 DISCUSSION.........88

5.5 CONCLUSION.........105

CHAPTER 6: Modeling Nearshore Habitat for Young-of-the-Year Sand Lance: A GIS Inductive Approach.........107

6.1 INTRODUCTION.........107

6.2 METHODS.........107

6.3 RESULTS.........112

6.4 DISCUSSION.........117

6.5 CONCLUSION.........120

CHAPTER 7: Conclusions.........122

REFERENCES.........127

APPENDICIES.........147

LIST OF TABLES

Table 2.1: Beginning and end sample dates for each year of sampling and the number of sample days and sample sites. Note that for some sample days, multiple sites were sampled. Also, a number of sample sites were sampled multiple times (not displayed here)……………………………………………………………………………10

Table 2.2: Descriptive statistics for tide height based on beach seine set for 2003……12

Table 2.3: Classification of *Beach Log Cover*……………………………………………16

Table 2.4: Qualitative substrate categories………………………………………………18

Table 2.5: Macro-vegetation categories used for underwater video analysis……………19

Table 2.6: Sieve mesh sizes used for dry sieving and the corresponding phi size class (-log2 mesh size) and particle description according to Wentworth classification…………21

Table 3.1: Frequency of surveys and detections for each survey method conducted in 2003. Note that due to logistical constraints, not all survey types could be conducted at each site. Total number of sites sampled was 60……………………………………………25

Table 3.2: Results from Chi-square analysis testing sampling methods against the normative standard…………………………………………………………………………28

Table 3.3: Comparison of detection methods for sites that were designated as sand lance present for at least one method of sampling (the normative standard, N=35)……………31

Table 3.4: Mann-Whitney U results comparing sand lance mean rank abundance grouped by visual survey presence/absence………………………………………………………31

Table 3.5: Logistic regression results using survey method detection (presence/absence) as the dependent and secchi depth (water clarity) as the independent variable…………………………………………………………………………………31

Table 3.6: Table 3.6: Comparison of marginal mean estimators for each tide height (controlling for the variation between sample sites)…………………………………………32

Table 4.1 Within-year consistency of sand lance occurrence for 15 sites with multiple sampling events in 2002. Values represent abundances based on the catch values of the beach seine at low tide. Dates with more than one sample event during the same time period are separated by commas……………………………………………………………39

Table 4.2: Comparison of sand lance abundance (*Sand lance Set Abundance*) for sites sampled in all three years of the study. Abundance is based on low tide beach seine data (sand lance per seine set)……………………………………………………………………40

Table 4.3: Comparison of sand lance occurrence among sites sampled in both 2003 and 2004. Presence/absence designation is based on low tide beach seine data................42

Table 4.4: Comparison of mean abundance for each year (sand lance per site, N=37)...42

Table 5.1: Dependent variable list considered for analysis.................................58

Table 5.2: Results from the ANOSIM for each dependent variable. *Sand lance Presence/Absence* was the only variable that was significant at the 0.05 level............60

Table 5.3: Mann-Whitney U test for each continuous independent variable grouped by each dependent variable. * Independent variable found to be significant at the 0.05 level..73

Table 5.4: Chi-square significance test for independent categorical variables grouped by four dependent variables. * Significant at the 0.05 level.......................................74

Table 5.5: Logistic regression modeling results based on the selected independent variables using *Sand lance Presence/Absence* as the dependent variable (N=55). *Independent variable significant at the 0.05 level (Wald statistic). **Model Chi-square value significant at the 0.05 level...78

Table 5.6: Logistic regression modeling results for sites without intertidal eelgrass (N=43). Modeling was based on the selected independent variables using *Sand lance Presence/Absence* as the dependent variable. *Independent variable significant at the 0.05 level (Wald statistic) **Model Chi-square value significant at the 0.05 level 79

Table 5.7: Cross-validation results for logistic regression models.........................80

Table 5.8: Classification rates for the two classification tree models......................82

Table 6.1: Results from the two-sample independent t-test analysis of independent variables grouped by the binary dependent variable *Sand lance Presence/Absence*......112

Table 6.2: Logistic regression modeling result using *Sand lance Presence/Absence* as the dependent variable (N=55). *Independent variable significant at the 0.05 level (Wald statistic). **Model Chi-square value significant at the 0.05 level.......................114

LIST OF FIGURES

Figure 2.1: The study area Barkley Sound relative to Vancouver Island (inset). Dots indicate sample sites from 2003……………………………………………………………7

Figure 3.1: Comparison of beach seine set detection frequencies based on tide height (N=16)……………………………………………………………………………29

Figure 3.2: Comparison of sand lance detection frequency using visual method survey techniques (N=29)……………………………………………………………………30

Figure 3.3: Comparison of sand lance detection frequency among intertidal digging, visual boat parallel, low tide beach seine, and snorkel surveys (N=20)…………………30

Figure 4.1: *Sand lance Presence/Absence* at 37 sites sampled in all three years (each dot is a site) for 2002 (A), 2003 (B), 2004 (C). Red= Present, Blue=Absent…………………41

Figure 4.2: Mean sand lance abundance (sand lance per seine set) for 2003. Sites were grouped based on whether they had sand lance present in one, two or all three years. Error bars represent the 95% confidence interval…………………………………………43

Figure 5.1 Histogram of the dependent variable *Sand lance Set Abundance* (sand lance per set, ln transformed)……………………………………………………………………57

Figure 5.2: ANOSIM for four dependent variables A: *Sand lance Presence/Absence,* B: *Sand lance High/Low*, C: *Sand lance Burying/Not burying*, D: *Sand lance Nil/Low/Med/High*. The data label for the sample statistic is not displayed for *Sand lance Presence/Absence* because the sample statistic was off the scale (global R=0.166)……………………………………………………………………………………61

Figure 5.3: Grain size distribution for combined sites with sand lance absent (black, N=27) and combined sites with sand lance present (blue, N=26)……………………………69

Figure 5.4: Grain size distribution for combined sites with intertidal eelgrass absent (blue, N=43) and combined sites with intertidal eelgrass present (black, N=12)…………70

Figure 5.5: Grain size distribution for combined sites with subtidal eelgrass absent (blue, N=27) and combined sites with subtidal eelgrass present (black, N=26)………………71

Figure 5.6: Comparison of the *UV Mud %* (transformed) between sites with intertidal eelgrass (N=5) and subtidal eelgrass (N=7). Sites with both intertidal and subtidal eelgrass were not included………………………………………………………………72

Figure 5.7: Eelgrass categories (*Intertidal Eelgrass Presence/Absence* and *Subtidal Eelgrass Presence/Absence*) combinations grouped by *Sand lance Presence/Absence.*

The four categories (Intertidal, Intertidal and Subtidal, None, Subtidal) are combinations of the binary eelgrass variables……………………………………………………………75

Figure 5.8: Mean plots of all independent variables selected for analysis. Bars indicate 95% confidence intervals……………………………………………………………………76

Figure 5.9: Comparison of resubstitution values (blue diamonds) to data splitting cross-validation values (red squares)………………………………………………………………81

Figure 5.10: Classification trees grown from misclassification pruning (A) and deviance pruning (B). For each node box, the node number is in the top left, the ratio of absence to presence in the top middle and the node class (sand lance absence = 0, sand lance presence = 1) in the top right. The histograms within the node boxes represent the class proportions with green solid representing absence and pink hatched representing presence. The splitting rule indicates the rule determining the cases belonging in the left node of the split with the remaining cases belonging in the right node…………………………………………………………………………………………83

Figure 5.11: 10-fold cross-validation cost of the misclassification pruning procedure (A) and the deviance pruning procedure (B). Tree Number indicates the number of splits occurring in the tree and error bars represent standard error. * the tree number selected by the pruning procedure……………………………………………………………………84

Figure 5.12: Importance of independent variables in classification tree analysis for the misclassification pruned tree (green bars) and the deviance pruned tree (blue bars) ranked on a scale of 0-100. Ranking is based on the overall change in node impurity summed over the entire tree and standardized by the largest sum……………………………………85

Figure 5.13: MDS ordination of fish communities where sand lance presence (green triangles) and absence (blue triangles) was used as the factor…………………………86

Figure 5.14: ANOSIM of fish community structure using sand lance presence/absence as the factor. The global R-value was not significant (p=0.12)……………………………87

Figure 5.15: *Grain Size Mean vs. Sand lance Set Abundance* for 2003. Lines represent regression fits (linear R^2=0.015, quadratic R^2=0.084, neither significant at the 0.05 level). Absent sites removed before analysis……………………………………………102

Figure 6.1: 10-fold cross-validation cost for misclassification pruning (A) and deviance pruning (B). Tree Number indicates the number of splits occurring in the tree and error bars represent standard error. * the tree number selected by the pruning procedure…..115

Figure 6.2: Classification tree grown using *Sand lance Presence/Absence* as the dependent variable and the five selected independent variables. Misclassification pruning and deviance pruning produced the same tree. For each node box, the node number is in the top left, the ratio of absence to presence in the top middle and the node class (sand lance absence = 0, sand lance presence = 1) in the top right. The histograms within the node boxes represent the class proportions with green solid representing absence and pink hatched representing presence. The splitting rule indicates the rule for the left node determining the cases belonging in the left node of the split with the remaining cases belonging in the right node……………………………………………………………….116

Figure 6.3: Importance of independent variables in regression tree analysis the deviance-pruned tree ranked on a scale of 0-100. Ranking is based on the overall change in node impurity summed over the entire tree standardized by the largest sum…………………117

LIST OF APPENDICIES

Appendix 2.1: Description of nearshore sub-regions...147

Appendix 3.1: Sampling method detection frequency in 2003. "1" indicates sand lance presence according to the method while "0" indicates sand lance absence. The "normative standard" designates sites where sand lance were found present by at least one method (1) or for none of the methods (0)...148

Appendix 5.1: Explanatory variables considered in analysis.................................150

Appendix 5.2: Values of independent variables and dependent variables by site. Sand lance abundance was ln transformed...154

Appendix 5.3: Summary of model classifications of sand lance presence (1) and absence (0). "Actual" represents the presence/absence designation based on beach seine sampling that was used for model construction...157

ACKNOWLEDGEMENTS

First, I would like to thank my primary supervisor Cliff Robinson. You gave me guidance and direction when it was much needed. Most importantly, you allowed me to enjoy myself throughout this whole experience and I am grateful to you for that. To my co-supervisor Phil Dearden and my supervisory committee Alan Burger and Dave Duffus, thank you for telling me where I went wrong and, just as importantly, where I went right. Your comments helped hone my thesis and improve my writing immensely. Thanks to Pippa Sheppard for all your help with the field seasons and your insightful suggestions and encouragement. A special thanks to my former professors Doug Facey and Alton Harestad for inspiring me to start graduate studies.

I would like to thank everyone who contributed to the fieldwork portion of this study. A huge thanks to Eric Baron who was probably the most competent and proficient field assistant that I could have hoped to work beside and was a major reason for the success of the first two field seasons. Thanks to Ryan "Tugger" Stevenson and Andrew "Mono" George for their contributions to the work in the field. Also, thanks to Andrea "Dre" Hunter for the amazing support in the third field season. A better neighbor I have never had. Thank you to Dan Vedova and Darren Salisbury for keeping an eye on us out there and putting up with our daily check-in/check-out calls.

Thanks and love to my family for supporting me in every way possible. Without you I could have never undertaken this thesis. I cannot thank my parents enough, who have always believed in me and given me the confidence to succeed throughout my life. Thanks to my sister and brother-in-law for being so supportive and always willing to take in their nomadic brother. Thanks to my little bro for all his help collecting field data (even though you had to get up early!) and to my grandparents who have always been so important in my life.

Funding for this project was primarily from the Species at Risk Interdepartment Recovery Fund administered by Environment Canada. Logistical support was provided by the Northern and Western Service Centre, Parks Canada Agency, and Pacific Rim National Park Reserve of Canada.

CHAPTER 1: General Ecology of the Pacific Sand Lance and Study Objective

The genus *Ammodytes* is recognized as being an ecologically important group of forage fish species. They are considered a critical link between secondary producers and the top predators (Springer and Speckmann 1997, Willson *et al.* 1999, Litzow *et al.* 2000). *Ammodytes* are considered extremely important to predator populations such as seabirds, marine mammals and piscivorous fishes because of their general high energetic content and their tendency to occur in high biomass (Springer and Speckmann 1997, Willson *et al.* 1999). The genus *Ammodytes* contains six species: *americanus*, *dubius*, *hexapterus*, *marinus* and *tobianus,* all whose ranges exist within the northern hemisphere (Robards and Piatt 1999). These fishes are commonly referred to as "sand lance" or in the United Kingdom "sand eels". Within the genus, there are variable amounts of literature available for each species. In some parts of the world, such as the North Sea and in regions of Asia, there are major fisheries for sand lance and therefore more information is available on their population dynamics. For example, *A. marinus* is one of the most heavily fished species in the North Sea (Furness 1999) with the catches ranging from 0.6×10^6 to 1.1×10^6 tonnes, (Bergstad and Høines 2001).

The Pacific sand lance *(A. hexapterus)* is distributed from the Northeast Pacific Ocean from California to the Beaufort Sea to the West Pacific in the Sea of Okhotsk and along the north tip of Hokkaido (Robards and Piatt 1999). As with other species in the genus, the Pacific sand lance is a primary prey item of many marine predators due to its importance as a high-energy protein food source (Robards *et al.* 1999a) and its relatively high seasonal abundance in the nearshore region (Robards and Piatt 1999). Predation on Pacific sand lance transfers energy efficiently from their zooplankton prey to the higher trophic levels (Anthony *et al.* 2000). Although it is considered an important forage fish, relatively little is known about its biology and ecology (Field 1988, Burkett 1995, Blackburn & Anderson 1997, Chikilev and Datskii 2000).

There are no large-scale fisheries for the Pacific sand lance and little is known about its population dynamics through its range. Besides the fact that there are no major fisheries, one reason for the general lack of research on this species is the difficulties associated with studying it. The ecological features of the Pacific sand lance, such as burrowing in the sediment and feeding in the water column, make its populations challenging and expensive to monitor and study (Chikilev and Datskii 2000, Litzow *et al.* 2000). However, there has been a recent increase in study of *A. hexapterus* in part due to the recognition of its importance in the trophic structure of the marine food web.

Adult Pacific sand lance have been recorded to reach a maximum of 260mm (Field 1988), however, they are more commonly found to be less than 180mm (Robards *et al.* 1999b). They can live up to seven years and usually show sexual maturity within their second year (Robards *et al.* 1999c). The Pacific sand lance shares morphology common to that of the rest of the genus, a feature of the genus that has caused taxonomic confusion in the past (Robards and Piatt 1999). A slender, elongated body characterizes *Ammodytes* morphology. *A. hexapterus* has a uniform body depth from the opercular region to the beginning of the anal fin. At the anal fin its body begins to taper towards to caudal peduncal, ending in a forked tail (Auster and Stewart 1986). Its "eel like" body is associated with a distinct undulating swimming style, characterized by a transverse wave that passes down the body in order to produce thrust. Its body shape and swimming style is likely an adaptation for burying in sediment.

The Pacific sand lance is a visual feeder distributed in the shallow coastal regions. Sand lance feed on zooplankton in the water column during daylight hours, generally in large schools. Although the diet of post-larval Pacific sand lance consists of a wide array of prey items, including epibenthic invertebrates, copepods are the dominant prey (O'Connell and Fives 1995). The Pacific sand lance utilizes a broad range of thermohaline habitats and therefore, compared to other forage fishes, they are less restricted in their distribution with regards to spatial patterns in temperature and salinity (Abookire and Piatt 2005).

Sand lance bury themselves in the benthic substrate at night (Hobson 1986), in the presence of predators during the day (Girsa and Danilov 1976) and for longer periods during winter months in a state of dormancy (Ciannelli 1997, Robards *et al.* 1999a). Sand lance can either bury in subtidal sediment (*e.g.* Hobson 1986, Wright *et al.* 2000, Holland *et al.* 2005) or intertidal sediment (*e.g.* Quinn 1999, Robards *et al.* 1999a) but the extent that they use each habitat type and the connectedness between the two is unknown.

Sand lance likely use burrowing habitat as a refuge to avoid high levels of predation (Dick and Warner 1982, Quinn 1999) and, due to a lack of a gas-filled swim bladder, to conserve energy (Girsa and Danilov 1976, Pearson *et al.* 1984, Quinn 1999). Because of this high reliance on benthic habitat as a refuge, adult and juvenile sand lance are strongly linked with a suitable substrate in which they can bury (Robards *et al.* 1999a). This substrate tends to be well-washed, well-drained sand/gravel with mud, silt and coarse gravel absent (Reay 1970, Dick and Warner 1982, Pinto *et al.* 1984, Wright *et al.* 2000). Adaptations such as lack of a swim bladder, a slender tapered body shape and the ability to respire interstitial water which generally has low dissolved oxygen, permit sand lance to bury in substrate (Quinn 1999) and remain buried in the exposed intertidal sediment during low tides (Quinn and Schneider 1991).

In early fall, Pacific sand lance use intertidal sand or fine gravel beaches to deposit demersal adhesive eggs in small excavated pits (Robards and Piatt 1999). Although intertidal spawning has been documented (Penttila 1997, Robards *et al.* 1999c), it is still unknown whether sand lance utilize subtidal areas for spawning as well. The abundance of larval sand lance peaks prior to the spring plankton bloom (March and April in Alaska, Haldorson *et al.* 1993). Post-larval sand lance are found in abundance in nearshore regions (defined as <1km from shoreline, Abookire and Piatt 2005) during the spring and summer months, generally in water less than 50 m in depth (Robards and Piatt 1999, Ostrand *et al.* 2005). Adult Pacific sand lance appear in high numbers the nearshore region (intertidal and subtidal) in the early summer, followed later by the appearance of juveniles (*i.e.* 0-year class) in the later summer as they recruit from the larval stage (Dick and Warner 1982). As the summer progresses, older sand lance become less abundant in nearshore waters, which instead become dominated by newly recruited juveniles (Robards and Piatt 1999). Both juvenile and adult sand lance likely bury themselves in substrate during the winter months in an inactive state and thus become less common in the nearshore region during the winter season (Dick and Warner 1982, Robards *et al.* 1999a).

As a forage fish, the Pacific sand lance is one of the most common food sources of many marine predators (Burkett 1995). Variability in sand lance abundance has been proposed to have major consequences for predator populations, with the majority of this research focusing on seabirds (*e.g.* Suryan *et al.* 2002, Litzow *et al.* 2002, Litzow and Piatt 2003). It is suggested that a decrease in sand lance abundance is closely linked to a decrease in the reproductive success of many species of seabirds (see Willson *et al.* 1999 for an overview). Because the reproductive success of predators has been shown to be associated with sand lance availability, sand lance recruitment variability may play a large role in determining the fitness of many predators. Also, the distribution of various life stages of sand lance may be key in explaining the predators foraging habitat use (*e.g.* Munk 2002, Zamon *et al.* 2003).

A main focus in ecology has been to explain the distribution of species. An organism's pattern of distribution is shaped by biotic and abiotic relationships as well as life history. Habitat associations can therefore provide insight to the ecology of a species. One method of investigating species habitat associations is through habitat modeling processes that model species occurrence relative to environmental characteristics. The study of a species relationship to its environment can reveal the characteristics of the environment that determine the species distribution (Heglund 2002). These relationships can be used to infer information on habitat requirements (*i.e.* ecological niche) or can be used to predict patterns of species distributions in unknown areas or for unstudied time periods.

There has been research on the relationship between sand lance and abiotic characteristics in laboratories (*e.g.* Winslade 1974a, Winslade 1974b, Pinto *et al.* 1984). However, there is a paucity of understanding of the relationship between sand lance and its environment (hereby known as the "species-environment relationship") and how that relationship affects the distribution of sand lance. Studies that have examined habitat selection in sand lance have focused only on subtidal habitat (Brown *et al.* 2000, Wright *et al.* 2000, Holland *et al.* 2005, Ostrand *et al.* 2005) even though the intertidal habitat is also thought to be important (Robards and Piatt 1999, Quinn 1999). Also, none of these studies distinguished between size-classes even though there are known differences in habitat use between juveniles and adults (Winslade 1974a, Robards and Piatt 1999, Robards *et al* 1999a, Chikilev and Datskii 2000, Robards *et al.* 2002).

In this study, my main objective is to describe the species-environment relationship for Young-Of-the-Year (YOY) Pacific sand lance in the intertidal and shallow subtidal environment. I use an inductive habitat modeling approach to determine what environmental characteristics significantly explain sand lance occurrence and assess the relative importance of each characteristic. In order to manage sand lance, we must have a good grasp on the species-environment relationship. This study aims to provide a better understanding of this relationship for sand lance, thus, providing valuable information required for management efforts for sand lance as well as species that predate upon sand lance.

This chapter (chapter one) outlines the basic biology of the Pacific sand lance and existing knowledge regarding possible habitat associations. Chapter two describes general field data collection including sand lance occurrence data as well as empirical environmental characteristics. Chapter three examines differences in methods for sampling sand lance in order to select the best measure of sand lance occurrence for prospective habitat models. Chapter four examines the temporal variation in sand lance distribution and abundance within the three years of this study in order to provide the context of temporal variation to the habitat modeling in later chapters

Studies conducted at several spatial scales can offer a greater insight into what environmental variables affect a species and at what scale(s) they operate on (Wiens 1989). Considering this, I examine sand lance distribution at two scales and two separate environmental data sets. In chapter five, I model sand lance occurrence using fine grain empirical data collected from the field (site-specific level). I examine the data from various modeling perspectives using an iterative modeling approach employing two separate statistical model types. In chapter six, I model sand lance occurrence using a larger grain size and environmental data that encompass the entire extent of the study (landscape level) and were generated in a Geographic Information System (GIS). I use the same statistical modeling approach employed in chapter five to create a sand lance occurrence model. Chapter seven contains the conclusions of this study.

CHAPTER 2: General Sampling Methods for Sand Lance and Environmental Data

2.1 STUDY SITE

Region

Barkley Sound is located on the southwest coast of Vancouver Island, British Columbia (Figure 2.1). It is approximately 25 km wide opening to a southwest Pacific exposure and 20 km long running in a southwest-northeast direction. Barkley Sound is broken into three major channels by two groups of islands: the Broken Group Islands (BGI) and the Deer Group Islands (DGI). This study focuses mainly on the BGI, however, I also chose study sites from the DGI and other areas of Barkley Sound. The BGI consists of an archipelago of over one hundred small islands and islets separated from other land masses by three channels: Loudon Channel, Sechart Channel, and Imperial Eagle Channel. The BGI has a diverse range of shoreline habitat types. For example, shoreline exposure ranges from completely exposed to the Pacific Ocean to almost completely sheltered by landmasses with only narrow channels providing access for tidal waters. The DGI is oriented more linearly, southwest to northeast and therefore generally has fewer sheltered areas.

Sites

I delineated the shoreline sampling unit by determining the length of shoreline that could be safely sampled using a beach seine. Beach seining generally requires a shoreline with an intertidal/subtidal region that is dominated by sandy or smooth benthos (Robards *et al.* 1999d). Sample sites necessarily had to be free of large obstacles that may have impeded the beach seine net. The delineation of shoreline to be sampled (*i.e.* the definition of a sample "site") was governed by natural hazards that broke up shoreline into "heterogeneous" segments. The mean length of sampled shoreline in 2003 and 2004 was 53.6 m, (S.D. = 75.6 m). Generally, rocky headlands were the most common feature that broke up the shoreline into discrete sample sites. At each site, I recorded a waypoint at the center of the shoreline using a Garmin handheld Global Positioning System (GPS) unit (multiple models were used: GPS 12XL, GPS eTrex, GPS 76).

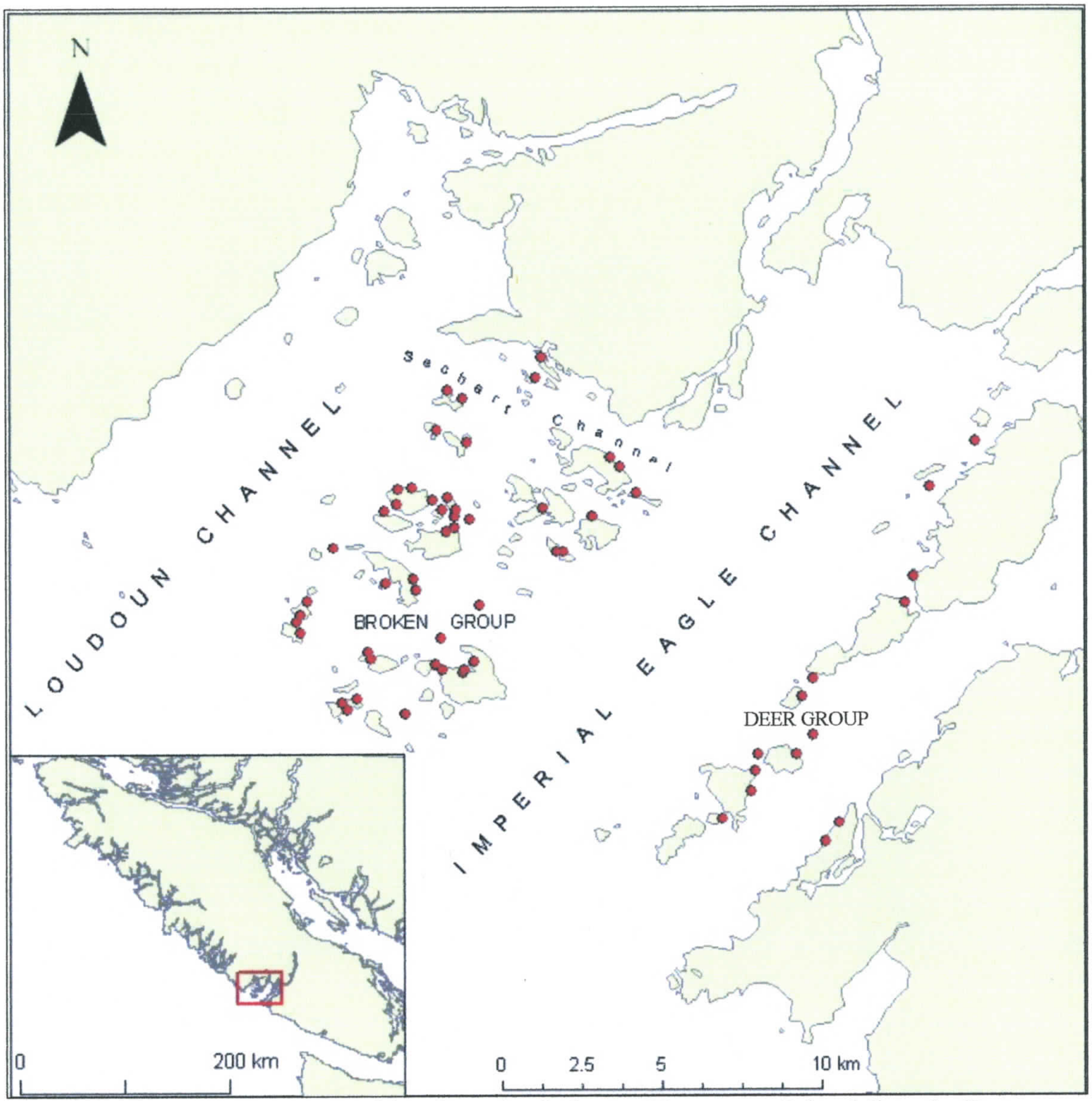


Figure 2.1: The study area Barkley Sound relative to Vancouver Island (inset). Dots indicate sample sites from 2003.

Scale of Study

Almost all ecological relationships vary with scale. As the scale of observation changes, so do the relationships that we observe. Given the significance of scale in ecological studies it is important to report the scale of study (Dungan *et al.* 2002). The term "scale" is marred by the fact that it has been used in many different contexts and thus has accumulated many different definitions (Corsi *et al.* 2000). Morrison and Hall

(2002) and Huston (2002) attempt to set a standardized terminology for "scale" for use in ecology. This is the definition that I employ in this study and will outline below.

Scale can be broken into two main components: 1)"grain" or sample resolution and 2)"extent" or the total area over which the data was collected (Morrison and Hall 2002). A study should explicitly define each of these components for the sake of clarity since defining spatial grain and extent, effectively defines the spatial scale of the study.

Grain is the smallest resolvable unit of a study (*e.g.* the minimum unit of space or time for which differences can be resolved). For the empirical portion of my study, the area sampled by each beach seine length is the true spatial grain size. However, since the sample units were observed to be relatively homogenous environments (*i.e.* superficially had similar environmental properties vertically along the shoreline), it is likely that there is a constant use of that habitat within the study site boundary. Given this, the grain size was assumed to be the total length of the shoreline of the site multiplied by the sampling length of the beach seine extending out from the shoreline. Given a 10m beach seine deployment the approximate grain size for a site is the shoreline length x 10m (mean area = 536 m^2, S.D. = 756 m^2, range = 100-4000 m^2).

Extent is the overall area in which the observations were made and to which the results can be applied without extrapolation. The size of the extent can affect how ecological patterns are sampled in two key ways (Huston 2002). First, the larger the extent, the more potential habitat and resources are included in the study. With regards to population dynamics and behavior, this can be a critical issue (*e.g.* including an entire home range of the study organism within the extent). Second, extent size influences the amount of environmental variability included in the study, with larger areas generally having greater variability. In this study, the extent is roughly defined as the BGI and the DGI of the Barkley Sound region. The BGI and DGI appear to be a reasonable extent for the study area since a broad range of environmental characteristics can be found in an area that can be feasibly sampled within a summer.

Nearshore

In this study I focus sampling on the nearshore region. The term "nearshore" has been defined as <1 km from the shoreline (Abookire and Piatt 2005). Here, I restrict this definition of nearshore to consider only the intertidal and shallow subtidal regions that can be sampled with a beach seine.

I divided the nearshore into sub-regions based on three tide heights (*i.e.* the height of the waterline): low tide, medium tide and high tide. I considered the low tide line as the lowest point in the tidal cycle for the day (slack tide). I considered the medium tide line as the tide height three hours after low tide and high tide line as the tide height 6.5 hours after low tide. I determined tide height values using Tides and Currents for Windows^{TM} (1997) using Stopper Islands (latitude 49°00'N, longitude 125° 21'W) as a reference point.

Based on these three tide heights, I divided the nearshore region into three sub-regions: 1) shallow subtidal, 2) low intertidal, and 3) high intertidal (Appendix 2.1). I used these sub-regions as categories for sampling environmental characteristics. The shallow subtidal region was defined as the zone from the low tide line to the 3 m depth contour. The sub-region approximates the subtidal zone that could be sampled by the beach seine used in this study. The low intertidal region was defined as the low tide line to the medium tide line. The high intertidal region was defined as the medium tide line to the high tide line.

2.2 SAMPLING SUMMARY

Overall Summary

I conducted sampling during the summer months of 2002-2004 (Table 2.1). Sampling objectives were two-fold: 1) to sample the sand lance distribution and abundance within the study extent and 2) to sample the environmental characteristics associated with each study site. For all years, beach seining at low tide was used as a sampling method for sand lance, however, I employed other sampling techniques for sand lance in 2003. The ability to sample with a beach seine at low tide was the major criterion for selecting sites for all years. There was variation in the quantity of environmental characteristics collected for each year with 2003 being the year where I collected the most environmental data.

Table 2.1: Beginning and end sample dates for each year of sampling and the number of sample days and sample sites. Note that for some sample days, multiple sites were sampled. Also, a number of sample sites were sampled multiple times (not displayed here).

Year	First Day of Sampling	Last Day of Sampling	Number of Sample Days	Number of Sample Sites
2002	June 16th	August 25th	42	70
2003	May 5th	September 9th	52	60
2004	May 17th	August 18th	34	60

Yearly Sampling Summary

In 2002, I sampled 70 sites across Barkley Sound while in 2003 I focused the study area to the BGI with less focus on the DGI and surrounding areas than in 2002. In 2003, I attempted to sample all sites in the BGI that were accessible for beach seining. Based on knowledge acquired during the field seasons, I consider the coverage obtained (approximately 90% of all possible sample sites) as adequate to assume complete coverage for the study area. Because sampling strategies were different for 2002 and 2003 in terms of total study area size and density of sample sites within the study area, not all sites sampled in 2002 were sampled in 2003 and *vice versa*. Also, there was a difference in sampling effort during 2003. I performed a more extensive sampling of each site by employing multiple sampling methods for sand lance and collecting a wider array of environmental characteristics. In 2004, I sampled all sites that were sampled in 2003 however I collected few environmental characteristics.

2.3 SAND LANCE SAMPLING

General Beach Seining Technique

The following is a description of beach seine sampling methods and equipment used in all three years of sampling.

Beach seining has shown to be an effective and unbiased sampling method for sand lance (Robards *et al.* 1999b). I used a knotless 4mm stretch mesh beach seine, 9.2m in length with a 3.1m centre that tapered to 1.1m at the wings. The seine had a lead-core bottom line and a foam float top line, both continuous down the length of the mesh. For sites with heavy eelgrass cover, I attached 1m of rope to a small weight and fastened it to the lead-line to help prevent the net from rolling into itself. At each wing, a 1m rope came off both the lead line and the float line. The ends of these two ropes coming off the wing were tied to a single rope that was 9m long, giving each wing a line approximately 10m in length attached to both the lead and float line. The beach seine was deployed using a 5m aluminum hull boat powered by an outboard engine. Beginning with the entire seine inside the boat, one of the wing lines was deployed by dropping off a person with the line in approximately 0.5m deep water directly off the shoreline. A second person then fed the net out as the boat was driven (by a third person) in a semicircle out from shore. Completing the semicircle, the second wing line was deployed approximately parallel with the first wing line thus aligning the net of the beach seine parallel to the shoreline. The wing lines were then hauled in from shore. Two replicates were done at each site, the second replicate offset adjacent and parallel to the first. Each replicate was completed in approximately 15 minutes. All fishes caught were identified to species, counted, and returned to the ocean alive.

I attempted to sample a standard unit volume of water by deploying the beach seine using the same method and distances for each replicate. Differences in physical structure of the sites (such as shoreline slope and navigation hazards) caused slight inconsistencies in the sampling volume for each beach seine replicate, however, these inconsistencies were assumed to be relatively slight and were ignored.

The ability to catch a certain species (frequency and abundance) using a beach seine is highly dependent on the sampling tide height (Robards *et al.* 1999a). Because preliminary results found that beach seine sampling is preferably done at an extreme low tide, I chose sample days that coincided with minimum low tide values during summer months. For example, the maximum low tide height sampled in 2003 using a beach seine was 144cm (Table 2.2). Sampling always commenced at the first low tide after sunrise since this was generally the lowest tide that could be sampled during that day. I conducted all sampling (2002-2004) and data collection on a flood tide or within one hour before low tide (slack tide).

Table 2.2: Descriptive statistics for tide height (cm) based on beach seine set for 2003.

Beach Seine Set	N	Minimum	Maximum	Mean	Std. Deviation
Set 1	123	34	144	74.4	23.8
Set 2	88	124	279	190.0	30.9
Set 3	57	244	339	303.6	19.6

Age Class Sampled

I measured fork length of random samples of sand lance for each site where they were captured. Two major size classes of sand lance could be distinguished in the field: 1) YOY and 2) adult (1+, one year or greater). YOY is the size class that recruited from the larval stage during the year of sampling. The YOY size class increases in length considerably throughout the summer season however YOY and 1+ sand lance were generally distinguishable as distinct size classes. In general, YOY were considered as sand lance <100mm in fork length (J. Beaubier, pers. comm.) which is similar to the values used by Robards *et al.* (1999a). Overall, the majority of sand lance I sampled were YOY. For example, 92% of the fish sampled in 2003 were YOY and there were no instances where 1+ sand lance were present at a site and YOY were not. Because of this, the results focus on the YOY size class and therefore, in all subsequent chapters it will be assumed that results are specific to the YOY size class only.

2.4 YEARLY SAMPLING

Summer 2002

Beach seining at low tide was the only sampling technique I employed in 2002. One goal of this sampling season was to explore the general distribution of sand lance within Barkley Sound to provide baseline information with which to focus sampling in subsequent years. I therefore attempted to sample as many sites as possible within the sampling season (Table 2.1). The number of sites sampled per day ranged from one to three depending on logistic considerations such as travel time between sites, time of initial low tide relative to sunrise, and ease of sampling at each site. In general, at each site, I conduced two beach seine replicates. However, during days when only one site was sampled I often conducted three or four replicates. Because there were often multiple sites done within the same day, many sites were not sampled directly at low tide as strived for in subsequent years. In general however, sites were sampled within two hours of low tide.

Summer 2003

In general, it is more difficult to prove a species is absent than present when collecting data for habitat modeling (Karl *et al.* 2002). Inadequate sampling effort or ineffective survey techniques can lead to false absences (Angermeier *et al.* 2002) that can confound habitat models. I sampled sites using multiple sampling methods in 2003 in order to conclude more definitively whether sand lance were absent from each site. These methods include beach seining at low, medium and high tide heights, snorkel survey, boat based and land based visual surveys and intertidal digging. Because I used a wider array of sampling methods to sample sand lance and more collected environmental variables, I was generally able to sample only one site per day. However, where sites were close together, I was able to sample two per day. I recorded the time of sampling for all methods and later determined the tide height for each sampling event using Tides and Currents for Windows^{TM}.

Beach Seining

For every sample site, I conducted two beach seine replicates at a low tide height. To determine how tide height affects the ability to catch sand lance, I also conducted a set of two replicates at a medium tide height and a high tide height. I conducted low tide beach seine sets at a total of 59 sites. Due to obstacles in the intertidal region at some sites, I conducted medium tide beach seining at 45 of the 59 sites, and high tide sets at 27 of the 59 sites.

Visual Surveys

Sites were surveyed for sand lance using three visual methods: two boat surveys (one parallel and one perpendicular to the shoreline), and a walk survey. For all visual surveys, I recorded positively identified sand lance and an estimation of the school size and age (YOY or 1+) when possible as well as the water visibility as measured by secchi depth.

I conducted visual boat surveys from the 5m boat at idle speed along each transect using an underwater viewer, made from a plastic cone with a glass bottom that is submerged. These boat transects were conducted at low tide (within 30 minutes o f slack tide) and ran 1) parallel to the shoreline and 2) perpendicular to the shoreline. The parallel transect followed an approximate 2.5m depth contour parallel to the shoreline for the length of an estimated 30m. The perpendicular transect ran directly out from the shoreline for an estimated 30m commencing at the shallow subtidal.

I conducted walk surveys just before the medium tide (three hours after low tide). Using chest waders I waded in approximately 1m of water, following this depth contour parallel to the shoreline for an estimated 30m. As I walked, I scanned the water for sand lance visually using polarized sunglasses to reduce the glare on the water.

Snorkel Survey

I conducted a snorkel survey at 36 of the 59 sites at approximately high tide when water clarity was generally best. Starting on one side of the shoreline at the waterline, I snorkeled a transect consisting of three sequential legs each approximately 20m in length: first perpendicular, second parallel, and third returning perpendicular to the shore. I made adjustments to the transects to compensate for navigation hazards but I made an effort to achieve the standard snorkeling distance total of 60m.

Intertidal Digging Survey

Sand lance can stay buried in the intertidal sediment after the waterline recedes beyond the intertidal. Thus, digging in the exposed intertidal sediment for buried sand lance is a viable sampling strategy (*e.g.* Dick and Warner 1982, Robards *et al.* 1999b). In 2002, intertidal digging was found to be successful in capturing sand lance. In 2003, I developed a standard method for intertidal digging sampling. Intertidal digging was conducted at approximately low tide (within 20 minutes of slack tide). Using a shovel (24.5 x 28.8 cm, square faced blade), I dug five 1 x 1 m holes just above the waterline. Because sand lance have been found to bury themselves 4-6cm deep (Quinn 1999) I dug holes approximately 8cm deep. The holes were staggered approximately 5m apart down the shoreline parallel to the waterline. For sites with less than 25m of shoreline, the holes were staggered closer together. For holes that would be over an area that prevented digging (*e.g.* bedrock), I chose the next closest suitable location along the shoreline.

Summer 2004

In the summer of 2004, I re-sampled each site for sand lance sampled in 2003. I conducted two beach seine replicates at low tide as conducted in 2002 and 2003. Generally, I sampled two sites per day in 2004. Sites sampled on the same day were chosen for their close proximity to each other. As in 2003, beach seine sampling occurred within one hour after low tide.

2.5 ENVIRONMENTAL SAMPLING

The second objective of the field data collection was to determine the environmental characteristics associated with the sampling sites. In 2003, I measured data on environmental characteristics that had the potential to play a role in sand lance habitat selection at each site. These characteristics ranged from qualitative categorical variables (*e.g.* substrate type, macro-vegetation cover) to quantitative variables (*e.g.* grain size of substrate). The following section outlines the environmental characteristics and collection methods.

Shoreline Length

I visually estimated shoreline length of the sample site (see "site" definition).

Beach log cover

I defined beach logs as coarse woody debris transported by the tide and accumulated at the high tide line. Beach logs may be an indirect indicator of exposure for each site as they are generally carried passively on ocean currents before being distributed onto the shoreline. Beach log cover was categorized into four groupings ranging from high beach log cover to no beach log cover (Table 2.3). The number of beach logs intersected by an imaginary perpendicular cross-section of the high tide line determined beach log cover for that site.

Table 2.3: Classification of *Beach Log Cover.*

Category	Description	Quantity
0	No beach logs	0 beach logs
1	Low beach log cover	1-2 beach logs
2	Medium beach log cover	3-4 beach logs
3	High beach log cover	>4 beach logs

Shoreline Slope

I measured the slope of the shoreline using a hand held abny level to the nearest half degree. Because high and low intertidal slope can differ markedly in magnitude, I took separate measures for each, using the low, medium and high tide lines as observation points.

Shoreline Aspect

Using a compass I measured aspect for each site as the direction perpendicular to the shoreline. The bearing was taken from the middle of the site shoreline. I measured aspect to the nearest degree and categorized it as follows: Northeast (1-90°), Southeast (91-180°), Southwest (181-270°), Northwest (271-360°).

Subtidal Characteristics

I sampled subtidal benthic characteristics using an underwater video ('UV' used as an acronym for variable naming in chapter five). Video transects allow for rapid capture of information for large areas of subtidal benthos (Dethier *et al.* 1993, Ninio *et al.* 2003). Video analysis is generally subjective since interpretation by an observer is required. Thus, there can often be a high discrepancy among observers and it is also often difficult to accurately classify down to fine taxonomic resolution (Meese and Tomich 1992, Ninio *et al.* 2003). These possible problems associated with video analysis were minimized by only using one observer to analyzed the video (T.H.) and algal species were classified into broad groups rather than down to species. The goal of this analysis was to determine the following subtidal benthic characteristics: 1) substrate composition, 2) macro-vegetation type and percent cover, 3) biogenic structure of the benthos (e.g. clam siphons, tubeworms), and 4) detritus percent cover.

The subtidal data were collected with an underwater video camera mounted in aluminum housing. An umbilical cable connected the underwater camera to a handheld video camera. The umbilical cable was connected through an overlay that stamped the video with the depth of the underwater camera and time. The real-time feed, viewed through the handheld video camera, enabled me to maintain a camera depth optimal for benthic viewing, avoid underwater obstacles and insure proper camera orientation. The underwater camera housing was attached to a fishing downrigger that spooled wire cable used to raise and lower the camera.

Underwater camera data were gathered between medium tide and high tide from a boat at idle speed. The transect consisted of three legs similar to that of the snorkel survey: first perpendicular out from shore for 40m beginning at the low tide line, second parallel to shore for 60m, and third perpendicular to shore for 40m returning to the low tide line. The camera was kept just above the seafloor (mean distance above the seafloor=65.9cm, S.D. =36.5cm).

I reviewed the video on a 70cm television monitor at real time speed estimating the benthic characteristics for each 9-second time segment. Segment characteristics were estimated over the entire segment, which required averaging of variables (such as percent cover) and the formation of "major" and "minor" categories for substrate type and macro-vegetation type. For each segment I recorded the time, the depth of the camera from surface and estimated the following variables:

Substrate composition

I classified substrate type into five groups (Table 2.4) determined from the video. For each 9-second segment, I recorded the major substrate category and the estimated the percentage relative to the overall substrate composition. The major substrate was defined as the category with the highest proportion relative to other substrate types present. I recorded the minor substrate category (the substrate category with the next highest proportion) in the same manner. Other substrates present for each 9-second segment beyond the first two for were ignored.

The substrate categories were further classified into possible burying habitat for sand lance after the review of the video was complete. The literature suggests that *Ammodytes* prefer sandy substrates for burying (Dick and Warner 1982, O'Connell and Fives 1995, Wright *et al.* 2000). In laboratory choice experiments, *A. hexapterus* showed a significant preference for either fine or coarse sand as opposed to gravel or silt dominated sediment (Pinto *et al.* 1984). Using this information on sediment preference, I generated a binary classification of the sediment for each segment: suitable for sand lance to bury in (1) or unsuitable to bury in (0). I classified segments with 75% of the sediment consisted of sand (as defined in Table 2.4) as suitable for burying. I then calculated the proportion of segments that were possibly suitable for burying relative to the overall number of segments and treated this as a separate variable.

Table 2.4: Qualitative substrate categories.

Category	Grain Size
Mud/Silt	<1/8mm
Sand	1/8 - 4mm
Gravel	4 - 30mm
Pebble	30 - 60mm
Cobble/Boulders/Bedrock	>60mm and/or consolidated rock

Estimate Percent Shell Content in Sediment

I estimated the overall percent of bivalve shell fragments in the sediment relative to the overall sediment composition. Shell fragments were usually easily distinguishable as white irregular shaped particles.

Macro-vegetation Types

I categorized macro-vegetation into eleven discernable groups (Table 2.5). Categories used were chosen by what could be clearly distinguished from the video and to match the categories of Druehl (2000). Algae or seagrass was the first major classification. Seagrass was further classified into eelgrass (*Zostera* sp.) or surf grass (*Phyllospadix* sp.). Algae was further classified into colour (green, red, brown) and then further into major morphological group (bushy branched, large bladed, or small bladed/globular). Crustose algae (crusts) were not considered in the analysis.

I used a semi-qualitative 2-dimensional approach to visually estimate macro-vegetation cover (Meese and Tomich 1992). I approximated the cover of the macro-vegetation canopy over the benthic area by estimating what percentage of the benthic surface area the canopy would cover if an observer were to look down at the seafloor from an observation point 1m above.

Table 2.5: Macro-vegetation categories used for underwater video analysis.

Algae or Seagrass	Algal colour	Major Category	Category
Algae	Colour (red, brown, green)	Bushy Branched (*e.g. Sargassum* sp.)	1
		Large Blades (*e.g. Nerocystis* sp.)	2
		Small Blades or Globular (*e.g. Fucus* sp.)	3
Seagrass	NA	Eelgrass	4
		Surfgrass	5

Biogenic Structure Percent Cover

Biogenic structure included all benthic invertebrates. The most common invertebrates that contributed considerably to the percent cover were: tube worms, anemones, bivalve siphons, orange sea pens and echinoderms (sea stars, sea cucumbers, sea urchins). I used the same approach to estimating percent cover of benthic invertebrates as what I employed for macro-vegetation percent cover (2-dimensional analysis of benthic invertebrate cover over the benthic surface viewed from 1m above).

Estimated Detritus Percent Cover

Detritus was defined as wood debris, drift algae and drift seagrass. I used the same method for estimating percent cover as for macro-vegetation percent cover and biogenic structure percent cover.

Intertidal Characteristics

Macro-vegetation

For many fish species, macro-vegetation can provide shelter from predators, slow water currents and provide access to prey (Ferreira *et al.* 2001). I estimated percent cover for both the low intertidal and shallow subtidal regions of the sample site according to the percentage of 2-dimensional cover as described by Meese and Tomich (1992). The high intertidal region generally has minimal amounts of macro-vegetation, therefore, I did not perform an estimation for this region.

Sediment

In order to assess the sediment preference of sand lance, I measured various substrate characteristics including the quantitative analysis of intertidal sediment properties (bulk mass, grain size distribution, organic content, carbonate content) and a qualitative visual classification of the intertidal region.

I took a bulk sediment sample at each site. Using a metal cylindrical tin (radius=14cm, height=26cm) I took a plug of sediment from the shoreline by plunging the cylinder into the sediment, digging around the cylinder and then slipping the shovel blade underneath in order to trap the sediment. I then weighed the cylinder containing the sediment to obtain a bulk sample mass for the site. The bulk mass is an indication of the sediment density. I retained the sediment collected for the bulk sample mass for further analyses. In the laboratory, I measured three characteristics of the sediment: 1) grain size distribution properties 2) organic content and 3) carbonate content.

I analyzed grain size properties using dry sieving techniques (Folk 1974). First, I dried the sediment in an oven for 48 hours at 100°C. The amount of sediment collected for the bulk sample was more than necessary for grain size analysis therefore I split the bulk sample into smaller subsamples. The bulk sample was partitioned using a laboratory sediment sample splitter, which divides the sample randomly into equal parts avoiding the biased selection of grain sizes. Each sample went through a series of splits until a desired subsample size was left for grain size analysis (mean=135.9 g, S.D.=20.9 g). The subsample was added to the top of a stack of 19 metal sieves ranging from 25,000 (top) to 38 microns (bottom) (Table 2.6). A mechanical shaker agitated the sieves for 15 minutes. After agitation, I weighed each grain size class to the nearest 0.001g.

I entered the results from sieving into the grain size distribution and statistical package GRADISTAT (Blott and Pye 2001). GRADISTAT is a Microsoft Excel based package that automates the calculation of various grain size distribution statistics such as mean, modes, sorting, skewness and kurtosis. I chose the Folk and Ward graphical technique, recommended by Blott and Pye (2001), for determining mean grain size and sorting values.

Table 2.6: Sieve mesh sizes used for dry sieving and the corresponding phi size class (-log2 mesh size) and particle description according to Wentworth classification (Wentworth 1922).

Sieve mesh size (microns)	Phi size class	Particle Description
>25000	>-4.6	Gravel/Cobble
25000	-4.6	Gravel
12700	-3.7	
6350	-2.7	
5660	-2.5	
4000	-2	
2830	-1.5	
2000	-1	
1400	-0.5	Sand
1000	0	
710	0.5	
500	1	
355	1.5	
250	2	
180	2.5	
125	3	
90	3.5	
75	3.7	
45	4.5	Silt
38	4.7	
<38	>4.7	Silt/Clay

Organic Matter and Carbonate

I measured the organic content and carbonate content of the sediment at each site using the Loss On Ignition (LOI) method (Dean 1974, Heiri *et al.* 2001). Using LOI, organic material and carbonate in the sediment are sequentially combusted in a muffle furnace.

LOI of organic matter and carbonate can be dependent on laboratory conditions such as sample size (initial dry weight used) and crucible position in the oven (Heiri *et al.* 2001). Heiri *et al.* (2001) determined that these factors increase with ignition time and generally affect LOI values of organic matter more than carbonate. However, samples with low organic content such as in this study (LOI organic matter mean=1.66%) are less prone to inconsistencies. I attempted to limit possible inconsistencies by using a narrow standard sample size between 1.5000-1.7000 g (mean=1.5988 g). I ground up the dry sediment with a mortar and pestle until the samples were a fine dust. I weighed the initial sediment mass to the nearest 0.0001 g, then transferred the sediment to a crucible and weighed it again to get the sample and crucible weight. Crucibles were then transferred to the muffle furnace. I arranged the ten crucibles in two rows of five so there was no "inside row" of crucibles and thus less chance of differential heating between samples.

As recommended by Heiri *et al.* (2001), I sequentially burnt the samples at 550°C for four hours (for organics) and 950°C for two hours (for carbonates). After the samples were heated to 550°C for four hours I removed the crucibles from the oven with tongs and fireproof gloves. I placed the crucibles in a dessicator until they were cooled to room temperature and then re-weighed each crucible. Organic matter LOI was calculated using equation (1)

Equation (1): $LOI_{550}=100(S_i-S_{550})/S_i$

Where LOI_{550} is the percent of the original sample mass lost after four hours at 550°C in the muffle furnace, S_i is the initial mass of the sample, S_{550} is the sample mass after four hours at 550°C in the muffle furnace.

Carbonate LOI was calculated using equation (2)

Equation (2): $LOI_{950}=100(S_{550}-S_{950})/S_i$

Where LOI_{950} is the percent of the original sample mass lost after two hours at 950°C in the muffle furnace, S_{950} is the sample mass after two hours at 950°C in the muffle furnace.

Qualitative Sediment Type Estimate

In a sediment preference study for *A. marinius*, Wright *et al.* (2000) used a visual assessment of sediment classification. Here, I also use a visual assessment method in order to compliment the quantitative data gathered through the grain size analysis. The high intertidal and low intertidal substrates were qualified at each site by categorizing the sediment into the same five major classes used for underwater video analysis (Table 2.4). This was a visual estimate based on the dominant sediment at each site. Because the

Pebble and *Cobble/Boulders/Bedrock* categories had very low frequencies (for both low intertidal and high intertidal), I later combined the two categories into one for analysis purposes.

CHAPTER 3: Comparison of Survey Methods for Sampling Sand Lance in the Nearshore

3.1 INTRODUCTION

Habitat models can be greatly influenced by the effectiveness of the sampling method used to determine the occurrence of a species. Although this possible source of error is often ignored, recently attention has been focused on developing efficient survey protocols (Stauffer *et al.* 2002). Sampling methods for presence/absence are generally compared based on the differences in detection efficiencies. Sampling methods generally have different sampling efficiencies since each method is affected by physical factors (*e.g.* substrate types) and behavioral factors (*e.g.* avoidance) differently (Charles-Dominique 1989).

Omission and commission errors can be considered in the context of assessing model predictions or in the context of assessing sampling methods. In the context of sampling methods, omission error is a site designated absent by a sampling method when the species is truly present, and commission error is when a site designated present by a sampling method when the species is truly absent. By assessing sampling methods against a known occurrence one can measure omission and commission errors of differing methods.

In this study I examine the differences in sand lance detection frequencies for eight survey methods. Specifically, I investigate 1) Which sampling methods can significantly determine sand lance presence/absence (occurrence) and 2) Whether there is a difference between omission error among methods. Also, I examine the effect of sampling at different tide heights on beach seine catch efficiency.

3.2 METHODS

Methods Analysis

Although it is generally not feasible to determine a species presence/absence in the field with complete certainty, a high sampling effort can greatly increase the confidence in sampling results. I used eight different methods in the summer of 2003 to sample sand lance (described in chapter two, Table 3.1, Appendix 3.1). Due to logistical constraints the number of sites sampled with each method varied. The most likely source of commission error is from false identification of other fish species as sand lance. Based on my experience in the field, I feel commission error is negligible in this study. False identification was unlikely since I only recorded sand lance presence if I felt certain that I had a positive identification and sand lance have a distinct body shape and swimming style (discussed in chapter one) easily distinguishable from other species. Omission error could likely arise from sampling methods failing to detect sand lance due to inefficient sampling. This is a more likely form of error in the context of this study.

Table 3.1: Frequency of surveys and detections for each survey method conducted in 2003. Note that, due to logistical constraints, not all survey types could be conducted at each site. Total number of sites sampled was 60.

Survey Type	Number of Surveys Conducted	Sand Lance Detections	% of Surveys With Detections
Low Tide Beach Seine	60	25	41.7
Medium Tide Beach Seine	46	14	30.4
High Tide Beach Seine	28	3	10.7
Visual Survey Boat - Parallel	52	14	26.9
Visual Survey Boat - Perpendicular	53	11	20.8
Visual Survey Walk	52	12	23.1
Digging Survey	59	15	25.4
Snorkeling Survey	36	17	47.2

In order to compare methods I determined a "normative standard" to represent the true occurrence of sand lance. Preferably, the normative standard is the actual true occurrence as designated by independent sources. Unfortunately, no such data were available therefore I used a proxy normative standard. This normative standard was based on the combination of the eight sampling methods (Appendix 3.1). If sand lance were detected at a site by any of the eight sampling methods, the site was given a designation of "sand lance present". Since commission error was likely negligible, these are likely accurate designations. If sand lance were not detected then the site was given a designation "sand lance absent". Because I used multiple detection methods at each site, omission error was likely minimal. Although the normative standard used here cannot be assumed to be the true occurrence, because the commission error and omission error are likely very low, it is assumed to have low error rate. Statistical tests were considered significant if $p < 0.05$.

Sand lance Presence/Absence

I performed a Chi-square test comparing each sampling method to the normative standard to ascertain whether each method could significantly determine sand lance presence/absence. I compared Kappa values (k) to quantify the relative agreement of each method to the normative standard. The Kappa value measures the relative agreement, corrected for chance, between two groups and has a range from –1, indicating perfect disagreement, to 1, indicating perfect agreement (see Cohen 1968 for calculation details). A Kappa value larger than 0.75 suggests an excellent agreement, while scores from 0.41 to 0.74 suggests moderate to good agreement, and a value less than 0.41 suggests weak agreement (Fleiss, 1973).

Omission Error

Commission errors were assumed to be negligible in this study however each method likely had some degree of omission error. To examine the omission error I compared detection frequencies of each method to sites that had sand lance present (as determined by the normative standard, $N=35$). I conducted a Cochran's Q test to determine whether there was an overall difference in omission error between all methods ($N=12$). I made 22 pair-wise comparisons between methods using McNemar tests (N varied according to comparison) and used a Holm correction to adjust for multiple comparisons (Holm 1979).

I also used simple graphical comparisons to evaluate omission error. I made three separate comparisons: 1) beach seining at low, medium and high tide height, 2) boat (parallel and perpendicular) and walk visual surveys and, 3) snorkel, parallel boat, low tide beach seine, and intertidal digging. Note, "3)" compared between major methods with the exception of perpendicular boat survey and walk survey, which had similar omission errors to that of parallel boat survey (all visual methods).

Detection* versus *Abundance

Detection frequency (*i.e.* the detectability) based on a survey method is often positively related to abundance (Pearce and Ferrier 2001). To test for this relationship I used separate Mann-Whitney U tests to compare the abundance of the 35 sites that were found to have sand lance present according to the normative standard grouped by each visual methods presence/absence. Abundance was measured by low tide beach seining catch value (*Sand lance Set Abundance*). If detection frequency of a visual survey is linked to abundance, the sites that had sand lance present according to that survey would likely have a higher abundance compared to sites that had sand lance absent according to that survey.

Water Clarity

To determine whether water clarity significantly affected visual surveys I performed separate logistic regression analyses for each visual survey method. The dependent variable was sand lance presence/absence as determined by the visual survey method and the independent variable was the secchi depth. I only used sites that had sand lance present according to the normative standard to test specifically if water quality affected the omission error for each survey method. If the logistic regression models were found to be significant then there is likely a relationship between omission error and secchi depth and comparisons between methods would have to account for water clarity.

Beach Seining at Different Tide Heights

I compared abundance among tide heights using a Frieman test for several related samples. I evaluated whether there is a difference in abundance (N=26) between low, medium and high tide beach seine sets. I ran the repeated measures Friedman test again using only low and medium tide catch values thus increasing the sample size (N=45). *Sand lance Set Abundance* was the dependent variable. This variable measures the average sand lance catch value of the two beach seine replicates at each tide height. Because *Sand lance Set Abundance* displayed extreme variation (strongly positively skewed), it was transformed to a natural logarithmic scale in order to reduce variability and approximate a normal distribution.

3.3 RESULTS

Presence/Absence

With the exception of the high tide beach seine survey, all methods had significant Chi-square values suggesting they all predict sand lance presence/absence (Table 3.2). Kappa values ranged from excellent agreement (Snorkel survey, $k = 0.780$) to poor agreement (High tide beach seine $k = 0.144$). Note that all methods showed general agreement ($k > 0.144$ with no negative values).

Table 3.2: Results from Chi-square analysis testing sampling methods against the normative standard.

		Sand Lance Presence			
Method	N	Detected by Method	Present According To Normative Standard	p-value	Kappa (*k*)
High Tide Beach Seine	28	3	17	0.14	0.144
Visual Survey Boat - Perpendicular	53	11	34	0.005	0.255
Visual Survey Walk	52	12	34	0.004	0.274
Digging Survey	57	15	35	<0.001	0.367
Visual Survey Boat - Parallel	52	14	32	0.001	0.374
Medium Tide Beach Seine	46	14	29	0.001	0.408
Low Tide Beach Seine	60	25	36	<0.001	0.645
Snorkel Survey	36	17	21	<0.001	0.780

Omission Error

The Cochran's Q test was significant ($p = 0.006$) suggesting there is an overall difference in omission error among methods. However, no pair-wise comparisons were significant after making a Holm's correction for multiple comparisons ($p>0.05$).

The graphical comparison among beach seining at different tide heights showed that detection frequency increased with decreasing tide height (Figure 3.1). It can be seen by this simple comparison that there appears to be a positive relationship between tide height and detection frequency, with low tide having the greatest detection frequency. The comparison among visual walk and boat-based methods suggests that the boat parallel survey is the most effective method out of the three for detecting sand lance, however the walk survey is close in detection frequency (Figure 3.2). Comparing among all major methods, the snorkel survey appears to be the most successful sampling method in detecting sand lance followed closely by low tide beach seining (Figure 3.3). Note that the graphical comparisons (Figure 3.1-3.3) have different detection frequency values to that of Table 3.3. This is since Table 3.3 shows the overall detection frequency for each method (all sites done with that method), while the graphical comparisons use sites common to all methods being compared.

All tests showed that sand lance abundance (as determined by low tide beach seining) were not significantly different between "present" and "absent" determined by each visual method (Table 3.4). This suggests that in this study abundance was not significantly related to the ability to detect sand lance visually.

None of the logistic regression models were significant suggesting that in the context of this study, water quality did not significantly affect omission error (Table 3.5) and thus, comparisons among visual methods do not need to consider water clarity.

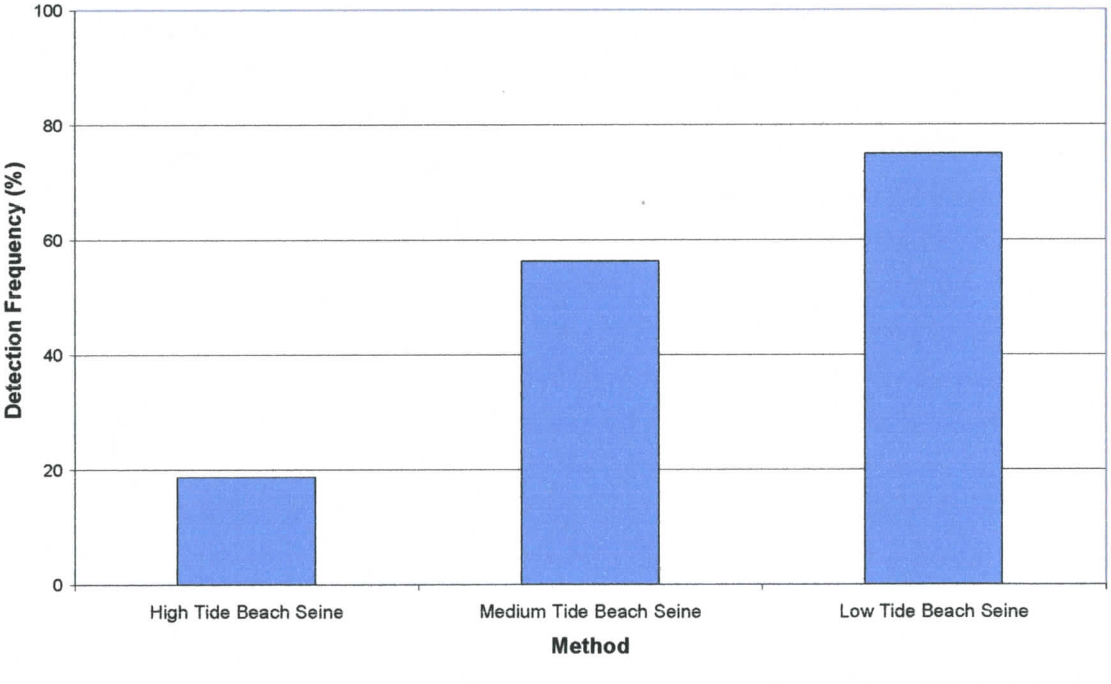


Figure 3.1: Comparison of beach seine set detection frequencies based on tide height (N=16).

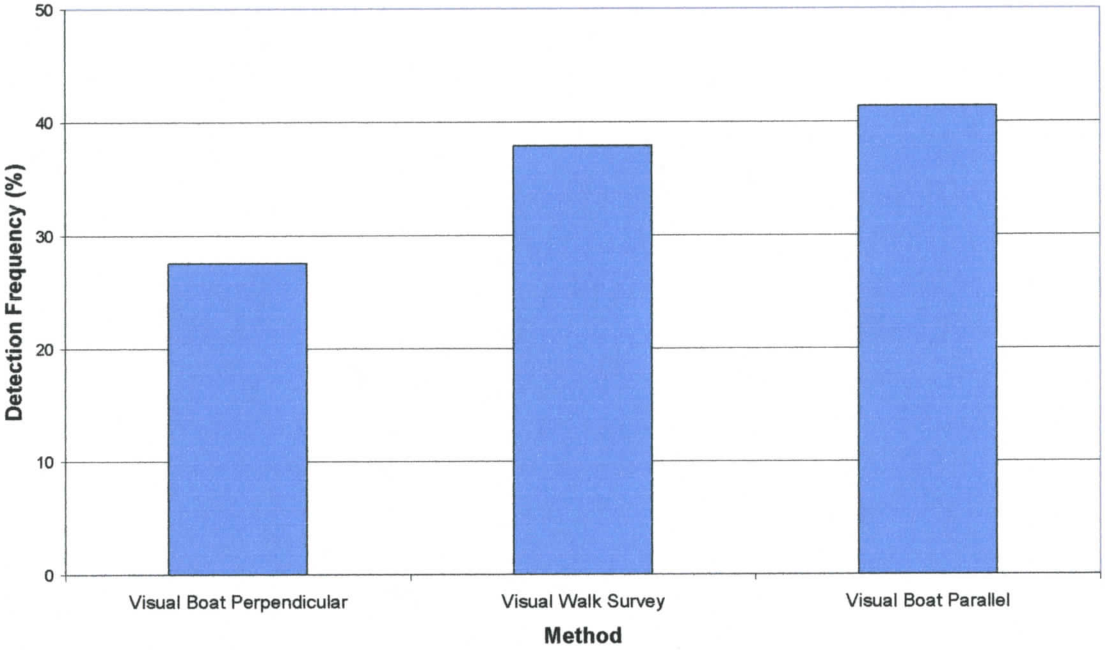


Figure 3.2: Comparison of sand lance detection frequency using visual method survey techniques (N=29).

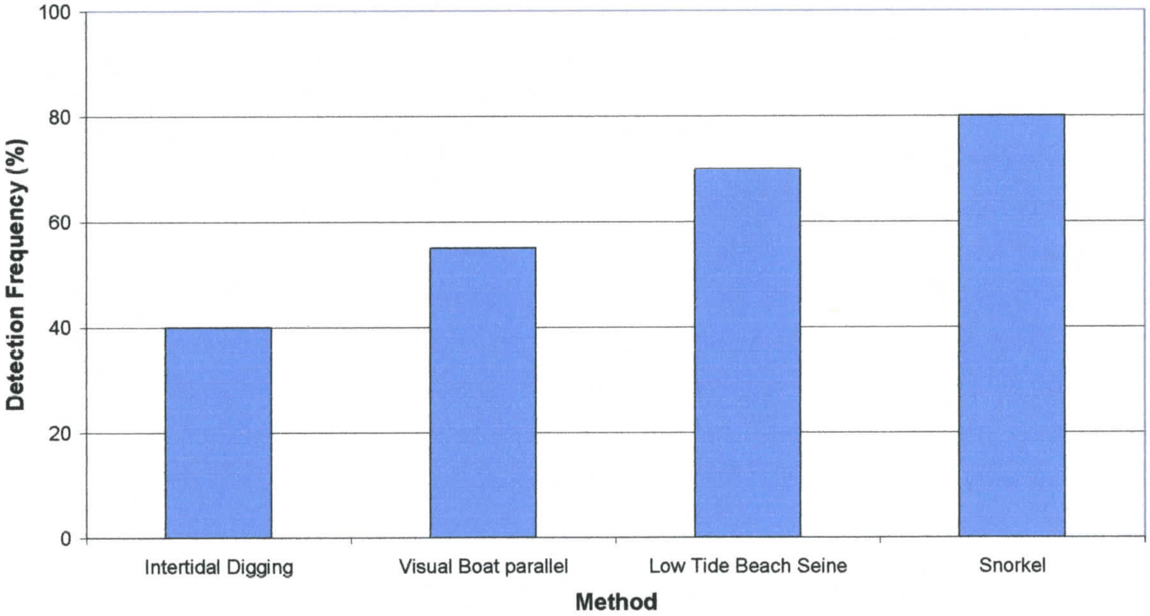


Figure 3.3: Comparison of sand lance detection frequency among intertidal digging, visual boat parallel, low tide beach seine, and snorkel surveys (N=20).

Table 3.3: Comparison of detection methods for sites that were designated as sand lance present for at least one method of sampling (the normative standard, N=35).

Survey Type	Number of Surveys	Number of Sand Lance Detections	Detection Frequency (%)
Low Tide Beach Seine	35	26	74.3
Medium Tide Beach Seine	29	15	51.7
High Tide Beach Seine	16	3	18.8
Visual Survey Boat - Parallel	31	13	41.9
Visual Survey Boat - Perpendicular	33	11	33.3
Visual Survey Walk	34	13	38.2
Digging Survey	34	15	44.1
Snorkel Survey	20	17	85.0

Table 3.4: Mann-Whitney U results comparing sand lance mean rank abundance grouped by visual survey presence/absence.

Visual Survey	Abundance Mean Rank "Present" [N]	Abundance Mean Rank "Absent" [N]	Mann-Whitney U	p-value
Snorkel	7.75 [4]	11.76 [17]	21	0.275
Boat-Parallel	19.86 [14]	13.89 [14]	79	0.077
Boat-Perpendicular	21.36 [11]	15.65 [23]	84	0.123
Walk	20.58 [12]	15.82 [22]	95	0.191

Table 3.5: Logistic regression results using survey method detection (presence/absence) as the dependent and secchi depth (water clarity) as the independent variable.

Method	N	Chi-square	Model Significance	Variable Significance
Visual Survey Boat - Parallel	27	3.206	0.073	0.113
Visual Survey Boat - Perpendicular	29	2.566	0.109	0.150
Visual Survey Walk	30	2.609	0.106	0.134
Snorkel Survey	19	0.386	0.534	0.538

Beach Seining at Different Tide Heights - Abundance

Repeated measures Friedman test statistic was significant (p = 0.002). *Post-hoc* comparison of the marginal means showed that low tide set had the highest sand lance abundance, followed by medium tide set and last high tide set (Table 3.6). Relative abundance was significantly lower at high tide compared to medium (p=0.045) and low tide (p=0.004), but there was no difference in abundance between medium and low tide (p=0.097). The Friedman test comparing only low tide and medium tide sets (increased

N) showed that relative abundance was not significantly higher at low than medium tide ($p = 0.061$).

Table 3.6: Comparison of marginal mean estimators for each tide height (controlling for the variation between sample sites).

Tide Height	Marginal Mean Estimator (Sand Lance)	Std. Error
Low Tide	1.437	0.481
Medium Tide	0.805	0.303
High Tide	0.232	0.19

3.4 DISCUSSION

Sand lance can be present at a sample site for two reasons. The first reason is that there are specific physical habitat characteristics of that site that fill sand lance habitat requirements (*e.g.* sediment for burying or sites for reproduction) and so sand lance are actively selecting the location. The second reason is that sand lance are there "incidentally" (*e.g.* searching for food in the water column or passing by the site) and are not present due to specific habitat conditions. If sand lance are present for reasons other than the habitat characteristics of the site, this could confound attempts to determine the species-environment relationship (Fielding 2002, Huston 2002). Because the different sampling methods focus sampling efforts on different regions of a site, one method may be more prone to sample sand lance that are present due to stochastic events. Beach seine surveys and intertidal digging focus on the intertidal region and therefore are more likely to sample sand lance that are using the physical burying habitat. Visual sampling is more likely to detect sand lance at the periphery of the sites and thus, visual methods may be more prone to sampling sand lance not strongly associated with the physical characteristics of the site.

As discussed above, the normative standard used here was not the true occurrence of sand lance but may represent a reasonable approximation. However, it is necessary to point out that this normative standard is not independent but rather is based on the unbalanced combination of all sampling methods that I am testing. Thus, the true sampling efficiencies of each method cannot be revealed by this study. Also, since the normative standard was partially based on visual methods, there is a possibility that it includes some sand lance present designations due to stochastic events.

The snorkel survey had the highest detection frequency out of all methods, however, due to logistical constraints I was able to conduct snorkel surveys at only 56% of all sites (31 of 55 sites). Snorkel surveys can be difficult to conduct in the field due to safety considerations and poor weather conditions. However, one advantage of snorkeling is that it can be conducted at sites that cannot be sampled using a beach seine. This advantage makes snorkel surveys a viable survey method for broad community surveys (Brind'Amour and Boisclair 2004). This advantage was not utilized in this study since the ability to sample with a beach seine was one of the main criteria for selecting sites. However, use of snorkel surveys could expand the areas that could be sampled for sand lance.

Snorkel and beach seine survey methods have been found to perform similarly in determining community descriptors in inland lake ecosystems (Brind'Amour and Boisclair 2004). Brind'Amour and Boisclair (2004) found that for some estimates, such as fish density and biomass, snorkel survey underestimated true values compared to beach seining while other values, such as abundance, snorkel survey outperformed beach seining. For simple presence/absence data of a non-cryptic species such as sand lance (when in the water column), it is expected that snorkeling detected sand lance more readily than beach seining. Snorkeling has the advantage of covering a broader area of the sample unit and sand lance can be readily seen when schooling in the water column or when entering and exiting the sediment. Although easily detected, I often found it difficult to estimate sand lance abundance visually for snorkeling and visual surveys. Sand lance often school in large numbers that make estimation difficult. Also, although it is usually easy to detect sand lance exhibiting burrowing behavior as they enter or exit the sediment, it is impossible to estimate the numbers already buried. Presumably without a marked out sample area it is difficult to standardize estimates of relative abundance, especially if water turbidity, and hence detectability of fish, varies.

Beach seining at low tide had a comparable Kappa value and a slightly lower but similar detection rate to that of snorkeling. The efficiency of beach seining as a nonbiased sampler for sand lance (Robards *et al.* 1999b) makes it an attractive choice for sampling. Although it requires more resources (*i.e.* boat, net, at least two individuals), it is an easily standardized technique that is not affected by water visibility. Beach seining also has the advantage of physically capturing the target species allowing for measurements and accurate counting as well as providing information on other fish species at the site.

Visual surveys that followed the bathymetric contour parallel to the shoreline (walk survey and parallel boat survey) had a better detection rate than surveys that took a cross-section perpendicular to the shoreline (perpendicular boat survey). This finding probably reflects the fact that the parallel surveys concentrate the survey effort on the shallow subtidal/low intertidal region, a region in which sand lance appear to be concentrated during low tide. The perpendicular survey only surveys this region briefly during the cross-section. The perpendicular survey progresses away from the shore line and therefore much of the area sampled by the survey is in depths in which sand lance would be required to venture away from intertidal burying sediment.

Though boat based or land based visual surveys are easy to conduct, they have some disadvantages. First, they are less proficient in detecting sand lance than beach seining at low tide and snorkel surveys. There is almost a 20% drop in detection frequency when compared to beach seining and almost a 30% drop compared to the snorkel survey. Second, like the snorkel surveys, it is difficult to accurately estimate abundances or year classes of the sand lance seen and there is a small chance of a false positive by visually misidentifying another fish species as sand lance by inexperienced observers. Also, as for snorkel surveys, they are likely less efficient when water clarity is poor, although a relationship was not detected in this study.

Intertidal digging was one of the least successful methods for detecting sand lance presence when compared to other major methods (Figure 3.3), but does have some advantages. First, it is probably logistically the easiest to conduct in terms of effort and requires minimal resources. Second, like the beach seine, intertidal digging is easily standardized and allows for the capture of sand lance for aging measurement. Third and perhaps most important, is the ecological information implied when sand lance are discovered by this survey technique. When sand lance are found in the intertidal sediment it is clear that they are utilizing the site specifically as habitat (*i.e.* as a refuge). The same is likely true for the beach seining methods, which focus sampling on the intertidal region. Finding sand lance buried in sediment suggests that sand lance are present due to habitat selection rather than stochastic events.

Tide Height

This study supported previous work suggesting that tide height affects fish catch efficiency of a beach seine (Gibson *et al.* 1996). When comparing frequency of occurrence, low tide beach seine sampling had a detection frequency approximately four times greater than that at high tides, with medium tides falling in-between. This suggests that the probability of detecting sand lance using beach seine sampling may be inversely functionally related to tide height. Also, as tide height increases there also is a decrease in abundance of sand lance.

These relationships may be explained by the effects of the tidal cycle on the distribution of sand lance and their availability to the beach seine. Because sand lance appear to remain close to the site throughout the tidal cycle, it may be that the tidal cycle acts to dilute sand lance abundance at high tides or concentrate at low tides within the intertidal and shallow subtidal region. Sand lance have been found to move into intertidal regions during high tide (Robards *et al.* 1999b). At high tide, sand lance distribution is unrestricted throughout the intertidal and shallow subtidal, thus their concentrations are diluted in these areas. When the tide is low, sand lance are concentrated in the shallow subtidal region with the exception of those that remain buried in the intertidal sediment. This in turn increases the probability of catching sand lance (frequency of occurrence) and the probability of a higher abundance.

In Alaska, Robards *et al.* (1999b) found that frequency of occurrence and relative abundance for juvenile sand lance sampled with a beach seine was lower at low tide than at high tide. Robards *et al.* (1999b) suggested that the juvenile sand lance may move from demersal habitat to the water column during the flood tide and return to the demersal during ebb tide. Dick and Warner (1982) and Blackburn and Anderson (1997) found similar results for the same region in Alaska. This is the complete opposite relationship to that I found in this study. It is difficult to postulate the reasons for the disparate results found between this study and those mentioned above. It is possible that there are major differences in general site characteristics between the two regions, such as physical structure and sediment distribution, which may cause differences in sand lance behavior in response to the tidal cycle.

3.5 CONCLUSION

Out of all sampling methods evaluated, snorkel and low tide beach seine surveys were the most effective in detecting sand lance. However, snorkeling was not conducted at all sites sampled because of logistical constraints and it was difficult to estimate sand lance abundance with this method. Therefore, I used the beach seine low tide survey to determine sand lance abundance and presence/absence for the subsequent habitat models presented later in this study.

CHAPTER 4: Temporal Variation in Sand Lance Distribution and Abundance

4.1 INTRODUCTION

Species occur "in space through time" (Henebry and Merchant 2002) and thus, the temporal scale of a study can strongly affect habitat selection (Heglund 2002). Many studies fail to replicate temporally, and therefore only provide a cross-sectional representation of habitat selection (Heglund 2002). Studies that fail to put habitat models in the context of temporal variation assume a species is in equilibrium with its environment (Hirzel and Guisan 2002), however, since population characteristics fluctuate over time, this is generally not the case (Van Horne 1983).

Here, I examine seasonal variation in sand lance occurrence in 2002 and the interannual variation in abundance and occurrence over the three summer sampling seasons 2002-2004. Although subsequent habitat models presented in later chapters do not use data from all three years, the variation detected here provides a temporal context to the habitat modeling process.

4.2 METHODS

Sampling

Sand lance abundance estimates in this study were based on sampling at low tide using a beach seine as described in chapter two and thus results are relevant only to YOY.

Seasonal Distribution

Since it was not logistically feasible to re-sample all sites throughout the season, it was important to test the assumption that sand lance usage at a site (*i.e.* occurrence) remained constant throughout the summer season. In 2002, 15 sites were re-sampled one to four times throughout the summer in order to test whether sand lance occurrence was constant.

Interannual Distribution and Abundance

I compared sand lance occurrence and abundance among the three years of sampling. I assessed the difference in occurrence (presence/absence) using a Cochran's Q test to determine whether there was an overall difference among the three years and a McNemar's Chi-square test to examine paired differences between years (McNemar 1947). I compared yearly abundance using a repeated measure Friedman test for *Sand lance Set Abundance*. Sites that were sampled in all three years were used for the comparison (N=37).

Sites that had sand lance present in all three years may be of particular importance as sand lance habitat. It is possible that these sites may also have a higher abundance of sand lance for any given year compared with sites that had sand lance present in only one or two years of sampling. Sand lance abundance values within 2003 (the year with the most sampling) were examined by grouping sites based on differences in detection frequencies from all three years. Sites sampled in all three years could be placed into four categories based on whether sand lance were present in: 1) one year (N=6), 2) two years (N=7), 3) in all three years (N=8), or 4) not present in any years (N=41). Using abundance data from 2003 I compared the first three groups both graphically and statistically, using a repeated measures Friedman test. I also compared between sites that had sand lance present in all three years (category 3, N=8) *versus* sites that had sand lance present in either one or two years (categories 1 and 2 combined, N=13) using an independent sample t-test.

4.3 RESULTS

Seasonal Distribution

In 2002 I found consistency of detection in 14 of the 15 re-sampled sites, with only one site showing an inconsistency in detection (Table 4.1). At the only inconsistent site (LYALL POINT N), I caught sand lance on the first sampling event on July 24th but not the second on August 25th.

Table 4.1: Within-year consistency of sand lance occurrence for 15 sites with multiple sampling events in 2002. Values represent abundances based on the catch values of the beach seine at low tide. Dates with more than one sample event during the same time period are separated by commas.

	Abundance						
Site	Early June	Late June	Early July	Late July	Early August	Late August	Consistency In Sand Lance Detection?
MAYNE BAY	0	0					y
REEKS		0, 0					y
TURRET-EELGRASS					0	0	y
HANDCAMPN					0	0	y
STOPPER N	0	0, 0			0		y
DICEBOX					22	46	y
DODD BAY					2	10	y
DODD W				2200	191	11300	y
LYALL POINT N				49		0	n
TAPALTOS					9	1	y
TRICKETT-14		69		6			y
DEMPSTERSE		25			18	7	y
EFF LIGHT				500	1800	3200	y
TRICKETT1		6	604		8500		y
CLARKE CAMP	289	15	27	3	742	45	y

Interannual Distribution and Abundance

Frequency

The summer of 2003 had the highest frequency of sites with sand lance present compared to 2002 and 2004 (Table 4.2, Figure 4.1). There were no instances where sand lance were present in either 2002 or 2004 without being present in 2003. The summers of 2002 and 2004 had similar frequencies of sand lance presence however the same sites were not always utilized by sand lance in these years (Table 4.2). The Cochran's Q test showed that there was a significant difference in occurrence between years (Chi-square=10.80, d.f.=2, p=0.005). Paired McNemar Chi-square comparisons showed 2002 and 2004 did not differ significantly from each other (p=1.00) however both had significantly lower frequencies compared with 2003 (p= 0.01 and p=0.04 respectively). When comparing between only 2003 and 2004 (N=60, a larger sample size) I found the same result of a higher frequency of sand lance presence in 2003 (Table 4.3).

Table 4.2: Comparison of sand lance abundance (*Sand lance Set Abundance*) for sites sampled in all three years of the study. Abundance is based on low tide beach seine data (sand lance per seine set)

Site	Abbreviation	2002	2003	2004	Frequency of occurrence (years)
CHALK	CHA1	8	70	1	3
CLARKE CAMP	CLCP	289	5000	500	3
DEMPSTERNE	DENE	1	9	1200	3
DEMPSTERSE	DESE	18	58	375	3
DODD BAY	DOBA	10	5	1	3
EFF LIGHT	EFFL	500	45	7	3
SANFORD NW	SANF	5	1500	12	3
TRICKETT1	TRIC	8575	10000	155	3
DICEBOX	DICE	22	800	0	2
DODD NORTH	DONO	1	27	0	2
DODD W	DOWE	1000	3000	0	2
HELBY	HELB	376	5	0	2
NETL-1	NET1	0	4	11	2
NETTLE 2	NET2	0	9	38	2
TZAR ROBB	TZAR	0	400	140	2
BENSON E	BENE	0	3000	0	1
BENSONCH	BENC	0	3500	0	1
DIANA E	DIAE	0	12500	0	1
FABER NE	FABE	0	1	0	1
OHIAT	OHIA	0	34	0	1
WOUWERBATELY	WOBA	0	1	0	0
BRABANT 55	BR55	0	0	0	0
ROSS ISLETS	ROSS	0	0	0	0
BRAB 53	BR53	0	0	0	0
BRADYS	BRAD	0	0	0	0
CLARK WB	CLWB	0	0	0	0
HAINES E	HAIN	0	0	0	0
HAND NORTH	HANO	0	0	0	0
HANDCAMPN	HACP	0	0	0	0
JAQUE-JAR	JAQU	0	0	0	0
PINK 48	PI48	0	0	0	0
REEKS	REEK	0	0	0	0
SCOTTS BAY	SCOT	0	0	0	0
TURRET-EELGRASS	TUEG	0	0	0	0
TURTJAQU	TRTJ	0	0	0	0
WILLISCAMP	WILL	0	0	0	0
WIZARD	WIZA	0	0	0	0
Overall % Presence		32	57	29	NA

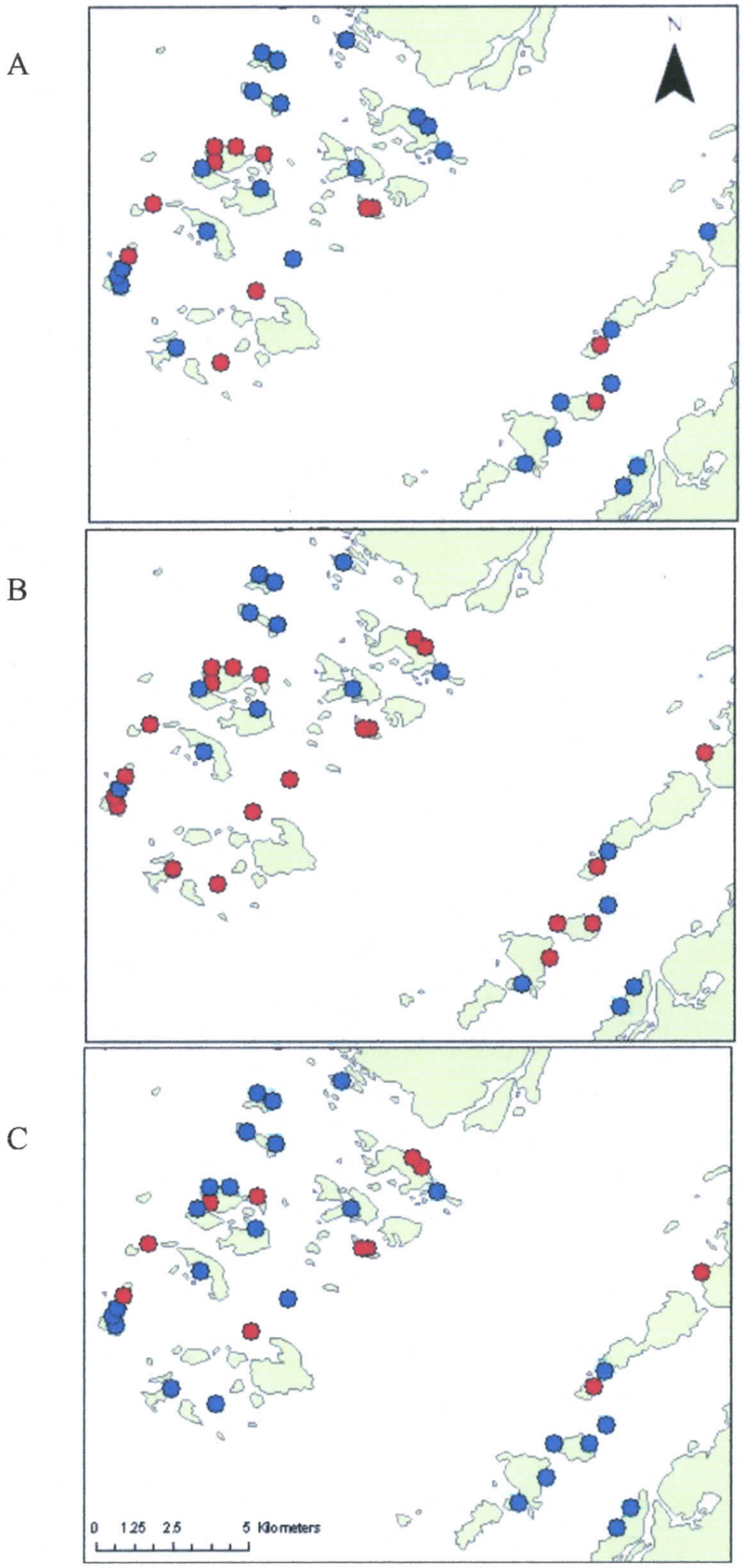


Figure 4.1: *Sand lance Presence/Absence* at 37 sites sampled in all three years (each dot is a site) for 2002 (A), 2003 (B), 2004 (C). Red= Present, Blue=Absent.

Table 4.3: Comparison of sand lance occurrence among sites sampled in both 2003 and 2004. Presence/absence designation is based on low tide beach seine data.

Site Presence/Absence	Number of Sites	Proportion of sites (%)
Absent in Both Years	30	50
Present in 2004 only	2	3
Present in 2003 only	16	27
Present in Both Years	12	20
Total	60	100

Abundance

The Friedman test statistic was significant (p = 0.001) suggesting that there was a significant difference among years for *Sand lance Set Abundance*. A *post-hoc* comparison of mean abundance showed that 2003 had a considerably higher mean (Table 4.4). The pair-wise comparison demonstrated a significant difference between 2003 and 2002 (p=0.042), 2003 and 2004 (p=0.030) but not between 2002 and 2004 (p=0.338). Therefore, 2003 had a significantly higher mean *Sand lance Set Abundance* than the other two years, which didn't differ significantly from each other.

Table 4.4: Comparison of mean abundance for each year (sand lance per site, N=37).

Year	Mean Sand Lance Abundance	Std. Dev,
2002	300	1431
2003	1110	2778
2004	67	221

When I compared abundance values of sites categorized by sand lance present in one, two or three years I found no trend or significant differences among the three categories (Figure 4.2; Friedman statistic p=0.335). When comparing the combined categories 1 and 2 (sand lance present in either 1 or 2 years) with category 3 using a t-test, abundance was still not significantly different (p=0.704).

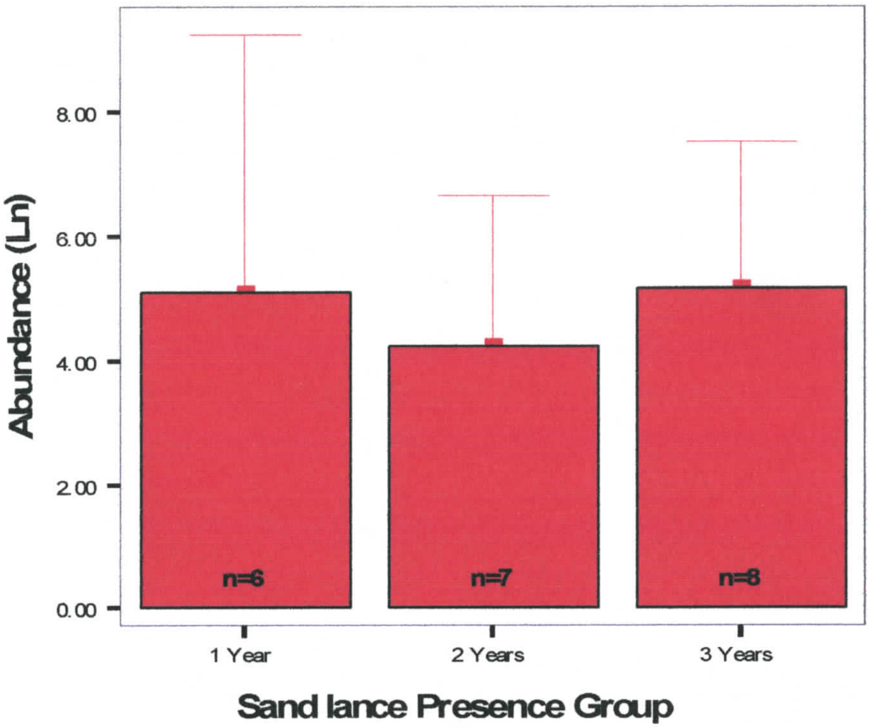


Figure 4.2: Mean sand lance abundance (sand lance per seine set) for 2003. Sites were grouped based on whether they had sand lance present in one, two or all three years. Error bars represent the 95% confidence interval.

4.4 DISCUSSION

Seasonal Distribution

Empirical models, such as those presented in subsequent chapters, assume that a species distribution is at equilibrium (*i.e.* static) such that it does not fluctuate during the temporal window for which the sampling is done (Hirzel and Guisan 2002). Results from this study strongly suggest that detection consistency (occurrence) could be assumed to be constant throughout the season. Sites where sand lance were caught had sand lance present throughout the sampling season in 2002, while sites where sand lance were not caught were not used by sand lance throughout the sampling season. This supports the assumption that sand lance habitat use remains constant during the sampling season (mid-May to early September) and that one sampling event per site per season is adequate to sample sand lance occurrence.

Interannual Distribution and Abundance

It is difficult to study the factors regulating marine fishes since the systems in which they exist are demographically open (*e.g.* high immigration and emigration; Robards *et al.* 2002, Webster *et al.* 2003). Further, the temporal scale of this study was too short to investigate possible mechanisms in the variation of distribution and abundance. However, it may be useful to compare results here to that of longer-term studies in order to draw inferences of possible mechanisms of variation. Also, the examination of interannual variation is important as it puts habitat models developed in later chapters in the context of three years of temporal variation.

Frequency

Wright and Begg (1997) suggest that because North Sea sand lance (*A. marinus*) are dependent on specific sediment types, the distribution of post-settled sand lance is "restricted and constant through time". This comment was based on data from Wright (1996), which used sampling units (scale grain) larger than 1 km^2. Although the statement may possibly be true at that scale for that species, the results from the three years of summer sampling at the finer scale of this study show that sand lance distribution can fluctuate significantly. Wright and Begg (1997) were referring to a North Sea population of sand lance (*A. marinus*), however, *A. hexapterus* has similar biology and ecology.

In support of Wright and Begg's (1997) statement, there were a group of sites in my study that had sand lance present in all three years. These eight sites (22% of the sites sampled in all three years) maintained their consistency in use as sand lance habitat in each year of sampling and therefore may be areas where sand lance can be found predictably each season. Sites that were used by YOY sand lance in all three years did not show a significantly higher abundance when compared to sites where sand lance use varied over the three years. This suggests that the sites that were used in all three years, although perhaps temporally important, did not support more sand lance than sites used only in one or two years. Density does not necessarily reflect habitat quality (see Van Horne 1983 for a detailed discussion) thus, even though sites may vary in temporal importance, that importance does not have to be related to the density that sites can support.

Although there was a group of sites where use was constant, the fact that there was interannual variation in habitat use for other sites implies that the distribution of sand lance is not "restricted" to certain sites. The yearly occurrence data show that sand lance appeared to have had a wider distribution in 2003 than 2002 and 2004 (Figure 4.1). As well, almost all of the sites that had sand lance present in 2002 or 2004 also had sand lance present in 2003. Sites that showed temporal variation in sand lance use may be less important temporally than sites that sand lance use each year however without proper study of the relative fitness of sand lance in relation to these two groups of sites, it is impossible to determine a "value" for each group.

The question then arises: What is the cause of the interannual variation of sand lance use of the sites? One possibility is that there may be changes in environmental characteristics of sites between years causing changes in habitat quality. Sediment characteristics are likely one of the more important habitat factors for *Ammodytes* as shown in the literature (*e.g.* Pinto 1984, Dick and Warner 1982, Wright *et al.* 2000). There are no studies of temporal changes in sediment parameters for the BGI area to my knowledge, and it is difficult to generalize from studies conducted in other areas on temporal variation in sediment. Because I did not assess sediment characteristics for each year, rather only for 2003, I cannot quantitatively assess the interannual variation, however I did not notice any large changes in sediment composition among years.

Another possible reason for interannual variation is dispersal of the sand lance. Dispersal is a largely stochastic process, being influenced heavily by chance (Huston 2002). A "false negative" may be a result of a situation where a species dispersal processes has not yet colonized a suitable area. The chance of a species occupying a suitable habitat rests heavily on the mode of dispersal, with higher mobility leading to an increase in the chance that the suitable habitat will be colonized (Huston 2002). Sand lance can disperse into new habitat as pre-settled larvae settling into the nearshore or with movement as YOY and 1+ sand lance. Little is known about the mechanisms behind sand lance larval recruitment to the nearshore or the daily movement (*e.g.* home-range) of post-settled sand lance thus, it is difficult to draw conclusions about dispersion capabilities. Pre-settled larvae *Ammodytes* have the potential to have a wide dispersion (Proctor *et al.* 1998) however, it has been suggested that post-settled sand lance (*A. marinus*) do not move far from their habitat in which they settle and the broad scale distribution is determined during the larval phase (Jensen *et al.* 2003). However, Kuhlmann and Karst (1967) describe daily movements of sand lance in the West Baltic to be within a range of a "few kilometers" (cited in Pinto *et al.* 1984) and Gauld (1990) reported tagged sand lance movements of 27-64 km over 1-3 years suggesting potentially high dispersion capabilities.

Lastly, interannual differences in sand lance distribution may be directly related to abundance. The increase in occurrence in 2003 coincided with a significantly higher abundance compared with 2002 and 2004 (which had abundances similar to each other). It is possible that the increase in distribution in 2003 may be associated with the increase in abundance with that year. Since abundance reflects the YOY population, it suggests that recruitment for 2003 was higher than in 2002 and 2004. Thus, the increase in YOY abundance in 2003 may have lead to a broadening of the distribution as predicted by the Fretwell-Lucas ideal-free distribution model (Fretwell and Lucas 1970). As population density increases a greater diversity of habitat types should be occupied. If population densities are extremely high then individuals may be forced to occupy "suboptimal" habitat (Heglund 2002).

Abundance

Forage fishes generally have short life spans, early maturation and high fecundity (Abookire and Piatt 2005). Forage fish populations can demonstrate quick fluctuations due to high predation and rapid population response to environmental change (Anderson and Piatt 1999, Abookire and Piatt 2005). *Ammodytes* have been found to have high interannual variation in abundance in other locations (*e.g.* Nelson and Ross 1991, Kishi *et al.* 1991, Arnott and Ruxton 2002, Arnott *et al.* 2002) including *A. hexapterus* (Robards *et al.* 1999b). Because the majority of the sand lance sampled in this study were YOY (see chapter two), the interannual fluctuation in sand lance abundance is most likely due to variation in recruitment. Sand lance interannual recruitment is known to be highly variable within the genus (Robards and Piatt 1999). Recruitment itself can be affected by multiple factors including broad scale environmental conditions and population dynamic effects.

Broad scale environmental variables such as seasonal water temperature can affect the abundance of prey available for larvae (Haroldson *et al.* 1993, Robards *et al.* 1999b, Arnott and Ruxton 2002), the development of fish eggs and the development and behavior of larvae (*e.g.* Johnston *et al.* 1998). By directly affecting the plankton production cycle, broad scale climate conditions in turn affect larval fish recruitment. However, the life history of sand lance causes it to be resistant to the effects of temporal variation in zooplankton peaks (Robards *et al.* 1999b) and therefore the mechanism of sand lance recruitment variation with regards to environment is poorly understood.

Population dynamics can affect larval recruitment through density dependent processes. Research in the North Sea by Arnott and Ruxton (2002) showed that *Ammodytes marinus* recruitment was significantly affected by density. They found a negative relationship between recruitment and first year sand lance densities. Similar results have been found for *A. personatus* in Japan (Hamada 1966, 1967, Kishi *et al.* 1991). All four of the above studies found a yearly alternation of high and low recruitment similar to the variation in abundance found here. It may be the case that the yearly fluctuation in abundance seen here may be linked to the yearly fluctuation in recruitment. Again, the mechanism is unclear. Arnott and Ruxton (2002) suggest that the negative relationship could arise from disturbance of eggs by burrowing behavior by first year sand lance, cannibalism after the eggs hatch by first year sand lance or direct competition for food and space between first year sand lance and the recruiting larvae.

Although the exact mechanism is unknown, it is likely a combination of both environmental conditions and population dynamics that affect recruitment of sand lance (Arnott and Ruxton 2002). In addition to recruitment, other factors are suspected to affect sand lance interannual variation such as overwintering mortality of adults (Kishi *et al.* 1991).

4.5 CONCLUSION

With only three years of sampling data it is difficult to draw conclusions about interannual variation in population dynamics, however, trends observed here in abundance match those shown in other studies on *Ammodytes*. In regards to variation in distribution, there was a marked interannual variation in sand lance occurrence that coincided with interannual changes in abundance with 2003 having a significantly higher frequency and abundance compared to 2002 and 2004. Habitat models built on data from populations at abnormally high levels may not perform well when predicting habitat selection for lower population densities (O'Connor 1986). Since it appears that the sand lance YOY population had high abundances in 2003, results from following habitat modeling should be interpreted with caution.

CHAPTER 5: Modeling Nearshore Habitat for Young-of-the-Year Sand Lance: An Empirical Inductive Approach.

5.1 INTRODUCTION

Species distributions are rarely uniform in time and space (Mackey and Lindenmayer 2001). Many ecological studies try to explain the variability in species distributions by modeling species-environment relationships. A fundamental premise in this modeling is that there is a predictable relationship between a species and relevant environmental variables, an assumption that is based in niche theory (for a discussion see Heglund 2002). Studies construct habitat models using biotic or abiotic variables that are either known or supposed to have causal influence on habitat use patterns or assumed to be correlated to a causal variable that was not measured. Resulting habitat models can then be used to describe and/or predict patterns of species distribution, ranging from single species models to complex population dynamics or species diversity models (Guisan and Thuiller 2005).

According to Corsi *et al.* (2000), the first major useful categorization of habitat models is whether the model uses a deductive (inferring from a general pattern to a particular instance) or an inductive approach (using particular facts to infer a general pattern). A deductive approach uses known information on the species-environment relationship to model how a species will be distributed. This approach has been termed elsewhere as "mechanistic" or "process based" modeling (Guisan and Zimmermann 2000 and references therein, Maurer 2002). Deductive models make inferences on the species-environment relationship from known relationships in the literature or from expert opinion and. This information, which is usually assumed to be cause-and-effect in nature, is then combined to describe the species-environment relationship. The constructed species-environment relationship can then be used to create a habitat model.

An inductive approach is used when there is a poor understanding of the species-environment relationship prior to modeling (Stauffer 2002). Inductive methods generally use quantitative statistical models. This approach has also been termed "empirical" or "phenomenological" modeling (Guisan and Zimmermann 2000 and references therein, Maurer 2002). Inductive models use a known species distribution (generally found through field data) to determine statistically the species-environment relationship (Kobler and Adamic 2000). The abundance or the probability of occurrence of the species is correlated with the environmental variables within the study area producing an indirect description of the species-environment relationship (*i.e.* not specifically using any information about the life history or physiological constraints). Rather, these models determine correlations between species distribution and habitat parameters. Where a deductive approach uses data to create the habitat model after the species-environment relationship has been deduced, an inductive model objectively uses data to first determine the species-environment relationship and then uses this relationship to create the habitat model. The strength of this approach lies in the objectivity in determining variables that are important in shaping a species distribution (Van Horne 2002). Models are not constrained to use only the known relationships based on our own limited understanding of a species ecology. Instead, the data "speak for themselves" with important variables being selected through iterative modeling procedures.

Inductive habitat models generally have one of two goals: 1) prediction or 2) description (De'ath and Fabricius 2000, Maurer 2002, O'Connor 2002). Predictive models focus on estimating the probability of finding a species at a point in time and space rather than description of important characteristics. The species-environment relationship is constructed in the same fashion as in the descriptive approach and is generally then backfitted over an unstudied region to provide a spatially explicit model predicting the species habitat use. Descriptive models represents the "systematic structure of the data as simply as possible" (De'ath and Fabricius 2000). These models are concerned with the exploration of variables that are important to a species habitat selection rather than focusing on predicting areas of suitable habitat. Habitat modeling through exploratory models is a useful way to discern information about an organism's ecology (Fielding 2002). Although predictive models can be useful for management purposes, Austin (2002a) cautions that descriptive goals should not be treated as secondary since description detects important species-environment relationships and can test ecological theory upon which mechanistic deductive models are based.

Selecting Modeling Procedures for Modeling Sand Lance Habitat

Selection of a modeling approach can depend on many factors such as data collection logistics and the overall goal of the modeling procedure (Maurer 2002). The amount of ecological information available on the species is another important consideration since it can restrict whether a deductive or inductive approach should be used. When there is a lack of information that prohibits a deductive approach, an inductive approach is appropriate. Here, I have selected two inductive analytical models for modeling sand lance habitat selection: logistic regression and Classification and Regression Tree (CART, the use of this acronym here is not to be confused with acronyms associated with proprietary software). Logistic regression has been one of the most popular ways to model presence/absence data for many years (Tyre *et al.* 2001, O'Connor 2002). Classification and Regression Tree is a class of models that have recently been used in ecological applications and have proven to be promising in habitat modeling.

Logistic Regression

Although many statistical approaches have been developed for habitat modeling, regression type analyses are the most common (Austin 2002b). Maximum likelihood logistic regression is a type of Generalized Linear Model (GLM) that has been extensively tested and has proven to be a robust technique (Reckhow *et al.* 1987, Brotons *et al.* 2004 and references therein). Logistic regressions assumptions can be met with relative ease (Reckhow *et al.* 1987) and it has a comparable performance to other models (Manel *et al.* 1999, Elith 2000). Binomial logistic regression is used for binary dependent variables (*e.g.* presence/absence) and assumes a binomial error structure (Hinsley *et al.* 1995). Logistic regression handles independent variables that range from negative infinity to positive infinity (Shriner *et al.* 2002) and also has the ability to input categorical independent variables, interaction terms, covariates and accounts for redundancy of input variables (Parasiewicz and Dunbar 2001).

Classification and Regression Trees

Ecological data commonly have complex and unbalanced structure (De'ath and Fabricius 2000). Relationships between variables are regularly nonlinear and involve interaction effects (Rejwan *et al.* 1999, De'ath and Fabricius 2000, Scott *et al.* 2002). Often, traditional statistical exploratory and modeling techniques do not adequately model these complexities and meaningful ecological relationships can be overlooked. In response to these difficulties, researchers are pursuing other avenues for modeling ecological data. One result of this is the rise in popularity of CART models.

CART is a class of models that are ideal for dealing with the complexities of ecological data. CART models are non-parametric and do not make assumptions about the relationships between explanatory variables or the functional relationships between explanatory variables and the dependent variable (Andersen *et al.* 2000). Also, CART automatically handles spatially autocorrelation among predictors (Anderson *et al.* 2000, Cablk *et al.* 2002), which is increasingly being recognized as a confounding issue with more classical approaches such as GLMs (Keitt *et al.* 2002). Because of these advantages, CART models have been shown out-perform traditional models (Rejwan *et al.* 1999, De'ath and Fabricius 2000, Dettmers *et al.* 2002, Olden and Jackson 2002 and Muñoz and Felicísimo 2004).

CART explains the variation of a single categorical or continuous dependent variable using one or more categorical or continuous independent variables. If the dependent variable is categorical, the model class is "classification tree". If the dependent variable is continuous, the model class is "regression tree". Trees are constructed by repeatedly dividing and subdividing the data until the stopping criterion has been satisfied. Each sequential binary split divides the observations into groups, attempting to maximize homogeneity within groups. The splits are based the on sum of squares for regression trees and proportions of categories (*e.g.* presences and absences) for classification trees (De'ath and Fabricius 2000). At each split in the tree, all variables are considered as the potential splitting variable. The variable with the highest predictive power is used to split the data, thus maximizing the homogeneity of the two nodes created by the split. For a more detailed description of how trees are grown see De'ath and Fabricius (2000) and Larson and Speckman (2004).

Graphically, a CART model is in the form of a hierarchical tree where data are represented by nodes resembling branches. Trees begin with the "parent" or "root" node that contains all the data. The root node is then recursively split into subsequent nodes until terminal nodes are reached. The terminal nodes, or "leaves", represent the final groupings. Since a tree can be grown to the extent of having only one observation for each terminal node, an *a priori* user-based decisions determine to what degree the tree is grown. Many different strategies have evolved to determine the optimal tree size including overfitting a tree and then pruning or shrinking back to a desired level (*e.g.* Huettmann and Diamond 2001), iterative cross-validation techniques to select or prune a tree (*e.g.* Moisen and Frescino 2002, Larsen and Speckman 2004), or *a priori* rules to determine minimum and maximum tree sizes (*e.g.* Yen *et al.* 2004). The final tree displays the hierarchical habitat selection process with each node split representing a habitat selection "decision" and with interactions between explanatory variables easily identifiable within the tree structure (Rejwan *et al.* 1999).

Overall Modeling Approach

The following two chapters use inductive exploratory habitat modeling using logistic regression and classification tree models. This chapter examines habitat selection at a finer scale grain using empirical environmental data from field measurements while chapter six examines sand lance habitat selection at a larger scale grain using GIS derived environmental data. Because of the limited extent of the study area and the paucity of information regarding sand lance habitat requirements, the models provided will focus on a description and exploration rather than prediction.

Biotic Interactions

Although it is argued that physical and chemical factors can be used to define habitat for the purpose of conservation (*e.g.* Zacharias and Roff 2000, Roff and Taylor 2000), these factors would be insufficient in describing patterns without the consideration of species interactions (Salomon *et al.* 2001). Biotic interactions are more difficult to study and have generally been underrated in past ecological models (Linnell and Strand 2000). In order to explore possible biotic relationships that shape sand lance habitat selection, I conducted a separate analysis examining for possible differences in fish communities between sites with and without sand lance. Differences in fish communities may arise due to species interactions with sand lance (*e.g.* predation or competition) and the detection of these differences would give insight into the biotic interactions affecting sand lance.

5.2 METHODS

In a data-driven approach the data are used in an exploratory fashion as opposed to the hypothesis-driven approach where the data are used in a confirmatory fashion. Although a data-driven approach for determining the species-environment relationship may have some pitfalls (see Welch and MacMahon 2005), for relatively unstudied species, models constructed *a priori* runs the risk of missing important environmental correlates which may yet be unknown (*e.g.* Dettki *et al.* 2003). Here I use a data-driven approach to model YOY sand lance habitat selection using site-specific environmental data measured in the field. The process of statistical analysis begins with exploration of the data followed by the modeling (De'ath and Fabricius 2000). I use a modeling approach involving the following steps: 1) the examination data for relationships and structure, 2) selection of variables that will appropriately represent the data, 3) the fitting of the model using a suitable modeling technique, 4) the examination of the model fit using various summaries and validation procedures, 5) the repetition of steps 3 and 4 until a satisfactory model has been reached.

The use of multiple modeling techniques is extremely valuable in habitat modeling (Maurer 2002). Many authors utilize one or more modeling techniques for data exploration and independent variable selection, then use a different technique for the final model (*e.g.* Yen *et al.* 2004). Also, different techniques can be used to independently construct multiple final models in order to compare and contrast. This often can lend greater support to results and conclusions (*e.g.* Norcross *et al.* 1999). Here, I use a univariate approach to select independent variables to be used in modeling from the suite collected. I select the dependent variable by using the ordination procedure ANOSIM (ANalysis Of SIMilarities). I used two separate independent modeling procedures, logistic regression and CART (classification tree), to construct the final habitat models. These two model types have been used in conjunction by other authors (*e.g.* Norcross *et al.* 1999, Clarke *et al.* 1999).

Data Set

I collected a wide array of environmental variables in 2003 and out of the 60 sites sampled, 55 had essentially complete environmental data. I dropped the five sites with incomplete data from modeling analysis.

Analysis started with a total of 30 explanatory variables (Appendix 5.1). Except for eelgrass (*Zostera* sp.) this list did not include any variables regarding specific macro-vegetation species because during preliminary screening I did not find any associations.

Due to the nature of the sampling design, I could not monitor environmental characteristics at each site throughout the sampling season. I therefore chose environmental variables that could be considered more or less constant throughout the summer and one measurement would be adequate. Sea surface temperature and salinity was recorded for each site only once at one point of one day during the season. Temperature and salinity have a relatively high temporal variation, thus, sample site temperature and salinity could not be readily compared among sites. Temperature and salinity were therefore not included in modeling.

Exploratory Analysis

Independent Variables

I examined each independent variable to determine the shape and nature of its distribution. I transformed variables that deviated from normality (tested by examining normality plots and Kolmogorov-Smirnov test values). Various transformations were used including sine, power and natural log transformations, however, even after transformation, some variables did not pass the Kolmogorov-Smirnov test for normality. Therefore, I used non-parametric univariate tests during exploration.

Even after transformation the *Intertidal Eelgrass % Cover* and *Subtidal Eelgrass % Cover* were still far from normally distributed due to a high frequency of zero counts and strongly skewed data. To alleviate this problem, I converted each of the % eelgrass variables to a binary classification of eelgrass presence/absence (both intertidal and subtidal).

Preliminary analysis showed that sediment grain size properties are likely important for sand lance habitat selection. I examined sediment grain size distribution using *Sand lance Presence/Absence* as a factor. I compared the grain size distribution for sites with sand lance present (N=26) *versus* sites with sand lance absent (N=27) using graphical methods. Eelgrass presence/absence was also shown to be important in preliminary analysis. Since grain size and eelgrass presence/absence are likely related, I did two similar analyses of grain size distribution using *Intertidal Eelgrass Presence/Absence* and *Subtidal Eelgrass Presence/Absence* as the factors.

Dependent Variables

As discussed in chapter two, sand lance sampled were primarily YOY thus the following dependent variables are specific to the YOY year class. YOY sand lance abundance data had an extremely high variation and was dominated by zero counts (Figure 5.1). Because they exist in patchy schools, high variability in abundance for beach seine samples was common. This is a common occurrence with schooling fish as schools are "encountered more contagiously than would be expected in a random distribution" (Kupschus 2003). Distributions of this nature present difficulties for statistical analysis since they generally violate assumptions for many statistical tests due to overdispersion (Nielson *et al.* 2005). There are various ways in dealing with this problem including using a specialized class of regression, however, these developments are relatively new and still have some weaknesses (Pearce and Ferrier 2001, Jones *et al.* 2002). Here, I chose to create categorical variables out of the abundance data. By creating semiquantitative dependent variables, the skewed abundance data is collapsed into a multinomial response, maintaining a semiquantitative estimate of abundance (Guisan 2002). This is the approach Litzow and Piatt (2003) took to avoid this problem specifically with sand lance catch data. As well, clustering literature suggests that categorical dependent variables may be more suitable in this type of situation (Zacharias *et al.* 1999). Here, I created three dependent binary categorical variables, *Sand lance*

Presence/Absence, Sand lance High/Low, Sand lance Burying/Not burying and one dependent ordinal categorical variable, *Sand lance High/Medium/Low/Nil* (Table 5.1). I chose what I considered biologically relevant breaks when creating categorical abundance dependent variables based on *Sand lance Set Abundance*.

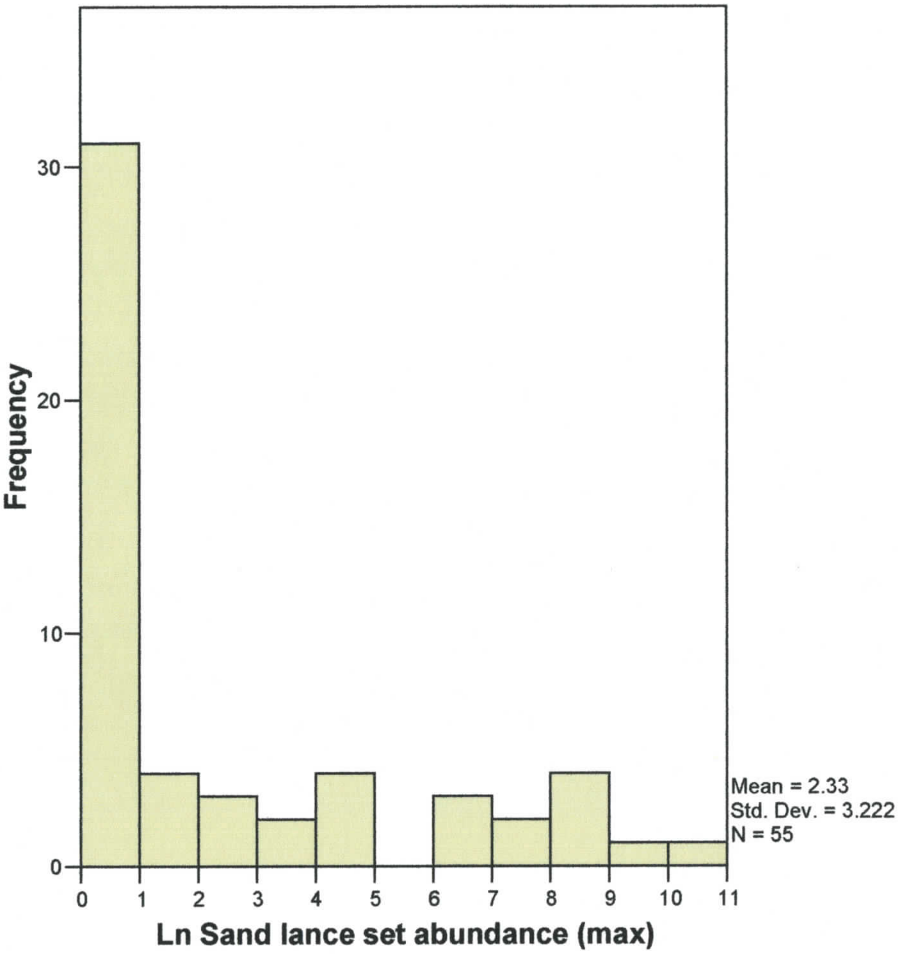


Figure 5.1: Histogram of the dependent variable *Sand lance Set Abundance* (sand lance per set, ln transformed).

Table 5.1: Dependent variable list considered for analysis.

Dependent Variable	Type	Description
Sand lance Set Abundance	Continuous	Mean sand lance abundance value (fish per seine) from 2 replicate beach seine sets conducted at low tide.
Sand lance Presence/Absence	Categorical - binary	Based on "*Sand lance Set Abundance*". If "*Sand lance Set Abundance*" >0 then sand lance are considered "present" (defined as "1"). If "*Sand lance Set Abundance*"=0 then sand lance are considered "absent" (defined as "0")
Sand lance High/Low	Categorical - binary	Based on "*Sand lance Set Abundance*". If "*Sand lance Set Abundance*" >10 then sand lance are considered "High" (defined as "1"). If "*Sand lance Set Abundance*" <10 then sand lance are considered "low" (defined as "0")
Sand lance Burying/Not burying	Categorical - binary	Based on whether sand lance were seen burying at the site during sampling. If sand lance were seen burying then "yes" (defined as "1"). If sand lance were not seen burying then "no" (defined as "0").
Sand lance High/Medium/Low/Nil	Categorical - ordinal	Based on "*Sand lance Set Abundance*". Nil: *Sand lance Set Abundance*=0, Low: 0<*Sand lance Set Abundance*<9, Med: 10<*Sand lance Set Abundance*<100, High: *Sand lance Set Abundance* >99

Univariate Testing

I conducted univariate analyses to explore the relationships among independent and dependent variables. Each independent variable was analyzed separately using each categorical dependent variable as the grouping variable. It is often not appropriate to run multiple individual univariate tests because the inflated chance of making a type I error (Clarke and Warwick 2001), thus, this analysis was treated as exploratory.

For the binary dependent variables I ran a Mann-Whitney U test. The Mann-Whitney U test is the non-parametric equivalent to the commonly used t-test and has been used previously for variable selection (see Johnson *et al.* 2002). The independent variable was grouped by each binary dependent variable to test the null hypothesis that the two independent groups are from the same population. The categorical dependent variable *Sand lance Nil/Low/Med/High* has four categories therefore I used non-parametric Kruskal-Wallis test for several independent samples. The Kruskal-Wallis test is an extension of the Mann-Whitney U test comparing among multiple independent samples rather than just two. I tested categorical independent variables (see Appendix 5.1) using Chi-square tests.

Dependent Variable Selection

The creation of categorical variables can be a qualitative process (*i.e.* subjective bins for abundance). To determine whether a dependent variable has a significant relation to the independent variables I ran an ANOSIM for each dependent variable. In this context, the ANOSIM tests the hypothesis that there is a significant difference in the environmental characteristics (independent variables) categorized by the dependent variable. I conducted a one-way ANOSIM (999 permutations) for each categorical dependent variable using PRIMER for Windows (V 5.2.9) on a normalized Euclidean distance similarity matrix. For each dependent variable, I used the independent variables found significant for that variable in the *Univariate Testing* section.

ANOSIM analysis suggests that *Sand lance Presence/Absence* is the best descriptor of independent variables found significant for each respective dependent (similar results were found when I included all independent variables). *Sand lance*

Presence/Absence was the only dependent variable that had significant relation to the environmental characteristics at the 0.05 level (Table 5.2, Figure 5.2), and thus, was the only dependent variable with categories the environmental characteristics can distinguish between. *Sand lance Presence/Absence* is therefore likely the most appropriate dependent variable for further analysis.

Table 5.2: Results from the ANOSIM for each dependent variable. *Sand lance Presence/Absence* was the only variable that was significant at the 0.05 level.

Dependent Variable	Sample statistic (Global R)	Pseudo p=value
Sand lance Presence/Absence	0.166	0.001
Sand lance High/Low	-0.017	0.575
Sand lance Burying/Not burying	0.025	0.229
Sand lance Nil/Low/Med/High	0.051	0.245

Correlations

In modeling procedures, multi-collinearity between independent variables can confound the analysis (Knapp *et al.* 2003). In order to screen for collinearity I ran a non-parametric bivariate correlation matrix using all the independent variables. I considered correlations above the cut off value of $|r| \geq 0.70$ to be highly correlated (Berry and Felman 1985). When two or more variables were highly correlated I kept the variable that was significantly related with *Sand lance Presence/Absence* variable as determined by the univariate analyses. Where more than one competing correlated variable was significantly related to a dependent I ran a preliminary logistic regression model for each variable using *Sand lance Presence/Absence* as the dependent. The variable that produced the best classification rate was kept for further analysis and the other variable(s) were dropped from the modeling process. I also generated scatter plots for each independent variable plotted *versus* all other independent variable to examine for non-linear correlations and possible outlier values.

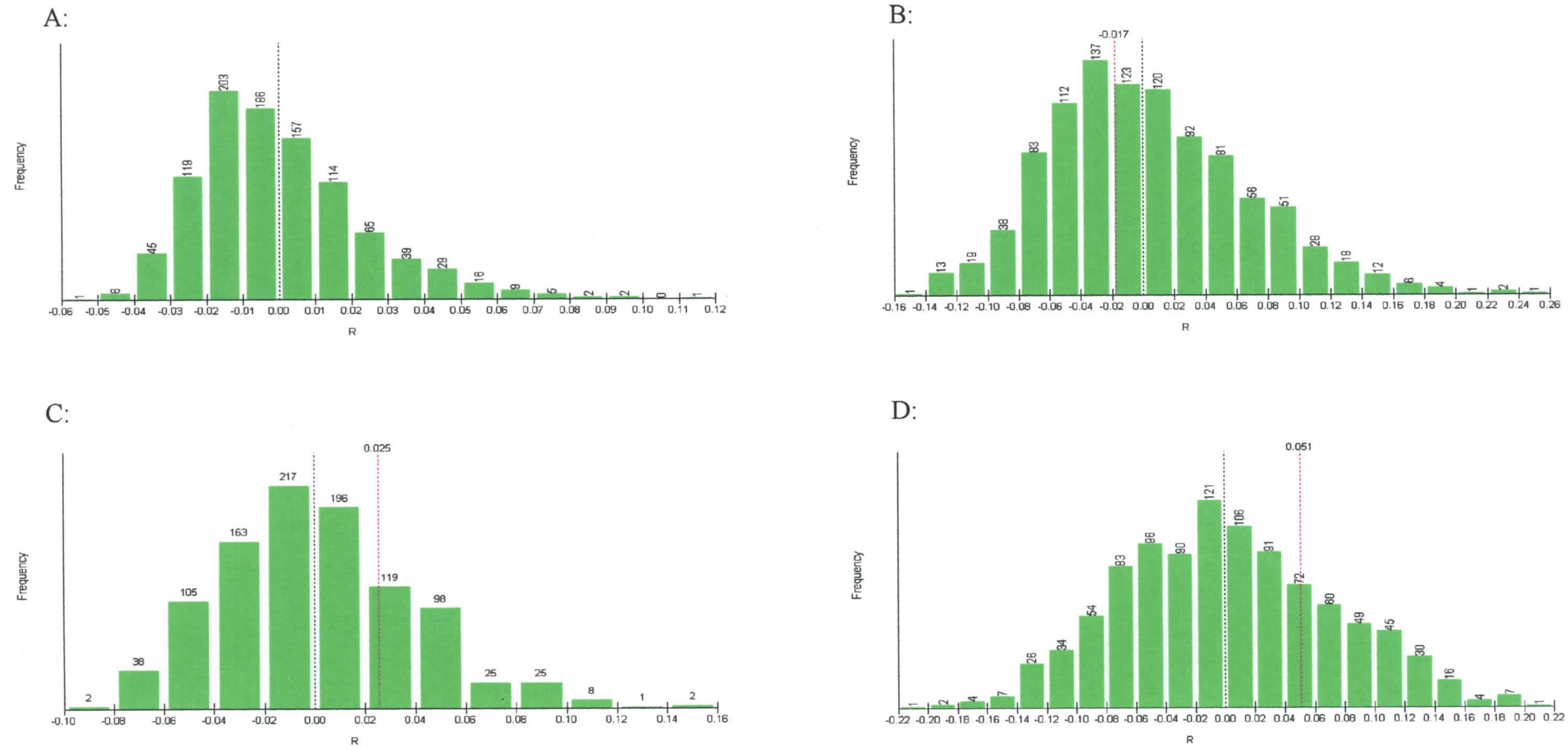


Figure 5.2: ANOSIM for four dependent variables A: *Sand lance Presence/Absence*, B: *Sand lance High/Low*, C: *Sand lance Burying/Not burying*, D: *Sand lance Nil/Low/Med/High*. The data label for the sample statistic is not displayed for *Sand lance Presence/Absence* because the sample statistic off the scale (global R=0.166).

Independent Variable Selection

One difficulty with using an empirical inductive approach is the potentially excessive number of independent variables that can be measured or created. Since only a limited number of variables influence a species distribution, the capacity to collect and create high numbers of independent variables leads to the necessity of independent variable selection (Van Horne 2002). I selected variables to include in modeling analysis based on the results of the exploratory univariate analysis, an approach used by numerous authors (Manel et al. 1999, Elith 2000, Franco et al. 2000, Parris 2002, Brotons et al. 2004). I considered variables that were significantly related to the dependent variable *Sand lance Presence/Absence* to be likely important. I designated these variables to be used in logistic regression and classification tree analyses.

Logistic Regression

Automated stepwise selection techniques are commonly used in habitat modeling to select explanatory variables for the final model (*e.g.* Manel *et al.* 1999, Pearce and Ferrier 2001, Parris 2002). I used a stepwise logistic regression modeling procedure to construct the habitat model for sand lance based on *Sand lance Presence/Absence*. The backwards elimination stepwise procedure begins with all variables in the model and then sequentially removes variables based on a chosen criterion. I began with the independent variables that I selected in the *Independent Variable Selection* section above. The stepwise selection removed variables based on the probability of the likelihood-ratio statistic, which is based on the maximum partial likelihood estimates. I used an exclusion/inclusion value of 0.05 for the variable selection at each step (classification cut-off =0.5, maximum iterations=20).

Because preliminary results showed that intertidal eelgrass had a perfect negative association with sand lance distribution (*i.e.* all sites with intertidal eelgrass had sand lance absent), I ran a separate logistic regression for sites with intertidal eelgrass absent. I used the same stepwise technique with the remaining independent variables. For this procedure, the sites with intertidal eelgrass present were filtered out leaving N=43.

Cross-validation

Examining resubstitution classification rates is often not adequate for model evaluation. Most habitat models are not adequately tested through cross-validation techniques (Lauver *et al.* 2002). Although the main purpose of the models constructed here is descriptive rather than predictive, model validation and the proper examination of model accuracy is essential in evaluating models. Cross-validation techniques can be used to select the preferred model from competing models (Death and Fabricius 2000). The difference between resubstitution and cross-validation is that cross-validation uses data independent from the model building process to evaluate the models predictive performance while resubstitution evaluates the model with the same data used to build the model (Reckhow *et al.* 1986). Since the model has been optimized to predict the data used to build it, cross-validation, which uses unbiased data, is considered a more accurate evaluator than resubstitution (Olden and Jackson 2002).

I used a form of cross-validation referred to as "data splitting" or "k-fold partitioning" (Fielding and Bell 1997). I randomly split the data into five equal subsets. The first subset was removed from the original data to be used later to validate the model leaving four fifths of the data with which to build the model. The logistic regression model was constructed with the remaining four fifths using the variables selected for in the stepwise logistic regression procedures in earlier steps. This model was then used to predict the remaining one-fifth subset that was removed prior to model building. The first subset was then added back to the original data. This procedure was repeated for each subset. This iterative process results in the prediction of each observation using a model that is independent of that observation (*i.e.* the observation was not used in model construction). The mean empirical error is the final estimation of the total error for the model (Muñoz and Felicísimo 2004).

Model Assessment

I used various methods to assess competing models. Classification error performance (resubstitution), pseudo R-squared values, and Pearson's chi-square are commonly reported and used in assessing model fit (Boyce *et al.* 2002). In addition, I used cross-validation results (as discussed above) as well as AIC and ROC, two popular quantitative model selection techniques.

The Akaike's Information Criterion (AIC, Akaike 1973) is a common way to select among competing models (Boone and Krohn 2002) and has been suggested as the one of the preferred ways for model assessment (Burnham and Anderson 1998). AIC uses a tradeoff between model fit and a penalty for model complexity to evaluate each model. The model with the lowest AIC value is suggested to be the simplest model that best explains the data.

The Receiver Operating Characteristic (ROC) method is a threshold-independent method of assessing model performance based on the sensitivity and specificity (Elith 2000, Tyre *et al.* 2001, Fielding 2002). *Sensitivity* is known as the ability to predict species presence while *specificity* is the ability to predict species absence. The ROC procedure plots the sensitivity *vs.* 1-specificity. The area under the ROC curve (commonly referred to as the AUC, Area Under the Curve) indicates the chances of correctly predicting a randomly selected presence/absence pair. The area under the curve ranges from 0-1 with a value of 0.5 suggesting no relationship (*i.e.* the model does not predict better than a random choice) and a value of 1 suggesting perfect discrimination. Elith (2000) suggests an ROC value >0.75 indicates a model that may discriminate sufficiently to be useful for management purposes.

Predicted vs. Abundance

It has been suggested that predicted values of occurrence models may be significantly related to abundance (*e.g.* Pearce and Ferrier 2001). If this is the case, occurrence models could provide further information on possible abundance rather than simply presence/absence data. To test this, I explored the relationship between the predicted values of the preferred logistic regression model (evaluated by model statistics) and *Sand lance Set Abundance* using a bivariate correlation for all sites (N=55) as well as only sites with sand lance present (N=26).

Classification Tree

Classification trees were built in Statistica 7.0 (StatSoft Inc.2004) using *Sand lance Presence/Absence* as the dependent variable and the same independent variables that were used for the logistic regression analysis. I constructed trees using the algorithm QUEST (Quick, Unbiased, Efficient, Statistical Tree; Loh and Shih 1997). QUEST has been shown to be less biased in variable selection and requires less computation compared to exhaustive search algorithms (Lim *et al.* 2000). In QUEST, the choice of the terminal node to split and the splitting predictor variable for that node is determined by a discriminant-based univariate evaluation. P-values based on a Chi-square test (for categorical variables) determine the univariate split selection. Further details on splitting criterion for QUEST can be found in Loh and Shih (1997).

I chose the criteria for predictive accuracy for the tree model as the *a priori* probability of sand lance presence/absence (based on the data gathered in the study area). I used the default minimum N of five, which determines the minimum number of cases present in a terminal node with more than one class for which the splitting will stop.

Selecting a Tree Size

The V-fold cross-validation procedure (Breiman *et al.* 1984) is one of the preferred ways of pruning back the original tree to a suitable size and is commonly used in habitat modeling (Hahn and O'Connor 2002, De'ath and Fabricius 2000, Moisen and Frescino 2002, Olden and Jackson 2002). V-fold cross-validation is useful when there is no test sample to validate the model or when the training data set is too small to remove a test sample set from (Statsoft Inc. 2006). Here I used a value of ten for the cross-validation pruning.

After the trees' growth had been terminated, the 10-fold cross-validation prunes the tree by randomly partitioning the data into ten groups of equal or similar size. A classification tree of a specified size is built using nine of the ten groups and evaluated with the withheld group. This is done iteratively with each of the ten groups being withheld. This is repeated for each of the considered tree sizes starting with the terminal tree and continuing until the parent node is reached. The resulting ten cross-validation costs for each tree size can be then averaged giving the final estimate of the 10-fold cross-validation cost for each size.

Two methods exist for pruning the tree using the prediction error of 10-fold cross-validation (Statsoft Inc. 2006). The first prunes on misclassification error and is termed the minimal cost-complexity cross-validation pruning. The second prunes on deviance and is termed minimal deviance-complexity cross-validation pruning. Each method is valid but use different error estimates and therefore can yield slightly different results. I grew two trees, one using each pruning method. For each pruning method, I used the one standard-error rule (Breiman *et al.* 1984) to determine the final tree. This selects the smallest sized tree within one standard error of the lowest cost or deviance found.

I evaluated the final two trees using a global cross-validation procedure which is useful in evaluating the final classification tree models (Breiman *et al.* 1984). This technique splits the data into V groups (in this case ten was used). As in other cross-validation techniques, each of the groups is sequentially withheld as the test sample while the V-1 groups are combined as the training sample. The entire analysis is run for the nine remaining groups and the tenth group is used as the test sample. This procedure is repeated ten times. The classification error and the cross-validation cost calculated from the global cross-validation can be used to evaluate the final models.

Biotic Interactions

I ran a Multi-Dimensional Scaling (MDS) analysis to determine whether there are differences in the fish communities between sites with sand lance present and sand lance absent. The MDS was run on a Bray-Curtis similarity matrix created using a site by species abundance (fourth root transformed) matrix, excluding sand lance. I ran an ANOSIM analysis to determine whether there is a significant difference between the fish communities for sites with sand lance present compared to sites with sand lance absent.

5.3 RESULTS

Exploratory analysis

Grain size distribution

The grain size distribution of sites with sand lance present is similar in shape to the distribution of sites with sand lance absent (Figure 5.3). Both distributions are bimodal with the smaller mode occurring in the "fine sand" region and the larger mode occurring in the "coarse gravel" region (classified according to Wentworth 1922). The major difference between the two distributions is the higher level of fine sediments in the sand lance absent distribution. The sand lance absent distribution has a higher first mode. The right tail of the sand lance absent distribution has extremely fine sediment (silt/clay) present where the right tail of the sand lance present distribution is approximately zero.

Comparison of the grain size distribution of intertidal eelgrass present sites compared to absent sites showed that there were differences in the sediment composition between the two groups (Figure 5.4). The major differences were seen in the finer sediment areas. Sites that had intertidal eelgrass present also had higher proportions of finer sediment compared to the sites with no intertidal eelgrass. Surprisingly, sites with subtidal eelgrass had less fine sediments in the intertidal compared to sites without subtidal eelgrass (Figure 5.5). This suggests that the subtidal eelgrass occurrence may not be related to the processes governing sediment composition in the intertidal region. Comparison between the subtidal mud % content (from the underwater video analysis) between sites with intertidal eelgrass and subtidal eelgrass seemed to show again that intertidal eelgrass was more likely associated with finer sediments compared to subtidal eelgrass (Figure 5.6) although the sample sizes were low.

Univariate Testing

Results from univariate exploratory tests are shown in Table 5.3 for continuous independent variables and Table 5.4 for categorical independent variables. A notable finding was that the binary eelgrass variables relationship to S*and lance*

Presence/Absence. *Sand lance Presence/Absence* was not found to be significantly related to subtidal eelgrass, however, the relationship to intertidal eelgrass was highly significant ($p<0.001$). For all sites where intertidal eelgrass was present (N=12), sand lance were absent (Figure 5.7).

Correlations

The correlation coefficient ranged from -0.999 to 0.966 and there were three groups of variables that were highly correlated. I dropped the following seven variables from the analysis based on the criteria outlined in the methods section: *Slope subtidal*, *UV % Fine Sand*, *UV Coarse Sand %*, *% Gravel*, *% Sand*, *% Mud*, and *Eelgrass Presence/Absence*. I did not detect any non-linear relationships and saw no obvious outliers in the scatter plots.

Independent Variable Selection

I selected eight independent variables for modeling based on the univariate analyses: *Slope Intertidal*, *Grain Size Mean*, *Grain Size Sorting*, *UV Possible Burying %*, *UV Shell %*, *UV Mud*, *Major Substrate Low Intertidal*, *Intertidal Eelgrass Presence/Absence*. Mean plots of selected independent variables are shown in Figure 5.8 and raw data can be found in Appendix 5.2.

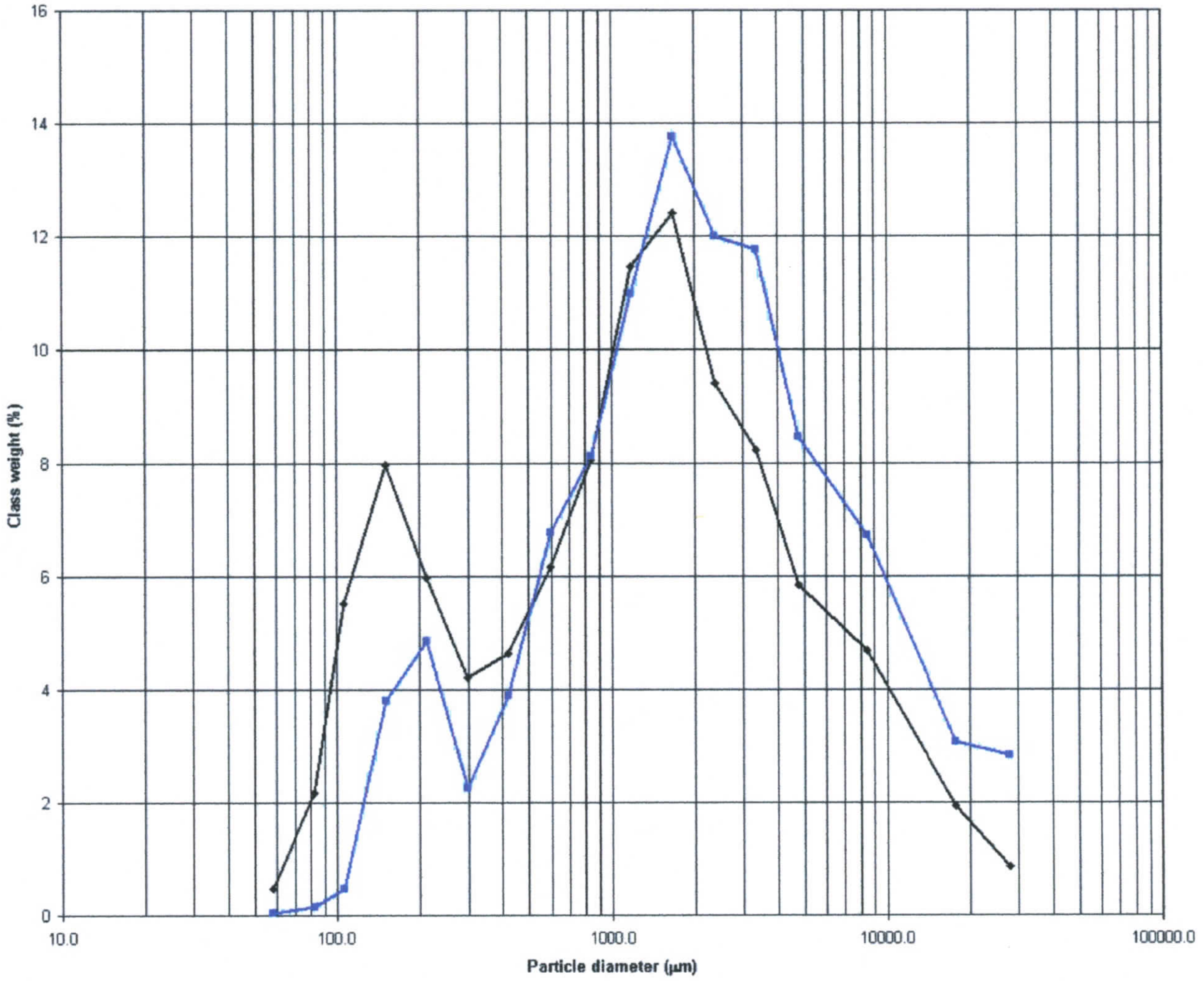


Figure 5.3: Grain size distribution for combined sites with sand lance absent (black, N=27) and combined sites with sand lance present (blue, N=26).

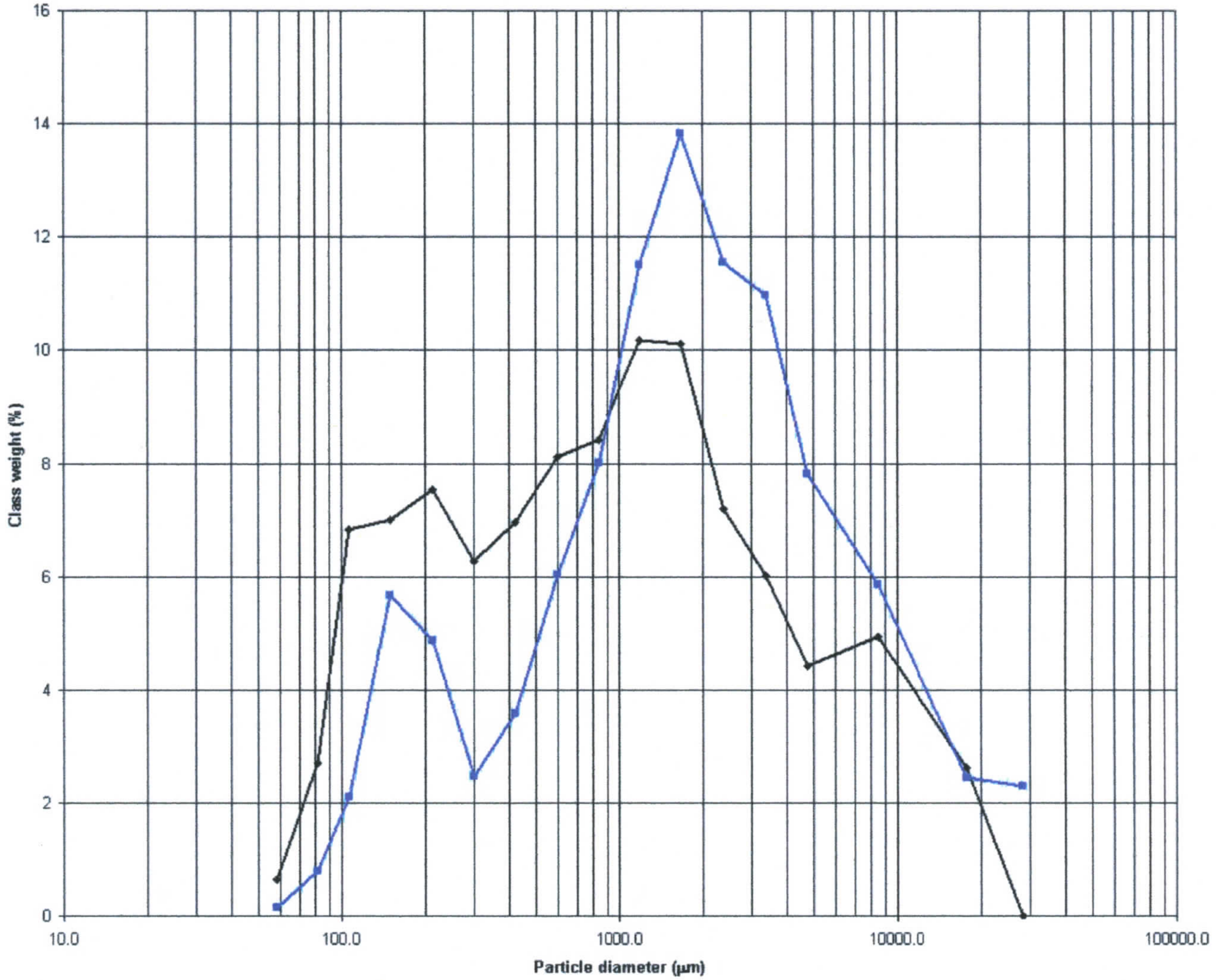


Figure 5.4: Grain size distribution for combined sites with intertidal eelgrass absent (blue, N=43) and combined sites with intertidal eelgrass present (black, N=12).

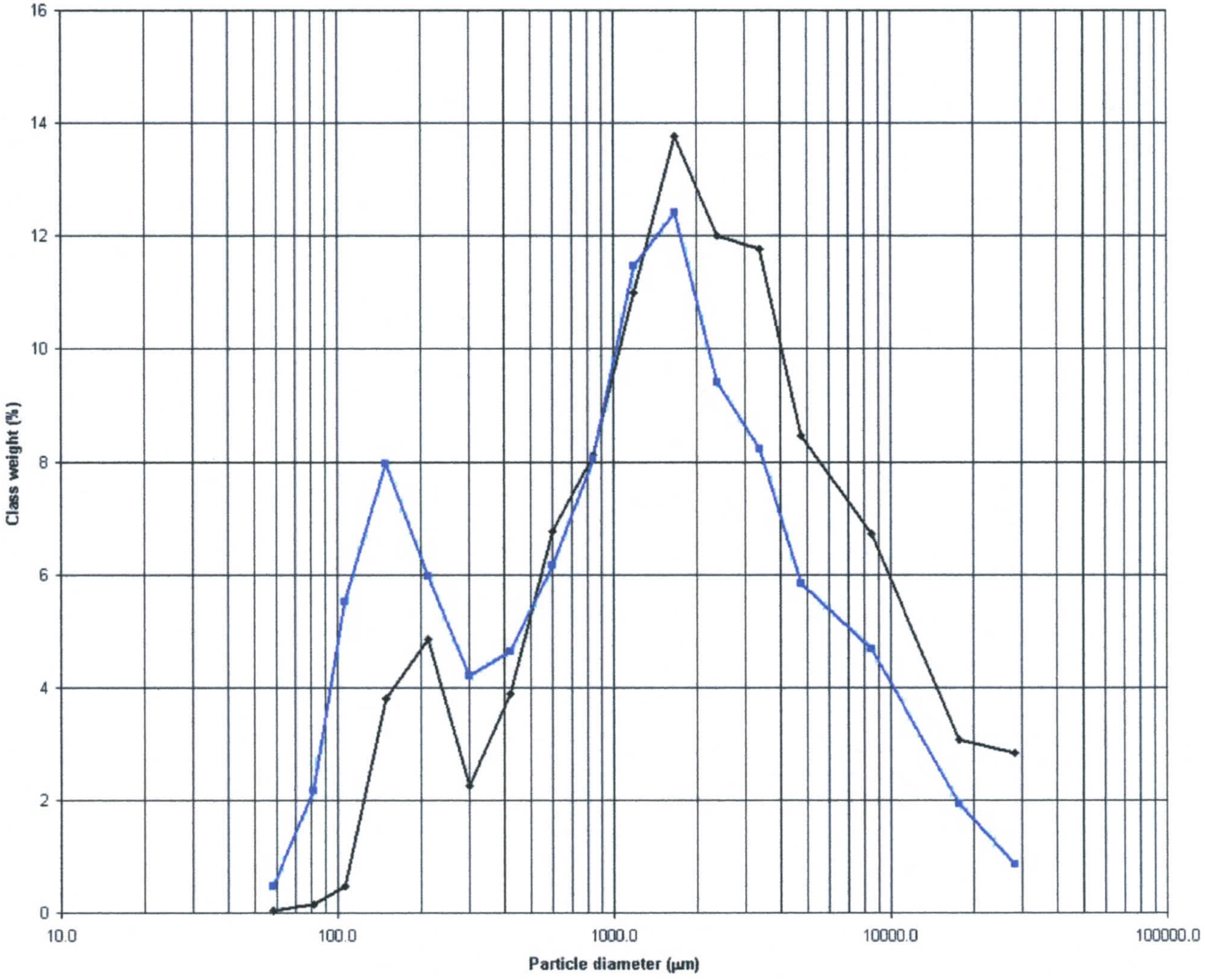


Figure 5.5: Grain size distribution for combined sites with subtidal eelgrass absent (blue, N=27) and combined sites with subtidal eelgrass present (black, N=26).

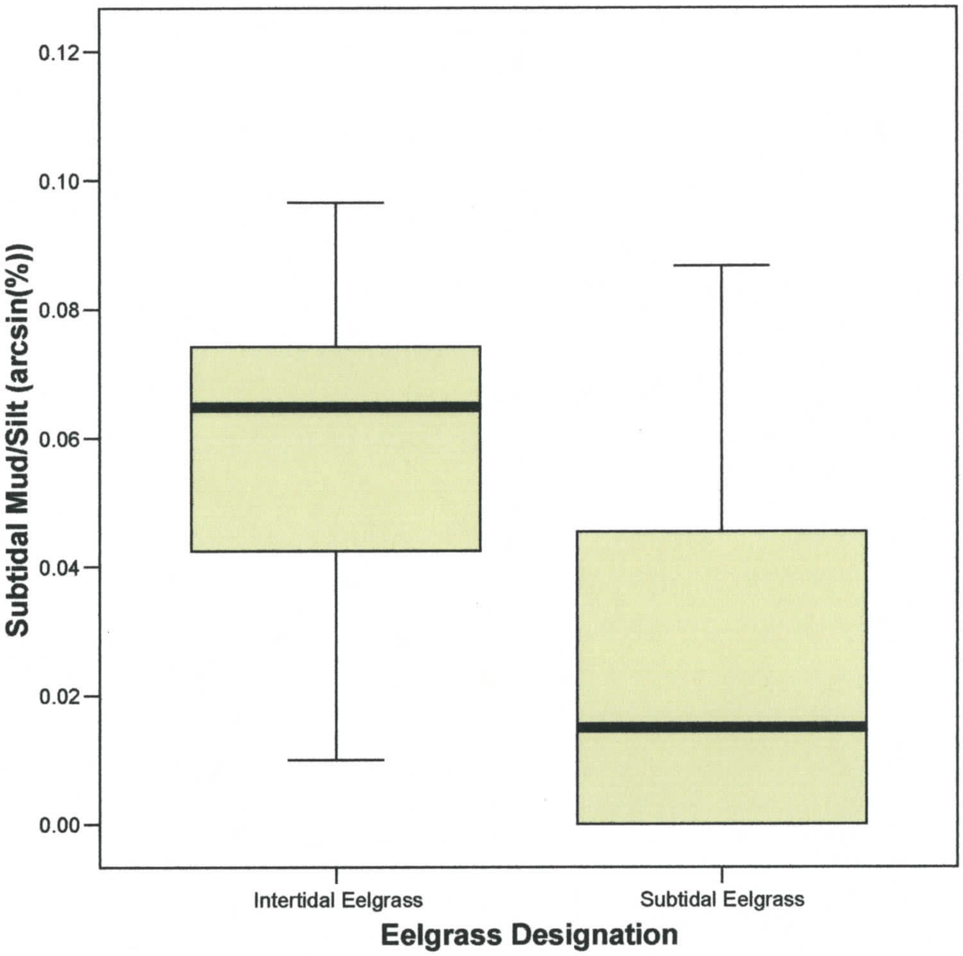


Figure 5.6: Comparison of the *UV Mud %* (transformed) between sites with intertidal eelgrass (N=5) and subtidal eelgrass (N=7). Sites with both intertidal and subtidal eelgrass were not included.

Table 5.3: Mann-Whitney U test for each continuous independent variable grouped by each dependent variable.
*Independent variable found to be significant at the 0.05 level.

	Mann-Whitney Asymp. Sig. (2-tailed)			Kruskal-Wallis (Asymp. Sig.)
Variable	*Sand lance Presence/Absence*	*Sand lance High/Low*	*Sand lance Burying/Not burying*	*Sand lance High/Medium/Low/Nil*
Slope Intertidal (°)	0.035*	0.088	0.295	0.027*
Slope Subtidal (°)	0.503	0.212	0.883	0.184
Bulk Substrate Mass (kg)	0.154	0.316	0.018*	0.447
Grain Size Mean	0.002*	0.069	0.006*	0.015*
Grain Size Sorting	0.006*	<0.001*	0.067	0.003*
Grain Size Skewness	0.067	0.387	0.226	0.293
Grain Size Kurtosis	0.946	0.863	0.738	0.931
Loss On Ignition 550	0.076	0.606	0.137	0.133
Loss On Ignition 950	0.133	0.247	0.235	0.358
UV Possible Burying %	0.045*	0.045*	0.034*	0.193
UV Macro-vegetation % mean	0.500	0.117	0.668	0.187
UV Shell %	0.043*	0.140	0.114	0.097
UV Mud %	0.005*	0.005*	0.020*	0.022*
UV Fine Sand %	0.443	0.163	0.306	0.527
UV Coarse Sand %	0.007*	0.110	0.015*	0.051
UV Pebble/Cobble %	0.150	0.418	0.156	0.541
UV Boulder/Bedrock %	0.273	0.030*	0.634	0.159
Intertidal Macro-vegetation Cover (%)	0.070	0.081	0.001*	0.739
Subtidal Macro-vegetation Cover (%)	0.786	0.421	0.711	0.407
% Mud	<0.001*	0.085	<0.001*	<0.001*

Table 5.4: Chi-square significance test for independent categorical variables grouped by four dependent variables.
*Significant at the 0.05 level.

	Dependent Variables			
	Chi-square (Pearson asymp. Sig. 2-sided)			
Independent Variables	*Sand lance Presence/Absence*	*Sand lance High/Low*	*Sand lance Burying/Not burying*	*Sand lance High/Medium/Low/Nil*
Major Substrate High Intertidal	0.178	0.175	0.576	0.036*
Major Substrate Low Intertidal	<0.001*	0.002*	0.012*	<0.001*
Eelgrass Presence/Absence	0.029*	0.135	0.174	0.186
Intertidal Eelgrass Presence/Absence	<0.001*	0.009*	0.011*	0.003*
Subtidal Eelgrass Presence/Absence	0.26	0.592	0.132	0.625
Aspect	0.298	0.169	0.259	0.524
Beach Log Cover	0.561	0.827	0.157	0.875

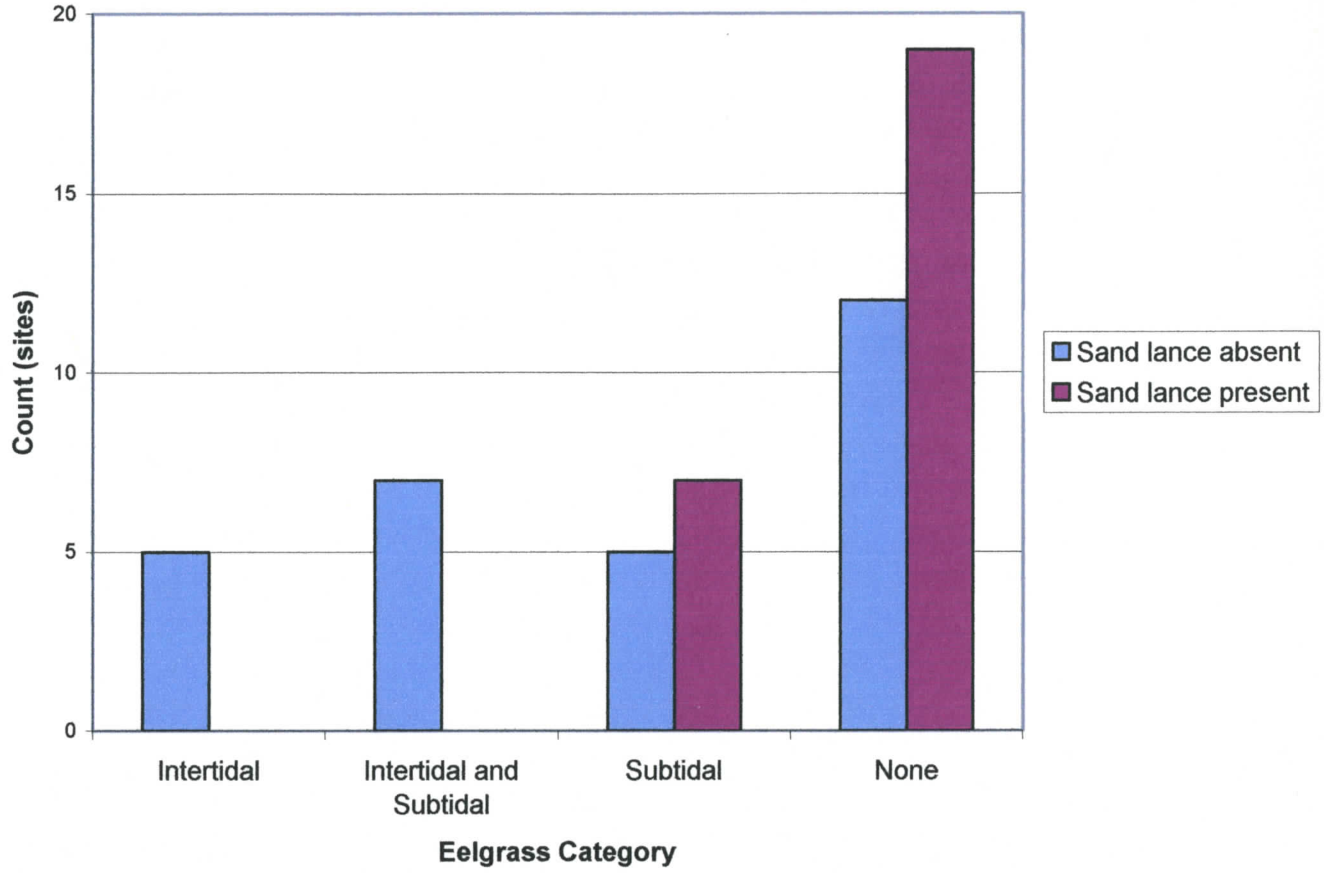


Figure 5.7: Eelgrass categories (*Intertidal Eelgrass Presence/Absence* and *Subtidal Eelgrass Presence/Absence*) combinations grouped by *Sand lance Presence/Absence.* The four categories (Intertidal, Intertidal and Subtidal, None, Subtidal) are combinations of the binary eelgrass variables

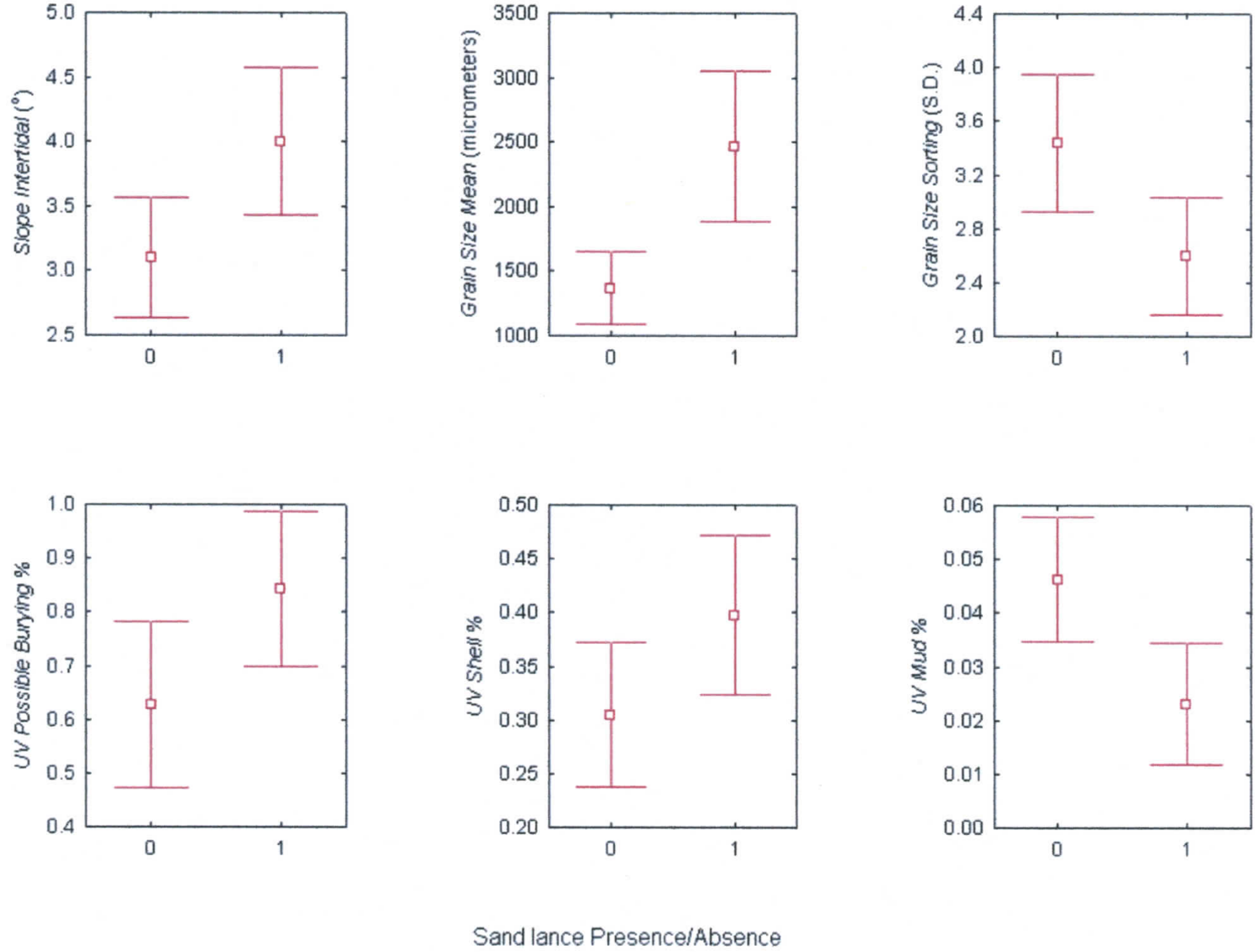


Figure 5.8: Mean plots of all independent variables selected for analysis. Bars indicate 95% confidence intervals

Logistic Regression

Model Selection

The backwards stepwise modeling procedure, starting with a model containing all eight variables, terminated after seven steps. At each step, I evaluated the model statistics to view of the performance of the model at that point.

From the seven possible model steps, two models outperformed the others based on classification error performance, pseudo R^2 values, and Pearson's Chi-square (Model 1 and 2, step five and step seven respectively, Table 5.5). Between Model 1 and Model 2, four out of the original eight variables were selected. These variables were: *Grain Size Mean*, *Grain Size Sorting*, *Major Substrate Low Intertidal*, *Intertidal Eelgrass Presence/Absence*. I constructed two more models for comparison again using backward elimination to determine whether removal of key variables would change the model performance. Model 3, the terminal step of four elimination steps, was built excluding *Major Substrate Low Intertidal*. Model 4, the terminal step of four elimination steps, excluded both *Major Substrate Low Intertidal* and *Intertidal Eelgrass Presence/Absence* (Table 5.5).

For sites with eelgrass absent the stepwise model selection procedure terminated after five steps, with the fifth step appearing to be the best model (Model 5, Table 5.6). I also constructed a model using the same process but excluded the *Major Substrate Low Intertidal* variable (Model 6, Table 5.6).

Table 5.5: Logistic regression modeling results based on the selected independent variables using *Sand lance Presence/Absence* as the dependent variable (N=55). *Independent variable significant at the 0.05 level (Wald statistic). **Model Chi-square value significant at the 0.05 level

	Model 1			Model 2			Model 3			Model 4		
Variable	Coefficient	S.E.	Wald	Coefficient	S.E.	Wald	Coefficient	S.E.	Wald	Coefficient	S.E.	Wald
Constant	-19.547	10042.8	<0.001	-20.196	10090.3	<0.001	-20.014	10914.7	<0.001	0.376	0.955	0.155
Grain Size Mean	0.001	0.001	2.456				0.001	<0.001	6.69*	0.01	<0.001	10.289*
Grain Size Sorting	-0.545	0.393	1.927				-0.888	0.366	5.886*	-0.905	0.343	6.948*
Intertidal Eelgrass Presence/Absence (1)	20.904	10042.8	<0.001	21.294	100090	<0.001	21.143	10914.7	<0.001			
Major Substrate Low Intertidal			4.708			10.157*						
Major Substrate Low Intertidal (1)	-3.117	1.826	2.912	-3.296	1.563	4.444*						
Major Substrate Low Intertidal (2)	-0.559	1.42	0.155	-0.492	1.261	1.152						
Major Substrate Low Intertidal (3)	-0.616	2.248	0.075	1.299	1.557	0.696						

Model Statistics	Model 1	Model 2	Model 3	Model 4
Model Chi-Square [df]	40.78 [6] **	36.12 [4] **	34.43 [3] **	22.20 [2] **
Sensitivity (% Correct Predictions)	82.8	72.4	82.8	75.9
Specificity (% Correct Predictions)	96.2	96.2	88.5	69.2
Overall % Correct Predictions	89.1	83.6	85.5	72.7
Nagelkerke pseudo R^2	0.699	0.643	0.621	0.443
AIC	47.94	50.24	49.66	59.89
Area Under the ROC Curve	0.923	0.897	0.902	0.834

Table 5.6: Logistic regression modeling results for sites without intertidal eelgrass (N=43). Modeling was based on the selected independent variables using *Sand lance Presence/Absence* as the dependent variable. *Independent variable significant at the 0.05 level (Wald statistic). **Model Chi-square value significant at the 0.05 level

	Model 5			Model 6		
Variable	Coefficient	S.E.	Wald	Coefficient	S.E.	Wald
Constant	1.358	1.424	0.909	1.129	1.078	1.097
Grain Size Mean	0.001	0.001	2.456*	0.001	<0.001	6.69*
Grain Size Sorting	-0.545	0.393	1.927*	-0.888	0.366	5.886*
Major Substrate Low Intertidal			4.708			
Major Substrate Low Intertidal (1)	-3.117	1.826	2.912			
Major Substrate Low Intertidal (2)	-0.559	1.420	0.155*			
Major Substrate Low Intertidal (3)	-0.616	2.248	0.075			
Model Statistics	Model 5			Model 6		
Model Chi-Square [df]	22.406 [5]**			15.057 [2]**		
Sensitivity (% Correct Predictions)	70.6			70.6		
Specificity (% Correct Predictions)	96.2			88.5		
Overall % Correct Predictions	86.0			81.4		
Nagelkerke pseudo R^2	0.406			0.312		
AIC	45.943			47.6563		
Area Under the ROC Curve	0.869			0.833		

Cross-Validation

Although the models used to predict the subsets of data are slightly different from the overall model (*i.e.* the model constructed from using the overall dataset), the variables used are the same. Cross-validation showed that models included here have a varying potential predictive accuracy (Table 5.7). Model 1 had the highest overall correct cross-validation classification rate with a value of 85.5%. The other models did not perform as well but were generally within a similar range with the exception of Model 4, which had an extremely low sensitivity. The overall classification rate was always lower for the cross-validation when compared to simple resubstitution values, however, individually, the sensitivity and specificity of each model were not both necessarily lower for the cross-validation (Figure 5.9).

Table 5.7: Cross-validation results for logistic regression models

	N	Sensitivity	Specificity	Overall
Model 1	55	88.5	82.8	85.5
Model 2	55	65.4	79.3	72.7
Model 3	55	69.2	86.2	78.2
Model 4	55	50.0	82.8	67.3
Model 5	43	84.6	76.5	81.3
Model 6	43	69.2	76.5	72.1

Model Assessment

Overall, the six logistic regression models all predict significantly better than the null model (*i.e.* they predicted better than random selection or selection of the most prevalent classification). Also, models have relatively good classification error performance, pseudo R^2 values, and Pearson's Chi-square values suggesting a relatively good model fit (Table 5.5, Table 5.6). All models had relatively high ROC values suggesting that the models have the potential to be used for management purposes (all models >0.75). Although the overall models were significant, the Wald statistic of many of the independent variables that were included in the final models were not found to be significant at the 0.05 level (see Table 5.5 and 5.7).

I examined model statistics in order to choose the best performing model. When considering models 1-4 (Table 5.5), Model 1 outperforms all other models with regards to every model statistic. Model 1 had the highest model statistic values for the chi-square, sensitivity, specificity, overall classification, pseudo R^2 value and ROC value. It also had the lowest AIC value signifying the most parsimonious model with the best fit. With regards to Models 5 and 6, Model 5 outperforms Model 6 in all comparisons. Thus, Model 1 best predicts the overall dataset while Model 5 best predicts sites with intertidal eelgrass absent.

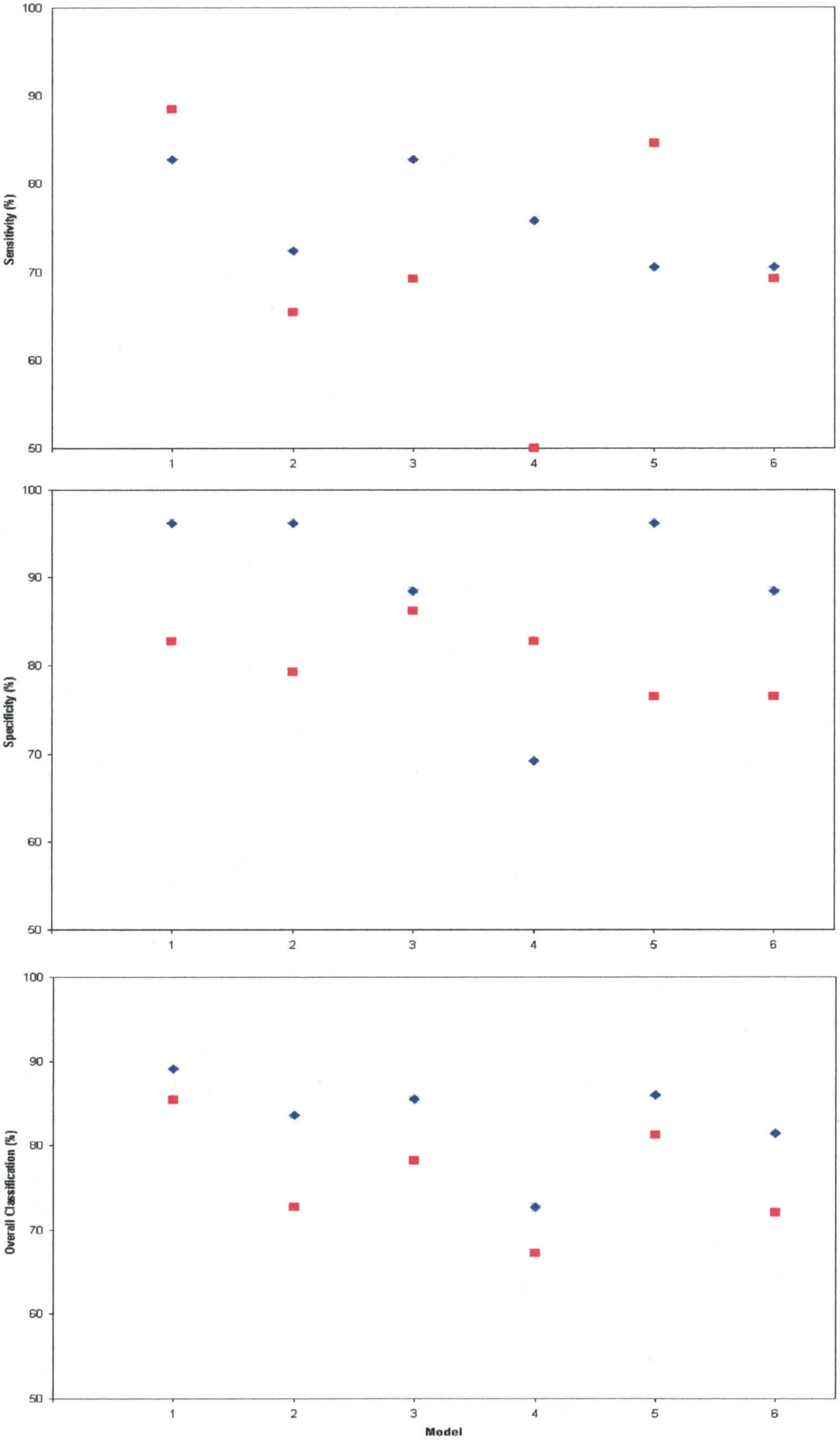


Figure 5.9: Comparison of resubstitution values (blue diamonds) to data splitting cross-validation values (red squares) for the six logistic regression models.

Predicted vs. Abundance

When all sites (presence and absence) were included, there was a significant positive correlation between the predicted values of Model 1 and abundance (Pearson's correlation coefficient = 0.656, $p<0.001$). When only sites with sand lance present were used, the relationship was not significant at the 0.05 level ($p=0.180$).

Classification Tree

Cost of the cross-validation varied between the two pruning methods resulting in the two different sized trees (Figure 5.10). The point at which the cross-validation cost reaches a minimum is the point for which the optimal tree size is reached (Figure 5.11). Misclassification pruning produced a tree with two splits and three leaves (Figure 5.10 A) while deviance pruning produced a tree with four splits and five leaves (Figure 5.10 B). The tree pruned on misclassification was pruned further towards the parent node than the tree pruned on deviance, therefore, the structure of the misclassification-pruned tree is identical to the first two splits of the deviance-pruned tree. The cross-validation classification rate used to compare models showed that for the global cross-validation (and the simple resubstitution) the deviance-pruned tree predicted better than the misclassification-pruned tree (Table 5.8) suggesting that the deviance-pruned tree is the preferred tree. The importance of each variable varied between the deviance-pruned tree and the misclassification-pruned tree (Figure 5.12).

Table 5.8: Classification rates for the two classification tree models.

				Observed		Overall Classification Rate (%)
	Model	Validation		0	1	
Predicted	Misclassification Pruning	Resubstitution	0		5	70.9
			1	11		
		Global Cross-validation	0		5	74.5
			1	9		
	Deviance Pruning	Resubstitution	0		1	83.6
			1	8		
		Global Cross-validation	0		1	90.9
			1	4		

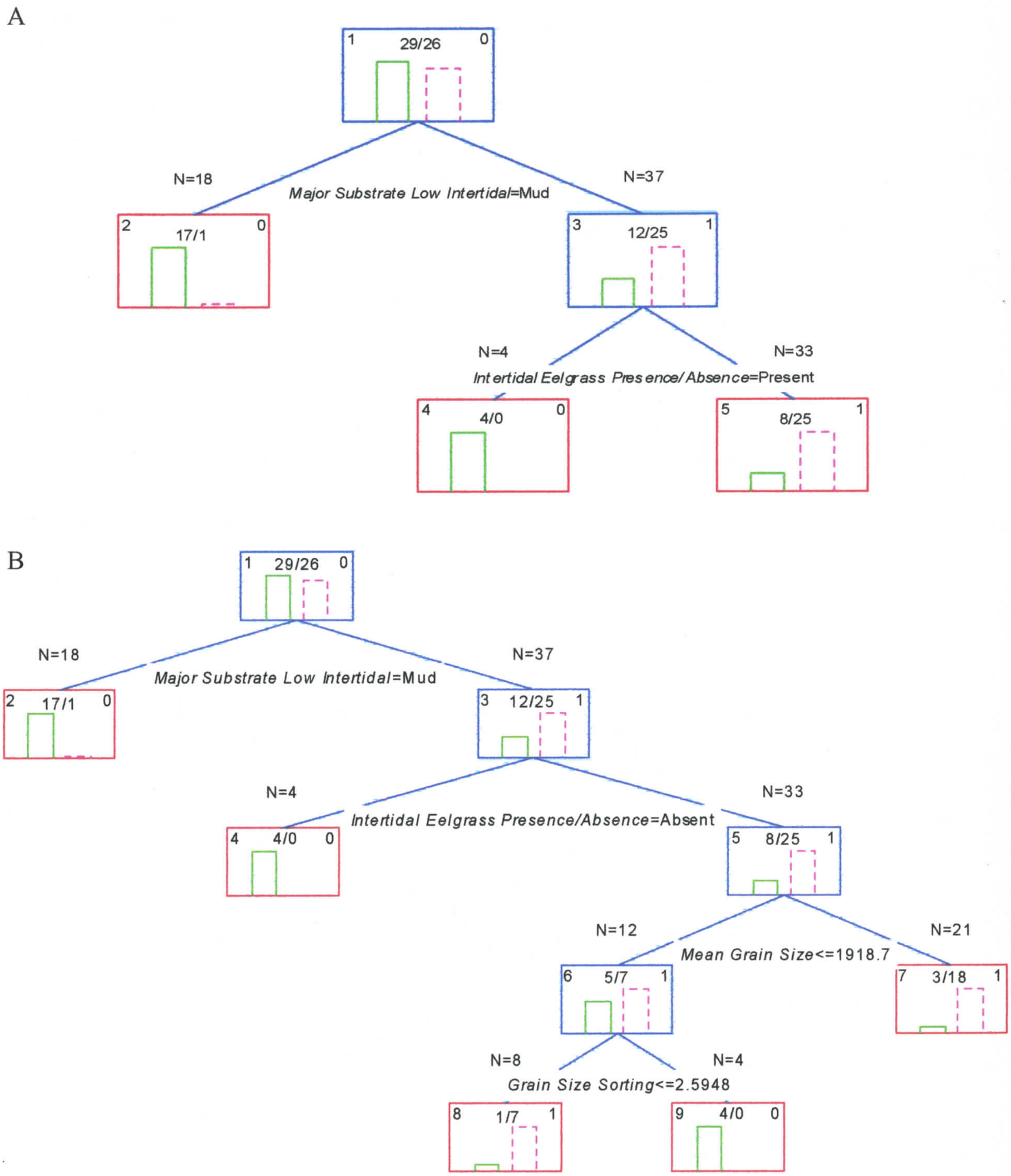
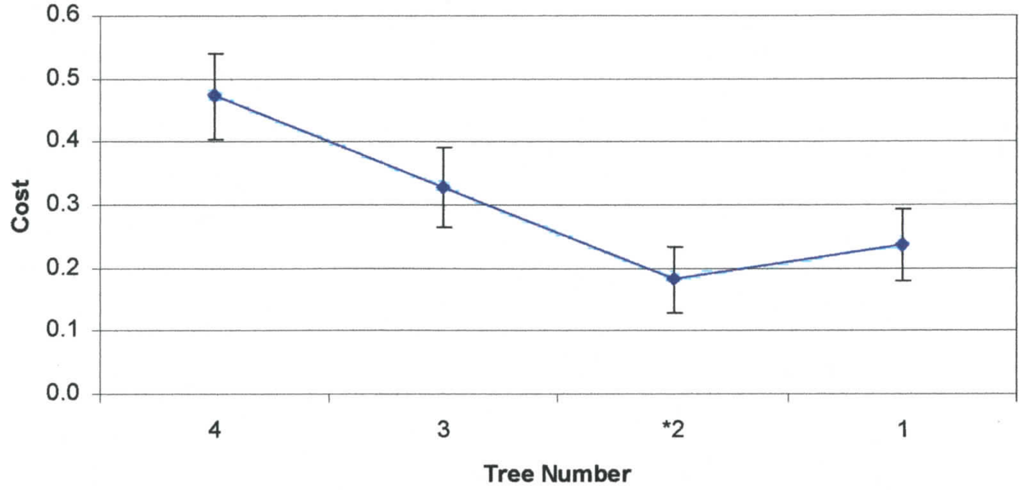


Figure 5.10: Classification trees grown from misclassification pruning (A) and deviance pruning (B). For each node box, the node number is in the top left, the ratio of absence to presence in the top middle and the node class (sand lance absence = 0, sand lance presence = 1) in the top right. The histograms within the node boxes represent the class proportions with green solid representing absence and pink hatched representing presence. The splitting rule indicates the rule determining the cases belonging in the left node of the split with the remaining cases belonging in the right node.

A



B

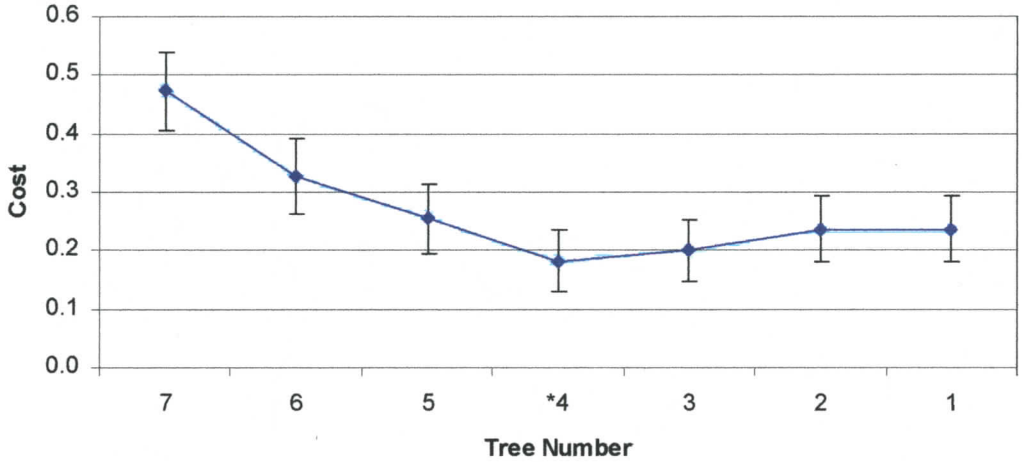


Figure 5.11: 10-fold cross-validation cost of the misclassification pruning procedure (A) and the deviance pruning procedure (B). Tree Number indicates the number of splits occurring in the tree and error bars represent standard error. *tree number selected by the pruning procedure.

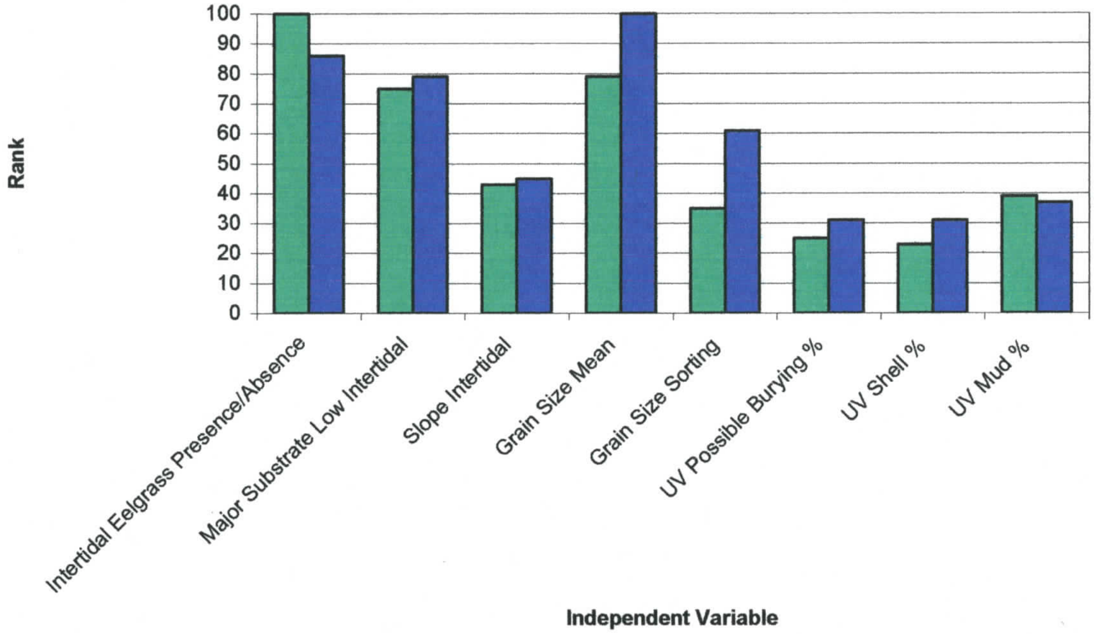


Figure 5.12: Importance of independent variables in classification tree analysis for the misclassification-pruned tree (green bars) and the deviance-pruned tree (blue bars) ranked on a scale of 0-100. Ranking is based on the overall change in node impurity summed over the entire tree and standardized by the largest sum.

Biotic Interactions

The MDS analysis had a high stress value and there was no clear visual separation between sites with sand lance present and sand lance absent sites (Figure 5.13) suggesting that the fish community structure of sites with sand lance present is not different from the community structure of sites with sand lance absent. The ANOSIM analysis did not find a significant difference between fish communities of sites with sand lance present compared to sites with sand lance absent (p=0.12, Figure 5.14). Similar results were found for both the MDS and the ANOSIM when species presence/absence rather than abundance was used to create the similarity matrix.

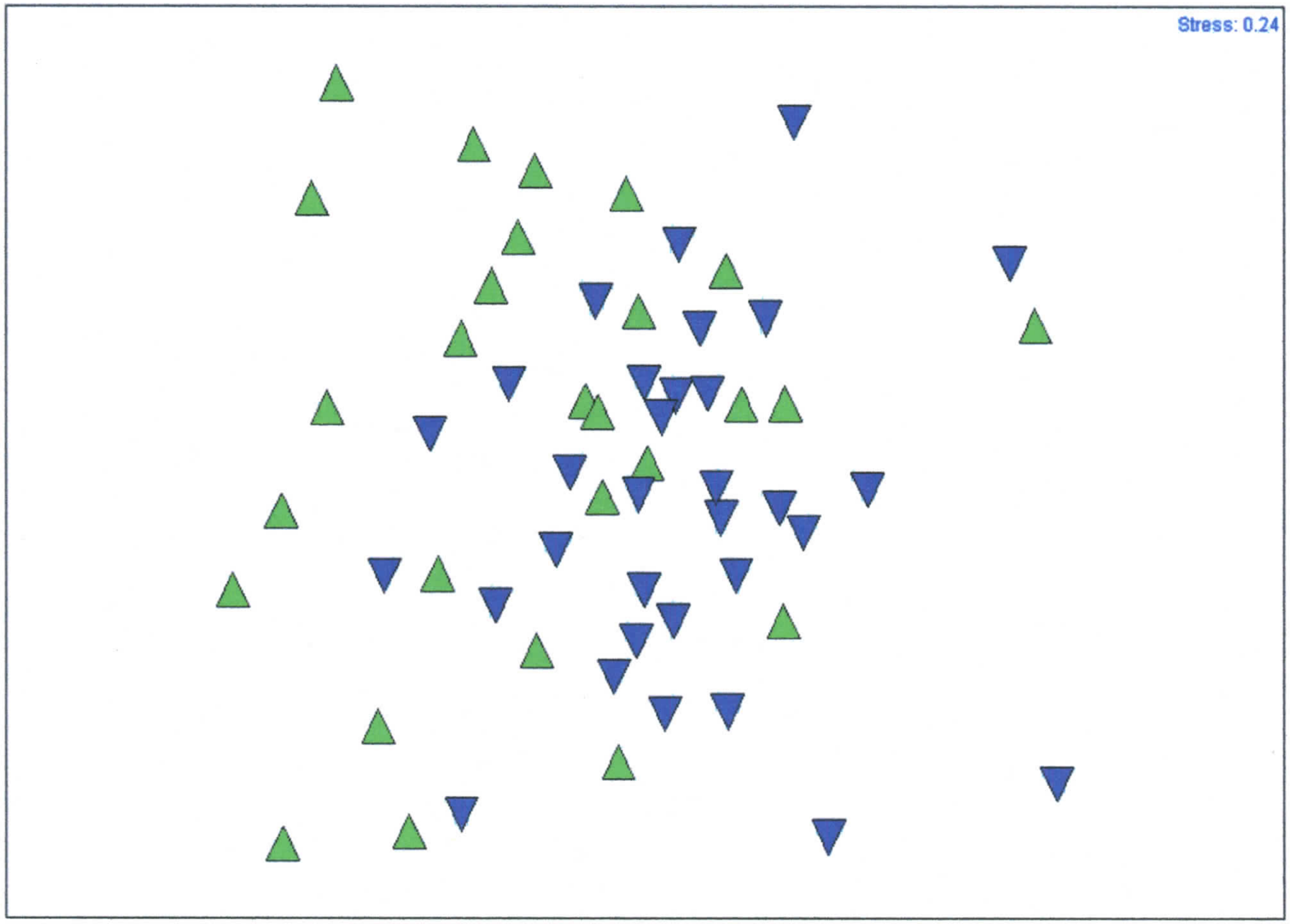


Figure 5.13: MDS ordination of fish communities where sand lance presence (green triangles) and absence (blue triangles) was used as the factor.

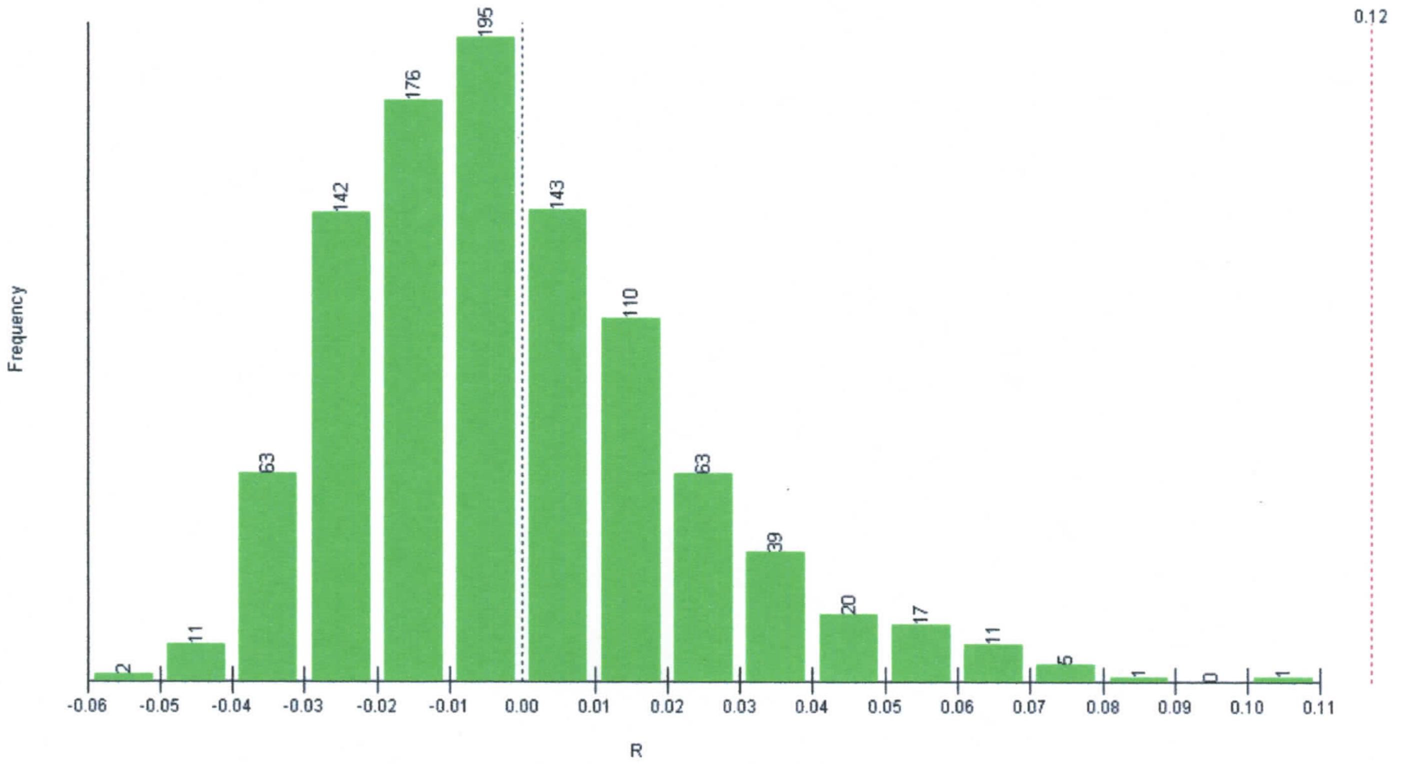


Figure 5.14: ANOSIM of fish community structure using sand lance presence/absence as the factor. The global R-value was not significant (p=0.12).

5.4 DISCUSSION

Logistic Regression

To assess the significance of the logistic regression models, each model was compared to the null model. Although each model was found to perform significantly better than the null, overall significance values for habitat models are of limited use and further exploration of model performance generally is necessary (Boyce *et al.* 2002). Although a significant value suggests that the species uses its habitat in a non-random manner, this information is rather trivial considering that it is accepted that organisms generally distribute themselves non-randomly within a given landscape (Mackey and Lindenmayer 2001).

More important than the overall significance of the models are the model evaluation statistics, which can be used to select the best model. The construction of a model is an iterative process in which model building and validation is repeated until an "optimal" model is achieved (Corsi *et al.* 1999). In this case, "optimal" refers to a model that includes the biologically relevant variables, that does not include the irrelevant variables, and the cross-validation shows the model to be able to perform with satisfactory accuracy (*i.e.* not overfit).

Here, the stepwise selection narrowed down possible independent variables to include in the model. Although widely used, some authors question the validity of stepwise selection procedures (De'ath and Fabricius 2000 and references therein, Guisan and Thuiller 2005). Because stepwise techniques use statistical rather than biological criteria to select a final model, it is important to critique the process to determine whether the selection procedure is meaningful. Here, I demonstrated four models that could predict the overall dataset (Table 5.5). From the model statistics, it was clear that the preferred model was Model 1, which was not the terminal step of the original stepwise model building process. Model 1 outperformed the model created from the terminal step of the stepwise procedure (Model 2). Thus, the stepwise procedure, in this case, did not provide the most preferred model although it did aid in narrowing down possible models.

The model building procedure used here is commonly used to inductively build predictive models. Although the models all had ROC values >0.75, suggesting their possible utility as predictive tools, I would caution the use of these models for predictive purposes. It is apparent from the interannual variation seen in chapter four that sand lance habitat selection can vary significantly annually. Thus, although these models may have the ability to predict well during years where the conditions (*i.e.* population densities, broad scale environmental influences, *etc.*) are similar to 2003, it is likely that the models would not perform adequately during years where conditions may be different and habitat use patterns change.

Cross-validation

For true unbiased estimators of model accuracy, independent data points should be used to cross-validate the model. In this case, external independent data were not available for validation. Although most authors suggest that proper cross-validation should involve external data, Van Horne (2002) de-emphasizes this point. She instead stresses that testing for the importance of independent variables, as they affect a species distribution, is far more important. Here, classification rates from the data splitting cross-validation provided validation for the overall models and aided in comparing competing models that used different combinations of independent variables.

When comparing cross-validation to resubstiution, the resubstitution rates either overestimated or underestimated the sensitivity or specificity compared to the data splitting cross-validation (Figure 5.9), however, the overall classification was always higher for the resubstitution method. This likely reflects the bias of the resubstitution evaluation method in overestimating classification rates.

Model cross-validation not only assesses how the model will perform given "new" data (Henebry and Merchant 2002) but also helps verify the models accuracy. For example, the cross-validation results showed that Model 4 had a high sensitivity error that was not seen with the simple resubstitution validation. The value of 50% for the sensitivity of Model 4 suggests that this model cannot adequately determine sites that should have sand lance present compared to a random choice. The specificity for Model

4 is acceptable suggesting that this model can still likely discriminate sites that should have sand lance absent.

Although stepwise procedures can "overfit" a model to the data, the cross-validation procedure helped validate this model suggesting that overfitting hasn't occurred. Resubstitution values were always higher than cross-validation, but they were not extremely different. If the model was overfit, there would likely be a larger difference between resubstituion and cross-validation results. For example, Model 1 had a high overall cross-validation classification rate similar to the high overall resubstitution rate (85.5% and 89.1% respectively).

Variables

The Wald statistic, a common way to determine independent variable significance, is based on the ratio of the Beta coefficient value to the standard error squared. If the standard error is high relative to the Beta coefficient, the variable would not be significant for the model. *Grain Size Mean* and *Grain Size Sorting* had very small coefficients yet were included in Model 1 stepwise selection method. Their inclusion suggests that there is a relationship between the dependent and *Grain Size Mean* and *Grain Size Sorting*. Possible non-linear relationships between each of these two variables and the dependent may have caused low Beta values. Another possibility is that there is a multicolinearity issue with the models. Although variables were screened for multicollinearity, the value of 0.70 as a cut off is somewhat subjective. *Intertidal Eelgrass Presence/Absence*, did not have a significant Wald statistic despite being extremely influential in all the models. This may be explained by the Hauk-Donner phenomenon. For logistic regression, or any binomial logit models, when the association of an independent variable is extreme, the Wald statistic can be inaccurate and not show variable significance (Hauk and Donner 1977).

Although all models presented here do not necessarily have significant Wald statistics for all the variables included, the relatively good performance of the models and the selection by the stepwise process supports their inclusion. For example, Model 1 did not have any significant variables despite being the preferred model for the overall dataset. Hinsley *et al.* (1995) points out that the emphasis should not be placed on p-values of individual variables since models should not be selected using statistical means alone.

Given there was perfect relationship with the sand lance absence and the presence of intertidal eelgrass (*i.e.* all sites with intertidal eelgrass did not have sand lance, N=12) sites without intertidal eelgrass were examined separately resulting in Model 5 and 6 (Table 5.6). It was important to examine these sites separately to determine whether there was a different set of independent variables that influenced sand lance presence/absence for sites without intertidal eelgrass. Model 5, which was preferred over Model 6 based on model statistics, utilized the same independent variables as Model 1 (with the exception of *Intertidal Eelgrass Presence/Absence* which was the filter variable). Model 1 and Model 5 both use the same independent variables and both models were the preferred models for their respective data sets. Given this, it would appear that the variables that are important for the entire dataset are also the variables important for sites without intertidal eelgrass.

Classification Tree

CART models have been found to be useful in identifying fish habitat (Norcross *et al.* 1997, Norcross *et al.* 1999, Olden and Jackson 2002, Rejwan *et al.* 1999) and may outperform more traditional techniques. However, when model assumptions are met there is no reason for traditional techniques such as logistic regression to be drastically different from a CART results (Olden and Jackson 2002). Here, results from classification tree analyses were similar to that found for logistic regression (Appendix 5.3). Similar variables were chosen by cross-validation procedures for final tree models as were chosen by the logistic regression stepwise selection procedure. The classification tree model selected based on deviance pruning (Figure 5.10 B) utilized the same independent variables as Model 1, the fifth step in the backwards stepwise procedure for logistic regression (Table 5.5). The classification tree model based on misclassification pruning (Figure 5.10 A) utilized the same independent variables as Model 2, the seventh step in the backwards stepwise procedure (Table 5.5). Comparing between the two classification trees, the tree pruned on deviance has the better global cross-validation results and is likely the preferred model with regards to description of sand lance habitat and predictive accuracy. The corresponding logistic regression model (Model 1) was also the preferred model compared to competing logistic regression models.

Although classification tree and logistic regression showed similar results, the classification tree approach has some notable advantages. Classification trees analyze data based on constraints rather than correlates (O'Connor 2002) which is more realistic in a habitat selection context (Huston 2002). Because of this, classification trees may be more robust compared with techniques that rely on linear correlations such as GLM's. In addition, classification trees deal automatically with non-pamametric data, interaction effects, non-linear predictors, and spatial autocorrelation. These issues can confound models such as GLM's that rely on correlates, causing the violation of assumptions (O'Connor 2002, Muñoz and Felicísimo 2004). Since relationships between environmental variables and a species distribution are generally nonlinear (Heglund 2002) it is likely that classification trees better represent the complex structures that may be present.

The results from the regression tree can be expressed graphically as a set of predictive rules. The hierarchical structure of the graphical representation of classification trees is easy to interpret and importance of variables and variable interactions can be readily seen. The splits near the top of the tree are thought to reflect the most general relationships between the independent and dependent variables while lower splits may be less generalizable and therefore less applicable to populations outside the spatial and temporal scale of sampling (Rejwan *et al.* 1999). Given this, it is possible that the misclassification-pruned tree is more generalizable to areas outside of the study. The better performance of the deviance-pruned tree suggest that the sediment properties grain size and sorting are likely important factors although perhaps less generalizable because they occur further down the tree.

Classification trees have the potential to uncover relationships that may not be detected by classical techniques and therefore allows the user to interpret the resultant models with an increased sense of confidence. This may be the case for *Mean Grain Size* and *Grain Size Sorting*. The hierarchical relationship between these two variables seen in the deviance-pruned tree likely reflects an interaction between the two not detected in the logistic regression analysis. Node 5 is split into fine and coarse grain sizes based on the

Mean Grain Size of 1918.7 microns. *Grain Size Sorting* further splits the fine grain size node (node 6, $\leq$1918.7 microns) suggesting that within finer grain sizes, *Grain Size Sorting* is important in determining sand lance occurrence.

Variables Important for Sand lance Habitat Structure

For inductive modeling procedures, whose goals are descriptive rather than predictive, the selection of relevant ecological terms is the focus (Brotons *et al.* 2004). Out of the 30 original variables, after univariate screening and both the logistic regression analysis and classification tree analysis, four variables appeared important: *Major Substrate Low Intertidal, Intertidal Eelgrass Presence/Absence, Grain Size Mean, Grain Size Sorting*. Because it is easier to examine variable structure and importance using the classification tree, I will discuss sand lance habitat structure based on tree results.

Major Substrate Low Intertidal

The first split in the tree is based on the qualitative *Major Substrate Low Intertidal*. Here, "Mud" was the important category with an almost perfect association with sand lance absence (1/17, presence/absence). This reflects *Ammodytes* known avoidance of silt/clay in the burying sediment (Inoue *et al.* 1967, Dick and Warner 1982, Pinto *et al.* 1984, O'Connell and Fives 1995, Wright *et al.* 2000, Holland *et al.* 2005).

The inclusion of a qualitative sediment variable where quantitative information exists may appear redundant, however, the results of the modeling procedure shows that the *Major Substrate Intertidal* was an important variable even when used in conjunction with quantitative sediment measurements. The samples for the quantitative grain size variables were taken from a small section of a shoreline, that could often have a large area, and at one particular tide height level within the low tide region. Thus, the quantitative data may only have provided a snapshot of one portion of the shoreline while the qualitative variable may better reveal the overall broad characteristics of the sediment for a sample site. Also, the qualitative variable may interpret the sediment characteristics differently than the quantitative variable. For instance, I believe that the qualitative variable stresses more on the mud/silt characteristics than does the quantitative variables. A site that appears "muddy" often still has a much higher proportion of sand/gravel compared the silt/clay portion. The sand/gravel proportion may obscure the detection of extremely fine sediments when examining quantitative characteristics such as *Grain Size Mean*. In this case, it appears pertinent to include both the qualitative description and quantitative measures of sediment characteristics in the same model.

The selection of this variable stresses the importance of the low intertidal in sand lance habitat selection. Conversely, high intertidal substrate (*Major Substrate High Intertidal*) was not significantly related to sand lance occurrence. Substrate properties of the shallow subtidal were included as selected variables for modeling (*UV Possible Burying %, UV Shell %, UV Mud*) but also were not found to be important in the models.

Intertidal Eelgrass Presence/Absence

Sediment properties have been known to affect sand lance distribution in subtidal regions of the marine environment (Wright *et al.* 2000, Holland 2005, Ostrand 2005) as they have here in the intertidal, however, this is the first time, to my knowledge, that eelgrass has been associated with sand lance habitat selection.

The avoidance of intertidal eelgrass reinforces the importance of the low intertidal environment for sand lance habitat selection in the nearshore. The presence of eelgrass in the intertidal, rather than the subtidal appears to strongly negatively influence sand lance habitat selection (a perfect negative association). Sand lance avoided all sites where eelgrass was found in the intertidal however did not significantly avoid sites where eelgrass was just found in the subtidal (in fact there was a slightly higher frequency of sand lance presence, Figure 5.7). Given that twelve sites (approximately 22% of total) had only subtidal eelgrass, it is important to make that distinction. One could easily overlook this relationship if sites were simply classified as eelgrass/non-eelgrass with no distinction made for the location of the eelgrass bed (whether it was subtidal or intertidal).

The perfect association between sand lance absence and sites having intertidal eelgrass present may be in part due to the finer sediments associated with intertidal eelgrass sites (Figure 5.4). Out of the twelve sites that had intertidal eelgrass present, eight were classified as "Mud" by *Major Substrate Low Intertidal* variable. The remaining four sites (split at node 4) were not classified as mud (three "fine sand", one "coarse sand"). Thus, the avoidance of intertidal eelgrass likely is not strictly due to the sediment properties of the site.

Grain Size Mean

The importance of *Grain Size Mean*, even after the division of "Mud" sites, shows that sediment characteristics beyond just the silt/clay fraction are important in determining sand lance habitat selection. In the deviance-pruned tree, node 5 was split on a mean grain size value of 1918.7 microns. This is close to the cut off between "coarse sand" and "very fine gravel" and thus sites in node 7 ranged from fine to coarse gravel. Holland *et al.* (2005) found that *A. marinus* (age class not specified) avoided sediment with a grain size mean greater than 2000 microns, however, here the node with *Grain Size Mean* greater than 1918.7 microns (node 7) was classified as sand lance present (18/3 presence/absence). Here, sand lance may not be avoiding the *Grain Size Mean* values in the gravel range as seen by Holland *et al.* (2005) but rather sand lance may be selecting it. Although the reason for this discrepancy is unclear, it may be attributed to the various differences between the studies such as different *Ammodytes* species being examined, differences between subtidal habitat selection measured by Holland *et al.* 2005 and intertidal habitat selection measured here or possible differences in age classes examined between the two studies.

For the tree models, the importance of the *Grain Size Mean* variable depends on whether the model is pruned on misclassification or deviance (Figure 5.12). For the deviance-pruned model, it is the most important variable. For the misclassification model it is less important, however, even though it is not used in the model, it still has a high rank importance (Figure 5.12).

The effects of sediment composition on sand lance distribution have been examined for other species in the genus. Wright *et al.* (2000), when examining the subtidal burying habitat of *A. marinus*, suggested some possible values for sediment that may constrain sand lance burying. They suggested that sand lance would not bury in sediment that had a median grain size larger than 2mm or sediment with a grain size fraction <0.063mm was >7% or when sediment with a grain size fraction of <0.125mm was >65%. In this study, the sediment composition of all sites fell within these boundaries so the restrictions outlined by Wright *et al.* (2000) cannot be compared to results found here.

Overall, it appeared that sand lance were selecting *Grain Size Mean* greater than approximately 1900 microns and were avoiding finer grain sizes (Figure 5.8).

Grain Size Sorting

The variable *Grain Size Sorting* further quantifies sediment properties. Its importance here suggests that regarding sand lance habitat selection, *Grain Size Mean* is not adequate in itself to physically describe sediment properties. Sorting refers to the range in grain size (the standard deviation). Well-sorted sediment has a grain size that is more uniform while poorly sorted sediment has a broader range of sediment sizes. The effect of sorting on physical sediment properties may directly affect sand lance burying ability. Also, sorting may be an indirect measure of oceanographic processes such as tidal current and exposure as these broad scale processes often affect sediment transport. As discussed above, there seems to be an interaction between *Grain Size Mean* and *Grain Size Sorting* as shown by the deviance-pruned tree.

Overall

Intertidal/shallow subtidal habitat is thought to be important for sand lance (Robards and Piatt 1999), thus, it is crucial to understand sand lance distribution in this zone if we are to understand the overall habitat selection of sand lance. The identification of major habitat requirements for ecologically important species, such as sand lance, is critical for development of management actions aimed at protecting these species. As mentioned in chapter one, existing habitat models for sand lance have focused on the subtidal region and spawning areas. Investigators have examining sand lance subtidal burying habitat (*e.g.* Wright *et al.* 2000, Holland *et al.* 2005), subtidal pelagic habitat (Ostrand *et al.* 2005), and winter spawning habitat (Robards *et al.* 1999c, Pentilla 2000) however, to my knowledge no one has empirically modeled YOY sand lance use of the intertidal and shallow subtidal habitat.

In all studies on subtidal habitat mentioned above, sediment was an important variable in modeling. Here, sediment was found to be important for YOY sand lance in the nearshore region. From the modeling procedures presented in this chapter, it is apparent that low intertidal sediment composition is one of the more important characteristics in determining sand lance habitat selection at this scale of study. This emphasizes how sand lance are dependent on refuge habitat. In general, refuge habitat is considered vital in regulating the predator-prey oscillations preventing the overexploitation and subsequent extinction of the prey species (Rosenzweig and Macarthur 1963, Pearson *et al.* 1984).

The preference of YOY sand lance for certain sediments as refuge habitat has implications for management and for the research on predators of sand lance. Sand lance are considered to prefer "sandy" habitat, however, this term is often not always clearly defined in the literature (*e.g.* Dick and Warner 1982, Høines and Bergstad 2000). The presence of shoreline considered "sandy" has been used as a proxy for sand lance habitat in predator studies (*e.g.* Yen *et al.* 2004) and "beach or not beach" has been used in modeling sand lance subtidal habitat (Ostrand *et al.* 2005). Although sand lance select sediments with average grain size in the "sand" category (according to the Wentworth scale), there is a range of sand/gravel preferred and an interaction with grain size sorting. Although sandy habitats appear uniform, they are spatially and temporally heterogeneous (Syms and Jones 2004). Classifying all sand/gravel or "beach" shoreline habitat as potential sand lance habitat will likely encompass all sand lance burying habitat, however, it will also include habitat that is likely not utilized by sand lance. The characteristics of intertidal sediment preferred by YOY *A. hexapterus* should be investigated in greater detail.

Pacific sand lance distribution in the nearshore did not appear to be affected greatly by subtidal characteristics. The extent of subtidal habitat with suitable sediment characteristics for burying is unknown. Ostrand *et al.* (2005) found that sand lance were found most commonly in shallow waters <40m. This may be due to a link to the intertidal and shallow subtidal burying habitat rather than deeper subtidal burying habitat. Also, anecdotal evidence suggests that the lower intertidal may play an important role as sand lance habitat. Although sand lance have been noted to travel over numerous sediment types during the day, they often return to sandy shorelines or "beaches" which are more well washed and contain less extremely fine sediments than the subtidal areas over which they forage (Kühlmann and Karst 1967, Pinto *et al.* 1984).

In Alaska, Dick and Warner (1982) suggested that sand lance were found less commonly on small islands than "on bights and coves". Here, in a study area dominated by "small" islands, sand lance seemed to be relatively frequent, especially in 2003 where they were present for almost half (47.3%) of all sites sampled. Also, Dick and Warner (1982) found that sand lance were more common "on moderately long or sloping beaches than on very tiny or flat ones." Ostrand *et al.* 2005 found slope to be one of the least important variables in describing sand lance relation to pelagic subtidal habitat. In this study, sites with sand lance present had a significantly steeper average intertidal slope compared to sites where sand lance were absent (Figure 5.8, Table 5.3). Although intertidal slope was significantly related to *Sand lance Presence/Absence* in the univariate analysis, it was not important in either the logistic regression or classification tree modeling procedures.

Model Limitations

General limitations

A correctly specified model is one in which the mathematical or statistical expressions on which the model is based are an accurate representation of the real world (Reckhow *et al.* 1986). Habitat models are simplified representations of the real world and therefore to some extent mispecified due to the inherent limitations concerned with modeling complex systems. Because of this, habitat models have a limited application. Many authors have pointed out various limitations with habitat modeling which are pertinent to outline here.

1) Complexity

The species-environment relationship is often too complex for accurate statistical modeling (Huston 2002, Railsback *et al.* 2003). Even when using simulated data (*i.e.* virtual populations) with no measurement uncertainty, high levels of unexplained variation can still remain (Hirzel *et al.* 2001, Tyre *et al.* 2001, Railsback *et al.* 2003).

The inability of habitat models to represent the true complexity can be seen in this study. For example, pseudo R^2 values for logistic regression models range from 0.406 to 0.524. The high levels of unexplained variation likely arise from the simplification of the complex species-environment relationship into a feasible statistical model. Even when a model explains a satisfactory level of variance, the complex interactions between limiting and non-limiting conditions make the true species-environment relationship unattainable (Huston 2002). Because habitat data are complex, predictive models are generally simplified representations of the real world using a low-dimensional representation of the high-dimensional set of relationships seen in nature (Scott *et al.* 2001).

2) Biotic Interactions

It is difficult to model ecosystem relationships such as biotic interactions (Railsback *et al.* 2003), however, these interactions may be important. Biotic interactions are generally difficult to measure and can be much more complex compared with abiotic relationships. Species can interact with hundreds to thousands of other species (Boone and Krohn 2002) and these interactions may affect population levels and thus distribution and abundance (*i.e.* habitat selection). Studies have shown that interactions such as competition can significantly affect a species distribution (*e.g.* Leatherwick and Austin 2001) and biotic variables may restrict habitat selection at the opposite end of the ecological gradient compared with abiotic variables (Brown *et al.* 1996). However, because biotic interactions are notoriously difficult to model, they are rarely incorporated into habitat models (Austin 2002b).

Specifically regarding sand lance, there are numerous biotic interactions that may affect their distribution, none of which were detected here with the fish community analysis. One of the more likely biotic interactions that may affect sand lance is predation. Sand lance are known to be heavily preyed upon (Willson *et al.* 1999) and it is possible that this would be reflected in the fish communities associated with sand lance habitat. This however was not the case, as the fish community analysis did not reveal any differences in community structure between sites with sand lance present and absent.

Little is known about the predator-prey relationships between sand lance and its predators (*i.e.* how predation affects sand lance habitat selection). Biotic interactions such as these are extremely difficult to measure, as mentioned above, and the lack of significant results likely means that the relevant biotic interactions were not detected rather than not being important. Although predation is likely important throughout the habitat range of sand lance, it may be that specifically at the refuge habitat it is not an important factor shaping sand lance habitat selection. Although, sand lance are known to be preyed upon when utilizing the burying habitat (Hobson 1986), much of the predation pressure may occur when sand lance are feeding in the water column away from the burying refuge. Qualitatively, marine birds, which are known to depend heavily on sand lance as prey, such as the marbled murrelet (*Brachyramphus marmoratus*), were rarely seen close to the sample beaches despite some beaches having extremely high abundance of sand lance. Because the sample sites were likely used as a burrowing refuge for sand lance to avoid predators, predation pressures may have been limited at each site and therefore not reflected in the fish community.

I did not attempt to include biotic variables in the habitat models because of the lack of information, the difficulties associated in modeling them. Rather, the community-based analysis was done to explore possible relationships between sand lance and other fishes. Although Austin (2002b) suggests that biotic processes should be given more consideration in habitat modeling, this requires a certain level of basic ecological information, which is not yet available for many species including the Pacific sand lance. However, because many of the biotic conditions vary with space and time, at the local scale an organism may be "cued" to possible suitable habitat by abiotic landscape features that correlate with biotic variables (Heglund 2002).

3) *Range of Environmental Sampling*

Ideally, a full range of environmental conditions should be sampled to build robust models (Huston 2002). Also, it is important to sample sufficient combinations of habitat variables to understand the complex underlying interactions that are generally present with ecological data. The broader the range of sampling and the higher the number of variable combinations included, the more likely relationships will be correctly described and important interactions will not be missed. The range of environmental variability sampled ultimately determines the accuracy, precision and generality of the model (Austin 2002a, Huston 2002). Because sampling all possible environmental gradients and combinations is logistically infeasible, extrapolation is usually unavoidable.

A broader range of environmental characteristics exists compared to what was sampled in this study. This adds a further point of caution to the interpretation of the modeling results as relationships that may hold true over the range of values used in the study may show disparate results when applied to those conditions that fall outside the range. Indeed, due to nonlinear relationships often seen in ecological data, the relationship may in fact not only differ in magnitude but also in direction.

An example from this study is the *Grain Size Mean* variable. Sand lance avoid small sized sediments such as mud and silt (*e.g.* Pinto *et al.* 1984, Wright *et al.* 2000). It is also likely that there is a maximum grain size for which sand lance cannot bury since at some point along the continuum, large sized grain would impede burying. Holland *et al.* (2005) demonstrated *A. marinus* avoidance of large sized grains as well as fine grains. Given the upper and lower restrictions of this habitat variable, we may expect to see a classic unimodal response to the spectrum of grain size if sampled over the entire continuum. In this study, the unimodal response was not seen for *Grain Size Mean* when compared to sand lance abundance (Figure 5.13). The lack of the unimodal response may be due to the lack of sampling at the extremes or because of the complexities of the interactions (Austin 2002b).

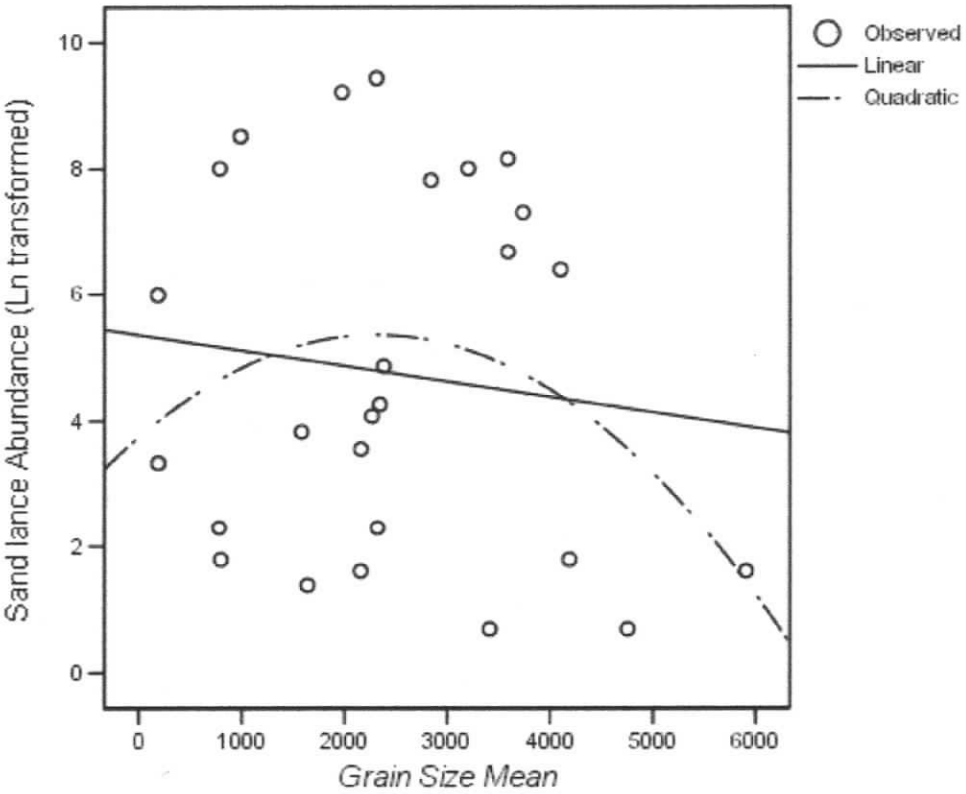


Figure 5.15: *Grain Size Mean vs. Sand lance Set Abundance* for 2003. Lines represent regression fits (linear R^2=0.015, quadratic R^2=0.084, neither significant at the 0.05 level). Absent sites were removed before analysis.

4) Temporal Considerations

The majority of empirical habitat models assume that a species distribution is at equilibrium (*i.e.* static) such that the distribution does not fluctuate during the sampling window (Hirzel and Guisan 2002). However, since resources or habitat can be available episodically, a species habitat selection is generally temporally dynamic, thus equilibrium is unlikely to exist (Henebry and Merchant 2002, O'Connor 2002). Despite the general prevalence of seasonal variation seen for most organisms, in chapter four I showed that throughout the summer, sand lance detection was consistent at the same sites.

Many habitat models do not consider interannual variation due to difficulties acquiring adequate time series data. Species that are *r*-selected (*i.e.* produce large numbers of undeveloped young), such as sand lance, are likely to have high interannual fluctuations in abundance. These fluctuations can cause the loss of predictability for habitat models built on one year and applied to other time periods (Boone and Krohn

2002). Results from the three years of sampling showed that sand lance habitat selection can vary drastically interannually (chapter four), however, the incorporation of temporal variation into the models is beyond the scope of this study. Incorporating the temporal component would require the measurement of all habitat variables over three years. Here, I was only able to measure environmental data in 2003 thus modeling was restricted to that year. Models, such as presented here, are likely robust for the time period of sampling but their ability to predict over other time periods is questionable (Johnson and Krohn 2002).

5) Stochastic Events

Stochastic events have the ability to mask true ecological relationships (Fielding 2002). It is generally assumed for inductive habitat modeling purposes, that the observed distributions used reflect the environmental selection of the species. This may not always be true since stochastic events, rather than a selection of location by the individuals, can cause individuals to be present at a location. Stochastic factors that are not correlated with environmental variables (*e.g.* dispersal) can confound habitat modeling procedures (Huston 2002). Further discussion on how stochastic events may affect sand lance distribution can be seen in chapter four. The disadvantage of sampling each site only once through the season is that there is an increasing the chance that a stochastic event causing individuals to be present at a site where they usually wouldn't reside may not be detected.

Habitat Preference* versus *Habitat Selection

I purposely avoided using terminologies such as "habitat preference" and "habitat quality" in this study. Since one cannot simply infer habitat quality from predicted probabilities of a habitat model (Nielsen *et al.* 2005), it is important to point out that the results presented here represent associations of habitat variables with sand lance distribution and not an intrinsic quality. Taking the extra step and assuming that these may represent habitat quality is filled with pitfalls (Railsback *et al.* 2003). Although the step from habitat associations to habitat preference may seem intuitive, it actually involves a further set of assumptions that have been criticized as flawed by many authors (*e.g.* Van Horne 1983, Hobbs and Hanley 1990, Garshelis 2000, Ralsback *et al.* 2003). Van Horne (1983) and Ralsback *et al.* (2003) outline numerous environmental and biological factors that can increase the probability that density will not be positively correlated with habitat quality. For example, sand lance appear to be temporally variable which could lead to high densities in low quality habitat due to overflow during high population levels (Van Horne 1983). Another example where the species density-habitat quality relationship breaks down is when unoccupied sites provide essential habitat that is used seasonally or during different life stages (Wiens 1996) or when suitable habitat is unoccupied since they have not been colonized yet (Ralsback *et al.* 2003). Because sand lance are found present at a site does not necessarily mean that it is of higher quality than unoccupied sites. In order to truly define habitat quality, the fitness associated with using the habitat must be measured (Van Horne 1983), an undertaking that is often prohibitively difficult.

Predicted **versus** ***Abundance***

One disadvantage to using logistic regression and classification tree models is that they do not provide any information with regards to abundance. When all sites were included, there was a significant positive correlation between the predicted values of logistic regression Model 1 and abundance. However, when only sites with sand lance present were included, there was no correlation. Thus, it is likely that "absent" sites are responsible for the correlation between abundance *versus* model predicted values using all sites. This suggests the model can discriminate between presence and absence locations, however, when sand lance are present, the abundances cannot be predicted by the logistic regression predicted probabilities. Similar results have been demonstrated in other studies (Pearce and Ferrier 2001, Nielsen *et al.* 2005). Nielsen *et al.* (2005) suggest that the lack of a functional relationship between the occurrence model and abundance implies that there may be different environmental factors that limit occurrence compared to what affects abundance (supported by Cushman and McGarigal 2004). Ecological factors such as competition and life history may have a larger effect on abundance than occurrence (Nielsen *et al.* 2005).

The choice of an occurrence model as opposed to an abundance model may be prudent considering the difficulties in modeling abundance. Occupancy models can be useful in determining possible environmental characteristics of suitable habitat and have been shown as useful predictive tools (Johnson and Krohn 2002) while abundance models have generally been less precise (Stauffer 2002). Although some authors maintain that it is important to eventually go beyond occurrence and integrate demographic models and biological interaction models (Scott *et al.* 2001, Nielsen *et al.* 2005), Stauffer (2002) suggests that due to the inherent noise of the systems, in most cases accurately predicting presence/absence is a more attainable goal.

Pearce and Ferrier (2001) examined the practical value of modeling abundance using both direct and indirect methods for a wide number of species. They found that models generally performed poorly suggesting that abundance models may be generally less robust than occurrence models. One issue with abundance models is that the assumptions are often difficult to meet. Robust abundance models require two major assumptions (Pearce and Ferrier 2001). First, there must be a known strong relationship between abundance and measured habitat quality. This assumption is not necessarily valid since abundance does not necessarily reflect habitat quality (Van Horne 1983, Ralsback *et al.* 2003) as discussed above. Second, there must be a strong relationship between the measured abundance and the true abundance. Reliable estimates of abundance are generally more difficult than simple presence/absence (Young and Hutto 2002) and may vary among habitat types, time of day, observer or other factors (Pearce and Ferrier 2001).

5.5 CONCLUSION

Here, the habitat modeling was necessarily descriptive rather than mechanistic due to the lack of information available in the literature to construct the conceptual model. In lieu of detailed biological information on *A. hexapterus* I constructed models that focused on variable selection rather than parameter estimation. These empirical models give insight into possible environmental characteristics that drive sand lance habitat selection and the response of sand lance to these environmental variables. However, because models are likely misspecified to some extent, it is important not to assume cause-effect relationships from statistical significance of parameters (Reckhow *et al.* 1986, Corsi *et al.* 2000). Models instead are summaries of data associations and describe functional relationships that may provide insight to possible cause-effect relationships. These insights are valuable for generating and supporting theory.

CHAPTER 6: Modeling Nearshore Habitat for Young-of-the-Year Sand Lance: A GIS Inductive Approach.

6.1 INTRODUCTION

GIS provides a tool for examining habitat data spatially and for producing habitat maps. Data utilized in a GIS generally has a coarser resolution than environmental data measured in the field (Thomas *et al.* 2002). Thus, GIS studies usually examine data at a broader scale than studies based solely on field data. It is often preferable to use habitat variables available over large areas because of the prohibitive cost of gathering fine scale field data (Shriner *et al.* 2002). Many species are affected by the broad scale physical structure of habitat and the structure of the surrounding landscape (Store and Jokimäki 2003). Abiotic variables based on these structures are often straightforward to implement in GIS (Eastwood *et al.* 2001) and the resulting spatially explicit models can be easily adapted for management purposes (*e.g.* Moses and Finn 1997, Corsi *et al.* 1999, Kobler and Adamic 2000, Rubec *et al.* 2001, Danks and Klein 2002, Ingram and Rogan 2002).

Here, I use environmental variables that are cover a larger extent compared to the fine grain data considered in chapter five. These variables were either modeled in or imported into the GIS environment. I use the same basic analytical approach I employed in chapter five to analyze the GIS variables in relation to YOY *Sand lance Presence/Absence* from 2003 to determine the species-environment relationship at this scale.

6.2 METHODS

Dataset

I used the same dependent variable as chapter five (*Sand lance Presence/Absence* for 2003) in order to be able to compare between the results of the two chapters. Thus, I used the same 55 study sites from 2003. I started with twelve independent variables that fell into three major groups: 1) Oceanographic, 2) Shoreline, 3) Benthic.

Oceanographic Variables

Six oceanographic variables were modeled in the GIS. Each of these variables was converted into a raster coverage that extended over the entire study area. I extracted the environmental data for each site from the continuous layers using a buffer with a radius of 200m. The buffer was created around the midpoint of each of the 55 study sites. Land features, defined as zero depth, and parts of the buffers on far sides of land barriers (buffer extended to water across a land barrier such as an opposite side of an island) were clipped out. The buffers therefore had variable areas (mean area = 66030m^2, S.D. = 26560m^2). Information on the daily home range of sand lance is not known, therefore, the choice of a 200m buffer radius was subjective, however, I chose this distance to capture the features at a larger grain than that of the empirical study. For each sample site, the average value of each continuous oceanographic variable within the buffer was extracted.

Depth

Depth was calculated as the average bathymetric depth extracted from a 25m-bathymetry grid. Bathymetry was based on Canadian Hydrographic Survey (CHS) mass point depth data, CHS chart 3671 contours and coastline data primarily from Terrain Resource Information Management (TRIM) basemaps (Collyer 2004c). Depth was extracted from the 25m grid using a Triangulated Irregular Network (TIN).

Tidal Velocity

Tidal Velocity was generated using the finite element numerical model described in Foreman and Thomson (1997). These data represented a depth integrated average tidal velocity and consisted of a point dataset that was interpolated using a TIN interpolation technique. Further details can be found in Collyer (2004c).

Stratification

Stratification was calculated using *Tidal Velocity* and *Depth* layers to measure the Simpson and Hunter stratification parameter. This parameter (Simpson and Hunter 1974) has commonly been used to calculate stratification (*e.g.* Yuasa and Ueshima 1992, Glorioso and Simpson 1994, Bisagni 1992). The stratification parameter (S) is calculated with the following formula:

Equation (1) $S=\text{Log}\,(H/U^3)$

Where H is the water depth and U is the amplitude of the tidal current. Using the bathymetric layer (*Depth*, H) and the *Tidal Velocity* (U) a continuous stratification layer was generated using Equation (1).

Slope

Slope was created from the bathymetric layer, using a TIN interpolation. The raster layer was computed by calculating the mean change in depth between a cell to the eight (or less) surrounding cells. *Slope* values represent the degree of slope for each cell location.

Complexity

Benthic complexity is defined by Ardron (2002) as the density of the change in slope (*i.e.* the slope of the slope) of an exaggerated bathymetry. *Complexity* was calculated using methods outlined by Ardron (2002). The original bathymetric layer (cell size of 25m) was exaggerated by a multiple of 20. A search radius of 500m was used to determine change in slope density. Further details can be found in Collyer (2004c).

Coastline Density

Coastline Density values were calculated as the linear density of coastline within a search radius of 200m to produce a continuous raster layer (*i.e.* the total linear length of coastline within the search radius).

Shoreline Variables

The two shoreline variables consisted of a classification of sediment type digitized from Lee and Bourne (1982) and a model of exposure based on methods from Keddy (1982). These variables were extracted from the categorical (*Sediment*

Classification) or continuous (*Shoreline Exposure*) values of each variable for segments of the coastline. *Sediment Classification* was missing data from 15 sites since the coverage of the Lee and Bourne study was only for the BGI.

Sediment Classification

Lee and Bourne (1982) partitioned shoreline into sample units for all shorelines in the BGI. Each shoreline unit was digitized for the purposes of this study along with associated sediment classification of the unit according to Lee and Bourne (1982). Description of digitization is found in (Collyer 2004b). *Sediment Classificaion* had four categories: mud, sand, gravel and bedrock. The sediment classification for each site was determined by the sediment classification of the shoreline unit that the site was located in.

Shoreline Exposure

Exposure was calculated by first calculating direct fetch, then effective fetch, and finally a relative exposure index as described by Keddy (1982). Direct fetch was defined as the distance across water in one direction to the next shoreline or the cut off value. Adey (1978) estimated that (in the Caribbean Sea) waves were saturated at 50-100 km, therefore, a cut off value of 100 km was used for this study.

Direct fetch was calculated by first, radiating lines from a point on the shoreline at every 11.5° starting at 0°. This produced 32 radiating lines, each 100 km long, for each point. These lines were then clipped by the landforms producing segments radiating from the original point with various lengths (*i.e.* direct fetch values). Broken line segments not connected to the point were discarded. Effective fetch was calculated by the vector addition of the direct fetch values within a bearing range (as described in Keddy 1982, Duarte and Kalff 1986). The relative index of exposure was calculated using methods from Keddy (1982) by adding prevailing wind data (details in Collyer 2004a).

Benthic Variables

The four benthic variables *Side-scan Mud, Side-scan Sand, Side-scan Gravel,* and *Side-scan Bedrock*, were based on side-scan sonar data collected in 2003 for specific regions in the study site. They were extracted using the 200m radius buffer used for oceanographic variables and represented the proportion of the specific sediment type within the buffer. Benthic variables were missing data from 28 sites since side-scan sonar was not conducted for the entire extent.

The side-scan sonar surveys were focused in the BGI and were conducted during August 18th-22nd, 2003. Survey lines were conducted and then ground-truthed using underwater drop video camera. The side-scan sonar used was a 100 kHz, fully scale and slant-range-corrected EdgeTech 260. A 100 m range (swath width, 200 m) was used for all surveys. One-second fixes were recorded for both the side-scan surveys and the drop camera ground-truthing surveys.

Exploratory Analysis and Univariate Testing

I examined all continuous variables for normality using the Kolmogorov-Smirnov test and transformed variables that had a distribution significantly different from the normal distribution. *Side-scan Mud* was dominated by zero counts and was therefore transformed into a categorical presence/absence variable. To explore for relationships between *Sand lance Presence/Absence* and the independent variables, I conducted an independent sample t-test for continuous independent variables, and a Chi-square for categorical independent variables.

Correlations

I screened for multicolinearity by constructing a correlation matrix. I considered variables with correlation values $|r| \geq 0.70$ as highly correlated. I used the criteria outlined in chapter five ("*Correlations*" section) for dropping highly correlated variables. I also examined bivariate scatterplots to screen for non-linear relationships between independent variables and outliers.

Logistic Regression and Classification Tree

I constructed a logistic regression model using the backwards stepwise approach outlined in chapter five ("*Logistic Regression*" section) using the selected independent variables.

Using the same modeling procedure outlined in chapter five ("*Classification Tree*" section). I constructed two classification trees, one based on misclassification pruning and one on deviance pruning. I used the independent variables I selected for the above logistic regression analysis.

6.3 RESULTS

Exploratory Analysis and Univariate Testing

The results showed no significant relationships between *Sand lance Presence/Absence* and independent variables with the exception of *Coastline Density*, which showed a significant relationship at the 0.05 level (Table 6.1). *Sediment classification* and the benthic variables were omitted from further analysis since there was no significant relationship between the *Sand lance Presence/Absence* and the addition of these variables in the model would greatly reduce the sample size due to missing data.

Table 6.1: Results from the two-sample independent t-test analysis of independent variables grouped by the binary dependent variable *Sand lance Presence/Absence*.

Variable	Z	DF	p-value
Stratification	-0.947	53	0.348
Slope	-1.164	53	0.25
Exposure	-1.575	53	0.121
Complexity	0.216	53	0.83
Coastline Density	2.271	53	0.027
Depth	-1.492	53	0.142
Tidal Velocity	0.204	53	0.839
Side-scan Sand	0.219	25	0.828
Side-scan Gravel	1.417	25	0.169
Side-scan Bedrock	0.325	25	0.748

Correlations

Based on the selection criteria outlined in chapter five, I dropped the variables *Slope* and *Tidal Velocity* from analysis. I did not detect any non-linear relationships between independent variables or obvious outlier values.

Logistic Regression and Classification Tree

Even though only one independent variable was significantly related to *Sand lance Presence/Absence* during the univariate analysis (*Coastline Density*), I started with all five independent variables, as this is not an unreasonable number of variables to begin analysis.

The backwards stepwise modeling procedure, starting with a model containing all five independent variables, terminated after five steps. At each step, I evaluated the model statistics to the view of the performance of the model at that point using ROC and AIC (Table 6.2). Based on the AIC statistic, the terminal step model (Model 5) is the preferred model with each earlier step having progressively higher AIC values. The ROC values show the opposite relationship compared to AIC. Based on the ROC values, Model 1 is the preferred model. Overall, Model 5 is likely the preferred model since it has a significant model Chi-square value, a significant independent variable and an ROC value comparable to the other four models. The other four models, although having better resubstitution classification rates and ROC values, are likely overfit as shown by the higher AIC values.

The regression tree produced similar results to the logistic regression Model 5. Both misclassification and deviance pruning selected a tree with only one split when using the 1-SE rule (Figure 6.1), and thus, the final tree contained only one independent variable for both pruning methods (Figure 6.2). As in Model 5 of the logistic regression analysis, the *Coastline Density* variable was the only independent variable selected for inclusion in the model. According to the tree model, lower *Coastline Density* values are associated with a higher proportion of sand lance present (node 2) while higher values are associated with sand lance absence (node 3). *Coastline Density* was ranked higher in importance compared to any other variables (Figure 6.3).

Table 6.2: Logistic regression modeling result using *Sand lance Presence/Absence* as the dependent variable (N=55). A backwards likelihood ratio stepwise selection method was used to construct the five models. Each model was one of the five steps with Model 5 being the terminal step. *Independent variable significant at the 0.05 level (Wald statistic). **Model Chi-square value significant at the 0.05 level.

	Model 1		Model 2		Model 3		Model 4		Model 5	
Variable	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
Complexity	0.000	0.002								
Stratification	0.129	0.598	0.131	0.594						
Depth	0.155	0.175	0.158	0.156	0.171	0.146				
Exposure	0.497	0.303	0.496	0.302	0.496	0.302	0.380	0.281		
Coastline Density	-0.002	0.001	-0.001	0.001	-0.002	0.001	-0.002*	0.001	-0.002*	0.001
Constant	-2.764	2.824	-2.766	2.820	-2.474	2.484	-0.696	1.976	1.729	0.899
Model Statistics	Model 1		Model 2		Model 3		Model 4		Model 5	
Model Chi-Square [df]	8.68 [5]		8.67 [4]		**8.63 [3]		**7.12 [2]		**5.23 [1]	
Sensitivity (% Correct Predictions)	75.9		75.9		72.4		62.1		58.6	
Specificity (% Correct Predictions)	65.4		65.4		65.4		69.2		53.8	
Overall % Correct Predictions	70.9		70.9		69.1		65.5		56.4	
Nagelkerke pseudo R^2	0.195		0.195		0.194		0.162		0.121	
AIC	81.82		79.83		77.95		76.44		74.85	
Area Under the ROC Curve	0.735		0.731		0.731		0.716		0.671	

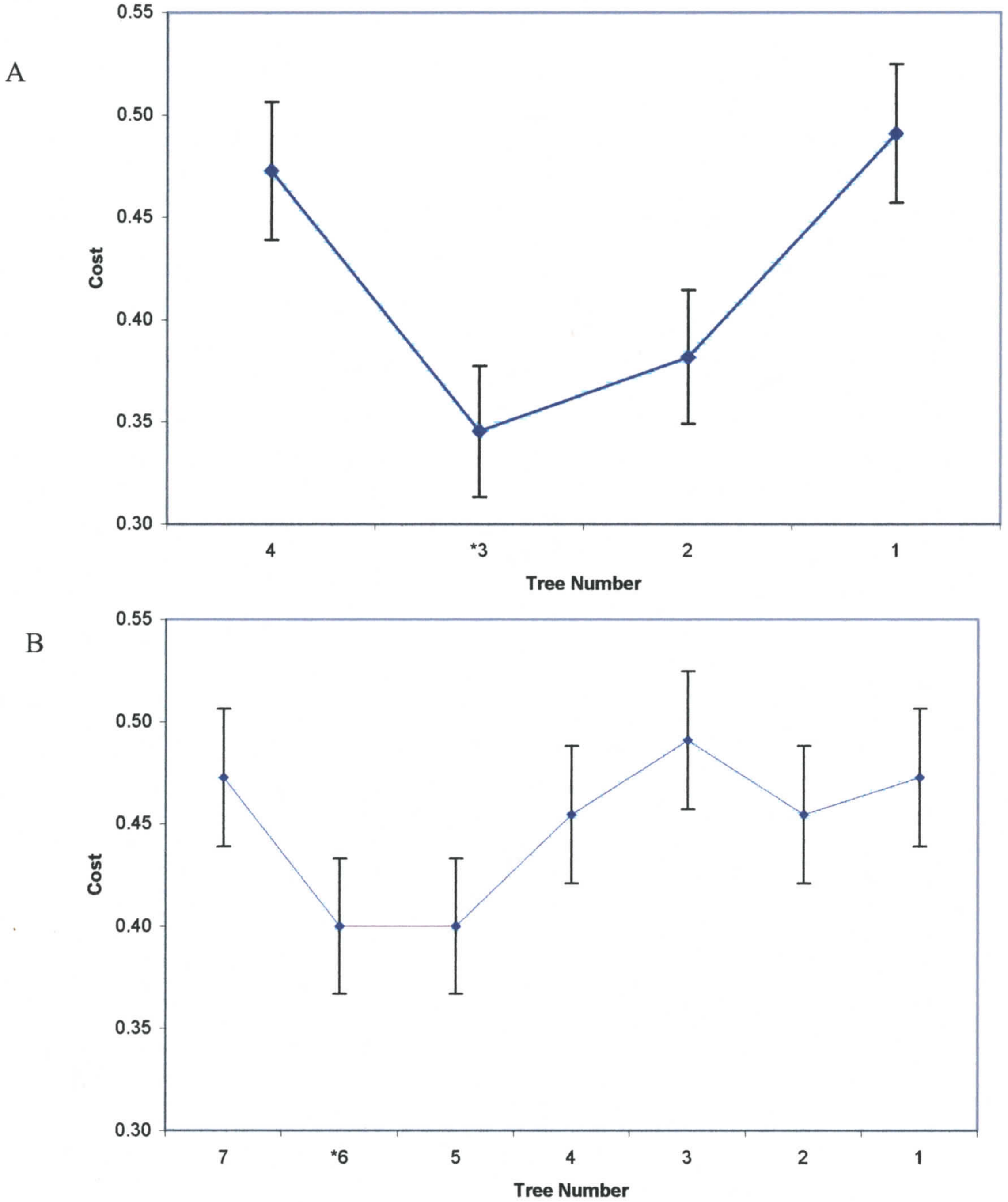


Figure 6.1: 10-fold cross-validation cost for misclassification pruning (A) and deviance pruning (B). Tree Number indicates the number of splits occurring in the tree and error bars represent standard error. *the tree number selected by the pruning procedure .

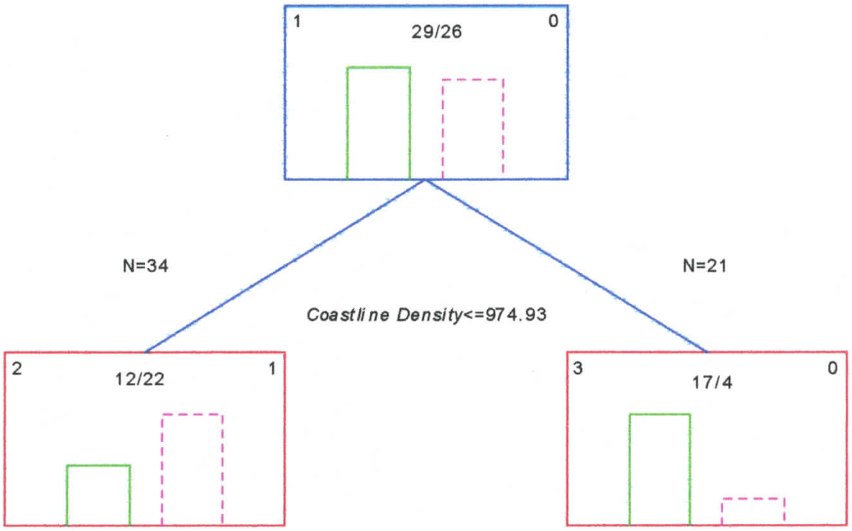


Figure 6.2: Classification tree grown using *Sand lance Presence/Absence* as the dependent variable and the five selected independent variables. Misclassification pruning and deviance pruning produced the same tree. For each node box, the node number is in the top left, the ratio of absence to presence in the top middle and the node class (sand lance absence = 0, sand lance presence = 1) in the top right. The histograms within the node boxes represent the class proportions with green solid representing absence and pink hatched representing presence. The splitting rule indicates the rule for the left node determining the cases belonging in the left node of the split with the remaining cases belonging in the right node.

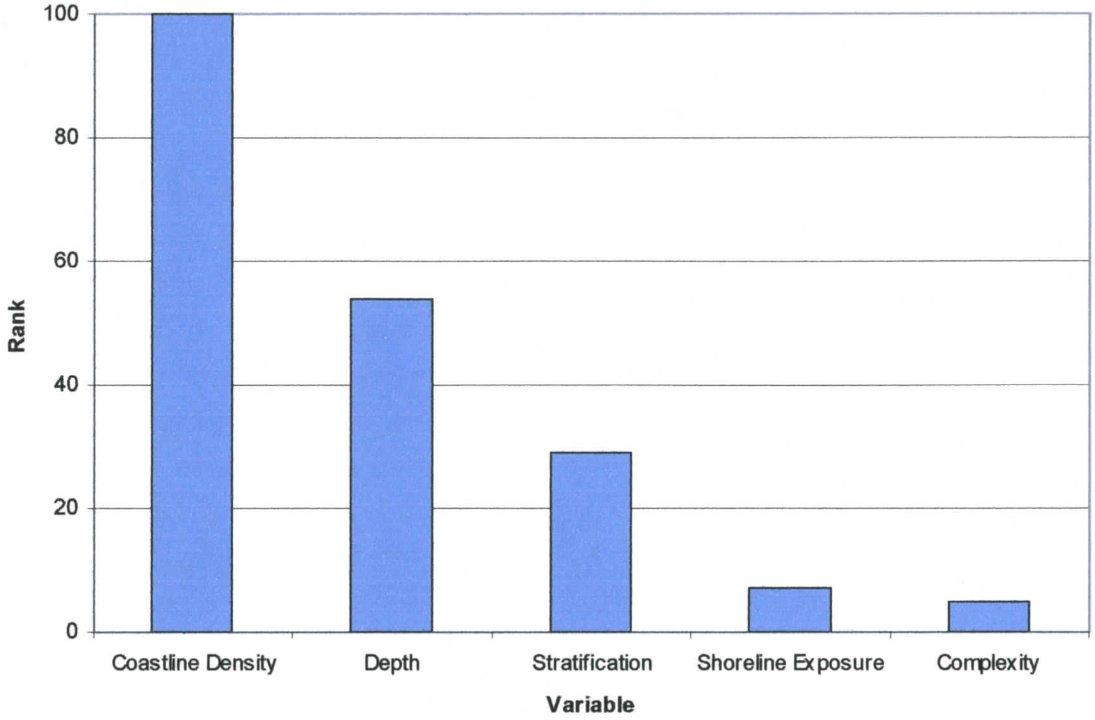


Figure 6.3: Importance of independent variables in regression tree analysis the deviance-pruned tree ranked on a scale of 0-100. Ranking is based on the overall change in node impurity summed over the entire tree standardized by the largest sum.

6.4 DISCUSSION

Other authors have examined habitat data on multiple scales and achieved a better understanding of factors that contribute to a species habitat selection (*e.g.* Johnson *et al.* 2002, Store *et al.* 2003). Also, Thomas *et al.* 2002 found that "fine" and "coarse" resolution environmental variables perform similar concluding that independently derived digital data has a comparable performance to fine grain field data. However, in this study, both logistic regression and classification tree procedures failed to predict sand lance presence/absence in the nearshore adequately using broad GIS derived data. This suggests that the variables chosen, with perhaps the exception of *Coastline Density*, have no biological relevance to sand lance at this scale of study.

The only variable that did prove to be significant was *Coastline Density*. *Coastline Density*, also known as "coastline complexity", is a physical abiotic measure that is related to the complexity of the local environment, water residence time, exposure

(Bartley *et al.* 2001) as well as potentially influencing terrestrial inputs such as groundwater discharge (Buddemeier 1996). *Coastline Density* has also been significantly related to the structure of ecological communities in the nearshore (*e.g.* Blanchart and Bourget 1999). However, despite its significance in the models presented in this study, it is difficult to explain its direct biologically relevance to sand lance habitat selection. Models based on variables measurable over large areas can identify potential habitat, however, they generally do not provide adequate information on what habitat features may directly affect the species (Shriner *et al.* 2002). *Coastline Density* could be considered an indirect distal variable, as it likely is an indirect measure of causal processes. Because the direct relevance to sand lance habitat selection is unknown, its significance should be interpreted with care.

Limitations

There are numerous general limitations with GIS modeling that may have contributed to the failure to adequately model sand lance habitat at this scale. I will discuss three possible limitations: 1) independent variable suitability and availability, 2) data accuracy, and 3) mismatches in scale.

Variable Suitability and Availability

GIS is limited by the availability of variables that can describe an organism's habitat. Generally, variables that are easily implemented in a GIS are physical variables that cover large extents such as landscape variables. Landscape ecology focuses on spatial patterns of habitat and the effect of these patterns on a species habitat selection (Levin 1992). The weakness of using this approach in this application is that important spatial variables or commonly used landscape variables may be unknown for sand lance. Landscape variables such as spatial heterogeneity, fragmentation and edge metrics are difficult to quantify for sand lance because much of its ecology, such as home-range and diel behavior, are poorly understood. Also, landscape variables are difficult to quantify because many of the marine habitat features are highly variable temporally and spatially compared to landscape features of terrestrial systems (Steele 1998).

Environmental variables I chose may not have been important in determining sand lance habitat but rather other unmeasured variables may be important at this scale of study. The variables I considered did have the potential to be important as habitat variables for sand lance. For instance, Ostrand *et al.* (2005) found that depth was significantly related to sand lance use of subtidal habitat. Here, however, *Depth* was not found significant in the modeling procedures. Also, two other potentially important subtidal variables, *Complexity* and *Stratification*, were not seen as important. This is in line with results from chapter five, where shallow subtidal characteristics were not important in explaining nearshore habitat use. Subtidal measurements considered in this chapter again appear not important in determining habitat use in the nearshore even though one of these measures has been related elsewhere to sand lance habitat selection in the subtidal. The *Shoreline Exposure* variable also potentially could have been important for sand lance habitat use. Exposure plays a major role in determining the physical characteristics of that coastline (Ekebom *et al.* 2002) and can influence habitat characteristics such as substrate type, macro-vegetation (Keddy 1982, Keddy 1983) as well as temperate fish communities (Thorman 1986). However, despite its theoretical potential, *Shoreline Exposure* was not significantly related to sand lance habitat selection.

Data Accuracy

A second possible reason for poor modeling results is the accuracy of the independent variables. Variables that are derived from interpolation and calculations, such as those derived in the GIS environment, are generally less accurate than the original data (Guisan and Zimmermann 2000). Here, the environmental variables were modeled in the GIS and thus may contain a higher degree of error compared to field measurements. Also, data that was digitized may be inaccurate as digitization itself introduces errors (Corsi *et al.* 2000). Errors in GIS data are difficult to assess and can propagate forward. It may be the case that a relationship exists between the sand lance occurrence and the environmental variables used here, however, the original modeling and interpolation of environmental characteristics in the GIS was imprecise. If the variables contain a high degree of error, true relationships between the independent variables and the dependent may be not be detected.

Mismatches in Scale

Lastly, the scale of observation may be inappropriate. Processes that influence species distributions must be properly quantified in order to understand a species habitat selection pattern (Huston 2002). Unless an appropriate scale (grain and extent) is used for analyzing ecological processes, these processes cannot be adequately described. There are often mismatches in scale (both spatial and temporal) at which data is modeling using GIS and field collections (Henebry and Merchant 2002). Preferably, the species data and the environmental data should be measured at the same scale (Guisan and Thuiller 2005). Here, field measurements of *Sand lance Presence/Absence* at sites were compared with broad environmental characteristics. Thus, it may be that there is a mismatch between these two scales. In other words, the broad environmental variables may not adequately describe the site-specific sand lance occurrence but rather may predict sand lance occurrence larger scale grain sizes. Because I did not have sand lance occurrence data on this scale I could test this possibility.

It is difficult to determine a suitable scale of study for sand lance since little is known about their daily use patterns of the intertidal and subtidal regions and the connectedness between the two. The value of a 200m radius buffer is a reasonable yet subjective starting point for a scale of study. It may be the case that the variables examined here are related to *Sand lance Presence/Absence* at a finer or coarser grain. One could test this by using multiple grain sizes (*e.g.* buffers of 100m, 150m, 200m, 250m…) however this runs the risk of producing spurious relationships (Henebry and Merchant 2002). This is known as the Modifiable Areal Unit Problem (MAUP). If the system being studied is well known, the scale of study can be set in an ecologically meaningful way and the MAUP can be avoided. If the system is not well understood, then the definition of scale becomes arbitrary.

6.5 CONCLUSION

Coastline Density was the only independent variable that I found significantly related to YOY sand lance use of the nearshore. The lack of relationships between the broad scale independent variables and *Sand lance Presence/Absence* suggests that at this scale of study, sand lance nearshore habitat patterns cannot be predicted given the environmental variables considered.

CHAPTER 7: Conclusion

The main objective of this study was to determine the general environmental factors that influence YOY sand lance use of the nearshore. At the site-specific level, I examined a total of 30 environmental variables and found four of these variables to be important in modeling habitat use: *Major Substrate Low Intertidal*, *Intertidal Eelgrass Presence/Absence, Grain Size Mean, Grain Size Sorting*. At the landscape-level, I examined twelve environmental variables and found only one of these to be important in modeling habitat use: *Coastline Density*. Unlike the site-specific models, the landscape-level models did not perform well in describing sand lance distribution. No important biotic interactions were detected with the fish community analysis.

The spatial scale considered in a study affects the detection of ecological patterns and processes that determine the species-environment relationship (Wiens 1989). There is no accepted correct scale (grain and extent) of study, but rather scales of study are chosen for their ability to accurately reflect the processes that influence the species habitat selection (Huston 2002). This being the case, a species distribution can best be predicted using a multi-scale approach (Johnson *et al.* 2002). In this study, I examined selection of nearshore habitat by YOY sand lance at the site-specific level and the landscape level (chapter five and six). Here, the site-specific level proved a better scale for determining processes important for habitat selection.

Studies that use fine-grained "microhabitat" data, such as the site-specific level used in chapter five, are often the most suitable for descriptive habitat modeling (Shriner *et al.* 2002). The main purpose of descriptive models is to determine habitat relations and refine patterns. However, because the data required are sampled at such a detailed scale grain size, data are generally not available to apply to large areas (Shriner *et al.* 2002). The fine-grained habitat models presented in chapter five provided important information with regards to habitat correlates but are of limited use to managers as predictive tools to map habitat. Here, fine-grained models suggested that intertidal properties are important for YOY sand lance in the nearshore. Sand lance avoided fine sediment in the intertidal and selected specific grain size and sorting values within the sand/gravel categories. Also, sand lance avoided intertidal eelgrass. Although many of the model statistics show potential predictive ability, these models are likely of limited use for prediction due to the nature of inductive modeling. Also, the interannual temporal variation in habitat use for some sites, seen in chapter four, suggests that one year of modeling would be inadequate for making accurate predictions. The true value of the site-specific modeling procedures is to identify variables that are important in shaping a species habitat (Young and Hutto 2002).

Studies that use a larger grain size, such as the landscape level analysis in chapter six, are often the most useful for prediction. In general, larger scale grain models use variables that cover large extents and can be readily implemented in a GIS. These models are spatially explicit and if they are found to perform well, they can be used to produce habitat maps for management purposes. However, the applicability of these models is limited by the availability of broad scale variables that can successfully predict a species habitat selection. Also, models of this nature are affected by data accuracy and scale mismatch issues. In chapter six, out of all the broad scale variables considered I found only *Coastline Density* was related to habitat use of sand lance. This suggests that the independent variables used were not important for sand lance at this scale, had problems in accuracy or were mismatched in scale with the dependent variable. Thus, because of poor modeling results, I was unable to construct habitat use maps that may have had utility for managing sand lance.

As pointed out earlier, it is important not to infer causation from inductive modeling such as those in chapter five and six. Because environmental factors generally show high covariance, it is often difficult to determine which predictors are the actual causal predictors and which are highly correlated to unmeasured causal variables (Huston 2002). An example of this would be *Coastline Density*. Although it was found to be significantly related to *Sand lance Presence/Absence* in chapter six, the direct biological relevance is unknown and it is likely correlated to one or more proximal variables that were unmeasured. For this reason, controlled experiments rather than field based studies are required to infer causation. However, for the purpose of resource management and conservation, distal variables can be used for application such as delineating habitat.

Management Considerations

Although single species inductive models have limited management applications, they are still required to further the understanding of the species of interest and the systems in which they live (James and McCulloch 2002). In order to manage special interest species, knowledge of habitat requirements is necessary. Sand lance habitat use is of particular interest because, given their importance in the marine food web, sand lance habitat selection and spatial and temporal distribution likely have important consequences for many other species. The spatial and temporal pattern of sand lance habitat use is likely linked to predator distribution and reproductive success (Willson *et al.* 1999). Knowledge of sand lance habitat selection is essential for managing sand lance as well as managing and researching species that predate on sand lance.

Here, models found that the intertidal sediment properties are important for sand lance habitat selection. Sand lance appear to select a very specific sediment type that can be described by the lack of fine grain sizes and the interaction between *Grain Size Mean*, and *Grain Size Sorting*. As discussed in chapter five, management or research efforts that do not properly quantify sediment properties when trying to predict sand lance habitat may provide inaccurate predictions. Sand lance burrowing habitat is important as refuge habitat and likely is key in regulating predator-prey dynamics (Rosenzweig and Macarthur 1963). It is important to include this habitat in management considerations, as it is an ecologically sensitive region susceptible to numerous anthropogenic impacts. Possible impacts include shoreline development, oil and heavy metal contamination, and upslope disturbances such as logging.

Shoreline development can affect sensitive intertidal areas that serve both as spawning and refuge habitat. Shoreline armoring such as bulkheads, removal of trees that provide shading and prevent erosion and habitat alteration due to aquaculture facilities could possibly be detrimental to sand lance intertidal/shallow subtidal habitat. These types of developments not only directly displace habitat but also interfere with sediment transport processes. In some areas, the effects of shoreline development on sand lance habitat are already being considered in shoreline management decisions (*e.g.* Penttila 1997, Penttila 2000).

Oil contamination of intertidal sediment has been shown to significantly affect sand lance burrowing behavior (Pinto *et al.* 1984, Pearson *et al.* 1984) and immune response to parasitism (Moles and Wade 2001). Fine grain littoral substrates (the sediment generally selected for by sand lance) are a major source of long-term sequestration of petroleum hydrocarbons (Moles and Wade 2001). It is possible that sand lance that come in contact with lipophilic hydrocarbons are capable of transferring these pollutants to predator populations thus spreading the pollutant impacts beyond the local area to regions that may not be otherwise affected (Moles and Wade 2001).

Sediment composition may also be altered by the artificial elevation of fine sediments and organics. Upland logging or other industrial activities can increase the amounts of fine sediments and organics that are carried to the marine environment by rivers and streams. These sediments may then be deposited on marine beaches near the mouths of freshwater outputs (*e.g.* Heikkila 1991, Tunnicliffe 2000). An increase in fine sediments and organics may lower the flow of interstitial water increase the oxygen demand from microbes resulting in lower interstitial oxygen available to burrowing sand lance (Quinn 1999).

Further Research

The results of this study statistically described the species-environment relationship, however, analysis was limited in both spatial and temporal scale. YOY use of the nearshore should be examined at different spatial and temporal scales to further build on the results presented here. This includes an increase in the spatial and temporal range of sampling for each environmental variable (*e.g. Grain Size Mean*) and longer-term studies of the temporal variation in YOY populations. This will help determine how spatial and temporal fluctuations in the environment or sand lance populations affect habitat use. Also, an integrated study is necessary to combine information on sand lance habitat use for various regions of the marine environment (*e.g.* subtidal, intertidal, pelagic) and for various life stages (*e.g.* spawning, larvae, YOY, 1+). This would provide an overall picture of the species-environment relationship and information on the connectedness between the use of different habitat types by different life stages.

Conclusion

Natural systems are generally too complex for accurate prediction at every temporal and spatial scale (O'Connor 2002). This study highlights the use of habitat modeling as descriptive tool that can enhance ecological knowledge rather than focusing on prediction. The models put forward here build on the understanding of sand lance habitat requirements by describing YOY use of the intertidal/shallow subtidal region. Although the models themselves likely have limited utility in other areas as accurate predictive tools, they did reveal the environmental factors that influence habitat use. This information can be incorporated into what is known about habitat use of sand lance in other regions of the marine environment to provide a better picture of the species-environment relationship.

REFERENCES

Abookire, A. A., and J. F. Piatt. 2005. Oceanographic conditions structure forage fishes into lipid-rich and lipid-poor communities in lower Cook Inlet, Alaska, USA. Marine Ecology Progress Series **287**:229-240.

Adey, W. H. 1978. Algal ridges of the Caribbean Sea and West Indies. Phycologia **17**:361-367.

Akaike, H. 1973. Information theory and extension of the maximum likelihood principle. Pages 267-281 *in* B. N. Petrov, editor. Proceedings of the 2nd International Symposium of Information Theory. Akademia Kiado, Budapest.

Andersen, M. C., J. M. Watts, J. E. Freilch, S. R. Yool, G. I. Wakefield, J. F. McCauley, and P. B. Fahnestock. 2000. Regression-tree modeling of desert tortoise habitat in the central Mojave desert. Ecological Applications **10**:890-900.

Anderson, P. J., and J. F. Piatt. 1999. Community reorganization in the Gulf of Alaska following ocean climate regime shift. Marine Ecology Progress Series **189**:117-123.

Angermeier, P. L., K. L. Krueger, and C. A. Dolloff. 2002. Discontinuity in stream-fish distributions: Implications for assessing and predicting species occurrence. Pages 519-527 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Anthony, J. A., D. D. Roby, and K. R. Turco. 2000. Lipid content and energy density of forage fishes from the northern Gulf of Alaska. Journal of Experimental Marine Biology and Ecology **248**:53-78.

Ardron, J. A. 2002. A recipe for determining benthic complexity: An indicator of species richness. Chapter 23. Pages 169-175 *in* J. Breman, editor. Marine Geography: GIS for the Oceans and Seas. ESRI Press, Redlands, CA, USA.

Arnott, S. A., and G. D. Ruxton. 2002. Sandeel recruitment in the North Sea: demographic, climatic and trophic effects. Marine Ecology Progress Series **238**:199-210.

Arnott, S. A., G. D. Ruxton, and E. S. Poloczanska. 2002. Stochastic dynamic population model of North Sea sandeels, and its application to precautionary management procedures. Marine Ecology Progress Series **235**:223-234.

Auster, P. J., and L. L. Stewart. 1986. Species profiles: life histories and environmental requirements of coastal fishes and invertebrates (North Atlantic)--sand lance. TR EL-82-4, U.S. Army Corps of Engineers.

Austin, M. P. 2002a. Case studies of the use of environmental gradients in vegetation and fauna modeling: theory and practice in Australia and New Zealand. Pages 73-82 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Austin, M. P. 2002b. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling **157**:101-118.

Bartley, J. D., R. W. Buddemeir, and D. A. Bennett. 2001. Coastline complexity: a parameter for functional classification of coastal environments. Journal of Sea Research **46**:87-97.

Bergstad, O. A., and Å. S. Høines. 2001. Effects of exploitation on age and size structure of sandeel, *Ammodytes marinus*, populations in the North Sea. Archive of Fishery and Marine Research **49**:3-18.

Berry, W. D., and S. Felman. 1985. Multiple regression in practice. Sage Publications, Beverly Hills, California, USA.

Bisagni, J. J. 1992. Differences in the annual stratification cycle over short spatial scales on Southern Georges Bank. Continental Shelf Research **12**:415-435.

Blackburn, J. B., and P. J. Anderson. 1997. Pacific sand lance growth, seasonal availability, movements, catch variability, and food in the Kodiak-Cook Inlet area of Alaska. Pages 409-426 *in* Forage Fishes in Marine Ecosystems. University of Alaska Sea Grant College Program, Anchorage, Alaska, USA.

Blanchard, D., and E. Bourget. 1999. Scales of coastal heterogeneity: Influence of intertidal community structure. Marine Ecology Progress Series **179**:163-173.

Blott, S. J., and K. Pye. 2001. GRADISTAT: a grain size distribution and statistics package for the analysis of unconsolidated sediments. Earth Surface Processes and Landforms **26**:1237-1248.

Boone, R., and W. B. Krohn. 2002. Modeling tools and accuracy assessment. Pages 265-270 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Boyce, M. S., P. R. Vernier, S. E. Nielsen, and K. A. Schmiegelow. 2002. Evaluating resource selection functions. Ecological Modelling **157**:281-300.

Breiman, L., J. H. Friedman, R. A. Olshen, and S. C.G. 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, California, USA.

Brind'Amour, A., and D. Boisclair. 2004. Comparison between two sampling methods to evaluate the structure of fish communities in the littoral zone of a Laurentian lake. Journal of Fish Biology **65**:1372-1384.

Brotons, L., W. Thuiller, M. B. Araújo, and H. A.H. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography **27**:437-448.

Brown, J. H., G. C. Stevens, and D. M. Kaufman. 1996. The geographic range: Size, shape, boundaries, and internal structure. Annual Review of Ecology and Systematics **27**:597-623.

Brown, S. K., K. R. Buja, S. H. Jury, and M. E. Monaco. 2000. Habitat suitability index models for eight fish and invertebrate species in Casco and Sheepscot Bays, Maine. North American Journal of Fisheries Management **20**:408-435.

Buddemeier, R. W. 1996. Groundwater flux to the ocean: Definitions, data, applications, uncertainties. Pages 16-21 *in* R. W. Buddemeier, editor. Groundwater Discharge in Coastal Zone: Proceedings of an International Symposium. LOICZ Reports and Studies No. 8. LOICZ, Texel, The Netherlands.

Burkett, E. E. 1995. Marbled murrelet food habits and prey ecology. Pages 223-246 *in* H. G. L. Ralph C. J., Raphael M. G. Jr., Piatt J. F., editor. Ecology and Conservation of the Marbled Murrelet. United States Department of Agriculture, Pacific Southwest Research Station, Forest Service, U.S. Dept. Agriculture, Albany, CA.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretical approach. Springer Verlag, New York.

Cablk, M., D. White, and A. R. Kiester. 2002. Assessment of spatial autocorrelation in empirical models in ecology. Pages 429-440 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Charles-Dominique, E. 1989. Catch efficiencies of purse and beach seines in Ivory Coast lagoons. Fisheries Bulletin **87**:911-921.

Chikilev, V. G., and A. V. Datskii. 2000. Pacific stout sand lance *Ammodytes hexapterus* (Ammodytidae) in the Gulf of Anadyr and adjacent waters. Journal of Ichthyology **40**:732-739.

Ciannelli, L. 1997. Winter dormancy in the Pacific sand lance (*Ammodytes hexapterus*) in relation to gut evacuation time. Pages 95-104 *in* Forage Fishes in Marine Ecosystems. University of Alaska Sea Grant College Program, Anchorage, Alaska, USA.

Clarke, K. R., and R. M. Warwick. 2001. Changes in marine communities: An approach to statistical analysis and interpretation, 2nd edition. Primer-E, Plymouth.

Clarke, W. R., R. A. Schmitz, and T. R. Bogenschutz. 1999. Site selection and nest success of ring-necked pheasants as a function of location in Iowa landscapes. Journal of Wildlife Management **63**:976-989.

Cohen, J. 1968. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin **70**:213-220.

Collyer, M. 2004a. Analysis of historical abalone density with oceanographic data for the Broken Group Islands. 5p429-04-020, Parks Canada.

Collyer, M. 2004b. Digitizing of coastal biodiversity data for Broken Group Islands. 5p437-04-0040, Parks Canada.

Collyer, M. 2004c. Marbled murrelet and sandlance ecology mapping and data management project March 2004 report. 5p437-04-0067, Parks Canada.

Corsi, F., J. de Leeuw, and A. Skidmore. 2000. Modeling species distribution with GIS. Pages 389-434 *in* L. Boitani and T. K. Fuller, editors. Research Techniques in Animal Ecology: Controversies and Consequences.

Corsi, F., E. Dupre, and L. Boitani. 1999. A large-scale model of wolf distribution in Italy for conservation planning. Conservation Biology **13**:150-159.

Cushman, S. A., and K. McGarigal. 2004. Patterns in the species-environment relationship depend on both scale and choice of response variables. Oikos **105**:117-124.

Danks, F. S., and D. R. Klein. 2002. Using GIS to predict potential wildlife habitat: a case study of muskoxen in northern Alaska. International Journal of Remote Sensing **23**:4611-4632.

De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology **81**:3178-3192.

Dean, W. E. J. 1974. Loss-on-ignition as an estimate of soil organic matter. Soil Science Society of America Journal **38**:150-151.

Dethier, M. N., E. S. Graham, C. S., and T. L.M. 1993. Visual versus random-point percent cover estimations: "Objective" is not always better. Marine Ecology Progress Series **96**:93-100.

Dettki, H., R. Löfstrand, and L. Edenius. 2003. Modeling habitat suitability for moose in coastal Northern Sweden: empirical *vs* process-oriented approaches. Ambio **32**:549-556.

Dettmers, R., D. A. Buehler, and J. B. Bartlett. 2002. A test and comparison of wildlife-habitat modeling techniques for predicting bird occurrence at a regional scale. Pages 607-615 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Dick, M. H., and I. M. Warner. 1982. Pacific sand lance, *Ammodytes hexapterus* Pallas, in the Kodiak Island group, Alaska. Syesis **15**:43-50.

Druehl, L. 2000. Pacific Seaweeds: A Guide To Common Seaweeds Of The West Coast. Harbour Publishing, Madeira Park.

Duarte, C. M., and J. Klaff. 1986. Littoral slope as a predictor of the maximum biomass of submerged macrophyte communities. Limnology and Oceanography **31**:1072-1080.

Dungan, J. L., J. N. Perry, M. R. T. Dale, P. Legendre, S. Citron-Pousty, M.-J. Fortin, A. Jakomulska, M. Miriti, and M. S. Rosenburg. 2002. A balanced view of scale in spatial statistical analysis. Ecography **25**:626-640.

Eastwood, P. D., G. J. Meaden, and A. Grioche. 2001. Modelling spatial variation in spawning habitat suitability for the sole *Solea solea* using regression quantiles and GIS procedures. Marine Ecology Progress Series **224**:251-266.

Ekebom, J., P. Laihonen, and T. Suominen. 2002. Measuring fetch and estimating wave exposure in coastal areas. Pages 155-160 *in* Littoral 2002: The Changing Coast. EUROCOAST/EUCC, Porto-Portugal.

Elith, J. 2000. Quantitative methods for modeling species habitat: Comparative performance and an application to Australian plants. *in* S. Ferson and M. A. Burgman, editors. Quantitative Methods in Conservation Biology. Springer, New York.

Ferreira, C. E. L., J. E. A. Gonççalves, and R. Coutinho. 2001. Community structure of fishes and habitat complexity on a tropical rocky shore. Environmental Biology of Fishes **61**:353-369.

Field, L. J. 1988. Pacific sand lance, *Ammodytes hexapterus,* with notes on related *Ammodytes* species. Pages 15-33 *in* N. J. Wilimovsky, L. S. Incze, and S. J. Westrheim, editors. Species synopses, life histories of selected fish and shellfish of the northeast Pacific and Bering Sea. Washington Sea Grant Program and Fisheries Research Institute, University of Washington, Seattle.

Fielding, A. H. 2002. What are the appropriate characteristics of an accuracy measure? Pages 271-280 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

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**:38-49.

Fleiss, J. L. 1973. Statistical methods for rates and proportions. Wiley, New York.

Folk, R. L. 1974. Pages 1-159 *in* Petrology of Sedimentary Rocks. Hemphill, Austin, TX.

Foreman, M. G. G., and R. E. Thomson. 1997. Three-dimensional model simulations of tides and buoyancy currents along the west coast of Vancouver Island. Journal of Physical Oceanography **27**:1300-1325.

Franco, A. M. A., J. C. Brito, and J. Almeida. 2000. Modelling habitat selection of common cranes *Grus grus* wintering in Portugal using multiple logistic regression. Ibis **142**:351-358.

Fretwell, S. D., and H. L. Lucas. 1970. On territorial behavior and other factors influencing habitat distribution in birds. Acta Biotheoretica **1970**:16-36.

Furness, R. W. 1999. Does harvesting a million metric tons of sand lance per year from the North Sea threatened seabird populations? Pages 407-424 *in* Ecosystem Approaches For Fisheries Management.

Garshelis, D. L. 2000. Delusions in habitat evaluation: Measuring use, selection, and importance. Pages 111-164 *in* L. Boitani and T. K. Fuller, editors. Research techniques in animal ecology, controversies and consequences. Columbia University Press, New York, New York, USA.

Gauld, J. A. 1990. Movements of lesser sandeels *Ammodytes marinus* Raitt tagged in the northwestern North Sea. Journal du Conseil, Conseil International Pour l'Exploration de la Mer **46**:229-231.

Gibson, R. N., L. Robb, M. T. Burrows, and A. D. Ansell. 1996. Tidal, diel and longer term changes in the distribution of fishes on a Scottish sandy beach. Marine Ecology Progress Series **130**:1-17.

Girsa, I. I., and A. N. Danilov. 1976. The defensive behavior of the white sea sand lance *Ammodytes hexapterus*. Journal of Ichthyology **16**:862-865.

Glorioso, P. D., and J. H. Simpson. 1994. Numerical modeling of the M2 tide on the northern Patagonian shelf. Continental Shelf Research **14**:267-278.

Guisan, A. 2002. Semi-quantitative response models for predicting the spatial distribution of plant species. Pages 315-326 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J.

B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: Offering more than simple habitat models. Ecology Letters **8**:993-1009.

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

Hahn, D. C., and R. J. O'Connor. 2002. Contrasting determinants of the abundance of an invasive species in its ancestral and colonized ranges. Pages 219-228 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Haldorson, L., M. Prichett, D. Sterritt, and J. Watts. 1993. Abundance patterns of marine fish larvae during spring in a southeastern Alaskan bay. Fishery Bulletin, U.S. **91**:36-44.

Hamada, T. 1966. Studies on fluctuation in the abundance of larval sand-lance in the Harima-nada and Osaka-Bay 1. Relation between the progeny-abundance and the age composition of parent fish. Nippon Suisan Gakkaishi **32**:393-398.

Hamada, T. 1967. Studies on fluctuation in the abundance of larval sand-lance in the Harima-nada and Osaka-Bay. 4. Relation between the number of eggs and the catch of 0-age fish. Nippon Suisan Gakkaishi **33**:410-416.

Hauck, W. W. J., and A. Donner. 1977. Wald's test as applied to hypotheses in logit analysis. Journal of the American Statistical Association **72**:851-853.

Heglund, P. J. 2002. Foundations of species-environment relations. Pages 35-41 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Heikkila, R. 1991. The influence of land-use on the sedimentation of the river Delta in the Kyronjoki Drainage-Basin. Hydrobiologia **214**:143-147.

Heiri, O., A. F. Lotter, and G. Lemcke. 2001. Loss on ignition as a method for estimating organic and carbonate content in sediments: Reproducibility and comparability of results. Journal of Paleolimnology **25**:101-110.

Henebry, G. M., and J. W. Merchant. 2002. Geospatial data in time: Limits and prospects for predicting species occurrences. Pages 291-302 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Hinsley, S. A., P. E. Bellamy, I. Newton, and T. H. Sparks. 1995. Habitat and landscape factors influencing the presence of individual breeding bird species in woodland fragments. Journal of Avian Biology **26**:94-104.

Hirzel, A. H., and A. Guisan. 2002. Which is the optimal sampling strategy for habitat suitability modeling? Ecological Modelling **157**:331-341.

Hirzel, A. H., V. Helfer, and F. Metral. 2001. Assessing habitat-suitability models with virtual species. Ecological Modelling **145**:111-121.

Hobbs, N. T., and T. A. Hanley. 1990. Habitat evaluation: Do use/availability data reflect carrying capacity? Journal of Wildlife Management **54**:515-522.

Hobson, E. S. 1986. Predation on the Pacific sand lance, *Ammodytes hexapterus* (pises: Ammodytidae), during the transition between day and night in Southeastern Alaska. Copeia **1**:223-226.

Høines, Å. S., and O. A. Bergstad. 2001. Density of wintering sand eel in the sand recorded by grab catches. Fisheries Research **49**:295-301.

Holland, G. J., S. P. R. Greenstreet, I. M. Gibb, H. M. Fraser, and M. R. Robertson. 2005. Identifying sandeel *Ammodytes marinus* sediment habitat preferences in the marine environment. Marine Ecology Progress Series **303**:269-282.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics **6**:65-70.

Huettmann, F., and A. W. Diamond. 2001. Seabird colony locations and environmental determination of seabird distribution: a spatially explicit breeding seabird model for the Northwest Atlantic. Ecological Modelling **141**:261-298.

Huston, M. A. 2002. Introductory essay: Critical issues for improving predictions. Pages 7-21 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Ingram, S. N., and E. Rogan. 2002. Identifying critical areas and habitat preferences of bottlenose dolphins *Tursiops truncatus*. Marine Ecology Progress Series **244**:247-255.

Inoue, A., S. Takamori, K. Kuniyuki, S. Kobayashi, and S. Nishina. 1967. Studies on the fishery biology of sand lance *Ammodytes personatus* Girard. Bulletin of the Naikai regional Fisheries Research Laboratory **25**:1-347.

James, F. C., and C. E. McCulloch. 2002. Predicting species presence and abundance. Pages 461-465 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Jensen, H., P. J. Wright, and P. Munk. 2003. Vertical distribution of pre-settled sandeel (*Ammodytes marinus*) in the North Sea in relation to size and environmental variables. ICES Journal of Marine Science **60**:1342-1351.

Johnson, C. M., L. B. Johnson, C. Richards, and V. Beasley. 2002. Predicting the occurrence of amphibians: An assessment of multiple-scale models. Pages 157-170 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Johnson, C. M., and W. B. Krohn. 2002. Dynamic patterns of association between environmental factors and island use by breeding seabirds. Pages 171-181 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Johnston, I. A., N. J. Cole, M. Abercromby, and V. L. V. Vieira. 1998. Embryonic temperature modulates muscle growth characteristics in larval and juvenile herring. Journal of Experimental Biology **201**:623-646.

Jones, M. T., G. J. Niemi, J. M. Hanowski, and R. R. Regal. 2002. Poisson regression: A better approach to modeling abundance data? Pages 411-418 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Karl, J. W., L. K. Scancara, P. J. Helglund, N. M. Wright, and M. J. Scott. 2002. Species commonness and accuracy of habitat-relationship models. Pages 573-580 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Keddy, P. A. 1982. Quantifying within-lake gradients of wave energy: interrelationships of wave energy, substrate particle size and shoreline plants in Axe Lake, Ontario. Aquatic Botany **14**:41-58.

Keddy, P. A. 1983. Shoreline vegetation in Axe Lake, Ontario - Effects of exposure on zonation patterns. Ecology **64**:331-344.

Keitt, T. H., O. N. Bjørnstad, P. M. Dixon, and S. Citron-Pousty. 2002. Accounting for spatial pattern when modeling organism-environment interactions. Ecography **25**:616-625.

Kishi, M. J., S. Kimura, H. Nakata, and Y. Yamashita. 1991. A biomass-based model for the sand lance (*Ammodytes personatus*) in Seto Inland Sea, Japan. Ecological Modelling **54**:247-263.

Knapp, R. A., K. R. Matthews, H. K. Preisler, and R. Jellison. 2003. Developing probabilistic models to predict amphibian site occupancy in a patchy landscape. Ecological Applications **13**:1069-1082.

Kobler, A., and M. Adamic. 2000. Identifying brown bear habitat by a combined GIS and machine learned method. Ecological Modelling **135**:291-300.

Kühlmann, D. H. H., and H. Karst. 1967. Freiwasserbeobachtungen zum Verhalten vos Tobias Fisch-schwarmen (Ammodytidae) in der westlichen Ostsee. Zeitschrift für Tierpsychologie **24**:282-297.

Kupschus, S. 2003. Development and evaluation of statistical habitat suitability models: An example based on juvenile spotted seatrout *Cynoscion nebulosus*. Marine Ecology Progress Series **265**:197-212.

Larsen, D. R., and P. L. Speckman. 2004. Multivariate regression trees for analysis of abundance data. Biometrics **60**:543-549.

Lauver, C. L., W. H. Busby, and J. L. Whistler. 2002. Testing a GIS model of habitat suitability for a declining grassland bird. Environmental Management **30**:88-97.

Leatherwick, J. R., and M. Austin. 2001. Competitive interactions between tree species in New Zealand old-growth indigenous forests. Ecology **82**:2560-2573.

Lee, and Bourne. 1982. Ecological (coastal) classification of the Pacific Rim National Park Reserve. Parks Canada.

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

Lim, T.-S., W.-Y. Loh, and Y.-S. Shih. 2000. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine Learning **40**:203-228

Linnell, J. D., and O. Strand. 2000. Interference interactions, co-existence and conservation of mammalian carnivores. Diversity and Distributions **6**:169-176.

Litzow, M. A., and J. F. Piatt. 2003. Variance in prey abundance influences time budgets of breeding seabirds: evidence from pigeon guillemots *Cepphus columba*. Journal of Avian Biology **34**:54-64.

Litzow, M. A., J. F. Piatt, A. A. Abookire, A. K. Prichard, and M. D. Robards. 2000. Monitoring temporal and spatial variability in sandeel (*Ammodytes hexapterus*) abundance with pigeon guillemot (*Cepphus columba*) diets. ICES Journal of Marine Science **57**:976-986.

Litzow, M. A., J. F. Piatt, A. K. Prichard, and D. D. Roby. 2002. Response of pigeon guillemots to variable abundance of high-lipid and low-lipid prey. Oecologia **132**:286-295.

Loh, W.-Y., and Y.-S. Shih. 1997. Split selection methods for classification trees. Statistica Sinica **7**:815-840.

Mackey, B. G., and D. B. Lindenmayer. 2001. Toward a hierarchical framework for modelling the spatial distribution of animals. Journal of Biogeography **28**:1147-1166.

Manel, S., J. M. Dias, S. T. Buckton, and S. J. Ormerod. 1999. Alternative methods for predicting species distribution: An illustration with Himalayan river birds. Journal of Applied Ecology **36**:734-747.

Maurer, B. A. 2002. Predicting distribution and abundance: thinking within and between scales. Pages 125-132 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler,

M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

McNemar, Q. 1947. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika **12**:153-157.

Meese, R. J., and P. A. Tomich. 1992. Dots on the rocks: a comparison of percent cover estimation methods. Journal of Experimental Marine Biology and Ecology **165**:59-73.

Moisen, G. G., and T. S. Frescino. 2002. Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling **157**:209-225.

Moles, A., and T. L. Wade. 2001. Parasitism and phagocytic function among sand lance *Ammodytes hexapterus* Pallas exposed to crude oil-laden sediments. Bulletin of Environmental Contamination and Toxicology **66**:528-535.

Morrison, M. L., and L. S. Hall. 2002. Standard terminology: Toward a common language to advance ecological understanding and application. *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

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

Munk, P. 2002. Larval sand lance (*Ammodytes sp.*) in the diet of small juvenile wolffish (*Anarhichas spp.*): predatory interactions in frontal water masses off western Greenland. Canadian Journal of Fisheries and Aquatic Science **59**:1759-1767.

Muñoz, J., and A. M. Felicísimo. 2004. Comparison of statistical methods commonly used in predictive modelling. Journal of Vegetation Science **15**:285-292.

Nelson, G. A., and M. R. Ross. 1991. Biology and population changes of northern sand lance (*Ammodytes dubius*) from the Gulf of Maine to the Middle Atlantic Bight. Journal of Northwest Atlantic Fishery Science **11**:11-27.

Nielsen, S. E., J. J. Johnson, D. C. Heard, and M. S. Boyce. 2005. Can models of presence-absence be used to scale abundance? Two case studies considering extremes in life history. Ecography **28**:197-208.

Ninio, R., S. Delean, K. Osborne, and H. Sweatman. 2003. Estimating cover of benthic organisms from underwater video images: variability associated with multiple observers. Marine Ecology Progress Series **265**:107-116.

Norcross, B. L., A. Blanchard, and B. A. Holladay. 1999. Comparison of models for defining nearshore flatfish nursery areas in Alaskan waters. Fisheries Oceanography **8**:50-67.

Norcross, B. L., F. Muter, and B. A. Holladay. 1997. Habitat models for juvenile pleuronectids around Kodiak Island, Alaska. Fisheries Bulletin **95**:504-520.

O'Connell, M., and J. M. Fives. 1995. The biology of the lesser sand-eel *Ammodytes tobianus* I. in the Galway Bay area. Biology and Environment **95b**:87-98.

O'Connor, R. J. 1986. Dynamical aspects of avian habitat use. Pages 235-240 *in* J. Verner, M. L. Morrison, and C. J. Ralph, editors. Wildlife 2000: Modeling habitat relationships of terrestrial vertebrates. University of Wisconsin Press, Madison.

O'Connor, R. J. 2002. The conceptual basis of species distribution modeling: Time for a paradigm shift? Pages 25-33 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Olden, J. D., and D. A. Jackson. 2002. A comparison of statistical approaches for modelling fish species distributions. Freshwater Biology **47**:1976-1995.

Ostrand, W. D., T. A. Gotthardt, S. Howlin, and M. D. Robards. 2005. Habitat selection models for Pacific sand lance (*Ammodytes hexapterus*) in Prince William Sound, Alaska. Northwestern Naturalist **86**:131-143.

Parasiewicz, P., and M. J. Dunbar. 2001. Phystical habitat modelling for fish - a developing approach. Archiv fur Hydrobiologie Supplement **135/2-4**:239-268.

Parris, K. M. 2002. The distribution and habitat requirements of the great barred frog (*Mixophyes fasciolatus*). Wildlife Research **29**:469-474.

Pearce, J., and S. Ferrier. 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. Biological Conservation **98**:33-43.

Pearson, W. H., D. L. Woodruff, P. C. Sugarman, and B. L. Olla. 1984. The burrowing behavior of sand lance, *Ammodytes hexapterus*: effects of oil-contaminated sediment. Marine Environmental Research **11**:17-32.

Penttila, D. E. 1997. Investigations of intertidal spawning habitats of surf smelt and Pacific sand lance in Puget Sound, Washington. Pages 395-407 *in* Forage Fishes in Marine Ecosystems. University of Alaska Sea Grant College Program, Anchorage, Alaska, USA.

Penttila, D. E. 2000. Intertidal spawning ecology of three species of marine forage fishes in Washington State. Journal of Shellfish Research **20**:1998.

Pinto, J. M., W. H. Pearson, and J. W. Anderson. 1984. Sediment preferences and oil contamination in the Pacific sand lance *Ammodytes hexapterus*. Marine Biology **83**:193-204.

Proctor, R., P. J. Wright, and A. Everitt. 1998. Modelling the transport of larval sandeels on the north-west European shelf. Fisheries Oceanography **7**:347-354.

Quinn, T. 1999. Habitat characteristics of an intertidal aggregation of Pacific Sandlance (*Ammodytes hexapterus*) at a North Puget Sound beach in Washington. Northwest Science **73**:44-49.

Quinn, T., and D. E. Schneider. 1991. Respiration of the teleost fish *Ammodytes hexapterus* in relation to its burrowing behavior. Comparative Biochemistry and Physiology **98A**:71-75.

Railsback, S. F., H. B. Stauffer, and B. C. Harvey. 2003. What can habitat preference models tell us? Tests using a virtual trout population. Ecological Applications **13**:1580-1594.

Reay, P. J. 1970. Synopsis of biological information on north Atlantic sand eels of the genus Ammodytes. *in* UN FAO Fisheries Synopsis No. 82, Rome, Italy.

Reckhow, K. H., W. B. Robert, B. S. Thomas, Jr., J. D. Vogt, and J. G. Wood. 1987. Empirical model of fish response to lake acidification. Canadian Journal of Fisheries and Aquatic Science **44**:1432-1442.

Rejwan, C., N. C. Collins, L. J. Brunner, B. J. Shuter, and M. S. Ridgway. 1999. Tree regression analysis on the nesting habitat of smallmouth bass. Ecology **80**:341-348.

Robards, M. D., J. A. Anthony, G. A. Rose, and J. F. Piatt. 1999a. Changes in proximate composition and somatic energy content for Pacific sand lance (*Ammodytes hexapterus*) from Kachemak Bay, Alaska relative to maturity and season. Journal of Experimental Marine Biology and Ecology **242**:245-258.

Robards, M. D., J. F. Piat, A. B. Kettle, and A. A. Abookire. 1999b. Temporal and geographic variation in fish communities of lower Cook Inlet, Alaska. Fisheries Bulletin **97**:962-977.

Robards, M. D., J. F. Piat, and G. A. Rose. 1999c. Maturation, fecundity, and intertidal spawning of Pacific sand lance in the northern Gulf of Alaska. Journal of Fish Biology **54**:1050-1068.

Robards, M. D., and J. F. Piatt. 1999. Biology of the genus *Ammodytes* - the sand lances. Pages 1-16 *in* M. D. Robards, M. F. Willson, R. H. Armstrong, and J. F. Piatt, editors. Sand lance: a review of biology and predator relations and annotated bibliography. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR, U.S.

Robards, M. D., J. F. Piatt, and G. Drew. 1999d. Distribution and abundance of small schooling fish in nearshore areas of Glacier Bay, Alaska during June, 1999. Preliminary Summary Alaska Biological Science Center, Anchorage.

Robards, M. D., G. A. Rose, and J. F. Piatt. 2002. Growth and abundance of Pacific sand lance, *Ammodytes hexapterus*, under differing oceanographic regimes. Environmental Biology of Fishes **64**:429-441.

Roff, J. C., and M. E. Taylor. 2000. National frameworks for marine conservation - a hierarchical geophysical approach. Aquatic Conservation: Marine and Freshwater Ecosystems **10**:209-223.

Rosenzweig, M. L., and R. H. MacArthur. 1963. Graphical representation and stability conditions of predator-prey interactions. American Naturalist **97**:209-223.

Rubec, P. J., S. G. Smith, M. S. Coyne, M. White, A. Sullivan, T. C. MacDonald, R. H. J. McMichael, and D. T. Wilder. 2001. Spatial modeling of fish habitat Suitability in Florida Estuaries. Pages 1-15 *in* Spatial Processes and Management of Marine Populations. Alaska Sea Grand College Program.

Salomon, A. K., J. L. Ruesink, B. X. Semmens, and R. T. Paine. 2001. Incorporating human and ecological communities in marine conservation: an alternative to Zacharias and Roff. Conservation Biology **15**:1452-1455.

Scott, M. J., P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson. 2002. Introduction. Pages 1-5 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting Species Occurrencs: Issues of Accuracy and Scale. Island Press, Washington, D.C.

Shriner, S. A., T. R. Simons, and G. L. Farnsorth. 2002. A GIS-based habitat model for wood thrush, *Hylocichla mustelina*, in Great Smoky Mountains National Park. Pages 529-535 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Simpson, J. H., and J. R. Hunter. 1974. Fronts in Irish Sea. Nature **250**:404-406.

Springer, A. M., and S. G. Speckman. 1997. A forage fish is what? Summary of the symposium. Pages 773-806 *in* Forage Fishes in Marine Ecosystems. University of Alaska Sea Grant College Program, Anchorage, Alaska, USA.

StatSoft, Inc. (2004). STATISTICA (data analysis software system), version 7. www.statsoft.com.

Stauffer, D. F. 2002. Linking populations and habitats: Where have we been? Where are we going? Pages 53-61 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Stauffer, H. B., C. J. Ralph, and S. L. Miller. 2002. Incorporating detection uncertainty into presence-absence surveys for marbled murrelet. Pages 357-365 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Steele, J. H. 1998. Regime shifts in marine ecosystems. Ecological Applications **8**:S33-S36.

Store, R., and J. Jokimäki. 2003. A GIS-based multi-scale approach to habitat suitability modeling. Ecological Modelling **169**:1-15.

Suryan, R. M., D. B. Irons, M. Kaufman, J. Benson, P. G. R. Jodice, D. D. Roby, and E. D. Brown. 2002. Short-term fluctuations in forage fish availability and the effect on prey selection and brood-rearing in the black-legged kittiwake *Rissa tridactyla*. Marine Ecology Progress Series **236**:273-287.

Syms, C., and G. P. Jones. 2004. Habitat structure, disturbance and the composition of sand-dwelling goby assemblages in a coral reef lagoon. Marine Ecology Progress Series **268**:221-230.

Thomas, K., T. Keeler-Wolf, and J. Franklin. 2002. A comparison of fine- and coarse-resolution environmental variables towards predicting vegetation distribution in the Mojave Desert. Pages 133-139 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Thorman, S. 1986. Physical factors affecting the abundance and species richness of fishes in shallow waters of the southern Bothnian Sea, Sweden. Estuarine, Coastal and Shelf Science **22**:357-370.

Tunnicliffe, V. 2000. A fine-scale record of 130 years of organic carbon deposition in an anoxic fjord, Sannich Inlet, British Columbia. Limnology and Oceanography **45**:1380-1387.

Tyre, A. J., H. P. Possingham, and D. B. Lindenmayer. 2001. Inferring process from pattern: Can territory occupancy provide information about life history parameters? Ecological Application **11**:1722-1737.

Van Horne, B. 1983. Density as a misleading indicator of habitat quality. Journal of Wildlife Management **47**:893-901.

Van Horne, B. 2002. Approaches to habitat modeling: The tensions between pattern and process and between specificity and generality. *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Webster, M. S. 2003. Temporal density dependence and population regulation in a marine fish. Ecology **84**:623-628.

Welch, E., and J. A. MacMahon. 2005. Identifying habitat variables important to the rare Columbian spotted frog in Utah (U.S.A.): An information-theoretical approach. Conservation Biology **19**:473-481.

Wentworth, C. K. 1922. A scale of grade and class terms for clastic sediments. Journal of Geology **30**:377-392.

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology **3**:385-397.

Wiens, J. A. 1996. Wildlife in patchy environments: Metapopulations, mosaics, and management. *in* D. R. McCullough, editor. Metapopulations and wildlife conservation. Island Press, Washington D.C.

Willson, M. F., R. H. Armstrong, M. D. Robards, and J. F. Piatt. 1999. Sand lance as cornerstone prey for predator populations. Pages 17-38 *in* M. D. Robards, M. F. Willson, R. H. Armstrong, and J. F. Piatt, editors. Sand lance: a review of biology and predator relations and annotated bibliography. United States Department of Agriculture, Portland, OR, U.S.

Winslade, P. R. 1974a. Behavioral studies on the lesser sandeel, *Ammodytes marinus* (Raitt) I.The effect of food availability on activity and the role of olfaction in food detection. Journal of Fish Biology **6**:565-576.

Winslade, P. R. 1974b. Behavioral studies on the lesser sandeel, *Ammodytes marinus* (Raitt) II: The effect of light intensity on activity. Journal of Fish Biology **6**:577-586.

Wright, P. J. 1996. Is there a conflict between sandeel fisheries and seabirds? A case study at Shetland. Pages 154-165 *in* S. P. R. Greenstreet and M. L. Tasker, editors. Aquatic predators and their prey. Fishing News Books, Blackwell Science, Oxford.

Wright, P. J., and G. S. Begg. 1997. A spatial comparison of common guillemots and sandeels in Scottish waters. ICES Journal of Marine Science **54**:578-592.

Wright, P. J., H. Jensen, and I. Tuck. 2000. The influence of sediment type on the distribution of the lesser sandeel, *Ammodytes marinus*. Journal of Sea Research **44**:243-256.

Yen, P. P. W., F. Huettmann, and F. Cooke. 2004. A large-scale model for the at-sea distribution and abundance of Marbled Murrelets (*Brachyramphus marmoratus*) during the breeding season in coastal British Columbia, Canada. Ecological Modelling **171**:395-413.

Young, J. S., and R. L. Hutto. 2002. Use of regional-scale exploratory studies to determine bird-habitat relationships. Pages 107-122 *in* M. J. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: Issues of accuracy and scale. Island Press, Washington, D.C.

Yuasa, I., and H. Ueshima. 1992. A tidal front in winter influenced by river discharge. Journal of Oceanography **48**:239-255.

Zacharias, M. A., M. C. Morris, and D. E. Howes. 1999. Large scale characterization of intertidal communities using a predictive model. Journal of Experimental Marine Biology and Ecology **239**:223-242.

Zacharias, M. A., and J. C. Roff. 2000. A hierarchical ecological approach to conserving marine biodiversity. Conservation Biology **14**:1327-1334.

Zamon, J. E. 2003. Mixed species aggregations feeding upon herring and sandlance schools in a nearshore archipelago depend on flooding tidal currents. Marine Ecology Progress Series **261**:243-255.

APPENDICIES

Appendix 2.1: Description of nearshore sub-regions.

		Sampling	
Nearshore Sub-Region	Range	Environmental	Sand lance
Shallow Subtidal	3m Depth Contour - Low Tide Line	Underwater Video Data Macro-vegetation Cover	Low Tide Beach Seine Visual Survey Methods
Low Intertidal	Low Tide Line - Medium Tide Line	Quantitative Intertidal Sediment Properties Qualitative Sediment Classification Macro-vegetation Cover	Medium Tide Beach Seine Intertidal Digging
High Intertidal	Medium Tide Line - High Tide Line	Qualitative Sediment Classification Beach Log Cover	High Tide Beach Seine

Appendix 3.1: Sampling method detection frequency in 2003. "1" indicates sand lance presence according to the method while "0" indicates sand lance absence. The "normative standard" designates sites where sand lance were found present by at least one method (1) or by none of the methods (0)

Site	Low Tide Beach Seine	Medium Tide Beach Seine	High Tide Beach Seine	Visual Survey Boat - Parallel	Visual Survey Boat - Perpendicular	Visual Survey Walk	Digging Survey	Snorkeling Survey	Normative Standard
BENSON E	1	na	na	1	1	0	1	na	1
BENSONCH	1	na	na	1	0	0	1	na	1
BRABANT 55	0	1	na	0	0	1	0	na	1
CAMBLAIN	1	na	na	na	1	1	0	1	1
CHALK	1	0	na	na	na	0	0	na	1
CHALK2	1	0	na	1	1	1	1	na	1
CLARK WB	0	0	0	0	1	na	0	na	1
CLARKE CAMP	1	na	na	1	1	na	1	1	1
COOPER W	0	0	0	0	0	0	1	0	1
DEMPSTERNE	1	1	na	na	1	1	1	na	1
DEMPSTERSE	1	0	na	0	0	0	1	1	1
DIANA E	1	0	0	1	1	1	1	1	1
DIANA NE	0	0	0	0	0	0	0	1	1
DICEBOX	1	1	0	1	0	0	0	1	1
DODD BAY	1	0	0	0	0	0	0	1	1
DODD NORTH	1	na	na	0	0	0	0	na	1
DODD W	1	1	1	0	0	1	0	1	1
EFF LIGHT	1	na	na	0	0	1	0	na	1
FABER NE	0	0	na	0	0	0	0	1	1
GEER W	1	0	na	na	1	0	0	na	1
GIBRA-CAMP	1	1	1	0	0	0	0	1	1
GILBERT CMP	0	1	na	0	0	0	1	1	1
HAND NORTH	0	0	0	1	0	0	0	0	1
HELBY	1	1	0	1	0	0	0	1	1
NETL-1	1	0	0	1	0	1	0	1	1
NETTLE 2	1	1	0	1	1	0	1	1	1
OHIAT	1	1	0	0	0	1	1	na	1
ROSS ISLETS	0	1	0	0	0	0	1	na	1
SANFORD NW	1	1	1	1	na	0	na	1	1
SCOTTS BAY	0	na	na	0	0	1	0	na	1
STUDISLETS	1	1	0	0	0	1	1	0	1
TINY GRP1	0	0	na	1	1	0	0	1	1
TRICKETT1	1	1	na	1	1	1	1	1	1
TZAR ROBB	1	0	na	0	0	0	0	na	1
WOUWER TP	0	1	na	1	0	0	0	0	1
WOUWERBATELY	1	0	0	0	0	0	1	na	1

Appendix 3.1 continued

Site	Low Tide Beach Seine	Medium Tide Beach Seine	High Tide Beach Seine	Visual Survey Boat - Parallel	Visual Survey Boat - Perpendicular	Visual Survey Walk	Digging Survey	Snorkeling Survey	Normative Standard
BRAB 53	0	0	na	0	0	0	0	0	0
BRADYS	0	0	0	0	0	0	0	0	0
CHALK3	0	0	na	0	0	0	0	0	0
DODDCAMP	0	0	0	0	na	na	0	na	0
EFFBAY-1	0	0	na	0	0	0	na	0	0
EFFI-GILB	0	0	na	0	0	0	0	na	0
EFFINGHAME	0	na	na	0	0	0	0	0	0
EFFINGHAMS	0	0	na	0	0	0	0	0	0
HAINES E	0	0	0	0	0	0	0	0	0
HANDCAMPN	0	0	0	0	0	0	0	na	0
JAQUE-JAR	0	na	na	na	na	na	0	na	0
PINK 48	0	na	na	na	na	na	na	na	0
PINKERTON1	0	na	na	0	0	na	0	na	0
REEKS 46	0	0	0	0	0	0	0	0	0
ROBBERPASS	0	0	0	0	0	0	0	0	0
TURRET E	0	na	na	0	0	0	0	0	0
TURRET NE	0	0	0	0	0	0	0	0	0
TURRET-EELGRASS	0	na	na	na	na	na	0	na	0
TURTJAQU	0	0	na	na	na	na	0	na	0
TURTLE E	0	0	0	0	0	0	0	0	0
WALSH N	0	na	na	0	0	0	0	na	0
WILLISCAMP	0	0	0	0	0	0	0	0	0
WIZARD	0	0	0	0	0	0	0	0	0
WOUWERNW	0	0	0	0	0	0	0	0	0

Appendix 5.1: Explanatory variables considered in analysis.

Variable	Type	Description
Intertidal Eelgrass Presence/Absence	Categorical - binary	A binary classification of eelgrass presence/absence in the intertidal. Any eelgrass in the intertidal region (as defined in chapter 2) was classified as "present".
Subtidal Eelgrass Presence/Absence	Categorical - binary	A binary classification of eelgrass presence/absence in the subtidal. Any eelgrass in the subtidal region (as defined in chapter 2) was classified as "present".
Major Substrate High Intertidal	Categorical - ordinal	For each site the major substrate type of the high intertidal was classified according to the code in Table MM3 (chapter 2). The categories for "Pebble" and "Cobble/Boulders/Bedrock" were later combined into one category.
Major Substrate Low Intertidal	Categorical - ordinal	For each site the major substrate type of the high intertidal was classified according to the code in Table MM3 (chapter 2). The categories for "Pebble" and "Cobble/Boulders/Bedrock" were later combined into one category.
Beach Log Cover	Categorical - ordinal	Beach log cover at high tide line was estimated by counting the number of beach logs bisected by an imaginary cross section of the high tide line (perpendicular to the beach). Categories used are shown in Table MM6 (chapter 2).
Aspect	Categorical - nominal	The aspect of the sample was recorded in degrees from 1-360° and then coded into 4 categories: Northeast (1-90°), Southeast (91-180°), Southwest (181-270°), Northwest (271-360°).
Slope Intertidal (°)	Continuous	Slope measured with an abny level from the low tide line to the high tide line.

Appendix 5.1 continued

Variable	Type	Description
Slope Subtidal (°)	Continuous	Slope measured with an abny level from the shallow subtidal to the low tide line.
Bulk Substrate Mass (kg)	Continuous	At each site a bulk sample of sediment was taken by plunging a cylindrical container (L=40cm, radius=10cm) into the sediment and then weighing the resultant sample.
Grain Size Mean	Continuous	Grain size mean was determined through sieving sediment into 20 size classes (ranging from >25000 microns to <38 microns). Linear interpolation is used to calculate statistical parameters by the Folk and Ward (1957) graphical method.
Grain Size Sorting	Continuous	Grain size sorting was determined through sieving sediment into 20 size classes (ranging from >25000microns to <38 microns). Linear interpolation is used to calculate statistical parameters by the Folk and Ward (1957) graphical method.
Grain Size Skewness	Continuous	Grain size skewness was determined through sieving sediment into 20 size classes (ranging from >25000 microns to <38 microns). Linear interpolation was used to calculate statistical parameters by the Folk and Ward (1957) graphical method.
Grain Size Kurtosis	Continuous	Grain size kurtosis was determined through sieving sediment into 20 size classes (ranging from >25000 microns to <38 microns). Linear interpolation was used to calculate statistical parameters by the Folk and Ward (1957) graphical method.

Appendix 5.1 continued

Variable	**Type**	**Description**
Intertidal Macro-vegetation Cover	Continuous - percent	Estimated percent cover of the major macro-vegetation type (defined as the macro-vegetation type that had the highest percent cover) in the low intertidal zone.
Subtidal Macro-vegetation Cover	Continuous - percent	Estimated percent cover of the major macro-vegetation type (defined as the macro-vegetation type that had the highest percent cover) in the shallow subtidal zone.
% Gravel	Continuous - percent	Proportion of gravel (grain size distribution determined through dry sieving) in the sediment. Gravel is defined as grain size class 2-64 mm.
% Mud	Continuous - percent	Proportion of mud (grain size distribution determined through dry sieving) in the sediment. Mud is defined as grain size class <63 microns.
% Sand	Continuous - percent	Proportion of sand (grain size distribution determined through dry sieving) in the sediment. Sand is defined as grain size class 63-2000 microns.
Loss On Ignition 550	Continuous - percent	Organic content of the sediment determined by the mass loss on ignition in a muffle furnace at 550°C for 4 hours. This was expressed as a percent of total original mass.
Loss On Ignition 950	Continuous - percent	Carbonate content of the sediment determined by the mass loss on ignition in a muffle furnace at 950°C for 2 hours. This was expressed as a percent of total original mass.
Intertidal Eelgrass % Cover	Continuous - percent	Estimated percent cover of eelgrass in the low intertidal zone.

Appendix 5.1 continued

Variable	Type	Description
Subtidal Eelgrass % Cover	Continuous - percent	Estimated percent cover of eelgrass in the shallow subtidal zone.
UV Mud %	Continuous - percent	The estimated mean percent mud (defined according to Table MM3, chapter 2) content in the sediment sampled using underwater video (UV).
UV Fine Sand %	Continuous - percent	The estimated mean percent fine sand (defined according to Table MM3, chapter 2) content in the sediment sampled using underwater video (UV).
UV Coarse Sand %	Continuous - percent	The estimated mean percent coarse sand (defined according to Table MM3, chapter 2) content in the sediment sampled using underwater video (UV).
UV Pebble/Cobble %	Continuous - percent	The estimated mean percent pebble/cobble (defined according to Table MM3, chapter 2) content in the sediment sampled using underwater video (UV).
UV Boulder/Bedrock %	Continuous - percent	The estimated mean percent boulder/bedrock (defined according to Table MM3, chapter 2) content in the sediment sampled using underwater video (UV).
UV Possible Burying %	Continuous - percent	The estimated mean percent of subtidal sediment that is possibly burying habitat for sand lance (UV).
UV Marcovegetation %	Continuous - percent	The estimated mean percent subtidal macro-vegetation cover sampled using underwater video (UV).
UV Shell %	Continuous - percent	The estimated mean percent shell content in the sediment sampled using underwater video (UV).

Appendix 5.2: Values of independent variables and dependent variables by site. Sand lance Abundance was ln transformed.

	Selected Independent Variables								Dependent Variables				
Site	*Slope Intertidal (°)*	*Grain Size Mean*	*Grain Size Sorting*	*UV Possible Burying %*	*UV Shell %*	*UV Mud %*	*Major Substrate Low Intertidal*	*Intertidal Eelgrass Presence/Absence*	*Sand lance Abundance*	*Sand lance Presence/Absence*	*Sand lance High/Low*	*Sand lance Burying/Not burying*	*Sand lance High/Medium/Low/Nil*
BENSON E	4.5	3203.8	2.51	0.74	0.38	0.05	3	0	8.01	1	1	1	2
BENSONCH	5.0	3586.6	1.87	1.22	0.42	0	3	0	8.16	1	1	1	2
BRAB 53	2.5	447.3	2.53	1.17	0.18	0.01	1	0	0	0	0	0	0
BRABANT 55	3.5	874.3	2.2	1.12	0.34	0.03	2	0	0	0	0	0	0
BRADYS	4.0	368.5	4.3	1.57	0.16	0	2	0	0	0	0	0	0
CAMBLAIN	2.0	1639.5	4.1	1.57	0.43	0.02	1	0	1.39	1	0	1	1
CHALK	3.5	2343.3	1.88	0.5	0.24	0.03	2	0	4.26	1	1	0	2
CHALK2	2.0	2381	2.11	1.25	0.88	0.03	4	0	4.87	1	1	1	2
CHALK3	2.5	1591	3.2	0.43	0.69	0.06	1	0	0	0	0	0	0
CLARK WB	1.0	120.2	1.5	0.52	0.32	0.04	2	1	0	0	0	0	0
CLARKE CAMP	4.5	985.7	1.67	0.85	0.35	0.02	2	0	8.52	1	1	1	2
COOPER W	3.5	1992.2	1.71	0.98	0.35	0	2	0	0	0	0	1	0
DEMPSTERNE	2.5	780.4	1.77	1.11	0.63	0.03	2	0	2.3	1	0	1	1
DEMPSTERSE	3.0	2268.1	1.9	0.89	0.27	0.01	2	0	4.08	1	1	1	2
DIANA E	4.0	2309.8	2.51	1.15	0.25	0	3	0	9.43	1	1	1	2
DIANA NE	4.5	782.7	4.99	1.04	0.23	0.03	2	0	0	0	0	0	0
DICEBOX	6.5	3589.4	2.47	1.01	0.35	0	3	0	6.69	1	1	0	2
DODD BAY	2.0	796.8	2.48	0.21	0.69	0.09	2	0	1.79	1	0	0	1

Appendix 5.2 continued

Site	*Slope Intertidal (°)*	*Grain Size Mean*	*Grain Size Sorting*	*UV Possible Burying %*	*UV Shell %*	*UV Mud %*	*Major Substrate Low Intertidal*	*Intertidal Eelgrass Presence/Absence*	*Sand lance Abundance*	*Sand lance Presence/Absence*	*Sand lance High/Low*	*Sand lance Burying/Not burying*	*Sand lance High/Medium/Low/Nil*
DODD NORTH	3.0	188.5	1.33	1.21	0.29	0	4	0	3.33	1	1	0	2
DODD W	4.5	785.9	1.73	0.75	0.64	0.01	2	0	8.01	1	1	1	2
DODDCAMP	4.0	1033.3	2.61	0	0.13	0.06	1	1	0	0	0	0	0
EFF LIGHT	2.5	1580.7	2.19	0.56	0.22	0	2	0	3.83	1	1	1	2
EFFBAY-1	2.5	776.2	3.6	0.33	0.34	0.09	1	1	0	0	0	0	0
EFFI-GILB	5.0	1067.3	2.87	0.5	0.3	0.07	4	0	0	0	0	0	0
EFFINGHAME	1.0	2792.2	3.05	0.32	0.32	0.09	1	0	0	0	0	0	0
EFFINGHAMS	1.0	555.4	3.24	0.5	0.33	0.07	1	1	0	0	0	0	0
FABER NE	7.0	4753.8	2.7	0.5	0.45	0	3	0	0.69	1	0	0	1
GEER W	3.0	4100.6	3.55	0.97	0.43	0.01	3	0	6.4	1	1	0	2
GIBRA-CAMP	6.5	2840.6	3.44	0.5	0.28	0.04	3	0	7.82	1	1	0	2
GILBERT CMP	4.0	2281.7	2.8	1.11	0.32	0.03	2	0	0	0	0	1	0
HAINES E	5.5	327.6	2.31	0.44	0.1	0.07	1	1	0	0	0	0	0
HAND NORTH	4.5	1229	5.7	0.75	0.29	0.04	1	1	0	0	0	0	0
HANDCAMPN	3.0	1634.1	4.35	0	0.15	0.1	1	1	0	0	0	0	0
HELBY	5.0	4186.9	5.66	1.3	0.13	0	3	0	1.79	1	0	1	1
NETL-1	3.5	5907.7	2.47	0.23	0.12	0.08	3	0	1.61	1	0	1	1
NETTLE 2	4.5	2321.1	3.56	0.41	0.22	0.09	2	0	2.3	1	0	1	1
OHIAT	3.0	2160.1	2.71	1.06	0.47	0.01	2	0	3.56	1	1	1	2
PINKERTON1	1.0	1537.7	2.53	0	0	0.1	1	1	0	0	0	0	0

Appendix 5.2 continued

Site	*Slope Intertidal (°)*	*Grain Size Mean*	*Grain Size Sorting*	*UV Possible Burying %*	*UV Shell %*	*UV Mud %*	*Major Substrate Low Intertidal*	*Intertidal Eelgrass Presence/Absence*	*Sand lance Abundance*	*Sand lance Presence/Absence*	*Sand lance High/Low*	*Sand lance Burying/Not burying*	*Sand lance High/Medium/Low/Nil*
REEKS	3.0	1684.5	6.31	0.44	0.19	0.05	1	0	0	0	0	0	0
ROBBERPASS	4.5	2706.4	5.63	0.44	0.33	0.06	1	1	0	0	0	0	0
ROSS ISLETS	3.0	1930.6	2.25	1	0.38	0.01	2	1	0	0	0	1	0
SANFORD NW	6.0	3735.7	1.9	0.74	0.38	0.05	3	0	7.31	1	1	1	2
STUDISLETS	4.0	2156.8	5.09	0.79	0.53	0	2	0	1.61	1	0	1	1
TINY GRP1	2.0	1373.8	1.93	0.75	0.38	0.03	2	1	0	0	0	0	0
TRICKETT1	4.0	1973.9	2.07	0.36	0.64	0	2	0	9.21	1	1	1	2
TURRET E	3.5	1990.5	3.81	1.18	0.42	0	1	0	0	0	0	0	0
TURRET NE	3.0	808.4	4.32	0.3	0.17	0.05	2	0	0	0	0	0	0
TURTLE E	2.0	1588.3	6.41	0	0.22	0.08	1	0	0	0	0	1	0
TZAR ROBB	3.5	184.8	1.34	1.1	0.28	0.02	4	0	5.99	1	1	0	2
WALSH N	4.0	1167.4	3.46	0.67	0.46	0.03	1	0	0	0	0	0	0
WILLISCAMP	3.0	893.7	3.29	0.48	0.23	0.04	1	0	0	0	0	0	0
WIZARD	4.0	2252.5	3.26	0.74	0.38	0.05	3	1	0	0	0	0	0
WOUWER TP	3.0	2512.4	2.31	0.84	0.92	0	3	0	0	0	0	0	0
WOUWERBATELY	4.5	3414	2.58	0.93	0.37	0.01	3	0	0.69	1	0	1	1
WOUWERNW	2.0	1335	3.16	0.57	0.2	0.05	1	0	0	0	0	0	0

Appendix 5.3: Summary of actual classification and model classifications of sand lance presence (1) and absence (0). "Actual" represents the the presence/absence designation based on beach seine sampling. Note that "Actual" 2003 was used for all model construction.

	Actual			Logistic Regression						Classification Tree	
Site	2002	2003	2004	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Misclassification	Deviance
BENSON E	0	1	0	1	1	1	1	1	1	1	1
BENSONCH	0	1	0	1	1	1	1	1	1	1	1
BRAB 53	0	0	0	0	0	0	0	0	0	0	0
BRABANT 55	0	0	0	1	1	1	0	1	1	1	1
BRADYS	0	0	0	0	1	0	0	0	0	1	0
CAMBLAIN	NA	1	0	0	0	0	0	0	0	0	0
CHALK	1	1	1	1	1	1	1	1	1	1	1
CHALK2	NA	1	0	1	1	1	1	1	1	1	1
CHALK3	NA	0	1	0	0	0	0	0	0	0	0
CLARK WB	0	0	0	0	0	0	0	NA	NA	0	0
CLARKE CAMP	1	1	1	1	1	1	1	1	1	1	1
COOPER W	NA	0	0	1	1	1	1	1	1	1	1
DEMPSTERNE	0	1	1	1	1	1	0	1	1	1	1
DEMPSTERSE	1	1	1	1	1	1	1	1	1	1	1
DIANA E	0	1	1	1	1	1	1	1	1	1	1
DIANA NE	NA	0	0	0	1	0	0	0	0	1	0
DICEBOX	1	1	0	1	1	1	1	1	1	1	1
DODD BAY	1	1	1	1	1	0	0	1	0	1	1
DODD NORTH	1	1	0	1	1	1	0	1	1	1	1
DODD W	1	1	0	1	1	1	0	1	1	1	1
DODDCAMP	NA	0	0	0	0	0	0	NA	NA	0	0
EFF LIGHT	1	1	1	1	1	1	1	1	1	1	1
EFFBAY-1	NA	0	0	0	0	0	0	NA	NA	0	0
EFFI-GILB	NA	0	0	1	1	0	0	1	0	1	0
EFFINGHAME	NA	0	0	0	0	1	1	0	1	0	0
EFFINGHAMS	NA	0	0	0	0	0	0	NA	NA	0	0
FABER NE	0	1	0	1	1	1	1	1	1	1	1
GEER W	NA	1	0	1	1	1	1	1	1	1	1
GIBRA-CAMP	NA	1	0	1	1	1	1	1	1	1	0

Appendix 5.3 continued

Site	2002	2003	2004	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Misclassification	Deviance
GILBERT CMP	NA	0	0	1	1	1	1	1	1	1	0
HAINES E	1	0	0	0	0	0	0	NA	NA	0	0
HAND NORTH	0	0	0	0	0	0	0	NA	NA	0	0
HANDCAMPN	0	0	0	0	0	0	0	NA	NA	0	0
HELBY	1	1	0	1	1	1	1	1	1	1	1
NETL-1	0	1	1	1	1	1	1	1	1	1	1
NETTLE 2	0	1	1	1	1	1	0	1	1	1	1
OHIAT	0	1	0	1	1	1	1	1	1	1	1
PINKERTON1	NA	0	0	0	0	0	0	NA	NA	0	0
REEKS	0	0	0	0	0	0	0	0	0	0	0
ROBBERPASS	NA	0	0	0	0	0	0	NA	NA	0	0
ROSS ISLETS	0	0	0	0	0	0	1	NA	NA	0	0
SANFORD NW	1	1	1	1	1	1	1	1	1	1	1
STUDISLETS	NA	1	0	1	1	0	0	1	0	1	1
TINY GRP1	NA	0	1	0	0	0	1	NA	NA	0	0
TRICKETT1	1	1	1	1	1	1	1	1	1	1	1
TURRET E	NA	0	0	0	0	0	0	0	0	0	0
TURRET NE	NA	0	0	0	1	0	0	0	0	1	0
TURTLE E	NA	0	0	0	0	0	0	0	0	0	0
TZAR ROBB	0	1	1	1	1	1	0	1	1	1	1
WALSH N	NA	0	0	0	0	0	0	0	0	0	0
WILLISCAMP	0	0	0	0	0	0	0	0	0	0	0
WIZARD	0	0	0	0	0	0	1	NA	NA	0	0
WOUWER TP	NA	0	0	1	1	1	1	1	1	1	1
WOUWERBATELY	0	1	0	1	1	1	1	1	1	1	1
WOUWERNW	NA	0	0	0	0	0	0	0	0	0	0