PREDICTING DISTRIBUTIONS OF ESTUARINE ASSOCIATED FISH AND INVERTEBRATES IN SOUTHEAST ALASKA

By

Katharine Bollinger Miller

RECOMMENDED:

Advisory Committee Chair

Department Head

APPROVED:

Dean, School of Fisheries and Ocean Sciences

Dean of the Graduate School

29 April 2013

Date

PREDICTING DISTRIBUTIONS OF ESTUARINE ASSOCIATED FISH AND

INVERTEBRATES IN SOUTHEAST ALASKA

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSPHY

By

Katharine Bollinger Miller, M.M.A

Fairbanks, Alaska

May 2013

Abstract

Estuaries in Southeast Alaska provide habitat for juveniles and adults of several commercial fish and invertebrate species; however, because of the area's size and challenging environment, very little is known about the spatial structure and distribution of estuarine species in relation to the biotic and abiotic environment. This study uses advanced machine learning algorithms (random forest and multivariate random forest) and landscape and seascape-scale environmental variables to develop predictive models of species occurrence and community composition within Southeast Alaskanestuaries. Species data were obtained from trawl and seine sampling in 49 estuaries throughout the study area. Environmental data were compiled and extracted from existing spatial datasets. Individual models for species occurrence were validated using independent data from seine surveys in 88 estuaries. Prediction accuracy for individual species models ranged from 94% to 63%, with 76% of the fish species models and 72% of the invertebrate models having a predictive accuracy of 70% or better. The models elucidated complex species-habitat relationships that can be used to identify habitat protection priorities and to guide future research. The multivariate models demonstrated that community composition was strongly related to regional patterns of precipitation and tidal energy, as well as to local abundance of intertidal habitat and vegetation. The models provide insight into how changes in species abundance are influenced by both environmental variation and the co-occurrence of other species. Taxonomic diversity in the region was high (74%) and functional diversity was relatively low (23%). Functional diversity was not linearly correlated to species richness, indicating that the number of species in the estuary was not a good predictor of functional diversity or redundancy. Functional redundancy differed across estuary clusters, suggesting that some estuaries have a greater potential for loss of functional diversity with species removal than others.

Dedication

This dissertation is dedicated to my husband who is my biggest fan, my strongest advocate, and my best friend.

Table of Contents

Table of Figures

viii

Table of Tables

Acknowledgements

Chapter 1: General Introduction

Understanding the mechanisms that determine species' distributions and the relationship between species composition and ecosystem functions are essential goals of ecology and ecosystem management (de Bello et al. 2009, Maxwell et al. 2009, Sundblad et al. 2009). Habitat includes both biotic and abiotic elements of the environment in which a species lives. Classical niche theory describes a species' habitat as a multidimensional hypervolume with each axis represented by an environmental variable (either biotic or abiotic) that is important for the species' survival and reproduction (Hutchinson 1957, Cushman and Huettmann 2010, Drew et al. 2011). The niche can be dynamic, defined by the physical environment and physiological tolerances of the species and constrained by interspecific competition (Polechová and Storch 2008). Co-occurrence of species has been attributed to deterministic processes, such as shared environmental tolerances and niche overlap, to stochastic processes like dispersal limitation and colonization/extinction dynamics, and combinations of these processes (Chase 2010, Weins 2011, Fiser et al. 2012).

Species distribution models attempt to find a statistical correlation between the biotic and abiotic environment and species' occurrences or abundances and to extrapolate species distribution patterns to unsampled areas (Austin 2002, Ferrier and Guisan 2006, Valavanis et al. 2008). Species distribution models are also called niche models because they describe the realized niche of a species or community (Elith and Leathwick 2009, Williams et al. 2009, Oppel and Huettmann 2010, Hardy et al. 2011). Although correlative models cannot determine the mechanisms by which species are distributed along environmental gradients, they can identify relationships to be investigated through theoretical analysis or additional research (Austin 2002, Elith et al. 2006, Magness et al. 2010). The more accurately a model can predict a species' spatial distribution, the less likely it is that relationships between species occurrence or abundance and habitat variables of importance in the model are spurious (Burnham and Anderson 1998, De'ath 2002).

Traditional approaches to the development of species-habitat models are based on an assumption that the data used in the model conform to a known stochastic distribution (Breiman 2001, Prasad et al. 2006, Magness et al. 2010). The goal of modeling then becomes one of hypothesizing plausible relationships between the predictor variables and the response and evaluating how well the data fit these hypothesized relationships (Barry and Welsh 2002). This

approach has several limitations, many of which are widely recognized by ecologists and modelers. Models that assume a linear relationship between predictor variables and the response variable very rarely conform to true species-habitat relationships where lags, thresholds and feedbacks are common (Barbier et al. 2008). As a result, models that provide good fits to the data used to build the model are often poor predictors when applied to new data. Spatial data are inherently collinear (Wedding et al. 2011, Dormann et al. 2012), but many traditional modeling approaches require independence between observations and the variables used in the model (Kutner et al. 2004). To address collinearity, modelers often arbitrarily limit the number of predictor variables, choosing those that have the least collinearity with other predictors in the data. This results in data loss and can potentially bias mode results (Strobl et al. 2008, Freckleton 2010). Species' distributions, especially of schooling species, are often skewed and rare, or under-sampled, species can cause large numbers of zero occurrences in the data (McArdle and Anderson 2001, Cunningham and Lindenmayer 2005, Gray et al. 2005, 2006, McGill et al. 2007). These zero-inflated data do not fit standard distributional assumptions and cannot be easily modeled with standard methods (Barry and Welsh 2002).

An even larger issue involves the selection of variables to include in the models. Ecosystems are complex with an enormous number of potential factors that may influence species occurrence and distribution. However, traditional models are limited in the number of predictor variables they can accommodate and cannot model situations where the number of predictor variables is much larger than the number of samples (n<<p). To satisfy statistical assumptions, the modeler is required to preselect the variables that are hypothesized to the be the most likely predictors of the response, which potentially biases the ecological results of the analysis (Brosse et al. 2001). Data reduction is often accomplished by using principal components analysis or factor analysis, but combining variables into a reduced number of components dilutes the effect of individual variables in the analysis (Strobl et al. 2009).

Ecosystem-based management requires quantitative models capable of predicting the composition of communities and identifying relationships between habitat variables and both individual species and community structure (Olden et al. 2006, Gutiérrez-Estrada et al. 2008). Models that can accurately predict species distributions individually and at the community level for several life stages will be extremely valuable as that is when research and management approaches move beyond emphasis on single-species and refocus on understanding species within an ecosystem context. Developing species and community models is hampered by the

complexity of species-habitat relationships. It is here that machine learning can help find patterns and signals in complex data (Craig et al. 2009). Species respond to biotic and abiotic environmental factors at a variety of spatial scales and these responses can be non-linear or discontinuous (Mayor et al. 2009). Environmental influences do not act in isolation, and interactions among environmental factors are known to be complex. Additionally, development of community models requires the ability to fit a number of environmental variables simultaneously to more than one response variable, representative of the suite of species in a location. All of these issues make it difficult to develop accurate predictive models using standard statistical approaches such as frequency (Zar 2009) and Bayesian (McCarthy 2007).

Methods that assume that the distribution of the data is unknown and rely on recursive algorithms to approximate the response surface are emerging as alternatives to traditional hypothesis-driven approaches (Breiman 2001, Stephens et al. 2005, Hobbs and Hilborn 2006). These data mining methods have been developed for use in disciplines where there is a need to elucidate patterns in high dimensional, complex data such as genetics research, financial modeling (hedge funds, investment banking, and computer science (face recognition, spam detection, internet usage/marketing). These methods also have been used extensively in terrestrial ecology (Phillips et al. 2004, Ohse et al. 2009, Williams et al. 2009, Cushman and Huettmann 2010, Magness et al. 2010, Oppel and Huettmann 2010) but have received less attention in the marine and fisheries sciences (Oppel and Huettmann 2010, Wei et al. 2010, Hardy et al. 2011, Palialexis et al. 2011). In addition to removing the constraint of preselecting an underlying distribution, data mining methods can also model nonlinear relationships and interactions between variables and can incorporate a larger number of predictor variables that are more representative of the complexity of natural ecosystems. The newer machine learning or data mining algorithms have increased model predictive accuracy over many standard statistical methods (De'ath 2002, Hochaka et al. 2007, Magness et al. 2010, Drew et al. 2011).

Here, I investigate the feasibility of developing predictive species distribution models with high accuracy for estuarine-associated fish and invertebrates at both the individual and community level. In Alaska, more than a third of the commercial fish species managed by the National Marine Fisheries Service (NMFS) have been documented in Southeast Alaska estuarine surveys, many of them as juveniles (Murphy et al. 2000, Johnson et al. 2003). However, for most species, very little is known about the relationship between habitat and particular life stages, especially early life stages (Ecosystem Principles Advisory Panel 1999, NPRB 2005).

Estuaries are known to provide important spawning, feeding and rearing habitat for many commercial, recreational and subsistence fish and invertebrates (Mueter and Norcross 1999, Norcross et al. 1999, Abookire et al. 2001, Johnson et al. 2003). Estuaries are complex systems characterized by widely varying hydrological, morphological, and chemical conditions (Ellis et al. 2006).The geology and climatology of Southeast Alaska contribute to a high level of physical variability among estuaries by affecting wind mixing, tidal flow, freshwater discharge, and nutrient transport (Weingartner et al. 2009). Several studies have demonstrated that nearshore Alaskan fishes exhibit nonrandom patterns of distribution in relation to salinity, depth and other physical factors (Mueter and Norcross 1999, Norcross et al. 1999, Abookire et al. 2001), but to date there has been no comprehensive analysis of the environmental factors that affect the spatial distribution of fish species and composition of fish assemblages with respect to their use of estuarine environments. The few published studies linking biotic and abiotic habitat variables to fish and invertebrate presence in Alaskan estuaries have focused on single species (Heifetz et al. 1989, Stone and O'Clair 2001, Stoner and Abookire 2002) or have been limited to a specific estuary or location (Mueter and Norcross 1999, Abookire et al. 2000).

The goal of this research is to describe for the first time, the multispecies habitats and community structure of estuarine fish and invertebrates in Southeast Alaska in relation to biotic and abiotic aspects of their environment. The research objectives are (i) to model the spatial distribution of estuarine species in Southeast Alaska estuaries in relation to environmental factors, (ii) to predict species occurrence to unsampled areas, (iii) to evaluate change in estuarine community composition along spatial and environmental gradients, and (iv) to evaluate how biotic and abiotic factors affect the ecological function of estuaries in Southeast Alaska.

In Chapter 2, I develop models for predicting the occurrence of estuarine-associated fish and invertebrates in relation to landscape-scale habitat features. The objective is to evaluate the feasibility of using remotely sensed environmental land and seascape data to accurately predict species occurrence from a small but efficient sample dataset. The hypothesis is that remotely sensed data can serve as suitable surrogates for data measured in situ and result in models with high predictive accuracy for most species. I construct a large spatial database in ArcGIS and use the random forest algorithm to model species distributions. Models are validated using an independent sample dataset.

In Chapter 3, I model species habitat relationships at the community level to evaluate how community composition changes in relation to environmental heterogeneity. This tests the

hypothesis that the influence of physical and biological factors results in non-random species assemblages with a defined species composition. This research uses a new multivariate random forest modeling technique that simultaneously models multiple species against several environmental variables.

In Chapter 4, I evaluate differences in taxonomic and functional diversity between estuaries to evaluate how change in the composition of species affects ecosystem processes. Functional diversity reflects the ecological complexity of the study area and can provide insight into the vulnerability of these estuarine ecosystems to change, including alterations in species composition as a result of anthropogenic disturbance or exploitation. I compare taxonomic and functional diversity across estuaries and between the estuary clusters defined in Chapter 3, to examine spatial patterns in species dominance and the distribution of functional traits, and to test the hypothesis that estuaries that have dissimilar species composition are also ecologically (functionally) dissimilar. This analysis uses indices and multivariate methods.

Chapter 5 concludes with a summary of the major findings and discusses applications of this research for estuarine conservation and management.

6

REFERENCES

Abookire, A. A., J. F. Piatt, and B. L. Norcross. 2001. Juvenile groundfish habitat in Kachemak Bay, Alaska during late summer. Fishery Bulletin **8**:45-56.

Abookire, A. A., J. F. Piatt, and M. D. Robards. 2000. Nearshore fish distributions in an Alaskan estuary in relation to stratification, temperature and salinity. Estuarine, Coastal and Shelf Science **51**:45-59.

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

Barbier, E. B., E. W. Koch, B. R. Silliman, S. D. Hacker, E. Wolanski, J. Primavera, E. F. Granek, S. Polasky, S. Aswani, L. A. Cramer, D. M. Stoms, C. J. Kennedy, D. Bael, C. V. Kappel, G. M. Perillo, and D. J. Reed. 2008. Coastal ecosystem-based management with nonlinear ecological functions and values. Science **319**:321-323.

Barry, S. C. and A. H. Welsh. 2002. Generalized additive modelling and zero-inflated count data. Ecological Modelling **157**:179-188.

Breiman, L. 2001. Random Forests. Machine Learning **45**:5-32.

Brosse, S., J. L. Giraudel, and S. Lek. 2001. Utilisation of non-supervised neural networks and principal component analysis to study fish assemblages. Ecological Modelling **146**:159-166.

Burnham, K. P. and D. Anderson. 1998. Model Selection and Multimodel Inference: A Practical Information Approach. Springer, New York.

Chase, J. M. 2010. Stochastic community assembly causes higher biodiversity in more productive environments. Science **328**:1388-1391.

Craig, E. H., T. H. Craig, F. Huettmann, and M. R. Fuller. 2009. Use of machine learning algorithms to predict the incidence of lead exposure in Golden eagles.*in* Ingestion of spent lead ammunition: Implications for wildlife and humans. The Peregrine Fund, http://www.peregrinefund.org/Lead_conference/PDF/0303%20Craig.pdf Boise, ID.

Cunningham, R. B. and D. B. Lindenmayer. 2005. Modeling count data of rare species: some statistical issues. Ecology **86**:1135-1142.

Cushman, S. and F. Huettmann. 2010. Spatial Complexity, Informatics and Wildlife Conservation. Springer, Tokyo, Japan.

De'ath, G. 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology **83**:1105-1117.

de Bello, F., T. Wilfried, J. Lepš, P. Choler, J.-C. Clément, P. Macek, M.-T. Sebastiá, and S. Lavorel. 2009. Partitioning of functional diversity reveals the scale and extent of trait convergence and divergence. Journal of Vegetation Science **20**:475-486.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, A. K. Skidmore, D. Zurell, and S. Lautenbach. 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography **36**:27-46.

Drew, C. A., Y. Wiersma, and F. Huettmann. 2011. Predictive Modeling in Landscape Ecology. Springer, New York, NY.

Ecosystem Principles Advisory Panel. 1999. Ecosystem-based fishery management: a report to Congress. U.S. National Marine Fisheries Service, Washington DC.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, M. O. Jacob, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmerman. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Elith, J. and J. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology and Systematics **40**:677-697.

Ellis, J., T. Ysebaert, T. Hume, A. Norkko, T. Bult, P. Herman, S. Thrush, and J. Oldman. 2006. Predicting macrofaunal species distributions in estuarine gradients using logistic regression and classification systems. Marine Ecology Progress Series **316**:69-83.

Ferrier, S. and A. Guisan. 2006. Spatial modelling of biodiversity at the community level. Journal of Applied Ecology **43**:393-404.

Fiser, C., A. Blejec, and P. Trontelj. 2012. Niche-based mechanisms operating within extreme habitats: a case study of subterranean amphipod communities. Biology Letters **doi: 10.1098/rsbl.2102.0125, http://rsbl.royalsocietypublishing.org**.

Freckleton, R. P. 2010. Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. Behavioral Ecology and Sociobiology **65**:91-101.

Gray, J. S., A. Bjørgesæter, and K. I. Ugland. 2005. The impact of rare species on natural assemblages. Journal of Animal Ecology **74**:1131-1139.

Gray, J. S., A. Bjørgesæter, and K. I. Ugland. 2006. On plotting species abundance distributions. Journal of Animal Ecology **75**:752-756.

Gutiérrez-Estrada, J. C., R. Vasconcelos, and M. J. Costa. 2008. Estimating fish community diversity from environmental features in the Tagus estuary (Portugal): Multiple linear regression and artificial neural network approaches. Journal of Applied Ichthyology **24**:150-162.

Hardy, S., M. Lindgren, K. Hanumantharao, and F. Huettmann. 2011. Predicting the distribution and ecological niche of unexploited snow crab (*Chionoecetes opilio*) populations in Alaska waters: A first open-access ensemble model. Integrative and Comparative Biology **51**:608-622.

Heifetz, J., S. W. Johnson, K. V. Koski, and M. L. Murphy. 1989. Migration timing, size, and salinity tolerance of sea-type sockeye salmon (*Oncorhynchus nerka*) in an Alaska estuary. Canadian Journal of Fisheries and Aquatic Sciences **46**:633-637.

Hobbs, N. T. and R. Hilborn. 2006. Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. Ecological Applications **16**:5-19.

Hochaka, W. M., R. Caruana, D. Fink, A. Munson, M. Riedewald, D. Sorokina, and S. Kelling. 2007. Data-mining discovery of pattern and process in ecological systems. Journal of Wildlife Management **71**:2427-2437.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology **22**:415-427.

Johnson, S. W., M. L. Murphy, D. J. Csepp, P. Harris, and J. Thedinga. 2003. A survey of fish assemblages in eelgrass and kelp habitats of Southeast Alaska. US Department of Commerce, NMFS-AFSC-139.

Kutner, M., C. J. Nachtsheim, J. Neter, and W. Li. 2004. Applied Linear Statistical Models. McGraw Hill Irwin, New York, NY.

Magness, D. R., F. Huettmann, and J. M. Morton. 2010. Using random forests to provide predicted species distribution maps as a metric for ecological inventory and monitoring programs. Page 464 p *in* S. Cushman, F. Huettmann, editor. Spatial complexity, informatics and wildlife conservation. Springer, New York.

Maxwell, D. L., V. Stelzenmuller, P. D. Eastwood, and S. I. Rogers. 2009. Modelling the spatial distribution of plaice (*Pleuronectes platessa*), sole (*Solea solea*), and thornback ray (*Raja clavataI*) in UK waters for marine management and planning. Journal of Sea Research **61**:258-267.

Mayor, S. J., D. C. Schneider, J. A. Schaefer, and S. P. Mahoney. 2009. Habitat selection at multiple scales. Ecoscience **16**:238-247.

McArdle, B. H. and M. J. Anderson. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology **82**:290-297.

McCarthy, M. C. 2007. Bayesian Methods for Ecology. Cambridge University Press, Cambridge UK.

McGill, B. J., R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, and E. P. White. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology Letters **10**:995-1015.

Mueter, F. J. and B. L. Norcross. 1999. Linking community structure of small demersal fishes around Kodiak Island, Alaska to environmental variables. Marine Ecology Progress Series **190**:37-51.

Murphy, M. L., S. W. Johnson, and D. J. Csepp. 2000. A comparison of fish assemblages in eelgrass and adjacent subtidal habitats near Craig, Alaska. Alaska Fishery Bulletin **7**:11-21.

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.

NPRB. 2005. North Pacific Research Board Science Plan. Page 198. North Pacific Research Board, Anchorage.

Ohse, B., F. Huettmann, S. Ickert-Bond, and G. Juday. 2009. Modeling the distribution of white spruce (*Picea glauca*) for Alaska with high accuracy: an open access role-model for predicting tree species in the last remaining wilderness areas. Polar Biology **32**:1717-1724.

Olden, J. D., M. K. Joy, and R. G. De'ath. 2006. Rediscovering the species in community-wide predictive modeling. Ecological Applications **16**:1449-1460.

10

Oppel, S. and F. Huettmann. 2010. Using a random forest model and public data to predict the distribution of prey for marine wildlife management. Pages 151-164 *in* S. Cushman and F. Huettmann, editors. Spatial Complexity, Informatics and Wildlife Conservation. Springer, Tokyo, Japan.

Palialexis, A., S. Georgakarakos, I. Karakassis, K. Lika, and V. D. Valavanis. 2011. Fish distribution predictions from different points of view: comparing associative neural networks, geostatistics, and regression models. Hydrobiologia **670**:165-188.

Phillips, S. J., M. Dudik, and R. E. Schapire. 2004. A maximum entropy approach to species distribution modeling. ICLM '04 Proceedings on the 21st International Conference on Machine Learning, New York, NY.

Polechová, J. and D. Storch. 2008. Ecological Niche. Pages 10888-11097 *in* S. E. J. a. B. D. Fath, editor. Encylcopedia of Ecology,Vol 2. Elsevier, Oxford.

Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems **9**:181-199.

Stephens, P. A., S. W. Buskirk, G. D. Hayward, and C. Martinez Del Rio. 2005. Information theory and hypothesis testing: a call for pluralism. Journal of Applied Ecology **42**:4-12.

Stone, R. P. and C. E. O'Clair. 2001. Seasonal movements and distribution of Dungeness crabs, *Cancer magister*, in a glacial southeastern Alaska estuary. Marine Ecology Progress Series **214**:167-176.

Stoner, A. W. and A. A. Abookire. 2002. Sediment and size-specific distribution of young-of-the-year Pacific halibut in an Alaska nursery. Journal of Fish Biology **61**:540-559.

Strobl, C., A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. Conditional variable importance for random forests. BMC Bioinformatics **9**:307-318.

Strobl, C., J. Malley, and G. Tutz. 2009. An introduction to recursive partitioning. University of Munich **Technical Report Number 55**.

Sundblad, G., M. Harma, A. Lappalainen, L. Urho, and U. Bergstrom. 2009. Transferability of predictive fish distribution models in two coastal systems. Estuarine, Coastal and Shelf Science **83**:90-96.

Valavanis, V. D., G. J. Pierce, A. F. Zuur, A. Palialexis, A. Saveliev, I. Katara, and J. Wang. 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. Hydrobiologia **612**:5-20.

Wedding, L. M., C. A. Lepcxyk, S. J. Pittman, A. M. Friedlander, and S. Jorgensen. 2011. Quantifying seascape structure: extending terrestrial spatial pattern metrics to the marine realm. Marine Ecology Progress Series **427**:219-232.

Wei, C.-L., G. T. Rowe, E. Escobar-Briones, A. Boetius, T. Softwedel, M. J. Caley, Y. Soliman, F. Huettmann, F. Qu, Z. Yu, C. R. Pitcher, R. L. Haedrich, M. K. Wickstein, M. A. Rex, J. G. Baguley, J. Sharma, R. Danovaro, I. R. MacDonald, C. C. Nunnally, J. W. Deming, P. Montagna, M. Lévesque, J. M. Weslawski, M. Wlodarska-Kowalczuk, B. S. Ingole, B. J. Bett, D. S. M. Billett, A. Yool, B. Bluhm, K. Iken, and B. E. Narayanaswamy. 2010. Global patterns and predictions of seafloor biomass using random forests. PlosOne **5**:e15323.

Weingartner, T., L. Eisner, G. L. Eckert, and S. Danielson. 2009. Southeast Alaska: oceanographic habitats and linkages. Journal of Biogeography **36**:387-400.

Weins, J. J. 2011. The niche, biogeography and species interactions. Philosophical Transactions of the Royal Society B: Biological Sciences **366**:2336-2350.

Williams, J. N., C. Seo, J. Thorned, J. K. Nelson, S. Erwin, J. M. O'Brien, and M. W. Schwartz. 2009. Using species distribution models to predict new occurrences for rare plants. Diversity and Distributions **15**:565-576.

Zar, J. H. 2009. Biostatistical Analysis. Pearson Prentice-Hall, Saddle River, NJ.

Chapter 2: Spatial multispecies models for predicting the occurrence of estuarine-associated fish and invertebrates[1]

ABSTRACT

We investigated whether multivariate, remotely sensed geomorphic and landscape data could be used to develop accurate predictive models of estuarine-associated fish and invertebrate occurrence. Species occurrence data were obtained from trawl and seine sampling conducted in 49 estuaries in Southeast Alaska from 1998 to 2005. Using a conceptual model of estuarine processes, we identified 107 predictor variables representing hydrodynamics, habitat, and estuarine and watershed structure. We used the Random Forest algorithm (randomForest in R), a method that has demonstrated good predictive accuracy with high-dimensional data, to construct predictive models for all species that occurred at 5% or more of the sample estuaries. Receiver operating characteristic (ROC) curves were used to evaluate model performance. Model results were validated using independent sample data collected from 1998 to 2004 in 88 additional Southeast Alaska estuaries. Seventy-six percent of the Random Forest models for fish species, and 72% of the models for invertebrates, had predictive accuracies of 70% accuracy or better. Depth and slope of the estuary, tidal height, substrate type, and watershed vegetation were important variables in all of the species models. The models also identified new species-variable relationships, such as varying species occurrence with respect to precipitation, that can be explored in future research. This study demonstrates the utility of using remotely sensed data in Random Forest models to predict fish and invertebrate occurrence and sets the stage for incorporation of geomorphic and landscape data into future studies of estuarine species.

INTRODUCTION

Effective management of marine ecosystems and fisheries requires the ability to understand and predict species occurrence in relation to environmental and habitat variables and to identify the environmental variables important in structuring species distributions (Maxwell et al., 2009; Sundblad et al., 2009). Quantitative species distribution models can assist in the identification of habitat for protection and spatial planning (Austin, 2002; Maxwell et al., 2009;

---

[1] Submitted to Estuarine, Coastal and Shelf Science. Miller K, Norcross B, Huettmann F, Lorenz M. Spatial multispecies models for predicting the occurrence of estuarine-associated fish and invertebrates.

Valavanis et al., 2008), can increase understanding of ecosystem structure and function, can guide research on productivity of habitat for specific species, and can help predict changes in species occurrence as a result of climate change or invasion (Mueter and Litzow, 2008; Stojlgren et al., 2010; Williams et al., 2009). Predictive species distribution models have been widely used in terrestrial and freshwater ecological studies (Lawler et al., 2011; Magness et al., 2010; Ohse et al., 2009), but only in the last decade have they begun to be applied to marine ecosystems (Elith and Leathwick, 2009; Hardy et al., 2011; Oppel and Huettmann, 2010; Palialexis et al., 2011; Robinson et al., 2011; Wei et al., 2010). Species distribution modeling is equivalent to ecological niche modeling and describes the range of environmental conditions under which species are predicted to occur (Magness et al., 2010; Wiens, 2011).

A significant challenge to regional marine species distribution modeling is the limited availability of sample data at relevant scales. Data collection in Southeast Alaska is hampered by remote sampling locations and hindered by difficult weather and oceanographic conditions that restrict sampling. Yet, this area is essential to the regional fisheries as a nursery ground for many economically important species (Johnson et al., 2003; Murphy et al., 2000). As a result, sample data generally need to be commingled from various data collection efforts using different gears and methods over multiple years. Environmental data also can be difficult to obtain. Remotely sensed data can sometimes be used to characterize the shoreline and uplands, but there is a paucity of regional-scale data on subtidal and bottom habitat. The spatial accuracy of these data can also be an issue.

A further difficulty in constructing accurate predictive distribution models comes from the complexity of fish-habitat relationships and the processes that affect those relationships on a variety of spatial scales (Anderson et al., 2009; Elith and Leathwick, 2009; Hardy et al., 2011; Oppel and Huettmann, 2010). Particularly in estuaries, species tend to have wide environmental tolerances, and their response to environmental factors can be inherently multivariate, nonlinear or discontinuous (Gutiérrez-Estrada et al., 2008; Mueter and Norcross, 1999). Furthermore, the effects of most environmental factors do not occur in isolation from effects of other factors. This makes it difficult for researchers to attribute simple causality in explaining variation in species distributions and assemblage composition (Oppel et al., 2009). These aspects make it difficult to model species-habitat relationships using standard statistical methods (Magness et al., 2010; Prasad et al., 2006).

Landscape-scale data such as elevation, vegetative cover, and precipitation patterns are commonly used to predict species occurrence and abundance in terrestrial ecosystems (Magness et al., 2010). Freshwater fisheries research has demonstrated that landscape and stream geomorphic variables (e.g., channel complexity, stream size) can accurately predict the occurrence of freshwater fish (Brenden et al., 2007; Cèrèghino et al., 2005; Elmendorf and Moore, 2008; Wilkins and Snyder, 2011). The idea that estuarine geomorphology influences patterns in estuarine biota is not new (Roy et al., 2001), but models linking geomorphic variables to predicted marine species occurrence or abundance are limited (Dye, 2006; Hicks et al., 2010; Valesini et al., 2010). Most research has focused on predicting the presence or abundance of marine fish and invertebrates with respect to local habitat variables, such as habitat complexity (Pittman et al., 2007; Wedding and Friedlander, 2008), or in relation to specific habitat features, such as subtidal vegetation (Whitlow and Grabowski, 2012), or mangroves (Jelbart et al., 2006). In Alaska, a number of studies have identified species-habitat relationships for individual fish and invertebrate species in estuaries or the nearshore (Abookire et al., 2001; Norcross et al., 1997; Stone and O'Clair, 2001; Stoner et al., 2007); however, these studies have been limited to a specific estuary or location and none of them have attempted to predict species occurrence or abundance in unsampled areas. All of these Alaska studies also have used environmental data collected in-situ, making it difficult to extend the results to areas where environmental data have not been collected.

In this research, we investigate whether multivariate remotely sensed geomorphic and landscape data could be used to develop accurate predictive models of marine fish and invertebrate occurrence in southeast Alaskan estuaries. We used the Random Forest algorithm (Breiman, 2001), because it has been shown to develop highly accurate predictive species' distribution models using a large number of predictor variables (Lawler et al., 2011; Magness et al., 2010). Random Forest can effectively model nonlinear and interacting variables and can identify the variables with the strongest influence on species' distributions (Cushman and Huettmann, 2010; Drew et al., 2011). Species' distributions, especially of schooling species, are often skewed and rare, or under-sampled species can cause large numbers of zero occurrences in the data (Cunningham and Lindenmayer, 2005; Gray et al., 2005, 2006; McArdle and Anderson, 2001). Random Forest models are not sensitive to violations of the normality assumptions required of linear models and they do not make any assumptions about the underlying distribution of the data. Monotonic transformation of the response variables does not affect model results,

allowing modeling of skewed data. The Random Forest algorithm fits a large number of simple models to predict the relationship between a response variable (species occurrence or abundance) and a set of predictor variables and then combines the resulting models to develop robust estimates of the response relationship (Oppel et al., 2009). In this way Random Forest models differ from traditional regression-based models, such as Generalized Additive Models (GAMs), which attempt to fit a single "best" model to the data (Cutler et al., 2007; Prasad et al., 2006). We used Random Forest to develop predictive models for the occurrence of fish and invertebrates from estuaries in Southeast Alaska and evaluated the predictive accuracy of the models against a set of independent data.

MATERIALS AND METHODS

<u>Study area and data</u>

The study area (Figure 2.1) was the Alexander Archipelago: a collection of approximately 1,000 mountainous islands in Southeast Alaska from Dixon Entrance at the Canadian border (54º 47'35", 130º 38'06"W) to Lance Point in Lynn Canal (58º 44' 141", 135º 13' 996"). The coastline is generally steep and the islands are separated by deep channels and fjords. The entire archipelago is a temperate rainforest: precipitation varies locally and regionally with a general gradient of lower precipitation in the northwest and higher precipitation in the southeast. Precipitation is also strongly influenced by the coastal geology and topography (Weingartner et al., 2009). Average annual precipitation in the region is in excess of 1000 mm/yr (Neal et al., 2002) with much of the precipitation being released directly into the marine waters via numerous small streams and wetlands. Stream flow is highly seasonal and influenced both by precipitation and by snow and ice melt. The highest stream flows tend to occur in autumn when precipitation rates are high. Flows decrease in winter as a result of freezing, and increase again in the late spring and summer from melting of snow and ice. The flow of freshwater affects not only the nearshore estuarine circulation, but is the driver for larger-scale oceanographic circulation within Southeast Alaska's interior channels and on the continental shelf (Weingartner et al., 2009). Stream and river temperatures are influenced both by air temperatures and by runoff from glaciers, snowmelt, and precipitation.

The estuaries in the study area differ in their hydrological and geomorphological characteristics. In many Southeast Alaska estuaries, tidal energy is often much higher than energy from freshwater inflow (Weingartner et al., 2009). Southeast Alaska has mixed semi-

diurnal tides with tidal height increasing as the tide moves from the continental shelf into the interior of the archipelago (Inazu et al., 2009). The difference in height between mean higher high water and mean lower low water is between 1.18 to 5.13 m in the study area (NOAA, 2012). Tidal velocities are strongly influenced by bathymetry and channel morphology, and these, in turn, affect estuarine circulation, nutrient fluxes, and sediment dynamics (Weingartner et al., 2009). Coastal geology also varies greatly across estuaries in the study area. Most estuaries have a mixture of soft and hard substrate shorelines, but the amount of each type of substrate varies depending on both oceanographic and terrestrial processes.

Previous research identified approximately 12,000 estuaries in Southeast Alaska using the intersection of fresh and marine waters as the defining criteria (Albert et al., 2010); however; this definition does not take into account the degree to which the estuary is enclosed and somewhat isolated from other coastal waters. This enclosure has important implications for estuarine circulation as well as the physical and chemical properties of the estuarine waters. For this research, we defined an estuary as: *a coastal indentation with a restricted connection to saltwater and an aquatic environment affected by the physical and chemical characteristics of both fluvial drainage and marine systems.* Using this definition, we delineated 541 estuaries between the high tide line and the 30 m depth contour in ArcGIS 10 ™ for model projection (Figure 2.2, blue dots). Glacier Bay was excluded from the research because circulation within the bay is constrained by the shallow sill at the entrance to this fjord system. As a result, the processes structuring fish and invertebrate communities within that bay are different than those in open estuaries in Southeast Alaska (Matthews, 1981).

Species Sampling

Fish and mobile benthic invertebrates were sampled in 49 Southeast Alaska estuaries between 1998 and 2005 using both trawl and seine gear (Figure 2.2, red dots). Sampling was conducted during daylight hours between February and September at high and low slack water. Fish and mobile benthic invertebrates were captured using an otter trawl (3 m x 1 m, with 6 mm square mesh in the cod end) deployed with a bridle scope of approximately 20 m. The trawl was towed at a speed of approximately 3 kn along a depth contour between 5 m and 10 m. The exact depth of individual tows varied within this range depending on benthic structure of the estuary. One tow in each direction was made along the same transect at high and low slack water. The latitude and longitude of the beginning and ending points of the trawl were recorded along with

the average depth of the tow. The length of the trawl was calculated by marking the beginning and ending coordinates on the GIS and measuring along the appropriate depth contour. Fish also were sampled with a 37-m long variable mesh beach seine that tapered from 5 m wide at the center to 1 m at the ends. Outer panels were each 10 m with 32 mm stretch mesh, intermediate panels were each 4 m with 6 mm square mesh, and the bunt was 9 m with 2.3 mm square mesh. The net was set as a round haul by fixing one end on the beach, backing the skiff while deploying the net, and bringing the other end to shore approximately18 m down shore from the first end. The latitude and longitude for each sample were recorded. Captured fish were identified to species and measured in the field to the nearest millimeter for total length. Mobile benthic invertebrates were identified to the lowest taxonomic level in the field and counted.

The length at 50% maturity for commercially harvested fish and some forage fish was obtained from the Alaska Fishery Science Center's (AFSC) Life History database (http://access.afsc.noaa.gov/reem/LHWeb/Index.cfm) and was used to classify fish as adults or juveniles. For fish species not in the AFSC database, a variety of published sources was used to obtain length at 50% maturity information. For species that occurred in the data as a mix of juveniles and adults, life stages were pooled and modeled together. Most sites were sampled only once. The presence or absence of fish and invertebrates at each site was recorded by the month in which the sampling occurred.

To verify the predictive models, we used the Alaska Nearshore Fish Atlas (Johnson et al., 2005) as an independent dataset. This dataset is a compilation of seine sampling at 279 sites in Southeast Alaska from 1998 to 2004 (Figure 2.2, yellow dots). To ensure independence of the data, we used sites from the Fish Atlas data that were not sampled in the data used to develop the predictive models. Captured fish from the Fish Atlas samples were identified to species in the field and were measured to the nearest millimeter for total length. Very few invertebrates occur in the Fish Atlas data, so these data were used only to validate fish species models. Because of differences in gear and sampling method between the Fish Atlas data and the sampled data used in this study, only presence/absence data were used to develop predictive models. All fish and invertebrates were collected by NOAA/NMFS under their collection permits.

Environmental variables

For each estuary, we identified 107 predictor variables using a conceptual model of estuarine system components (Figure 2.3) and using remotely sensed landscape data as surrogates

for environmental data measured in situ. The conceptual model identifies four estuary components, and the predictor variable categories linked to those components. Although they are depicted as discreet inputs, the estuary components (watershed structure, estuary structure, habitat and hydrodynamics) are strongly interrelated. For example, tides affect not only hydrodynamics and circulation, but also sedimentation and habitat processes. Variables were selected based both on their ability to describe one or more aspects of the four estuary components, and on their availability at the spatial scale of this research.

GIS layers were acquired from the Southeast Alaska GIS library, the National Oceanic and Atmospheric Administration (NOAA), and the Alaska ShoreZone database (Table 2.1). The predictors had varying units of measure, so all variables except the ShoreZone variables were standardized by subtracting the mean of the variable across all samples from each variable value and dividing by the standard deviation. Random Forest models are invariant to monotonic transformation of the predictor variables (Shi and Horvath, 2006), but standardization makes it easier to plot variables for comparison purposes. Variables that described the structure of the estuary included estuary length, width at mouth, area, perimeter, intertidal area, depth, bathymetric slope, and open water area. Estuary length was measured as the distance from the high tide line to the estuary mouth. The ArcGIS Spatial Analyst extension was used to calculate average, minimum and maximum parameters for bathymetric slope, and average and range parameters for depth. Intertidal area was obtained directly from the Tongass National Forest High and Low Tidelines dataset (USDA, 2002), which contains polygons for intertidal areas. Open water area is the amount of open water at low tide and is the difference between the estuary area and the intertidal area. Each estuary was assigned a classifier for the type of water body into which the estuary drains (bay, inlet, channel, or open ocean). This was the only categorical variable in the predictor variable dataset. For those estuaries draining into a bay or inlet, we measured the distance from the estuary mouth to open water, either a major channel or the open ocean.

We also included variables describing the watershed surrounding the estuaries. These variables included the size and slope of the watershed, the type and amount of land cover (e.g. vegetation, bare land, development, glaciers), and the degree of land cover fragmentation. Watershed size was derived from 12-digit hydrologic units (USGS, 1995). We measured watershed slope at two distances from the estuary by placing buffers at 1 km and 5 km distance around the estuary and using a digital elevation model and the ArcGIS Spatial Statistics to

calculate maximum and average slopes within each buffer. The most recent National Land Cover Dataset for Alaska (2001) was used to extract percent area of vegetation, bare land, development, and glacial ice in the 1 km buffer. We used FRAGSTATS (McGarigal et al., 2002) to calculate five measures of land cover patchiness: total area, number of patches, patch density, largest patch index, and landscape shape index.

Freshwater inflow into Southeast Alaska estuaries is difficult to calculate. Much of the study area is remote and undeveloped and there is a paucity of stream flow data even for large rivers. To capture the influence of freshwater on estuarine communities, we compiled minimum monthly precipitation over the study period from the PRISM climate model for Alaska (SNAP, 2011).We calculated fluvial flow after the method of Digby et al. (1998) by multiplying the catchment area with the average annual rainfall and a runoff coefficient and dividing by the open water area of the estuary:

$$\text{Fluvial Flow} = \frac{(catchment\ area\ X\ average\ annual\ precipitation\ X\ runoff\ coefficient)}{open\ water\ area}$$

The runoff coefficient (RV) is based on the impervious fraction of the drainage area (I) (Schueler, 1987):

$$RV = 0.05 + 0.9(I)$$

We calculated the variable "I" as the non-vegetated, non-ice portions of the watershed from the land cover analysis. At the scale of this analysis, the variable "I" was sufficiently small that the runoff coefficient was essentially a constant (0.95) across all watersheds. We measured the total length of streams within the 1 km buffer around the estuary and in the intertidal area using the USFS streams data layer for the Tongass National Forest (USGS). The total length of streams within the buffer and the length and percent of streams in the intertidal area were included as predictor variables. Surface salinity data for Southeast Alaska are not available at the spatial scale of this research. Precipitation and fluvial flow variables were included in the analysis to capture differences in salinity and buoyancy-driven circulation between estuaries.

The great diurnal tide range and mean tide range for each estuary was compiled from NOAA tide data (NOAA, 2012). Estuaries without measured tidal data were attributed the tidal ranges from the nearest estuary with tidal data. Estuary depth and slope, and the depth and

location of bars or sills were included as predictor variables to capture the influence of bathymetry in tidal energy and flow. Bars and sills were defined as constrictions outside of the estuary that limited the flow of water into the estuary and where minimum depths were half or less of the average depth of the estuary. Bathymetric contours were evaluated to identify bars/sills and we individually measured the distance between the bar/sill and the estuary mouth for each estuary in which they occurred.

Geomorphological and biological characteristics of the intertidal portion of each estuary were obtained from the Alaska ShoreZone dataset (http://fakr.noaa.gov/shorezone/default.htm). ShoreZone is a mapping and classification system that uses oblique, low altitude aerial video and still images to classify segments of the shoreline according to natural breaks in geomorphic, sedimentary, and biological features (Harney et al., 2008). Our analysis used variables in three elements of the ShoreZone dataset: coastal class, bioband, and habitat class. Because shoreline segments are divided by natural breaks, they differ in length so the percentage of each variable within the estuary was used for analysis.

Coastal Class: In the ShoreZone system (Harney et al., 2008), *coastal class* is determined by substrate type, sediment type, across-shore width and slope. Coastal segments are first grouped into primary substrate categories: rock, rock and sediment, sediment, anthropogenic, channel, and glacier. The rock and sediment and sediment categories are subdivided into classes by the type of sediment: gravel, sand and gravel, sand, and organics. The rock, rock and sediment, and sediment classes are split into either wide (≥30 m) or narrow (<30 m) across-shore widths. Finally, each shore segment is organized according to three slope classes: steep (>20°), inclined (5° to 20°) or flat (<5°). Anthropogenic, channel, and glacier primary coastal categories are not divided by sediment type or slope. The result is a single number for each combination of features, for example, coastal class 22 is described as: sediment, gravel, narrow, inclined. The ShoreZone dataset has a total of 35 individual coastal classes; however, we only used classes that were represented in 5% or more of the estuaries for a total of 23 classes (Appendix A.3).

Biobands (Harney et al., 2008) are assemblages of sessile coastal biota observed in the supra-, inter-, and sub-tidal zones of the shoreline segments. Biobands are associated with a specific shore unit and are recorded as either "continuous" coverage or "patchy" coverage. We used the subtidal biobands for red algae (e.g. *Neorhodomela sp.*), *Alaria*, soft brown kelps (i.e. *Saccharina latissima*), dark brown kelps (i.e. stalked *Laminaria* sp.), and eelgrass (*Zostera*

marina). Predictor variables (Table 2.1) are the percentage of each type of coverage for each biota with respect to the total perimeter of the estuary.

The *habitat class* is a summary classification that combines the physical and biological information for a shoreline unit into a single variable that describes the intertidal biota together with the geomorphology (Harney et al., 2008). Shorelines are initially classified according to the dominant structuring process: wave energy, fluvial/estuarine, current energy, glacial, human modified, and lagoon. Wave dominated environments are further refined by substrate mobility (immobile, partially mobile, mobile) and coastal geology. All classes are finally grouped by biological wave exposure (very exposed, exposed, semi-exposed, semi-protected, protected, and very protected) using the presence and abundance of biota in the sampling unit as a proxy for energy conditions (Harney et al., 2008). Of the 48 potential habitat classes described in ShoreZone, this analysis used 13 habitat classes that were present in 5% or more of the sites. Predictor variables are the percentage of each class with respect to the total perimeter of the estuary (Appendix A.4).

Model development

The Random Forest algorithm (Breiman, 2001) was used to develop predictive models for occurrence of fish and mobile benthic invertebrates. These models have been shown to be highly accurate in their predictive ability (Magness et al., 2010; Prasad et al., 2006; Stojlgren et al., 2010) in a wide variety of applications. Random Forests are collections of classification or regression trees constructed by first randomly drawing with replacement from the entire dataset to create separate training and testing data, and then randomly drawing with replacement from the predictor variables of the training dataset to construct individual decision trees. Several hundred decision trees are grown in this manner using the dual randomization approach. The best split for each predictor variable is determined by averaging the results across all trees. The testing data not used in constructing the model, out-of-bag (OOB) data, were used to calculate an unbiased error rate and variable importance ranking (Breiman, 2001; Magness et al., 2010; Prasad et al., 2006). For classification, the prediction accuracy was assessed by plotting the number of correctly classified samples against those that are incorrectly classified.

Correlation between individual trees is reduced by randomizing both the samples and the predictor variables, which decreases the error estimate of the entire ensemble. Predictor variables are considered individually when constructing the decision trees, and the data are unconstrained

by the assumptions of the underlying distribution. This makes Random Forest models suitable for situations where the number of predictors is much larger than the number of samples (p >> n) (Oppel and Huettmann, 2010; Strobl et al., 2009) and where the data are skewed or over-dispersed. Random Forest models are able to incorporate both nonlinearities and interactions between predictor variables (Prasad et al., 2006).

Random Forest models were built for the occurrence of 25 fish species, and 18 mobile benthic invertebrate species and families using the randomForest package for R (Liaw and Wiener, 2002). In the Random Forest algorithm, the number of variables selected for prediction (mtry) and, to a lesser extent, the number of trees grown in a Forest (ntree) can be used to tune the models to obtain better results. For each species, we grew forests with 100, 500, 1000 and 1500 trees using from 1 to 30 variables (mtry) per tree. The receiver operating characteristic (ROC) was used to evaluate model performance for each species. ROC curves graphically represent model performance by plotting the rate of true positive classification (sensitivity) for both presences and absences against the false positive rate (specificity) (Figure 2.4). Each point along the curve represents a decision threshold indicating a tradeoff between true positive and false positive classification (Fawcett, 2006; Hamel, 2008). The area under the ROC curve (AUC) can be used to select between models for the same species. The AUC is comparable to a non-parametric Wilcoxon test of ranks, and as such can be used to select models that, on average, have a higher probability of accurately predicting species occurrence in random draws from the data (Lobo et al., 2008; Termansen et al., 2006). ROC curves were created using the pROC package in R for each model using the relative index of occurrence predicted by the models. The model with the highest AUC was chosen as the best model for that species. For all of the mobile benthic invertebrates, these predictions were made on the OOB dataset, which tends to result in more optimistic predictions than models tested on independent data. The validation dataset did not include invertebrates. For most of the fish species, predictions were assessed against the sampled presence/absence of species from the independent Fish Atlas dataset. The exceptions were six fish species that were captured entirely or mostly by trawl gear, for which the independent seine dataset was inappropriate for verification. These six fish species were predicted on the OOB data.

Although the AUC also is commonly used to assess model predictive accuracy (Cutler et al., 2007; Magness et al., 2010; Williams et al., 2009), its use has been questioned in species distribution models where uncertainty regarding zeros or absences in the data is high (Lobo et al.,

2008). Species may be absent from a recorded sample for reasons unrelated to their true occurrence at a site, including: inability or limited ability of the sampling gear to consistently sample the species, sampling during a time period or season when the species is absent, or low occurrence of the species in the habitat being sampled (rarity). The AUC gives equal weight to misclassification of presences and absences, which may not be appropriate in situations where confidence in species' absences is low. In both the modeling and verification datasets, the methods for collecting fish and mobile benthic invertebrate data were not designed to comprehensively sample the entire estuary, and sampling was not conducted in all estuaries in the same season. As a result, there is substantial uncertainty as to whether a species' absence from the data indicates a true absence from the habitat sampled. The AUC summarizes model performance over the entire ROC curve, but the upper right and lower left portions of the curve, which correspond to high false-negative rates and high false-positive rates, respectively, should not be considered as reasonable models since the objective is to maximize true predictive outcomes.

An alternative to the AUC is to assess model accuracy by comparing the number of presences and absences predicted by the model against the true presences and absences of species in the data. Several approaches can be used to calculate the percent of correctly predicted presences and absences, or the probability threshold above which a species will be considered to be present. We used the point on the ROC curve where the slope of the curve is equal to 1, or the highest sum of specificity and sensitivity (Jimenez-Valverde and Lobo, 2007). This "best" threshold (Figure 2.4) was applied to the relative index of species occurrences from the models to obtain the percent of correctly classified presences and absences, and the model results were sorted by percentage of presences accurately predicted. Species distribution maps for each species were developed by importing the relative index of occurrence (ROI) output from the individual models into the GIS.

Predictor variable importance in Random Forest is calculated by permuting the variables individually in the OOB data and measuring the decrease in prediction accuracy for models computed with and without the permuted data. To explore relationships between species occurrence and the predictor variables, partial dependence plots can be constructed for the most influential variables. Partial dependence plots are calculated by fixing the values of an individual variable, and averaging the prediction function over the remaining variables in the data (Cutler et al., 2007). For each species, we constructed partial dependence plots for the 30 most important

predictor variables. These plots can be used to explore the relationships between the predictors and response in a multivariate setting (Oppel and Huettmann, 2010).

RESULTS

The modeling dataset contained 57 species of fish from 23 families. Chum salmon (*Oncorhynchus keta*) was the most numerous fish species in the data, comprising 12% of the total catch across all sites and years, followed closely by pink salmon (*Oncorhynchus gorbuscha*) with 11% of the catch. Three species were captured at over 60% of the sites: Pacific staghorn sculpin (*Leptocottus armatus*), rock sole (*Lepidopsetta sp.*), and crescent gunnels (*Pholis laeta*).We developed predictive models for species captured in 5% or more of the sites with a result that 24 fish species were modeled (Appendix A.1). The invertebrate data contained 53 species and 43 families. Shell shrimp (*Crangon alaskensis*), and spot shrimp (*Pandalus platyceros*) made up the largest percentage of the total catch (16% and 12%, respectively), but 90% of the spot shrimp catch occurred at a single site. Two species and one family were captured at over 60% of the sites: shell shrimp, helmet crab (*Telmessus cheiragonus*),and gammarid amphipods (*Gammaridae*). We developed predictive models for mobile benthic invertebrate species and families occurring at 5% or more of the sample sites (18 invertebrates) (Appendix A.2).

Several fish species occurred more frequently or entirely as juveniles in both the sample and independent data. Species occurring only as juveniles included all species of salmon, Pacific herring (*Clupea palasii*), Pacific cod (*Gadus macrocephalus*), lingcod (*Ophiodon elongatus*), kelp greenling (*Hexagrammos decagrammus*), silverspot sculpin (*Blepsias cirrhosus*), butter sole (*Iopsetta isolepis*), and great sculpin (*Myoxocephalus polyacanthocephalus*). Species whose abundance was predominantly composed of juveniles were yellowfin sole (*Limanda aspera*), rock sole (*Lepidopsetta sp.*), and Pacific sand lance (*Ammondytes hextaperus*). Species with mixes of juveniles and adults were starry flounder (*Platichthys stellatus*), Pacific staghorn sculpin, and shiner perch (*Cymatogaster aggregata*). Species for which life stage could not be determined from the literature were the snake prickleback (*Lumpenus sagitta*), tube-snout (*Aulorhynchus flavidus*), and sturgeon poacher (*Podothecus accipenserinus*).

The AUCs of the Random Forest models for the individual species ranged from between 0.94 and 0.63 for the fish species and 0.94 to 0.62 for the invertebrates. For fish species, prediction accuracy ranged from 100% to 42% for species' presences and 87% to 15% for species' absences. For invertebrates, prediction accuracies ranged from 88% and 60% for

presences and from 89% to 50% for absences (Table 2.2). We defined strong models as those with an AUC of 0.80 or higher, moderate models as those with AUCs between 0.70 and 0.80, and weak models as those with AUCs below 0.70. Seventy-six percent of the fish species and 72% of the invertebrates had moderate to strong models. Model strength was not related to how common or rare a species was. The predictive accuracy of the models was generally equally strong for predicting species' presence and absence. Sixty-eight percent of the Random Forest models for fish species, and 67% percent of the models for the invertebrates, were able to predict both species' presence and absence in the independent data with an accuracy of 60% or better. The models for Pacific herring and sturgeon poacher showed much poorer absence prediction than presence prediction. These models had moderate prediction accuracy overall based on AUC. Presence and absence prediction accuracy for the invertebrate models was more even.

Five of the top strong model species shared month as their most influential variable. These species were shiner perch, lingcod, pink salmon, chum salmon and Pacific cod. Figure 2.5 shows the top portion of the variable importance plots for each of these species. The x-axis in these plots is the change in model performance when the variable is permuted and is different for each model. The y-axis is the code for the predictor variable. Each species had a unique set of important predictor variables with month as the highest or most influential. For shiner perch and lingcod, month was first among other influential variables, but for pink salmon, chum salmon and Pacific cod, month was the most influential variable by a wide margin. The relationship between month and species occurrence for each model can be visualized by examining the partial dependence plots for each species (Figure 2.6). Shiner perch, lingcod and Pacific cod were all more prevalent in the estuaries during the latter part of the sampling period, from June to October. Pink and chum salmon had higher prevalence prior to June.

All four of the flatfish species (butter sole, starry flounder, yellowfin sole and rock sole) had strong or moderate models (Table 2.2). Precipitation was among the most important variables in models for butter sole and starry flounder, which showed opposite trends for all months. Starry flounder were less likely to occur in areas with high precipitation and butter sole were more likely to occur in those locations. The partial dependence plots in Figure 2.7 shows the marginal effect of precipitation on the relative index of occurrence for each species. The scale of the vertical axis is half the logit of probability of occurrence (Cutler et al., 2007) and the x-axis is the value range of the variable from the data. The relationship can be evaluated by examining the shape of the curves. Butter sole predicted occurrence increases above 200 mm

minimum precipitation, while starry flounder predicted occurrence decreases at minimum precipitation values above 160 mm. When minimum February precipitation for the study period is projected onto the basemap of Southeast Alaska, both species show a fairly strong spatial separation in the predicted habitat distribution, which is closely associated to precipitation patterns (Figure 2.8). Variables most important to all flatfishes were those related to sediment and beach slope. Yellowfin sole in addition were associated with a high total length of watershed and estuarine streams. Large intertidal area was an important variable for rock sole.

All of the salmonids except for coho salmon (*Oncorhynchus kisutch*) were predicted by strong or moderate models (Table 2.2) with important variables reflecting the specific habitat requirements of the individual species. The AUC for the coho salmon model was 0.68, just short of the moderate model cutoff. Coho, pink, chum, and sockeye salmon (*Oncorhynchus nerka*) in both datasets were juveniles, parr, or smolt, while the majority of Dolly Varden (*Salvelinus malma*) were adults. Sampling month was the single most important variable for pink and chum salmon. Although the other salmonids also exhibited strong seasonal occurrence, other physical habitat variables, such as mean tidal height, estuary slope and geomorphology, and precipitation were more important in determining species' presence. Coho salmon occurrence was higher in estuaries with greater stream lengths.

For the other species with strong and moderate models, kelp greenling occurrence was associated with lower than average tidal exchanges and deep estuary slopes. Silverspot sculpin were associated with low January and February precipitation and shallow depths. Bay pipefish (*Syngnathus leptorhyncus)* occurrence was associated with high percentages of continuous and patchy eelgrass, and high minimum precipitation. Occurrence of three-spine sticklebacks (*Gasterosteus aculeatus)* was positively associated with a large amount of estuarine and total watershed stream channel and negatively associated with higher estuarine slopes. Tubesnout occurrence was associated with the ShoreZone habitat class for protected sediment substrates with shallow depths.

Subtidal vegetation and substrate type were important variables for most of the invertebrates. Cancer crabs (*Cancridae*), sunflower sea stars (*Pycnopodia helianthoides),* red rock crabs (*Cancer productus*), broken back shrimp (*Hippolyte sp.),* shell shrimp (*Crangon alaskensis*), dock shrimp *(Pandalus danae)*, and helmet crabs (*Telmessus cheiragonus*) were all positively associated with high percentages of patchy and continuous eelgrass. The graceful decorator crab (*Oregonia gracilis*) was positively associated with higher percentages of patchy

red algae, and the northern kelp crab (*Pugettia producta*), coonstripe shrimp (*Pandalus hypsinotus*), and broken back shrimp were positively associated with higher percentages of soft brown kelp. Most invertebrates were positively associated with the sediment and rock and sediment ShoreZone primary coastal classes, with subclass associations varying by species. Several invertebrates showed strong correlation to estuaries with low tidal exchanges, including dock shrimp, Dungeness crab (*Metacarcinus magister*), red rock crab, mottled sea star (*Evasterias troschelii)*, and graceful decorator crabs.

Precipitation was an important variable for most invertebrate species, with many species exhibiting non-linear and seasonal relationships with minimum precipitation. The graceful decorator crab was positively associated with low August and June precipitation but also with high April and November precipitation. Other species, including the red rock crab and mottled sea star also were related to seasonal precipitation patterns. Depth and slope were important variables for most species. Figure 2.9 shows partial dependence plots for maximum estuary slope for Dungeness crab, red rock crab, and graceful decorator crab. Dungeness crab occurrence was higher in estuaries with steeper slopes and red rock and graceful decorator crabs occurred in estuaries with shallower slopes and depths. Dungeness crab presence was also strongly associated with higher amounts of intertidal area.

DISCUSSION

The majority of the models developed in this study had good ability to predict species' presence as validated by the independent and OOB data. Seasonality in the data may explain why some models were better at predicting presence than absence. For example, both chum salmon and pink salmon occur in the estuaries seasonally as juveniles with highest abundances occurring between March and June. Dungeness crab also have been shown to exhibit seasonal patterns of occurrence in shallow estuarine waters. Female crabs are found in shallow water (<8 m) from April to June, and male crabs are found in water less than 25 m from April to July (Stone and O'Clair, 2001). Graceful decorator crabs have seasonal abundance patterns in relation to kelp density (Daly and Konar, 2010), and juvenile Pacific herring exhibit seasonal movement between shallow and deep water during their estuarine residence (Penttila, 2007). Because most of the estuaries were sampled only once, species many not have been present when the estuary was sampled. For estuaries that contain habitat appropriate for these seasonal species, the models may

predict the species to be present even where sample and validation data do not have a record of species' presence. This would result in a false-positive prediction.

For the salmonids, an additional factor limiting the accuracy of predicted absences may be the restricted amount of stream information in the data. Each salmon species has different stream velocity and morphology preferences (Keeley and Slaney, 1996; Vadas, 2000), but the stream variables used in the models do not include information on stream width or flow. Model results are based on variables predominantly associated with the estuary itself. The model for sockeye salmon was considerably weaker than for the other salmonids. Sockeye spawn primarily in lacustrine systems and the presence of lakes was not included as a predictor variable in the analysis. It is likely this is partially responsible for the relatively small percent of correct presence occurrence predictions for sockeye salmon. Estuaries with habitat variables important to sockeye salmon, but without lakes, were misclassified as sites where sockeye salmon should be present.

Species distribution models predict the potential distribution of a species, not necessarily the realized distribution (Jiménez-Valverde et al., 2008; Lawler et al., 2011; Magness et al., 2010; Wei et al., 2010). Species may not occur in suitable habitats for a variety of reasons, including biotic interactions such as competition or predation, or dispersal limitations, and this can result in false negatives in models where these factors are not considered. While it is generally agreed that reliable species' absence data can substantially improve species distribution modeling (Gu and Swihart, 2004; Lobo, 2008), determining that a species is absent from a habitat is much more difficult than determining that a species is present and requires a substantial sampling effort which, over large spatial scales, is often cost prohibitive. In the absence of comprehensive absence data, model results could be improved by incorporating additional information on species interactions and co-occurrence, which would provide insight into the importance of biotic interactions on species distributions. Including predator or prey species as predictor variables (Leathwick and Austin, 2001) and incorporating dispersal vectors (Boulangeat et al., 2012) are two approaches that have been used in species distribution models to account for biotic interactions, and could be included in Random Forest models in the future.

Interpreting the importance of individual predictor variables to species-environment relationships is challenging in multi-variable models (Knudby et al., 2010; Magness et al., 2010; Oppel and Huettmann, 2010). Unlike most linear models, Random Forest models take into account interactions among variables (Lunetta et al., 2004), and variables that do not have high

variable importance rankings may be influencing and increasing the importance rankings of other variables. Random Forest also does not assess the statistical significance of variables to the response in the same manner as frequentist statistical approaches (Cutler et al., 2007); its strength lies in the ability to identify a suite of variables with strong influence on the predictive power of the model from a larger number of potential predictor variables. As a result, relationships between species distribution patterns and individual variables must be evaluated in the context of the other variables in the model .

For all of the fish and invertebrate species, many of the variables that were identified by the models as being most important in determining species occurrence have been identified in other studies to be important habitat characteristics. Examples include a positive association between a high proportion of intertidal area and species occurrence for Dungeness crab (Holsman et al., 2003), and a positive association between depth and sediment type for occurrence of the flatfishes (Abookire et al., 2001; Moles and Norcross, 1995; Norcross et al., 1997).

Other variables linked to occurrence of many of the species in this study may point to new relationships that other modeling approaches have not detected, a common feature in machine learning models (Huettmann and Diamond, 2001) . One such variable is minimum precipitation, which was among the most important variables for many of the invertebrates and fishes, with individual species exhibiting positive or negative associations. However, the mechanism by which precipitation influences species occurrence is not entirely clear and requires further investigation. Precipitation and fluvial flow variables are included in the analysis to capture differences in salinity and buoyancy-driven circulation between estuaries, since remotely-sensed data for these processes are not available at the spatial scale of this analysis. Precipitation also may be a proxy for other oceanographic processes. High precipitation can freshen the surface water layer and result in stratification, which can enhance primary productivity. High freshwater discharge also is associated with the development of tidal fronts, areas of mixing that occur at the interface between stratified water and well mixed saline water as a result of tidal inflow into the estuary. Constrictions, such as a narrowing of the estuary mouths, act as hydraulic controls that can enhance the formation of these fronts (Wedding et al., 2011). Nutrients are drawn into the stratified surface layer of the front by diapycnal mixing at the frontal boundary, and over a period of time can enhance phytoplankton production (Macreadie et al., 2010). At the same time, convergent flows along the frontal boundary advect and concentrate plankton, which attracts grazers and higher trophic level predators. These fronts may also act as barriers to larval

transport helping to retain and distribute planktonic larvae within the estuary (Genuer et al., 2010; Svetnik et al., 2002). It is likely that the precipitation variable is influencing individual species differently with some responding to increased freshening of the water and lower salinity, while others may be more strongly associated with productivity or other oceanographic frontal processes.

Precipitation also may be influencing an aspect of the habitat, such as eelgrass abundance, rather than the fish or invertebrates themselves. Variation in precipitation between estuaries is strongly influenced by Southeast Alaska's complex geography, which also influences wind and air temperatures (Weingartner et al., 2009). As a result, the relationship may not be between species and precipitation, but could be between the species and independent variables not included in the analysis for which precipitation is acting as a proxy. This points out that, while accurate predictive models can be developed using landscape scale variables, any mechanisms underlying species-variable relationships need to be evaluated with additional data or research (see Drew et al. 2011).

Predicted species occurrences from the models can be mapped using GIS to evaluate spatial patterns. Mapping one or more species together can provide insight into habitat partitioning, especially for species with overlapping habitat ranges. For example, Dungeness crab and red rock crab both occur in estuaries along the Pacific coast. Mapping Dungeness crab and red rock crab together (Figure 2.8) showed areas of predicted occurrence overlap and areas where one species is predicted to occur without the other. Dungeness crab are widely distributed in estuaries along the inner channels of Southeast Alaska, while red rock crab tend to have higher predicted occurrence along the coast. Dungeness crab are common inhabitants of large estuaries throughout the Pacific coast of North America (Fisher and Velasquez, 2008). These crab are thought to prefer waters deeper than 15 m but regularly forage in shallower and intertidal waters as both adults (Curtis and McGaw, 2012) and juveniles (Holsman et al., 2003). Our predictive models indicated a high occurrence of this species in estuaries with steep slopes, which could make it easier for the crabs to move between shallow and deep waters. Red rock crab also are a common estuarine species, but as osmoconformers (Carroll and Winn, 1989) they do not tolerate low salinity water. Surface salinities in the study area tend to decrease with distance from the ocean, and this may be a factor affecting red rock crab predicted distributions in our models. Our model results predicted relatively low overlap between the two species. This is likely the result of different habitat preferences discussed above, but may be partially attributed to sampling bias.

Red rock crab tend to prefer harder substrates with more complex structure than Dungeness crab (Holsman et al., 2006). In Southeast Alaska, estuaries with steep slopes and sandy-mud substrates, the habitat type preferred by Dungeness crab, tend to have rocky, complex habitat toward the estuary mouth and outside the sampling area, making it less likely that red rock crab would be captured in the same samples as Dungeness crab.

CONCLUSIONS

Species distribution models that accurately predict species occurrence in relation to habitat characteristics provide an important tool for conservation and ecosystem management; however, the ability to develop accurate models is often constrained by a lack of *in-situ* habitat data. Although remotely-sensed environmental data are commonly used to develop distribution models for terrestrial species, the application of these data to marine species has been limited. This research demonstrated that ecologically useful predictive models can be developed for estuarine-associated fishes and invertebrates using regional-scale spatial datasets as a surrogate for environmental data collected in the field. The strongest models in this study were able to accurately predict between 100% and 63% of species' presences with respect to independent data. Variable importance measures from the models confirmed some species habitat relationships documented in previous research and highlighted new relationships that, with additional investigation, could further refine our knowledge of the factors influencing spatial distribution patterns. Model results for individual species can be plotted in a GIS to reveal distribution patterns with respect to both biotic and abiotic factors, which can be used to inform management decisions. In our data, where several species occurred more frequently as juveniles than adults, these maps can assist in identifying nearshore areas important to fishes and invertebrates at early life stages. By incorporating variables such as precipitation and vegetation cover, the models can also be used to evaluate the effects of habitat alteration, including climate change, on species distributions.

In this research, we focused on species occurrence to be able to evaluate the models against independent data collected using different gear, but predictive models for species abundance also could be developed using similar methods and data. The influence of local landscape variables, such as bottom type and proximity to seagrass beds, on fish abundance has been demonstrated (Gratwicke and Speight, 2005; Peltonen et al., 2007), and landscape-scale variables have been used to predict total fish abundance and guild abundance on coral reefs

(Kendall et al., 2011). In Alaska, *in-situ* variables such as temperature and salinity, have been evaluated against relative abundance of species in nearshore environments (Abookire et al., 2000; Mueter and Norcross, 1999), but there are no studies linking remotely-sensed landscape variables to nearshore species abundance. Application of these methods to species abundance is planned in future research.
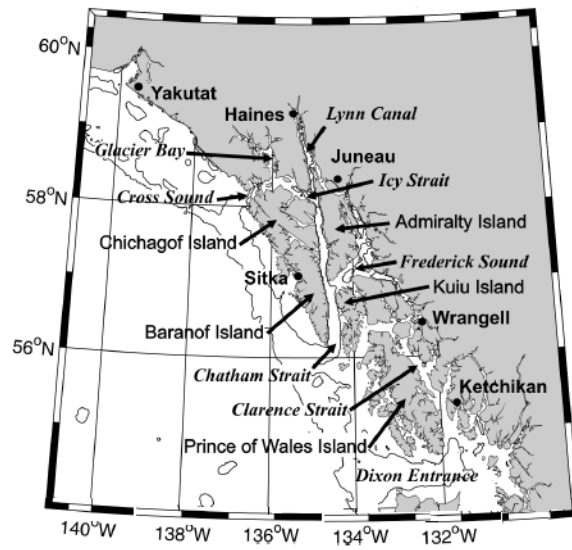
Figure 2.1: Map of the study area with major channels in italics and towns in bold (adapted from Weingartner et al. 2009)
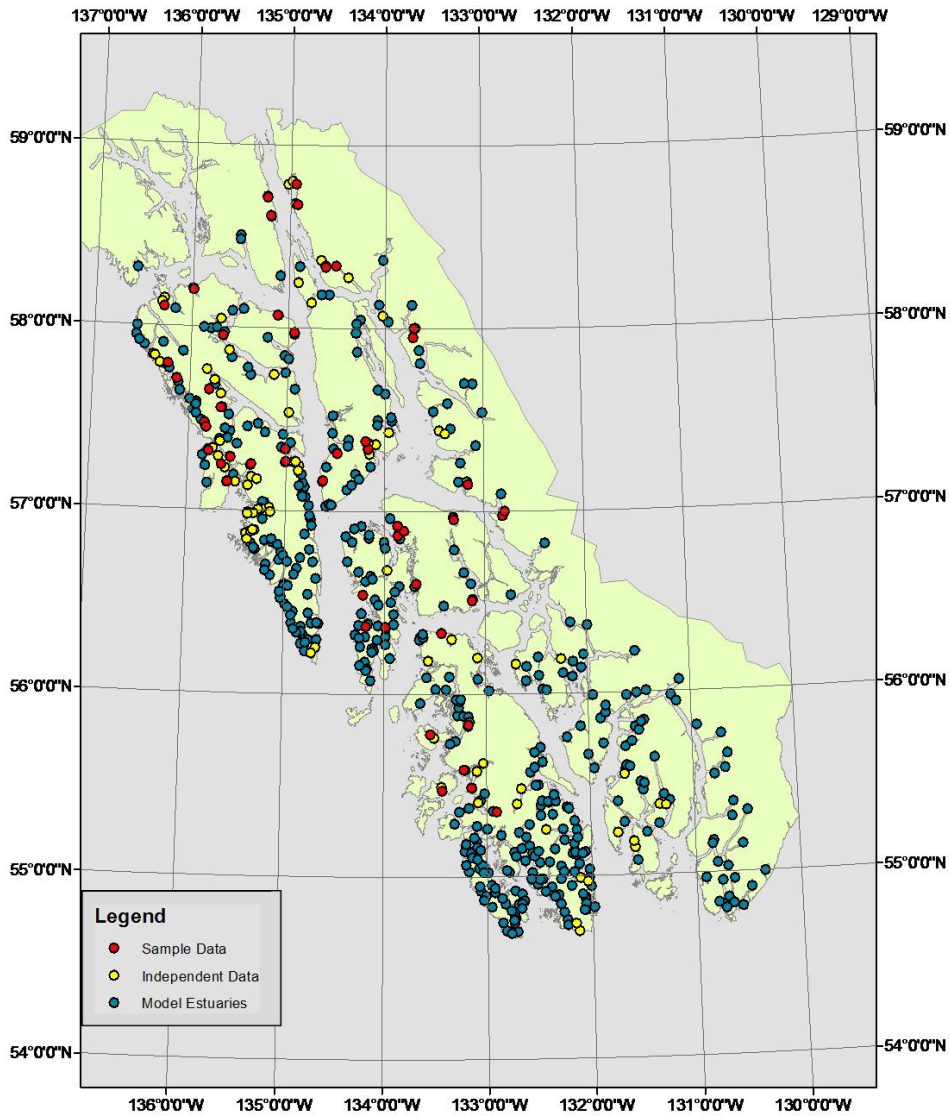
Figure 2.2: Estuaries in southeast Alaska. Red dots are sampling sites used in model development, yellow dots are sample sites for independent data, blue dots are the estuaries to which model results were predicted
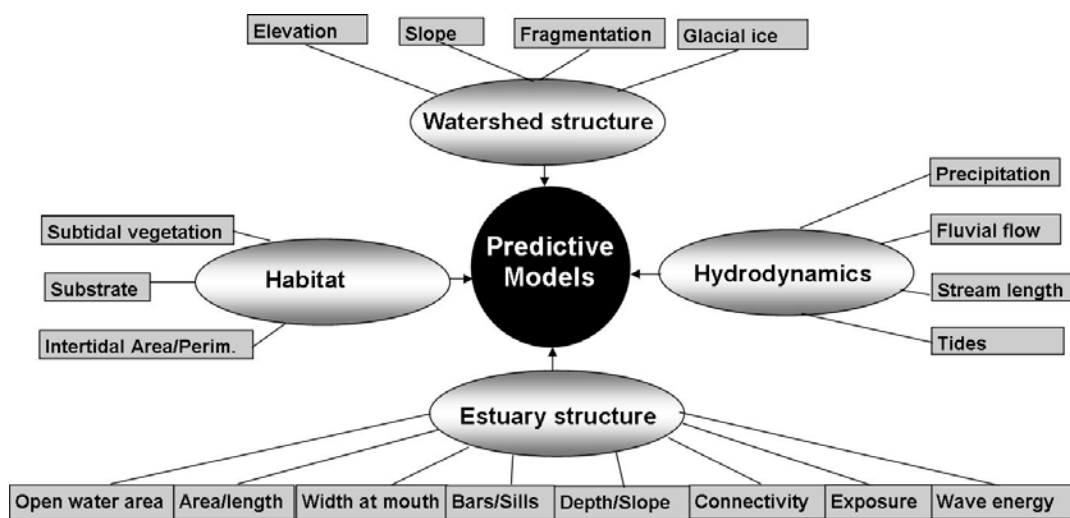
Figure 2.3: Conceptual model of a typical estuary showing relationship of model predictor variables to estuarine system components: structure, habitat, and hydrodynamics.
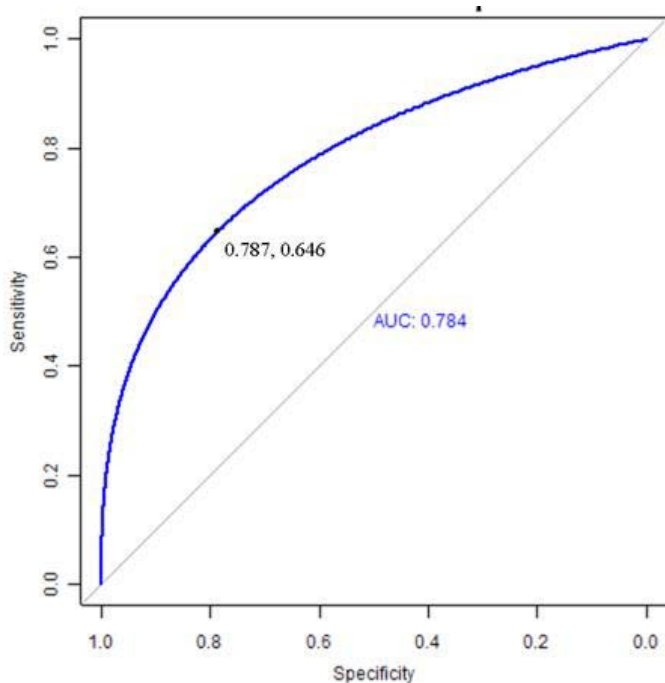
Figure 2.4: Example of a receiver operator characteristic (ROC) curve. The curve is the decision threshold between true positive classification (specificity), and false positive classification (sensitivity) for both presence and absence. The area under the curve (AUC) is the average predictive ability of the model. The optimal decision threshold is the point where the slope of the curve is equal to 1. Numbers along the curve identify the "best" threshold: or the highest sum of sensitivity and specificity.
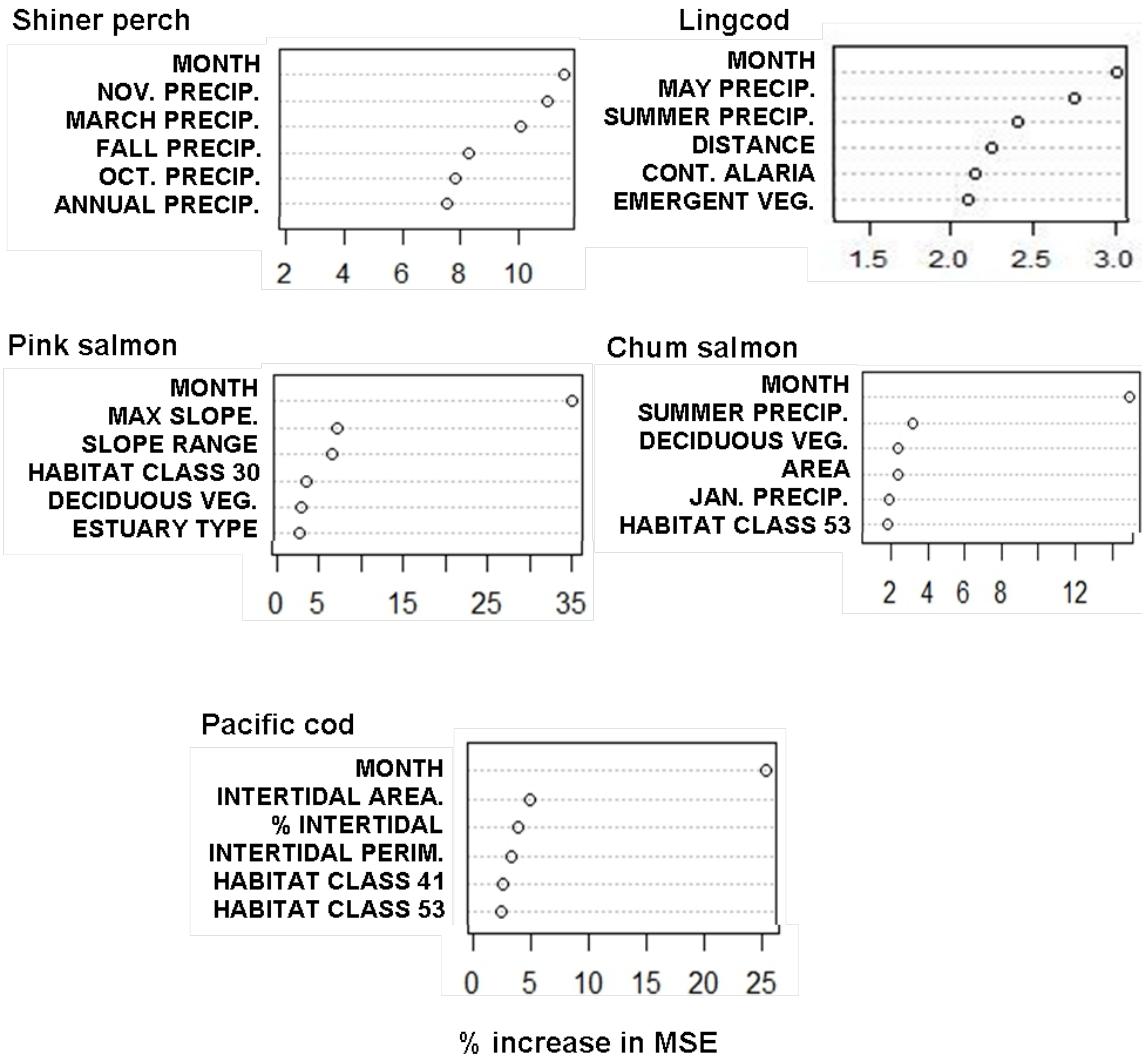
Figure 2.5: Top portion of the variable importance plots for species from the top five fish species' models X-axis is the change in model performance when the variable is permuted and the y-axis is the variable code.
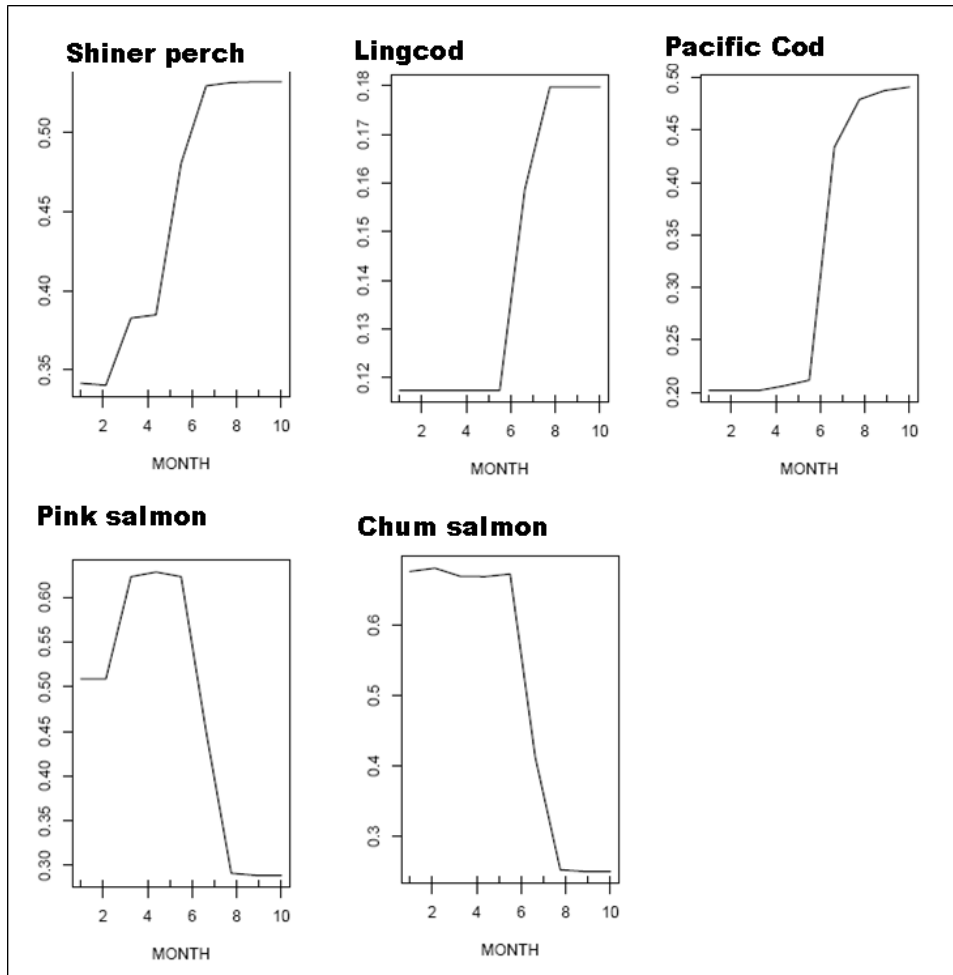
Figure 2.6: Partial dependence plots for the top five fish species models for the variable "month." X-axis is the numerical month and y-axis is the average trend of the variable in the model.
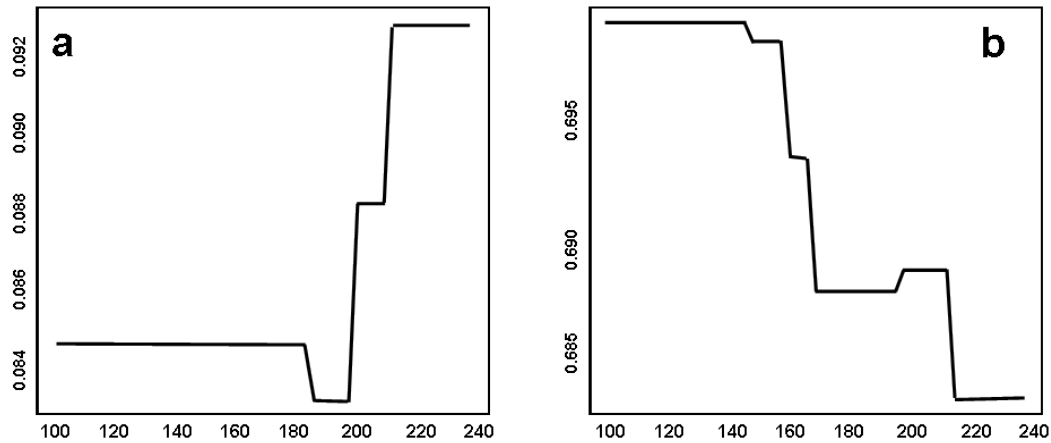
Figure 2.7: Partial dependence plots of the occurrence of butter sole (a) and starry flounder (b) against minimum February precipitation in mm. The x-axis is the range of the variable in the data. The y-axis is the average trend of the variable in the model. The shapes of the curves can be compared to visualize relationships between predicted species occurrence and variable values.
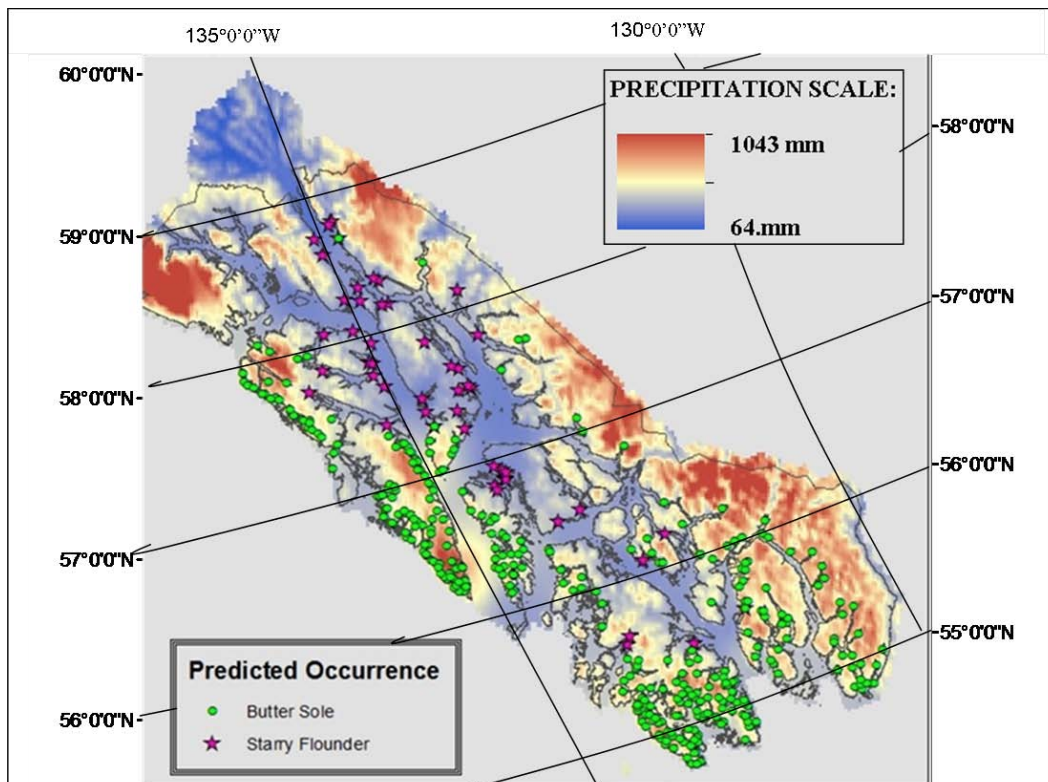
Figure 2.8: Predicted occurrence of starry flounder (red stars) and butter sole (green circles) with respect to average annual precipitation during the study period with warmer (reds) colors representing higher precipitation.
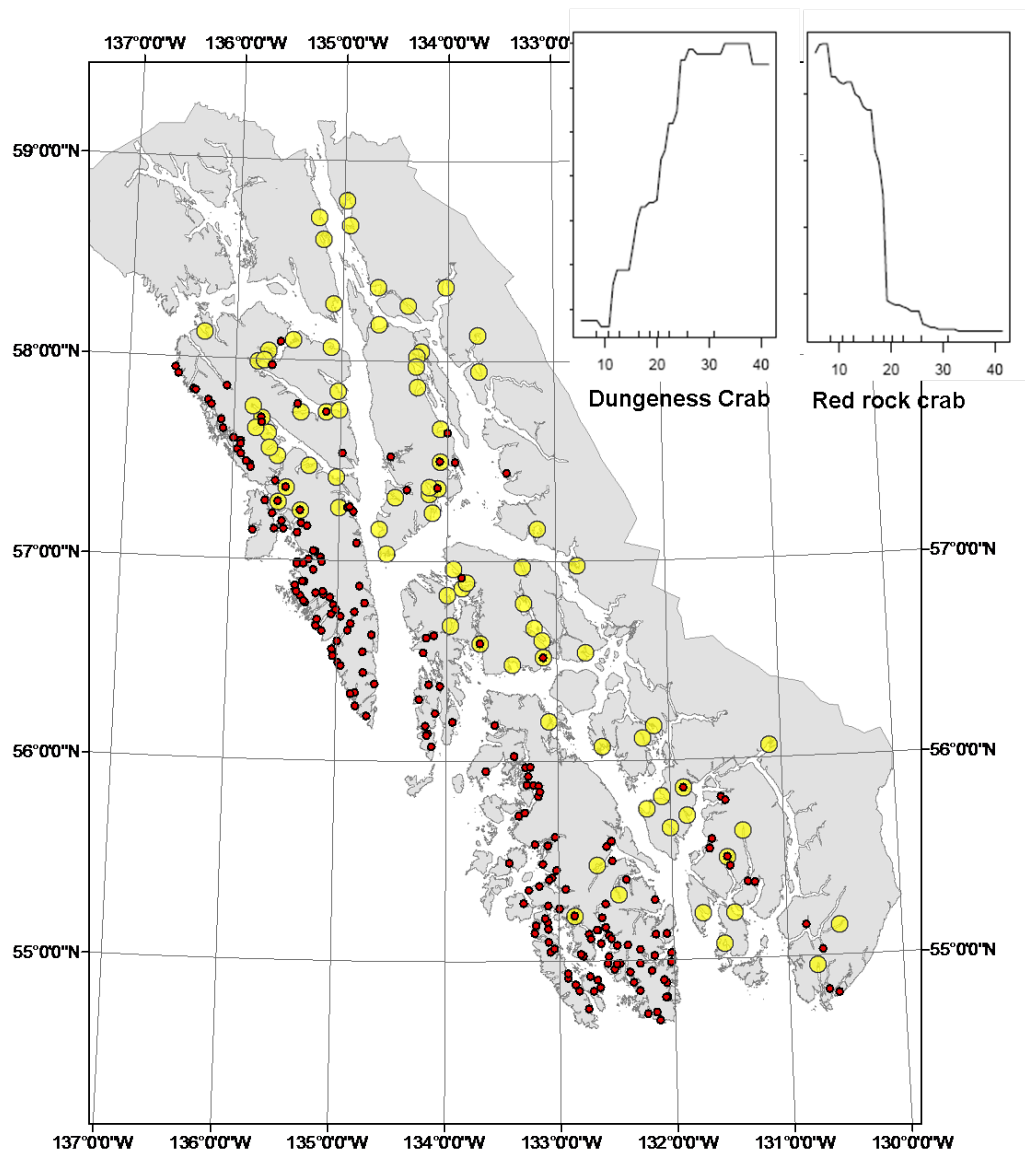
Figure 2.9: Partial dependence plot of Dungeness crab and red rock crab in relation to maximum estuary slope, and predicted distribution for Dungeness crab (yellow) and red rock crab (red) in the study area. Yellow circles with red centers indicate estuaries where the species are predicted to overlap

Table 2.1: Independent variables used in predictive models. NA= data without a spatial scale.

Type	No.	Unit	Time Scale	Spatial Scale of data source	Source
Estuary Area	1	Sq. Meters	NA	1:63,360	USFS Tongass GIS - derived
Estuary Perimeter	1	Meters		1:63,360	USFS Tongass GIS - derived
Intertidal Area	1	Sq. Meters		1:63,360	USFS Tongass GIS
Intertidal Perimeter	1	Meters		1:63,360	USFS Tongass GIS
Open water	1	Sq. Meters		1:63,360	USFS Tongass GIS - derived
Watershed Area	1	Sq. Meters		1:63,360	USGS Hydrologic Unit Maps

Streams in watershed
  Estuarine streams
  Total streams
  Percent estuarine
	3	Meters		1:63,360	USFS Tongass GIS

Tidal range	2	Feet		NA	NOAA
Type of waterbody	2	Category		NA	Derived
Distance to waterbody	1	Meters		1:63,360	Derived
Width	1	Meters		1:63,360	measured
Length	1	Meters		1:63,360	measured
Depth of bar/sill	1	Meters		5 m	NMFS AKR Bathymetry -
Width of bar/sill	1	Meters		1:63,360	measured

Estuary slope
  Mean slope
  Maximum slope
  Range of slope
	3	Degrees		5 m	NMFS AKR Bathymetry - derived

Depth
  Mean depth
  Maximum depth
  Range of depth
	3	Meters		5 m	NMFS AKR Bathymetry - derived

Land cover patchiness
  Total area
  Number of patches
  Patch density
  Largest patch index
  Landscape shape index
	5	Varies		30m	2001 National Land Cover Dataset

Annual precipitation	1	Millimeters	1998-2005	2 km	PRISM Climate Model
Monthly precipitation	12	Millimeters	1998-2005	2 km	PRISM Climate Model
Fluvial flow	1	Flow/sq m	1998-2005	NA	Derived

Land cover
  Ice
  Developed
  Barren
  Deciduous
  Evergreen
  Mixed vegetation
  Dwarf
  Scrub-shrub
  Woody wetlands
  Emergent herbaceous
	10	Percent		30 m	2001 National Land Cover Dataset

Slope of watershed	2	Degrees		300 m	USGS Digital Elevation Model
Habitat class	17	Percent		Coastal unit	Alaska ShoreZone dataset
Geology class	25	Percent		Coastal unit	Alaska ShoreZone dataset

Inter/subtidal vegetation (continuous/patchy)
  Red algae
  Aleria
  Soft brown kelp
  Dark brown kelp
  Eelgrass
	8	Percent		Coastal unit	Alaska ShoreZone dataset

Table 2.2: Occurrence model results for fishes and invertebrates sorted by area under the curve (AUC). For life stage, capital letters indicate higher occurrence in the data than lower case letters. A = adults, J=juveniles, A/J = equal numbers adults and juveniles, A(j) = more adults than juveniles, J(a) = more juveniles than adults, and M=species for which life stage could not be determined based on length of fish in samples. An * indicates that the model for that species was validated from OOB rather than independent data.

FISH SPECIES	Life Stage	% Present Correct	% Absence Correct	AUC	INVERTEBRATES	% Presence Correct	% Absence Correct	AUC
Shiner perch					Dock shrimp*			
Cymatogaster aggregata	A/J	100	87	0.94	*Pandalus danae*	88	89	0.94
Lingcod					Graceful decorator crab*			
Ophiodon elongatus	J	100	85	0.93	*Oregonia gracilis*	83	79	0.84
Pink salmon					Gammarid isopods*			
Oncorhynchus gorbuscha	J	86	77	0.85	*Gammaridae*	76	75	0.80
Chum salmon					Pandalid shrimp*			
Oncorhynchus keta	J	82	81	0.85	*Pandalidae*	73	66	0.80
Pacific cod					Red rock crab*			
Gadus macrocephalus	J	78	78	0.84	*Cancer productus*	80	65	0.79
Butter sole					Sunflower starfish*			
Isopsetta isolepis	J	75	74	0.84	*Pycnopodia hellianthoides*	84	63	0.77
Kelp greenling					Broken back shrimp*			
Hexagrammos decagrammus	J	81	70	0.82	*Hippolytidae*	71	64	0.75
Starry flounder					Coonstripe shrimp*			
Platichthys stellatus	A/J	86	63	0.77	*Pandalus hypsinotus*	70	64	0.75
Rock sole					Spot shrimp			
Lepidopsetta sp.	J(a)	72	63	0.77	*Pandalus platyceros*	71	64	0.72
Dolly varden					Cancer crabs*			
Salvelinus malma	A(j)	72	63	0.76	*Cancer sp.*	60	59	0.71
Yellowfin sole					Stiletto shrimp*			
Limanda aspera	J(a)	72	63	0.75	*Heptacarpus stylus*	68	52	0.70
Silverspot sculpin*					Northern kelp crab*			
Blepsias cirrhosus	J	67	69	0.75	*Pugettia producta*	75	57	0.70
Sockeye salmon					Dungeness crab*			
Oncorhynchus nerka	J	69	66	0.74	*Metacarcinus magister*	71	72	0.70
Pacific herring					Isopods*			
Clupea pallasi	J	54	20	0.72	*Ispodoa*	70	60	0.67
Tubesnout					Crangons*			
Aulorhynchus flavidus	M	76	60	0.71	*Crangonidae*	81	60	0.66
Bay Pipefish					Helmet crab*			
Syngnathus leptorhynchus	A/(j)	71	62	0.70	*Telmessus cheiragonus*	63	50	0.65
Sturgeon Poacher					Mottled sea star*			
Podothecus accipenserinus	M	68	15	0.70	*Evasterias troschelii*	71	60	0.64
Pacific staghorn sculpin					Pacific lyre crab*			
Leptocottus armatus	A/J	50	20	0.70	*Hyas lyratus*	60	56	0.62
Threespine stickleback								
Gasterosteus aculeatus	A	63	65	0.70				
Coho salmon								
Oncorhynchus kisutch	J	68	30	0.68				
Pacific sand lance								
Ammodytes hexapterus	J(a)	47	34	0.68				
Snake prickleback*								
Lumpenus sagitta	M	61	59	0.67				
Buffalo sculpin								
Enophrys bison	M	42	40	0.64				
Crescent gunnel								
Pholis laeta	A(j)	60	50	0.63				
Great sculpin								
M. polyacanthocephalus	J	72	63	0.63				

REFERENCES

Abookire, A.A., Piatt, J.F., Robards, M.D., 2000. Nearshore fish distributions in an Alaskan estuary in relation to stratification, temperature and salinity. Estuarine, Coastal and Shelf Science 51, 45-59.

Abookire, A.A., Piatt, J.F., Norcross, B.L., 2001. Juvenile groundfish habitat in Kachemak Bay, Alaska during late summer. Fishery Bulletin 8, 45-56.

Albert, D., Shanley, C., Baker, L., 2010. A preliminary classification of bays and estuaries in Southeast Alaska: A hierarchical framework and exploratory analysis, Coastal Ecological Systems in Southeast Alaska. The Nature Conservancy (http://conserveonline.org/library/a-preliminary-classification-of-bays-and-estuaries/view.html), Juneau, AK.

Anderson, T.J., Syms, C., Roberts, D.A., Howard, D.F., 2009. Multi-scale fish-habitat associations and the use of habitat surrogates to predict the organization and abundance of deep-water fish assemblages. Journal of Experimental Marine Biology and Ecology 379, 34-42.

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

Boulangeat, I., Gravel, D., Thuiller, W., 2012. Accounting for dispersal and biotic interactions to disentangle drivers of species distributions and their abundances. Ecology Letters 15, 584-593.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5-32.

Brenden, T.O., Wang, L., Clark Jr., R.D., Seelbach, P.W., 2007. Comparison between model-predicted and field-measured stream habitat features for evaluating fish assemblage-habitat relationships. Transactions of the American Fisheries Society 136, 580-592.

Carroll, J.C., Winn, R.N., 1989. Species Profiles: Brown rock crab, red rock crab and yellow rock crab. Biological Report 82(11.117).

Cèrèghino, R., Santoul, F., Compin, A., Figuerola, J., Mastrorillo, S., 2005. Co-occurrence patterns of some small-bodied freshwater fishes in southwestern France: implications for fish conservation and environmental management. Ambio 34, 440-444.

Cunningham, R.B., Lindenmayer, D.B., 2005. Modeling count data of rare species: some statistical issues. Ecology 86, 1135-1142.

Curtis, D.L., McGaw, I.J., 2012. Salinity and thermal preference of Dungeness crabs in the lab and in the field: Effects of food availability and starvation. Journal of Experimental Marine Biology and Ecology 413, 113-120.

Cushman, S., Huettmann, F., 2010. Spatial Complexity, Informatics and Wildlife Conservation. Springer, Tokyo, Japan.

Cutler, D.R., Edwards Jr., T.C., Bears, K.H., Cutler, A., Hess, K., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783-2792.

Daly, B., Konar, B., 2010. Temporal trends of two spider crabs (*Brachyura, Majoidea*) in nearshore kelp habitats in Alaska, USA. Crustaceana 83, 659-669.

Digby, M.J., Saenger, P., Whelan, M.B., McConchie, D., Eyre, B., Holmes, N., Bucher, D., 1998. A physical classification of Australian estuaries. Centre for Coastal Management, Southern Cross University,Urban Water Research Association of Australia, (http://au.riversinfo.org/library/nrhp/estuary_clasifn/).

Drew, C.A., Wiersma, Y., Huettmann, F., 2011. Predictive Modeling in Landscape Ecology. Springer, New York, NY.

Dye, A.H., 2006. Is geomorphic zonation a useful predictor of patterns of benthic infauna in intermittent estuaries in New South Wales, Australia. Estuaries and Coasts 29, 455-464.

Elith, J., Leathwick, J., 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology and Systematics 40, 677-697.

Elmendorf, S.C., Moore, K.A., 2008. The use of community-composition data to predict the fecundity and abundance of species. Conservation Biology 22, 1523-1532.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters 27, 861-874.

Fisher, W., Velasquez, D., 2008. Management recommendations for Washington's priority habitats and species: Dungeness crab (*Cancer magister*). Washington Department of Fish and Wildlife.

Genuer, R., Poggi, J.-M., Tuleau-Malot, C., 2010. Variable selection using random forests. Pattern Recognition Letters 31, 2225-2236.

Gratwicke, B., Speight, M.R., 2005. The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. Journal of Fish Biology 66, 650-667.

Gray, J.S., Bjørgesæter, A., Ugland, K.I., 2005. The impact of rare species on natural assemblages. Journal of Animal Ecology 74, 1131-1139.

Gray, J.S., Bjørgesæter, A., Ugland, K.I., 2006. On plotting species abundance distributions. Journal of Animal Ecology 75, 752-756.

44

Gu, W., Swihart, R.K., 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. Biological Conservation 116, 195-203.

Gutiérrez-Estrada, J.C., Vasconcelos, R., Costa, M.J., 2008. Estimating fish community diversity from environmental features in the Tagus estuary (Portugal): Multiple linear regression and artificial neural network approaches. Journal of Applied Ichthyology 24, 150-162.

Hamel, L., 2008. Model assessment with ROC curves, in: Wang, J. (Ed.), The Encyclopedia of Data Warehousing and Mining, 2nd Edition. Idea Group Publishers, Hershey, PA, p. 2542.

Hardy, S., Lindgren, M., Hanumantharao, K., Huettmann, F., 2011. Predicting the distribution and ecological niche of unexploited snow crab (*Chionoecetes opilio*) populations in Alaska waters: A first open-access ensemble model. Integrative and Comparative Biology 51, 608-622.

Harney, J., Morris, M., Harper, J., 2008. Shorezone coastal habitat mapping protocol for the Gulf of Alaska. Coastal & Ocean Resources, Inc. (http://www.fakr.noaa.gov/shorezone).

Hicks, A., Barbee, N.C., Swearer, S.E., Downes, B.J., 2010. Estuarine geomorphology and low salinity requirement for fertilisation influence spawning site location in the diadromous fish, *Galaxias maculatus*. Marine and Freshwater Research 61, 1252-1258.

Holsman, K., McDonald, P.S., Armstrong, D.A., 2006. Intertidal migration and habitat use by subadult Dungeness crab, *Cancer Magister*, in a NE Pacific estuary. Marine Ecology Progress Series 308, 183-195.

Holsman, K., Armstrong, D.A., Beauchamp, D.A., Reusink, J.L., 2003. The necessity for intertidal foraging by estuarine populations of subadult Dungeness crab, *Cancer Magister*: Evidence from a bioenergetics model. Estuaries 26, 1155-1173.

Huettmann, F., Diamond, A.W., 2001. Using PCA scores to classify species communities: an example using seabird classifications at sea. Journal of Applied Statistics 28, 843-853.

Inazu, D., Sato, T., Miura, S., Ohta, Y., Nakamura, K., Fujimoto, H., Larsen, C.F., Higuchi, T., 2009. Accurate ocean tide modeling in Southeast Alaska and large tidal dissipation around Glacier Bay. Journal of Oceanography 65, 335-347.

Jelbart, J.E., Ross, P.M., Connolly, R.M., 2006. Edge effects and patch size in seagrass landscapes: an experimental test using fish. Marine Ecology Progress Series 319, 93-102.

Jimenez-Valverde, A., Lobo, J.M., 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecologica 31, 361-369.

Jiménez-Valverde, A., Lobo, J.M., Hortal, J., 2008. Not as good as they seem: the importance of concepts in species distribution modelling. Diversity and Distributions 14, 885-890.

Johnson, S.W., Neff, A.D., Thedinga, J., 2005. An atlas on the distribution and habitat of common fishes in shallow nearshore waters of southeastern Alaska. U.S. Department of Commerce NOAA Technical Memo NMFS-AFSC-157, 89.

Johnson, S.W., Murphy, M.L., Csepp, D.J., Harris, P., Thedinga, J., 2003. A survey of fish assemblages in eelgrass and kelp habitats of Southeast Alaska. US Department of Commerce, NMFS-AFSC-139.

Keeley, E.R., Slaney, P.A., 1996. Quantitative measures of rearing and spawning habitat characteristics for stream-dwelling salmonids: Guidelines for habitat restoration, Watershed Restoration Project Report No. 4. Watershed Restoration Program, Ministry of Environment, Lands and Parks and Ministry of Forests. URL: env.gov.bc.ca/wld/documents/wrp/wrpr_4.pdf, Vancouver, British Columbia.

Kendall, M.A., Miller, T.J., Pittman, S.J., 2011. Patterns of scale-dependency and the influence of map resolution on the seascape ecology of reef fish. Marine Ecology Progress Series 427, 259-274.

Knudby, A., Brenning, A., LeDrew, E., 2010. New approaches to modelling fish-habitat relationships. Ecological Modelling 221, 503-511.

Lawler, J.J., Wiersma, Y., Huettmann, F., 2011. Designing predictive models for increased utility: Using species distribution models for conservation planning, forecasting and risk assessment, in: Drew, C.A., Wiersma, Y., Huettmann, F. (Eds.), Predictive Modeling in Landscape Ecology. Springer, New York, NY, p. 314.

Leathwick, J., Austin, M.P., 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. Ecology 82, 2560-2573.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2/3, 18-22.

Lobo, J.M., 2008. More complex distribution models or more representative data? Biodiveristy Informatics 5, 14-19.

Lobo, J.M., Jimenez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17, 145-151.

Lunetta, K.L., Hayward, L.B., Segal, J., Van Eerdewegh, P., 2004. Screening large-scale association study data: exploiting interactions using random forests. BMC Genetics 5, 32-45.

Macreadie, P.I., Hindell, J.S., Keough, M.J., Jenkins, G.P., Connolly, R.M., 2010. Resource distribution influences positive edge effects in a seagrass fish. Ecology 91, 2013-2021.

Magness, D.R., Huettmann, F., Morton, J.M., 2010. Using random forests to provide predicted species distribution maps as a metric for ecological inventory and monitoring programs, in: Cushman, S., F. Huettmann (Ed.), Spatial complexity, informatics and wildlife conservation. Springer, New York, p. 464 p.

Matthews, J.B., 1981. The seasonal circulation of the Glacier Bay, Alaska fjord system. Estuarine, Coastal and Shelf Science 12, 679-700.

Maxwell, D.L., Stelzenmuller, V., Eastwood, P.D., Rogers, S.I., 2009. Modelling the spatial distribution of plaice (*Pleuronectes platessa*), sole (*Solea solea*), and thornback ray (*Raja clavataI*) in UK waters for marine management and planning. Journal of Sea Research 61, 258-267.

McArdle, B.H., Anderson, M.J., 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology 82, 290-297.

McGarigal, K., Cushman, S., Ene, E., 2002. FRAGSTATS: Spatial pattern analysis program for categorical maps.Available at the following website: http://www.umas.edu/landco/research/fragstats/fragstats.html. University of Massachusetts, Amherst.

Moles, A., Norcross, B.L., 1995. Sediment preference in juvenile Pacific flatfishes. Netherlands Journal of Sea Research 34, 177-182.

Mueter, F.J., Norcross, B.L., 1999. Linking community structure of small demersal fishes around Kodiak Island, Alaska to environmental variables. Marine Ecology Progress Series 190, 37-51.

Mueter, F.J., Litzow, M.A., 2008. Sea ice retreat alters the biogeography of the Bering Sea continental shelf. Ecological Applications 18, 309-320.

Murphy, M.L., Johnson, S.W., Csepp, D.J., 2000. A comparison of fish assemblages in eelgrass and adjacent subtidal habitats near Craig, Alaska. Alaska Fishery Bulletin 7, 11-21.

Neal, E.G., Walter, M.T., Coffeen, C., 2002. Linking the Pacific decadal oscillation to seasonal stream discharge patterns in Southeast Alaska. Journal of Hydrology 263, 188-197.

NOAA, 2012. NOAA Tides and Currents. National Oceanic and Atmospheric Administration,Center for Operational Oceanographic Products, http://tidesandcurrents.noaa.gov.

Norcross, B.L., Mueter, F.J., Holladay, B.A., 1997. Habitat models for juvenile pleuronectids around Kodiak Island, Alaska. Fishery Bulletin 95, 504-520.

Ohse, B., Huettmann, F., Ickert-Bond, S., Juday, G., 2009. Modeling the distribution of white spruce (*Picea glauca*) for Alaska with high accuracy: an open access role-model for predicting tree species in the last remaining wilderness areas. Polar Biology 32, 1717-1724.

Oppel, S., Huettmann, F., 2010. Using a random forest model and public data to predict the distribution of prey for marine wildlife management, in: Cushman, S., Huettmann, F. (Eds.), Spatial Complexity, Informatics and Wildlife Conservation. Springer, Tokyo, Japan, pp. 151-164.

Oppel, S., Strobl, C., Huettmann, F., 2009. Alternative methods to quantify variable importance in ecology. Department of Statistics, University of Munich, Technical Report Number 65, Munich, pp. 6 (http://epub.ub.uni-muenchen.de/10992/10991/Oppeletal_techreport.pdf).

Palialexis, A., Georgakarakos, S., Karakassis, I., Lika, K., Valavanis, V.D., 2011. Fish distribution predictions from different points of view: comparing associative neural networks, geostatistics, and regression models. Hydrobiologia 670, 165-188.

Peltonen, H., Luoto, M., Pääkkönen, J.-P., Tuomaala, A., Pönni, J., Viitasalo, M., 2007. Pelagic fish abundance in relation to regional environmental variation in the Gulf fo Findland, northern Baltic Sea. ICES Journal of Marine Science 64, 487-495.

Penttila, D., 2007. Marine forage fishes in Puget Sound. Washington Department of Fish and Wildlife, www.pugetsoundnearshore.org/technical_papers/marine_fish.pdf.

Pittman, S.J., Christensen, J.D., Caldow, C., Menza, C., Monaco, M.E., 2007. Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. Ecological Modelling 204, 9-21.

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181-199.

Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, M.A., Possingham, H.P., Richardson, A.J., 2011. Pushing the limits in marine species distribution modeling: lessons from the land present challenges. Global Ecology and Biogeography 20, 789-802.

Roy, P.S., Williams, R.J., Jones, A.R., Yassini, I., Gibbs, P.J., Coates, R.J., Scanes, P.R., Hudson, J.P., Nichol, S., 2001. Structure and function of south-east Australian estuaries. Estuarine, Coastal and Shelf Science 53, 351-384.

Schueler, T., 1987. Controlling urban runoff: a practical manual for planning and designing urban BMPs. Department of Environmental Conservation New York State, http:\\www.dec.ny.gov/docs/water_pdf/simple.pdf.

48

Shi, T., Horvath, S., 2006. Unsupervised learning with Random Forest predictors. Journal of Computation and Graphical Statistics 15, 118-138.

SNAP, 2011. Scenarios Network for Alaska and Arctic Planning. University of Alaska, (http://www.snap.uaf.edu/data.php).

Stojlgren, T.J., Ma, P., Kumar, S., Rocca, M., Morisette, J.T., Jarnevish, C.S., Benson, N., 2010. Ensemble habitat mapping of invasive plant species. Risk Analysis 30, 224-235.

Stone, R.P., O'Clair, C.E., 2001. Seasonal movements and distribution of Dungeness crabs, *Cancer magister*, in a glacial southeastern Alaska estuary. Marine Ecology Progress Series 214, 167-176.

Stoner, A.W., Spencer, M.L., Ryer, C., 2007. Flatfish-habitat associations in Alaska nursery grounds: Use of continuous video records for multi-scale spatial analysis. Journal of Sea Research 57, 137-150.

Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning. University of Munich Technical Report Number 55.

Sundblad, G., Harma, M., Lappalainen, A., Urho, L., Bergstrom, U., 2009. Transferability of predictive fish distribution models in two coastal systems. Estuarine, Coastal and Shelf Science 83, 90-96.

Svetnik, V., Liaw, A., Tong, C., 2002. Variable selection in random forest with application to quantitative structure-activity relationship, in: Intrator, N., Masulli, F. (Eds.), 7th Course on Ensemble Methods for Learning Machines, September 22-28, 2002. Springer-Verlag, Salterno, Italy

Termansen, M., McClean, C.J., Preston, C.D., 2006. The use of genetic algorithms and Bayesian classification to model species distributions. Ecological Modelling 192, 410-424.

USDA, 2002. High and low tidelines. USDA Forest Service, Tongass National Forest, Southeast Alaska GIS Library, (http://seakgis.alaska.edu).

USGS, Alaska 300 m digital elevation model. U.S. Geological Survey EROS Alaska Field Office (http://agdc.usgs.gov/data/usgs/erosafo/300m/300m.html), Year accessed: 2009.

USGS, 1995. Alaska hydrologic units. U.S. Geological Survey (http://nhd.usgs.gov/data.html) accessed: 2009.

Vadas, R.L., 2000. Instream-flow needs for anadromous salmonids and lamprey on the Pacific coast, with special reference to the Pacific southwest. Environmental Monitoring and Assessment 64, 331-358.

Valavanis, V.D., Pierce, G.J., Zuur, A.F., Palialexis, A., Saveliev, A., Katara, I., Wang, J., 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. Hydrobiologia 612, 5-20.

Valesini, F.J., Hourston, M., Wilsdsmith, M.D., Coen, N.J., Potter, I.C., 2010. New quantitiative approaches for classifying and predicting local-scale habitats in estuaries. Estuarine, Coastal and Shelf Science 86, 645-664.

Wedding, L.M., Friedlander, A.M., 2008. Determining the influence of seascape structure on coral reef fishes in Hawaii using a geospatial approach. Marine Geodesy 31, 246-266.

Wedding, L.M., Lepcxyk, C.A., Pittman, S.J., Friedlander, A.M., Jorgensen, S., 2011. Quantifying seascape structure: extending terrestrial spatial pattern metrics to the marine realm. Marine Ecology Progress Series 427, 219-232.

Wei, C.-L., Rowe, G.T., Escobar-Briones, E., Boetius, A., Softwedel, T., Caley, M.J., Soliman, Y., Huettmann, F., Qu, F., Yu, Z., Pitcher, C.R., Haedrich, R.L., Wickstein, M.K., Rex, M.A., Baguley, J.G., Sharma, J., Danovaro, R., MacDonald, I.R., Nunnally, C.C., Deming, J.W., Montagna, P., Lévesque, M., Weslawski, J.M., Wlodarska-Kowalczuk, M., Ingole, B.S., Bett, B.J., Billett, D.S.M., Yool, A., Bluhm, B., Iken, K., Narayanaswamy, B.E., 2010. Global patterns and predictions of seafloor biomass using random forests. PlosOne 5, e15323.

Weingartner, T., Eisner, L., Eckert, G.L., Danielson, S., 2009. Southeast Alaska: oceanographic habitats and linkages. Journal of Biogeography 36, 387-400.

Whitlow, W.L., Grabowski, J.H., 2012. Examining how landscapes influence benthic community assemblages in seagrass and mudflat habitats in southern Maine. Journal of Experimental Marine Biology and Ecology 411, 1-6.

Wiens, J.J., 2011. The niche, biogeography and species interactions. Philosophical Transactions of the Royal Society B: Biological Sciences 366, 2336-2350.

Wilkins, B.C., Snyder, N.P., 2011. Geomorphic comparison of two Atlantic coastal rivers: toward an understanding of physical controls on Atlantic salmon habitat. River Research and Applications 27, 135-156.

Williams, J.N., Seo, C., Thorned, J., Nelson, J.K., Erwin, S., O'Brien, J.M., Schwartz, M.W., 2009. Using species distribution models to predict new occurrences for rare plants. Diversity and Distributions 15, 565-576.

Chapter 3: Multivariate Random Forest Models of Estuarine-associated Fish and Invertebrate Communities[1]

ABSTRACT

Models that evaluate species habitat relationships at the community level have been gaining attention with increasing interest in ecosystem management. Developing models that can incorporate both a multiple response variable of individual species occurrence or abundance, as well as a large number of predictor variables is challenging. One promising new approach is Multivariate Random Forests (MRF), a method that combines multivariate regression trees with bootstrap resampling and predictor subsampling from traditional Random Forests. Random Forest models have been shown to be highly accurate and powerful in their predictive ability in a wide variety of applications. They can effectively model nonlinear and interacting variables. Our research evaluated change in estuarine assemblage composition along habitat gradients in Southeast Alaska using landscape-scale habitat variables and MRF. For 541 estuaries, we identified 24 predictor variables describing the geomorphic and habitat environment on land and in the estuary. MRF models were constructed in R for combined fish and invertebrate assemblages. Cluster analysis of model proximities revealed strong spatial variation in community composition in relation to differences in tidal height, precipitation, percent of eelgrass, and amount of intertidal habitat. This research presents a new science-based management template that can be used to inform and assess species management and protection strategies, as well to guide future research on species distributions.

INTRODUCTION

The shift in focus from single-species management to ecosystem methods has increased interest in community or assemblage-based models. A common approach to multi-species modeling is to model each species independently and evaluate the overlaps in species distributions. An assumption of these single species models is that species respond to environmental differences in an individualistic manner (Ferrier & Guisan 2006, Baselga & Araújo 2010), which may not be the case where biotic interactions and dispersal pathways limit the species' realized niche (Wiens 2011, Boulangeat et al. 2012). Methods that do not

---

incorporate some information on species interactions, such as the use of metrics like species richness or diversity, have the severe limitation that they do not capture the identity of the species in sampled areas and therefore can provide no information on how communities are structured.

Species-habitat relationships are affected by biotic and abiotic processes that occur on a variety of spatial scales (Anderson et al. 2009, Elith & Leathwick 2009). In estuaries species tend to have wide environmental tolerances that make them adaptable to many different environments and to high environmental variability, and their response to environmental factors can be nonlinear or discontinuous (Mueter & Norcross 1999, Gutiérrez-Estrada et al. 2008). Furthermore, the effects of most environmental factors do not occur in isolation from effects of other factors, which makes it difficult for researchers to attribute simple causality in explaining variation in assemblage composition (Oppel et al. 2009). In Alaska, a number of studies have identified species habitat relationships for individual species in estuaries or the nearshore (Norcross et al. 1997, Abookire et al. 2001, Stone & O'Clair 2001, Stoner et al. 2007) but the key patterns and processes that influence structure in these estuarine communities are still undefined.

One reason so much attention has focused on single species models is the challenge of developing models that can incorporate both a large number of environmental predictor variables and a response composed of individual species occurrence or abundance (Olden 2003). These situations require fitting a model of environmental predictors simultaneously to dependent variables representing the presence or abundance of each species in the data. A relatively new approach to multi-species distribution models is Multivariate Random Forests (Segal & Xiao 2011), a method that combines multivariate regression trees (De'ath 2002) with bootstrap resampling and predictor subsampling from traditional Random Forests (Breiman 2001). Random Forest models have been shown to be highly accurate in their predictive ability in a wide variety of applications (Prasad et al. 2006, Magness et al. 2010, Stojlgren et al. 2010). They can effectively model nonlinear and interacting predictor variables and can identify the predictor variables with the strongest influence on community composition patterns. By modeling the response of a community of species to environmental variables, models such as Multivariate Random Forests (MRF) incorporate information on species co-occurrence that can be used to evaluate the influence of species interactions on community composition (Baselga & Araújo 2010). Results from these models can be used to extrapolate beyond the sampled assemblages to predict community composition in unsampled areas (Elith et al. 2006, Drew et al. 2011).

A challenge for community modeling is the large number of species that occur in only a few samples or occur in low numbers (Cao et al. 1998, Gray et al. 2005, Park et al. 2006, Ellingsen et al. 2007). In some cases, these species can account for up to half the species in the dataset (Clarke & Warwick 2001, Cunningham & Lindenmayer 2005). Species may be rare in samples because they are found in only a few of the habitats sampled, or they may be present at a broad number of sites and either occur in low numbers or are unable to be captured consistently with the sampling gear (Cunningham & Lindenmayer 2005). Similarly, life-history or behavioral traits, such as schooling, may result in spatial clumping of species (McArdle & Anderson 2001, Cunningham & Lindenmayer 2005, Gray et al. 2005, 2006). In most single-species models, species with low occurrence in the data are often excluded because they do not exhibit good statistical properties. Modeling methods that use similarity matrices also tend to exclude rare species. Commonly used similarity metrics, such as the Bray-Curtis measure, are strongly affected by species abundance (Clarke et al. 2007) and spurious grouping of species can occur as the number of individuals of a species within a sample becomes more sparse (Field et al. 1982, Cao et al. 1997, Clarke et al. 2006). Because Random Forest models are robust to violations of the normality assumptions required of linear models, they can provide accurate results even with sparse or zero-heavy data (Breiman 2001).

Our research evaluated change in estuarine assemblage composition along habitat gradients in Southeast Alaska. This region has approximately 22,500 km of shoreline divided among 1,100 islands in an area known as the Alexander Archipelago. The area's large size and remoteness make it difficult to comprehensively sample for habitat characteristics and species. Therefore, we evaluated the influence of landscape scale variables on changes in community composition in estuaries. Models predicting community composition over large spatial scales are common in terrestrial ecology (Huettmann & Diamond 2006, Magness et al. 2010, Oppel & Huettmann 2010), but their application to marine environments has been limited. Most marine landscape models have focused on specific environments or habitat types, such as coral reefs (Wedding & Friedlander 2008), mangroves (Jelbart et al. 2006), or seagrasses (Whitlow & Grabowski 2012). Few studies have investigated the relationship between landscape structure and composition of estuarine communities. In Australia, a comparison of fish assemblages among tropical estuaries found that estuary-level variables, such as tidal range, intertidal area and distance to closest estuary explained more variation in fish assemblages than site-specific physical variables such as salinity, substrate and turbidity (Sheaves & Johnston 2009). These

results support other research indicating that environmental variables at intermediate scales may explain spatial patterns in species assemblages better than either site-specific or large-scale variables (Townsend et al. 2003, Peres-Neto 2004, Wiens 2011). Digital datasets of environmental variables are becoming increasingly available and are much less expensive and easier to obtain than site-specific environmental parameters, especially in remote and challenging environments like Southeast Alaska. If these variables can be used to develop models for detecting changes in fish and invertebrate community composition, they could become important tools in marine conservation and research.

STUDY REGION

This research was conducted in the Alexander Archipelago, a collection of approximately 1,000 mountainous islands in Southeast Alaska, USA. The study area (Figure 3.1) extends from Lance Point in Lynn Canal (58º 44' N, 135º 13' W) to Cape Chacon in Dixon Entrance at the Canadian border (54º 41' N, 132º 01' W). The coastline is generally steep and the islands are separated by deep channels and fjords. The entire archipelago is a temperate rainforest; precipitation varies locally and regionally with a general gradient of lower precipitation in the northwest and higher precipitation in the southeast. Average annual precipitation in the region is in excess of 1000 mm/yr (Neal et al. 2002) with much of the precipitation being released directly into the marine waters via numerous small streams and wetlands. Stream flow is highly seasonal and influenced both by precipitation and by snow and ice melt. The highest stream flows tend to occur in autumn when precipitation rates are high. Flows decrease in winter as a result of freezing, and increase again in the late spring and summer from melting of snow and ice. The flow of freshwater affects not only nearshore estuarine circulation, but is the driver for larger scale oceanographic circulation within Southeast Alaska's interior channels and on the continental shelf (Weingartner et al. 2009). Stream and river temperatures are influenced both by air temperatures and by runoff from glaciers, snowmelt, and precipitation.

The estuaries in the study area differ in their hydrologic and geomorphologic characteristics. In many Southeast Alaska estuaries, tidal energy is often much higher than energy from freshwater inflow. Southeast Alaska has mixed semi-diurnal tides with tidal height increasing as the tide moves from the continental shelf into the interior of the archipelago (Inazu et al. 2009). Tidal velocities are strongly influenced by bathymetry and channel morphology, and these, in turn, affect estuarine circulation, nutrient fluxes, and sediment dynamics (Weingartner et

al. 2009). Coastal geology varies greatly across estuaries in the study area. Most estuaries have a mixture of soft and hard substrate shorelines, but the amount of each type of substrate varies depending on both oceanographic and terrestrial processes.

Previous research identified approximately 12,000 estuaries in Southeast Alaska using the intersection of fresh and marine waters as the defining criteria (Albert et al. 2010); however, this definition does not take into account the degree to which an estuary is enclosed and somewhat isolated from other coastal waters. The degree of enclosure has important implications for estuarine circulation as well as the physical and chemical properties of the estuarine waters. For this research, we developed the following definition for an estuary: *a coastal indentation with a restricted connection to saltwater and an aquatic environment affected by the physical and chemical characteristics of both fluvial drainage and marine systems.* Using this definition, we delineated manually 541 estuarine polygons, including 49 polygons for estuaries for which we had biological data, between the high tide line and the 30 m depth contour (Figure 3.1) in ArcGIS 10 ™ (ESRI 2011). Glacier Bay was excluded from the research because circulation within the bay is constrained by the shallow sill at the entrance to this fjord system. As a result, the processes structuring fish and invertebrate communities within that bay are different than those in open estuaries in Southeast Alaska (Matthews 1981). Northern Lynn Canal also was excluded due to limited spatial data availability.

MATERIALS AND METHODS

Biological sampling

Forty-nine estuaries were sampled for fish and mobile invertebrates between April and September from 1998 and 2005 (Figure3.1). Sampling was conducted during daylight hours using an otter trawl (3 m x 1 m, with 6 mm square mesh in the cod end) deployed with a bridle scope of approximately 5:1. The trawl was towed at a speed of approximately 3 kn along a depth contour between 5 m and 10 m. The depth of individual tows varied within this range depending on bottom structure of the estuary. One tow in each direction was made along the same transect at high and low slack water equaling four replicates at each station. The latitude and longitude of the beginning and ending points were recorded along with the average depth of the tow. Captured fish were identified to species and were measured to the nearest millimeter for total length in the field. Invertebrates were identified to the lowest taxonomic level possible in the field and counted.

Life stage, measured as the length at 50% maturity, was obtained from the Alaska Fishery Science Center's (AFSC) Life History database (http://access.afsc.noaa.gov/reem/LHWeb) for commercially harvested fish and some forage fish. This information was used to classify fish as adults or juveniles. For fish not in the AFSC database, other published sources (Froese & Pauley 2012, MSAP 2012) were used to obtain length at maturity information. Several species occurred in the data entirely as either juveniles or adults and could be analyzed according to life stage. For species with a mix of juvenile and adult life stages, the amount of available data was generally insufficient to analyze each life stage separately, so the data were pooled and modeled together. Catch-per-unit-effort (CPUE) was calculated as the number of species caught divided by tow length and then standardized to number of fish per 100 m. Most estuaries were sampled only a single time during the study period. For estuaries sampled more than once, CPUE was calculated as the total combined catch divided by the average trawl length and standardized to number of fish per 100 m. CPUE for both fish and invertebrates was transformed to relative abundance using the Hellinger transformation (Legendre & Gallagher 2001) by dividing the count of species in each estuary by the total abundance of species present in the estuary and taking the square root of the ratio. This transformation is widely used in multivariate analysis (Legendre & Gallagher 2001) including multivariate regression tree approaches (Matabos et al. 2011, Ouellette et al. 2012, Wehrly et al. 2012) for species data containing many zeros. All fish and invertebrates were collected by NOAA/NMFS under their collection permits.

Environmental data

For each of the 541 estuaries in this research, we identified 24 predictor variables describing the geomorphic and habitat environment of the estuaries. Data were compiled from GIS layers from the Southeast Alaska GIS library, the National Oceanic and Atmospheric Administration, and the Alaska ShoreZone database (Table 3.1). All variables except the ones for intertidal vegetation were standardized to a mean of zero and standard deviation of one. The vegetation variables were recorded as percent of the vegetation type within each estuary polygon. The great diurnal tide range and mean tide range for each estuary were compiled from NOAA tide data (NOAA 2012). Estuaries without tide stations were attributed the tidal ranges from the nearest estuary with tidal data. Variables that were used to describe the structure of the estuary included the open water area, intertidal area, length of the intertidal perimeter, width at the estuary mouth, and bathymetric slope and depth. Open water area was the surface area of open water at low tide and

was the difference between the estuary area and the intertidal area. Minimum depth, and maximum and average bathymetric slope, in each estuary polygon were calculated using the ArcGIS Spatial Analyst (ESRI 2011) extension. The width of the estuary mouth was measured for each estuary along a line between the landmasses on each side of the estuary entrance, or at the 30 m depth contour.

We included information on the intertidal environment, which provides important foraging habitat for subtidal species, such as juvenile Dungeness crab (*Metacarcinus magister)* (Holsman et al. 2003), as well as habitat for eelgrass (*Zostera sp)* communities. The intertidal area and perimeter length were obtained from the Tongass National Forest High and Low Tidelines dataset (USDA 2002) for the intertidal areas within each estuary polygon. We calculated the intertidal ratio: the ratio of the intertidal perimeter to the intertidal area. This variable is an index of the shape and complexity of the intertidal environment. Area/perimeter ratios are widely used in landscape analyses to study species distributions and densities with respect to habitat size and edge effects (Jelbart et al. 2006, Martins et al. 2010), and more recently are being applied in studies of the marine environment (Wedding et al. 2011).

We also included variables describing the size and slope of the watershed surrounding the estuaries. Watershed size was derived from 12-digit hydrologic units depicting watershed boundaries (USGS 1995). We measured watershed slope within a 5 km buffer around the estuary and used a digital elevation model (USGS 1997) and ArcGIS Spatial Statistics (ESRI 2011) to calculate maximum and average slopes within each buffer.

Freshwater inflow into Southeast Alaska estuaries is difficult to calculate. Much of the study area is remote and undeveloped and there is a paucity of stream flow data even for large rivers. To capture the influence of freshwater on estuarine communities, we compiled minimum monthly precipitation over the study period from the Scenarios Network for Alaska and Arctic Planning (SNAP) climate model for Alaska (SNAP 2011) into five seasonal variables: Spring (Feb – Apr), Summer (May – Jul), Autumn (Aug – Oct), Winter (Nov – Jan), and Annual. We calculated fluvial flow by multiplying the catchment area with the average annual rainfall and a runoff coefficient and dividing by the open water area of the estuary (Digby et al. 1998):

$$\text{Fluvial Flow} = \frac{(catchment\ area\ X\ average\ annual\ precipitation\ X\ runoff\ coefficient)}{open\ water\ area}$$

The runoff coefficient (RV) is based on the impervious fraction (I) of the drainage area (ADEC 2004):

$$RV = 0.05 + 0.9(I)$$

We calculated the variable I as the non-vegetated, non-ice portions of the watershed from the 2001 National Land Cover Dataset for Alaska (Homer et al. 2004). At the scale of this analysis, the variable I was sufficiently small that the runoff coefficient was essentially a constant (0.059) across all watersheds.

Sediment characteristics have been used in a number of studies to classify estuaries and coastal areas (Digby et al. 1998, Edgar et al. 1999) and to predict species assemblages (Schoch & Dethier 1996). In Southeast Alaska, nearly the entire shoreline has been classified using the ShoreZone mapping and classification protocol (Harney et al. 2008). This method uses oblique, low altitude video and still images to classify the shoreline according to natural breaks in geomorphic, sedimentary and biological features. Shoreline segments are classified according to substrate type, sediment type, across-shore width and slope. Assemblages of sessile coastal biota present within a shoreline segment are given a categorical descriptor of either continuous (>50% cover within the unit) or patchy (< 50% coverage). Shorelines are further defined by their habitat class, which is an index that combines geology, wave exposure, and biota into a single variable. For this analysis we used shoreline classes and habitat classes that were represented in 5% or more of the sampled estuaries for a total of 23 shorelines classes and 13 habitat classes (Table 3.2). Variables were calculated as the percentage of each class with respect to the total perimeter of the estuary polygon. We also included percent of continuous or patchy canopy kelp (*Alaria sp.*), eelgrass (*Zostera marina*) and soft brown kelps (*Saccharina latissima*) within the estuaries.

Data Analysis

MRF models were constructed in R 2.13.2 (Segal & Xiao 2011) for combined fish and invertebrate assemblages to investigate spatial variation in species relative abundance in relation to the environmental predictor variables. MRFs are a modification of multivariate regression trees which can fit models simultaneously to a number of different response variables. In our MRF model, the response is the occurrence of each species at each sampled estuary. An ensemble of multivariate regression trees (Breiman et al. 1984) is grown using variable subsampling and bootstrap resampling of the data as in traditional Random Forests (Segal & Xiao

2011). The prediction error for each tree is calculated using the data omitted from the bootstrap sample for that tree. The prediction error for the forest is the average prediction error of the individual trees. Variable importance is calculated in the same manner as for traditional Random Forest, by randomly permuting the values of the variables, running them through the model and evaluating the change in the mean squared error (MSE). Variables having the greatest effect on MSE have more influence on model accuracy (Breiman 2001). We applied the MRF algorithm to construct 300 trees.

In the proximity matrix, Random Forest models also provide a measure of site similarity based on both the physical and biological variables at an estuary. This matrix is constructed by comparing the location of estuaries in the terminal nodes of each tree in the forest and giving higher proximity values to estuaries in the same node. Proximity values for each tree are summed and normalized by dividing by the number of trees in the forest. Subtracting 1 from the proximity values of the matrix converts the data to squared Euclidean distances (Segal & Xiao 2011). We used multidimensional scaling and partition around the mediod (PAM) clustering (Kaufman & Rousseeuw 1990) on the distance matrix to classify the 49 sampled estuaries. The optimal number of clusters (k) was determined by selecting the k with the maximum average silhouette width, which is a measure of the difference between intra-cluster similarity and similarity with the next closest cluster. Silhouette widths close to 1 indicate perfectly assigned clusters. To predict the class membership of the unsampled estuaries, we used the clusters from the MRF PAM clustering as the response variable in a traditional Random Forest Model. The resulting model provided the splitting rules for assigning unsampled estuaries to each cluster and provided a misclassification rate that we used to evaluate model performance. All modeling and analysis was done in R (Hinchley et al. 2008).

RESULTS

The modeling dataset contained 22 species of fish from 12 families, and 14 species of invertebrates from 11 families. Snake prickleback (*Lumpenus sagitta*), yellowfin sole (*Limanda aspera*), and starry flounder (*Platichthys stellatus*) were the most numerous fish in the data, comprising 15%, 14% and 12% of the total catch over all samples and years, respectively. Four fish species were captured at 50% or more of the estuaries: Pacific staghorn sculpin (*Leptocottus armatus*), crescent gunnel (*Pholis laeta*), starry flounder, and rock sole (*Lepidopsetta sp.*). Shell shrimp (*Crangon alaskensis*) and spot shrimp (*Pandalus platyceros*) were the most numerically

abundant invertebrate species comprising 24% and 19% of the total catch, respectively, but 90% of the spot shrimp catch occurred at a single estuary. Three species and one family of invertebrates were captured at over 50% of the estuaries: sunflower sea star (*Pycnopodia helianthoides*), shell shrimp, helmet crab (*Telmessus cheiragonus*) and gammarid amphipods (*Gammarididae*) (Table 3.3).

Several fish species occurred more frequently or entirely as juveniles in the data. Species occurring only as juveniles included Pacific herring (*Clupea pallasi*), Pacific cod (*Gadus macrocephalus*), lingcod (*Ophiodon elongatus*), kelp greenling (*Hexagrammos decagrammus),* butter sole (*Isopsetta isolepis*), and great sculpin (*Myoxocephalus polyacanthocephalus*). Species whose abundance was predominantly composed of juveniles were yellowfin sole, rock sole, and Pacific sand lance (*Ammodytes hexapterus*). Species with mixes of juveniles and adults were starry flounder, Pacific staghorn sculpin, and shiner perch (*Cymatogaster aggregata*). Species for which life stage could not be determined from the literature were the snake prickleback, tube-snout (*Aulorhynchus flavidus*), and sturgeon poacher (*Podothecus accipenserinus*).

PAM clustering of estuaries using the MRF proximities identified three clusters (Figure 3.2) with silhouette widths of 0.31, 0.33, and 0.42, which explained approximately 55% of the point variability in the data. Using these clusters as dependent variables, the traditional Random Forest model had a classification error rate of 4% with one estuary in Cluster 1 (19 estuaries), zero estuaries in Cluster 2 (20 estuaries), and one estuary in Cluster 3 (10 estuaries) being misclassified. Tidal height, minimum precipitation, and percent of continuous eelgrass were the most influential variables in the model (Figure 3.3). Variables describing the amount of intertidal habitat, open water area, and characteristics of the watershed surrounding the estuary also were influential. Variable importance was measured by the percent increase in the mean squared error of the model when the variable is permuted or randomized (Figure 3.4). The data for all variables except vegetation had been standardized to a mean of zero. This means a zero value indicated an average value for the variable, with values above average plotted to the right of center and values below average plotted to the left. Vegetation variables were measured in percent of estuary perimeter and all had positive values.

Estuaries in Cluster 1 had intermediate tidal height values, small open water and intertidal areas, but high estuary slopes and depths and high minimum precipitation throughout the year. Despite having relatively small watersheds and open water areas these estuaries had high fluvial flow, likely as a result of the much higher than average precipitation. Cluster 1 estuaries had the

most even mix of fish species (Figure 3.5) with all species represented except the white spotted greenling (*Hexagrammos stelleri*). Dock shrimp (*Pandalus danae*) and the graceful decorator crab (*Oregonia gracilis*) also were absent from this cluster (Figure 3.6). Lingcod, Pacific cod, Pacific sand lance, rock greenling (*Hexagrammos lagocephalus*), northern sculpin (*Icelenus borealis*), and buffalo sculpin (*Enophrys bison*) were found in this cluster but occurred at low relative abundances or were absent from other clusters. Average continuous and patchy kelp coverage in these estuaries was 60% compared with 40% for estuaries in Clusters 2 and 3. Northern kelp crab (*Pugettia producta*), an herbivore that feeds on kelp and utilizes kelp pigments to maintain its shell color similar to it surrounding habitat (Lunetta et al. 2004), were abundant in this cluster. Other species whose high relative abundance may be explained by environmental variables were crescent gunnels and northern sculpin, which are regularly captured in kelp and eelgrass habitats in Southeast Alaska (Johnson et al. 2003), and juvenile lingcod and juvenile Pacific cod, which prefer structured habitats that include kelp and eelgrass beds (Petrie & Ryer 2006, Laurel et al. 2007).

Estuaries in Cluster 2 were characterized by low tidal ranges, large open water and intertidal values, and low minimum precipitation and fluvial flow. Estuaries in this cluster had the highest relative abundance of flatfishes and Dungeness crab (*Metacarcinus magister*), species typically found in estuaries with large intertidal areas (Holsman et al. 2003, Holsman et al. 2006), relatively shallow depths (Norcross et al. 1997), and steeper than average bathymetric slopes characteristic of low gradient stream systems and stream derived sediments. Cluster 2 estuaries also had high relative abundances of tube-snouts, a species associated with mixed habitats of sand, kelp, and eelgrass (Ishwaran 2007). Dock shrimp and graceful decorator crabs were absent from this cluster, and abundances of other shrimp were generally lower than in the other two clusters.

Cluster 3 estuaries had higher than average precipitation for all seasons except summer. Intertidal area in Cluster 3 was lower than average for the dataset, but both intertidal perimeter and intertidal ratio were higher than average. This was a result of the deeper depths and steeper slopes in these estuaries compared with estuaries in the other two clusters. Estuaries in Cluster 3 had the lowest number of fish taxa but the highest number of invertebrate species, particularly shrimp. In this cluster, estuaries averaged between 39% and 93% patchy or continuous eelgrass coverage. This cluster was characterized by high relative abundances of bay pipefish (*Syngnathus leptorhynchus*) and three-spine sticklebacks (*Gasterosteus aculeatus*), both eelgrass-

associated species (Johnson et al. 2003), but low abundances of Dungeness crab, a species that relies on the intertidal area for foraging at high tides (Holsman et al. 2003). Cluster 3 estuaries had high abundances of rock sole and starry flounder similar to estuaries in Cluster 2, but low relative abundances of yellowfin sole. The majority of yellowfin sole and rock sole occurred in the data as juveniles. Previous research has found that juvenile yellowfin sole prefer shallow locations at the head of bays while juvenile rock sole tend to prefer deeper water (Norcross et al. 1997).

The majority of the ShoreZone coastal class and habitat class variables had very low or negative variable importance, indicating low direct predictive value for the MRF model. Variable predictor scores in Random Forest models take into account interactions among variables (Lunetta et al. 2004), and variables that do not have high individual prediction scores may influence and increase or decrease the prediction scores of other variables. As a result, step-wise removal of variables with low or negative importance can result in deleting interacting variables that are otherwise important to the scores of variables that are retained resulting in unstable models. A number of methods have been proposed to select the most relevant variables in Random Forest models (Svetnik et al. 2002, Sandri & Zuccolotto 2006, Genuer et al. 2010), but there is no concurrence on how or whether variable removal should be done. Rather than implementing a step-wise variable removal method that might introduce bias into the model from interacting variables, we chose to remove all ShoreZone coastal and habitat class variables from the analysis and run the models on the remaining 24 environmental variables.

The spatial distribution of clusters reflected the influence of the highest performing variables: tidal height and winter precipitation (Figure 3.7a). Estuaries in Cluster 1 (black) were located on both the outer coast and along the deeper channels where precipitation from the interaction of ocean storms and coastal mountains is high. Cluster 2 estuaries (red) were located among the inland waters where tidal height is lowest and the rain shadow of the coastal mountains results in lower precipitation than the other clusters. Estuaries in Cluster 3 (green) were located on the outer coast adjacent to the open ocean. These estuaries tended to have the lowest tidal heights due to their proximity to the shelf. These general patterns of cluster distribution held when cluster membership was predicted for the 492 other estuaries in the study area, but more overlap occurred as a result of the other variables in the model (Figure 3.7b).

DISCUSSION

       Fish and invertebrate communities in Southeast Alaska demonstrated strong spatial variation related to differences in precipitation, tidal height, percent of eelgrass, and amount of intertidal habitat. One of the strongest patterns in the environmental data is the difference in precipitation and tidal exchange between the clusters, but it can be difficult to tease apart the influence of individual variables in models where the importance of any variable may be influenced by other interacting variables in the data (Lunetta et al. 2004, Knudby et al. 2010). Both precipitation and tidal exchange vary substantially across the study area at small spatial scales. Precipitation patterns are strongly influenced by watershed elevation with ocean storms releasing moisture as air is adiabatically cooled by the high mountainous landscape (Weingartner et al. 2009). This process results in areas of high precipitation near the coast, and in rain shadows along the interior passes. Tidal height increases as water moves inland away from the continental shelf and into the interior of the archipelago. Tidal currents are strongly influenced by channel width, bathymetric structure, and depth resulting in substantial variations in tidal current strengths at the scale of several kilometers. The interaction of these variables results in a spatial distribution of estuary clusters that generally aligns with patterns of precipitation (Figure 3.7b), but with deviations based on estuary-specific differences in bathymetry, estuary size, and amount of intertidal habitat. These complex interactions between ecosystem components would likely not have been identified using a linear model.

       The variables used in this analysis were those that could be extrapolated from available spatial datasets, rather than environmental conditions measured *in-situ*. This approach increases the value of the models for predicting community composition to unsampled areas, but can make it more difficult to relate mechanisms of community assemblages to environmental data. Precipitation and fluvial flow variables were included in the analysis to capture differences in salinity, and buoyancy-driven circulation between estuaries, but the relationship between community composition and salinity is complex. Several studies on juvenile groundfish in Alaska found only small or insignificant correlations between salinity and abundance (Norcross et al. 1997, Norcross et al. 1999, Abookire et al. 2001). In Kachemak Bay, AK Pacific herring and sand lance are substantially more abundant in the interior of the estuary in less-saline surface waters (Abookire et al. 2000). However, in the Skagit River estuary in Puget Sound, WA there is no correlation between Pacific herring annual abundance and river discharge (Sandri & Zuccolotto 2006).

Precipitation may also be a proxy for other oceanographic processes. Higher precipitation can freshen the surface water layer and result in stratification, which can enhance primary productivity (Weingartner 2007). Higher freshwater discharge is also associated with the development of tidal fronts: areas of mixing that occur at the interface between stratified water and well mixed saline water as a result of tidal inflow into the estuary. Constrictions, such as a narrowing of the estuary mouth, act as hydraulic controls that can enhance formation of these fronts (Largier 1992). Nutrients are drawn into the stratified surface layer of the front by diapycnal mixing at the frontal boundary, and over a period of time can enhance phytoplankton production (Largier 1992, 1993, Johnson & Costello 2002). At the same time, convergent flows along the frontal boundary advect and concentrate plankton (Dustan & Pinckney 1989, Franks & Chen 1996), which attract grazers and higher trophic level predators (Largier 1993, Kingsford & Suthers 1994). These fronts may also act as barriers to larval transport helping to retain and distribute planktonic larvae within the estuary (Eggleston et al. 1998). Freshwater discharge into estuaries in Cluster 1 was higher year round than for the other two clusters, and estuaries in this cluster have narrow average mouth widths. The oceanographic conditions in these estuaries may favor enhanced productivity that could explain the more even mix of fish species and the higher relative abundance of species such as sand lance and Pacific herring (Arimitsu et al. 2004); however, additional research is needed to further evaluate species-environment relationships.

Cluster results explained a little over half the variability in the data, and the average silhouette widths for the clusters were below 0.50, implying only moderate cluster assignment. This suggests that variables not included in the model are influencing species distributions and community composition. One such variable might be substrate type. Several studies have demonstrated strong associations between fish (Norcross et al. 1999, Abookire et al. 2001) and invertebrates (Schoch & Dethier 1996, Lunetta et al. 2004, Hovel & Wahle 2010) and substrate type. Although we did not find inclusion of subtidal geology variables from the ShoreZone dataset to be informative in our models, this may be a result of the way the variables were derived. These variables were calculated as the percent of each substrate occurring within the estuary, a method that is not sensitive to the patch size of the habitat and sediment types. Therefore, an estuary with 30% continuous coverage of sand, for example, is equivalent to an estuary with three small and well-spaced patches of sand of 10% each. While inclusion of the absolute area of each substrate type within the estuary might improve the performance of these variables in the model, it would be preferable to have some measure of both the size and

separation of habitats within each estuary. Size and shape of habitat patches are important factors affecting species abundance, diversity and habitat use (Barrantes & Sandoval 2009, Baselga & Araújo 2010, Morin 2011).

The spatial arrangement of habitat in estuaries may also be important in influencing community structure (Whitlow & Grabowski 2012). ShoreZone variables were extracted for the entire estuary polygon and may not correspond directly to the habitat in the area sampled. Trawl sampling is constrained to occur in areas with minimal rocks or hard structures, and these may represent only a portion of the substrates in the estuary. Including substrate variables directly under sampling transects along with information on adjacent habitats, may enhance the performance of ShoreZone substrate variables in the model. Our future research will attempt to include this type of finer scale substrate information in the models.

Biotic factors, such as competition, dispersal limitation and predation, are known to constrain species distributions (Dunson & Travis 1991, Boulangeat et al. 2012, Wisz et al. 2012), and some of the variance that is not captured in our model could be the result of biotic interactions. Unfortunately, incorporating relevant biotic factors into species distributions models is complicated by lack of data on species interactions at the scale of the analysis, and confounding effects of abiotic and biotic variables on individual species (Manly et al. 2002, Elith & Leathwick 2009). By using the species assemblage as the response variable, the multivariate Random Forest model used in this research implicitly incorporates patterns of species co-occurrence that could be reflections of species interactions, but the model cannot identify specific relationships with biotic factors or whether such factors are having a direct or indirect effect on the model. Methods for incorporating biotic interactions into species distribution models include adding competing or predator/prey species as explanatory variables (Kissling et al. 2011), including estimates of habitat productivity (Wisz et al. 2012), and including data on dispersal (Boulangeat et al. 2012). Incorporating these data into multispecies models is especially challenging due to the complexity of species interactions within a community. For marine species, incorporation of dispersal data is also hampered by lack of oceanographic data on a scale relevant to the analysis. Our ongoing work in this area involves evaluating the strength of biotic interactions through the development of niche models that use behavioral, functional, or phylogenetic traits (Miller 2013). We will be evaluating ways to incorporate insights on the importance of biotic factors to estuarine communities gained from these null models into our community analyses.

In this study, field data were collected over a period of seven years with most estuaries sampled only once during that period. As a result, not all estuaries were sampled in the same month or season. Seasonality in the data may have introduced uncertainty into some aspects of the analysis. Several Alaskan estuarine species have strong seasonal patterns. An example is the tube-snout, which is found in a variety of sandy and rocky habitats with adjacent eelgrass or kelp. Tube-snouts occur in the nearshore to depths of 30 m (Hunter-Thomson 2011). In nearshore seine net fish sampling in Southeast Alaska from 1998 to 2000, tube-snouts were captured in low numbers, but consistently, throughout Southeast Alaska (Johnson et al. 2003). In this analysis, tube-snouts were entirely absent from Cluster 3, although species that frequently co-occur with them, such as the bay pipefish, were abundant in that cluster and Cluster 3 has the highest percentage of eelgrass. This may be explained by the season in which the sampling occurred. Although tube-snouts are year-round residents of estuaries and nearshore areas, sampling in Prince William Sound, Alaska in 2006 and 2007 (Johnson et al. 2011) captured substantially more tube-snouts in September (331) than in April (21) or July (between 67 and 95). Estuaries in Cluster 3 were sampled primarily in April with some samples occurring in May and June. In contrast, estuaries in Cluster 2, which had the highest relative abundance of tube-snouts, were sampled relatively uniformly between April and September. Unfortunately, the September Prince William Sound surveys (Johnson et al. 2010) were conducted in a single year, giving no information on interannual differences, and no other information on seasonality in tube-snouts is available. Similarly, seasonal movement of species from nearshore areas to offshore areas in Alaska has been documented in other studies (Abookire & Norcross 1998, Stone & O'Clair 2001, Johnson & Thedinga 2006) but, as with the present research, sampling was not consistent between months and years. Interannual differences in temperature and oceanographic conditions affect timing of migratory behavior and make it difficult to compare species composition at estuaries sampled in different months and years. Most sampling in Southeast Alaska occurs from March to August and sometimes September, but there has been no research comparing species abundances for all of these months across more than one year at the same location. Until such research can be conducted, it is unclear whether changes in the seasonal abundance of species have an impact on the results of this analysis.

CONCLUSIONS

Understanding the mechanisms that structure communities is fundamental to ecology and a precursor for ecosystem management. Multi-species management strategies require knowledge not only of abiotic factors affecting species distributions (fundamental niche), but also of the relationship between species within a community. To this end, methods that can quantitatively model multiple species will provide additional insight into species functional roles and interactions. This is the first implementation of a MRF model to marine fish and invertebrate communities, and the first research to evaluate the relationship between landscape structure and estuarine community composition in Southeast Alaska. At the regional scale, estuaries clustered from the model show strong association with spatial patterns of precipitation and tidal height. At a more local scale the amount of intertidal habitat and availability of kelp and eelgrass habitats influenced the relative of abundance of individual species within the communities. Evaluating both large- and fine-scale patterns in community composition can inform species management and protection strategies, as well guide future research on species co-occurrence.
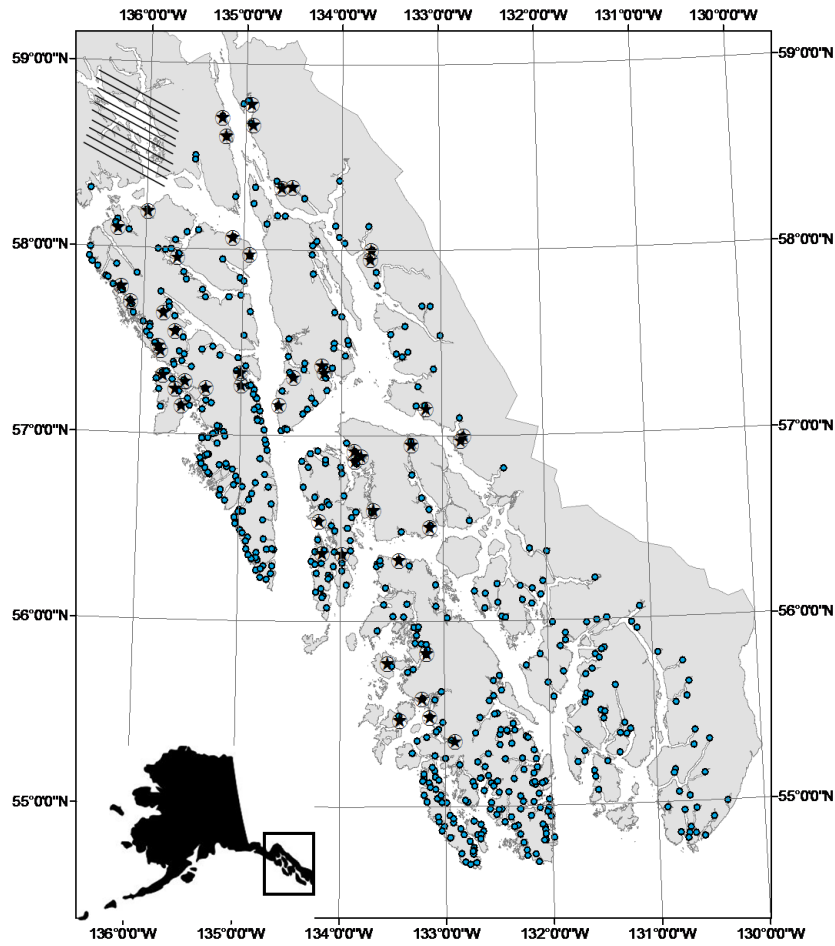
Figure 3.1: Study area in Southeast Alaska showing sampling locations (★, and estuaries to which model results were predicted (⦿). Hash marks indicate exclusion of Glacier Bay and North Lynn Canal.
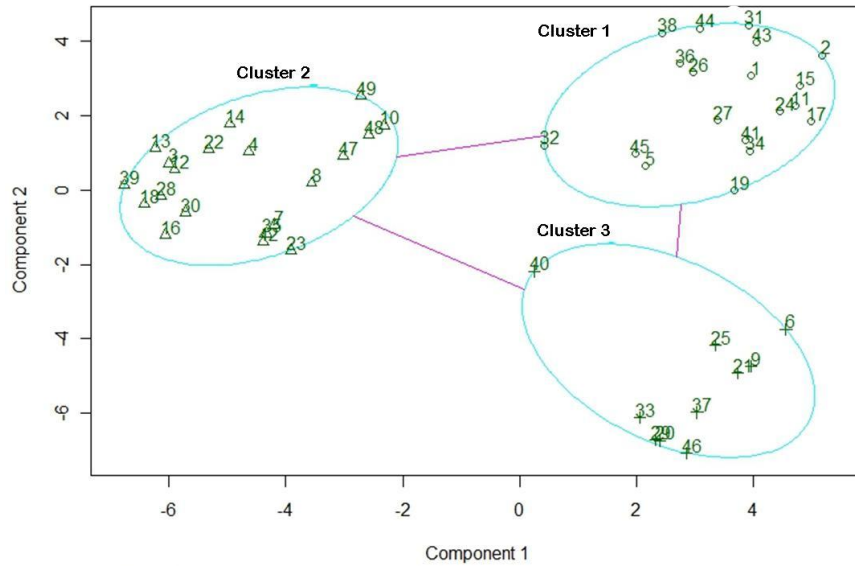
Figure 3.2**:** Multidimensional scaling plot of MRF proximity values.Ellipses delineate PAM cluster membership, and numbers are sample estuaries. The two components explain 54.85 % of the variation in the data. Lines indicate the distance between clusters.
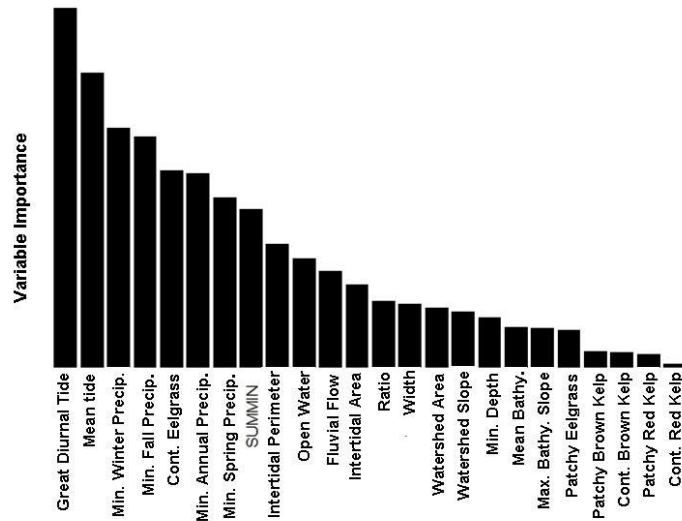
Figure 3.3: Relative importance of the predictor variables scaled by the decrease in the mean squared error of the model when the variable is permuted. Higher bars equate to higher variable importance.
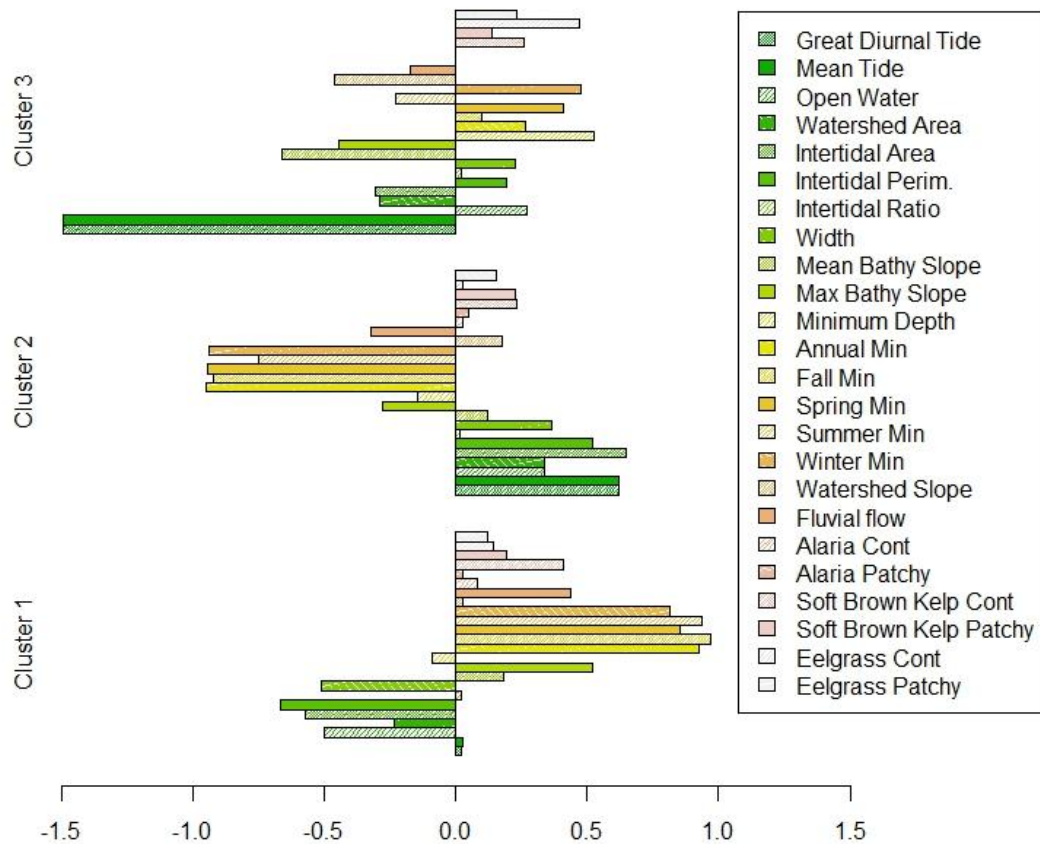
Figure 3.4: Average value of predictor variables for sampled estuaries by cluster.
The data for all variables except vegetation were standardized to a mean of zero, so a zero value indicates an average value for the variable, with values above average plotted to the right of center and values below average plotted to the left. Vegetation variables are measured in percent of estuary perimeter and all have positive values.
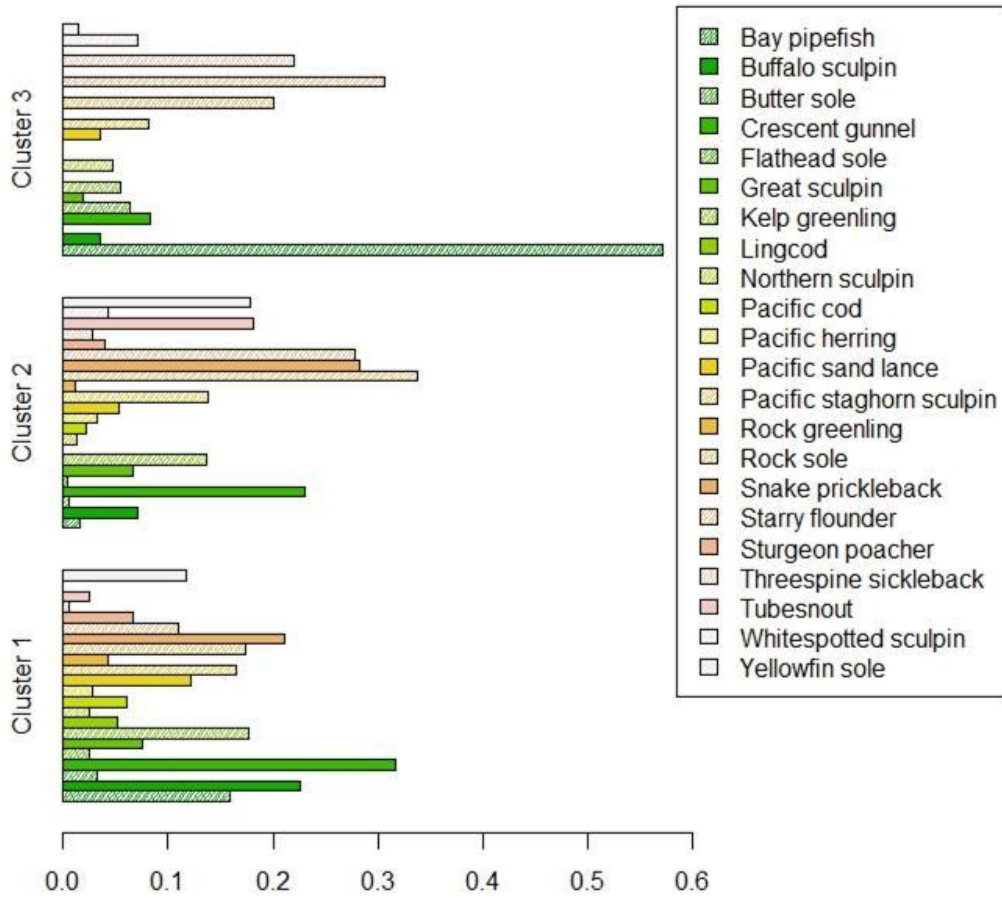
Figure 3.5: Average relative abundance of fish species for sampled estuaries by cluster.
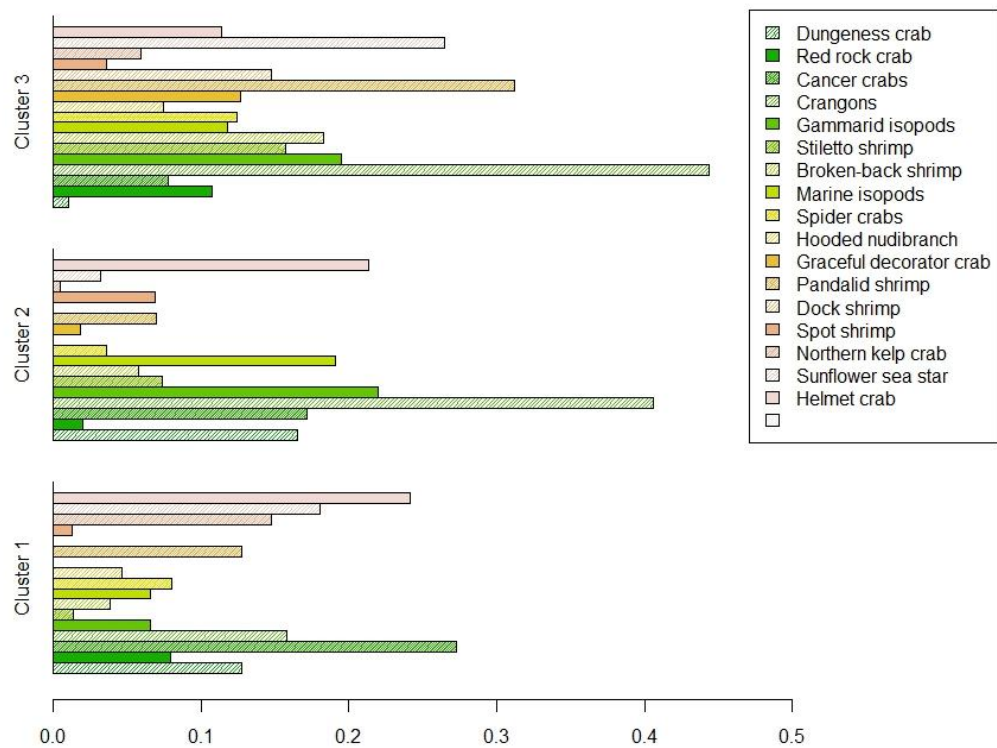
Figure 3.6: Average relative abundance of invertebrates by cluster for sampled estuaries.

a)

b)
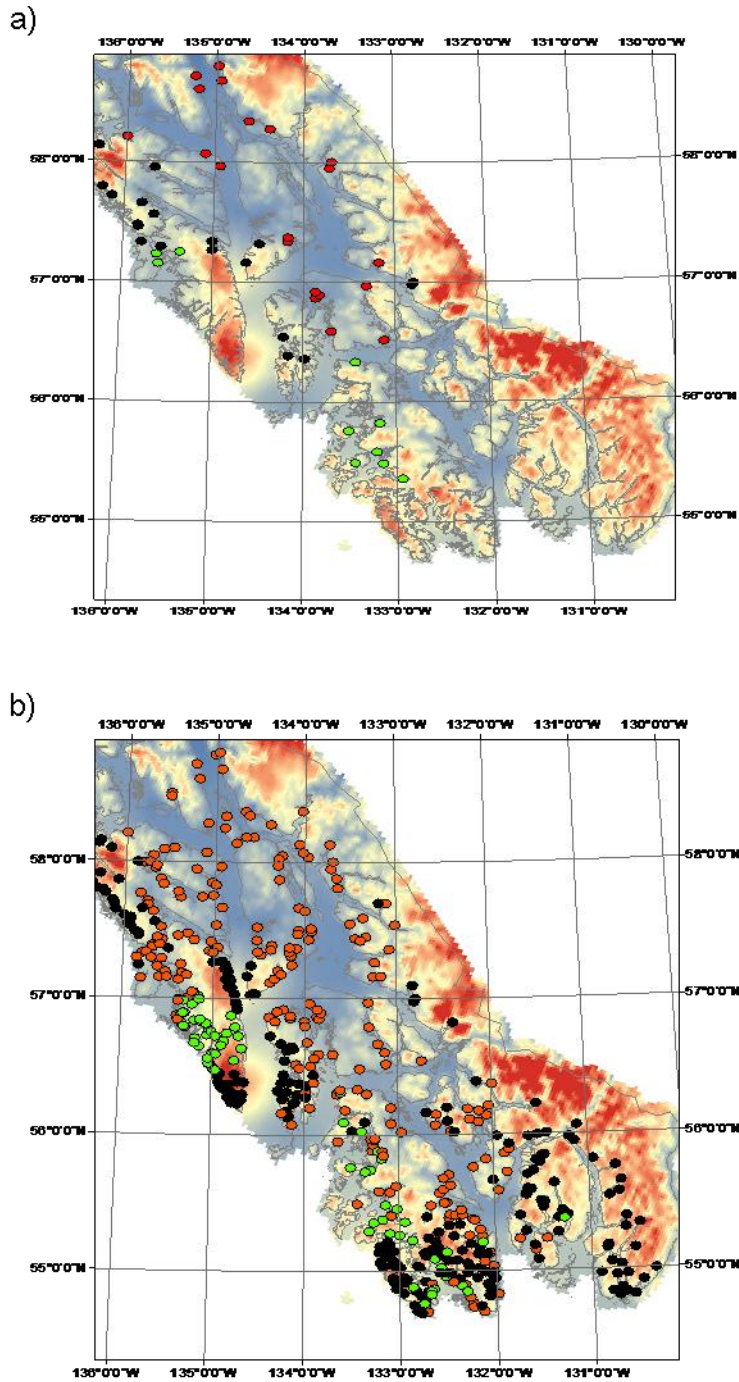
Figure 3.7: (a) Spatial distribution of 49 sample estuaries by cluster from MRF model and PAM analysis, and (b) predicted cluster membership of 492 delineated estuaries. Cluster 1 =●, Cluster 2 =●, and Cluster 3 =●. Estuaries colored by cluster membership and overlaid on winter minimum precipitation.

Table 3.1: Predictor variables used in MRF model. NA indicates variables without a spatial scale.

Variable	No.	Unit	Time Scale	Spatial Scale of data source	Source
Intertidal Area	1	Sq. Meters		1:63,360	USFS Tongass GIS
Intertidal Perimeter	1	Meters		1:63,360	USFS Tongass GIS
Intertidal ratio	1	Meters		1:63,360	Derived
Open water	1	Sq. Meters		1:63,360	USFS Tongass GIS - derived
Watershed Area	1	Sq. Meters		1:63,360	USGS Hydrologic Unit Maps
Watershed Slope	1	Degrees		300 m	USGS Digital Elevation Model
Tidal range	2	Feet		NA	NOAA
Width	1	Meters		1:63,360	measured
Estuary slope	2	Degrees		5 m	NMFS AKR Bathymetry - derived
Depth	1	Meters		5 m	NMFS AKR Bathymetry - derived
Minimum seasonal precipitation	4	Millimeters	1998-2005	2 km	PRISM Climate Model
Minimum annual precipitation	1	Millimeters	1998-2005	2 km	PRISM Climate Model
Fluvial flow	1	Flow/sq m	1998-2005	NA	Derived
Continuous/patchy subtidal vegetation	6	Percent		NA	Derived

Table 3.2: ShoreZone coastal class variables used in the MRF model. No. is the numeric
ShoreZone designation for the variable.

SUBSTRATE	SEDIMENT	WIDTH	SLOPE	COASTAL CLASS	NO.
ROCK	N/A	NARROW (<30m)	STEEP (>20°)	Rock cliff	3
			INCLINED (5-20°)	Rock Ramp, narrow	4
ROCK & SEDIMENT	GRAVEL	WIDE(>30m)	INCLINED (5-20°)	Ramp with gravel beach, wide	6
			FLAT (<5°)	Platform with gravel beach, wide	7
		NARROW (<30m)	STEEP (>20°)	Cliff with gravel beach	8
			INCLINED (5-20°)	Ramp with gravel beach	9
			FLAT (<5°)	Platform with gravel beach	10
	SAND & GRAVEL	WIDE(>30m)	INCLINED (5-20°)	Ramp w/ gravel & sand beach, wide	11
			FLAT (<5°)	Platform w/ gravel & sand beach, wide	12
		NARROW (<30m)	STEEP (>20°)	Cliff w/ gravel & sand beach	13
			INCLINED (5-20°)	Ramp w/ gravel & sand beach	14
			FLAT (<5°)	Platform w/ gravel & sand beach	15
SEDIMENT	GRAVEL	WIDE(>30m)	FLAT (<5°)	Gravel flat, wide	21
		NARROW (<30m)	INCLINED (5-20°)	Gravel beach, narrow	22
	SAND & GRAVEL	WIDE(>30m)	FLAT (<5°)	Sand & gravel flat or fan	24
		NARROW (<30m)	INCLINED (5-20°)	Sand & gravel beach, narrow	25
			FLAT (<5°)	Sand & gravel flat or fan	26
	SAND/MUD	WIDE(>30m)	FLAT (<5°)	Sand flat	28
			FLAT (<5°)	Mudflat	29
	ORGANICS	n/a	n/a	Estuaries	31
ANTHROPOGENIC	Man-made	n/a	n/a	Man-made, permeable	32
				Man-made, impermeable	33
CHANNEL	Current	n/a	n/a	Channel	34
GLACIER	Ice	n/a	n/a	Glacier	35

Table 3.3: CPUE of fish and invertebrates from trawl samples

FISH		INVERTEBRATES	
Family		**Family**	
Species	**CPUE**	*Species*	**CPUE**
Clupeidae (herrings)		Pychnopodiidae	
Clupea pallasi (Pacific herring)*	87	*Pycnopodia helianthoides (sunflower sea star)*	251
Gadidae (cods)		Tethyididae	
Gadus macrocephalus* (Pacific cod)	21	*Melibe leonina (hooded nudibranch)*	396
Aulorhynchidae (tubesnouts)		Crangonidae	
Aulorhynchus flavidus* (tubesnout)	146	*Crangon alaskensis (shell shrimp)*	1996
Gasterosteidae (sticklebacks)		Gammarids	68
Gasterosteus aculeatus* (three-spine stickleback)	59	Hippolytidae	
Sygnathidae (pipefishes)		*Heptacarpus stylus (stiletto shrimp)*	51
Syngnathus leptorhynchus* (Bay pipefish)	192	Cancridae	
Hexagrammidae (greenlings)		*Metacarcinus magister (Dungeness crab)*	331
Ophiodon elongates* (lingcod)	9	*Cancer productus (red rock crab)*	87
Hexagrammos stelleri* (whitespotted greenling)	7	*Cancer sp.*	152
Hexagrammos decagrammus* (kelp greenling)	104	Cheiragonidae	
Hexagrammos lagocephalus* (rock greenling)	47	*Telmessus cheiragonus (helmet crab)*	384
Cottifdae (sculpins)		Oregoniidae	
Icelinus borealis* (northern sculpin)	7	*Oregonia gracilis (graceful decorator crab)*	179
Leptocottus armatus* (Pacific staghorn sculpin)	122	Epialtinae	
Enophrys bison* (buffalo sculpin)	87	*Pugettia producta (northern kelp crab)*	339
Myoxocephalus polyacanthocephalus* (great sculpin)	24	Hyppolitidae	
Agonidae (poachers)		*Hippolyte sp.*	821
Podothecus accipenserinus* (surgeon poacher)	31	Pandalidae	
Stichaeidae (pricklebacks)		*Pandalus danae (dock shrimp)*	87
Lumpenus sagitta* (snake prickleback)	397	*Pandalus platyceros (spot shrimp)*	1536
Pholidae (gunnels)		*Pandalus sp.*	85
Pholis laeta* (crescent gunnel)	241		
Ammodytidae (sandlances)			
Ammodytes hexapterus* (Pacific sandlance)	99		
Pleuronectidae (righteye flounders)			
Hippoglossus stenolepis * (flathead sole)	13		
Platichthys stelatus* (starry flounder)	302		
Lepidopsetta sp* (rock sole)	218		
Isopsetta isolepis* (butter sole)	11		
Limanda aspera* (yellowfin sole)	376		

REFERENCES

Abookire AA, Norcross BL (1998) Depth and substrate as determinants of distribution of juvenile flathead sole (*Hippoglossoides elassodon*) and rock sole (*Pleuronectes bilineatus*), in Kachemak Bay, Alaska. Journal of Sea Research 39:113-123

Abookire AA, Piatt JF, Norcross BL (2001) Juvenile groundfish habitat in Kachemak Bay, Alaska during late summer. Fishery Bulletin 8:45-56

Abookire AA, Piatt JF, Robards MD (2000) Nearshore fish distributions in an Alaskan estuary in relation to stratification, temperature and salinity. Estuarine, Coastal and Shelf Science 51:45-59

ADEC, Alaska Department of Environmental Conservation (2004) Total maximum daily load (TMDL) for fecal coliform in the waters of Little Survival Creek in Anchorage, Alaska. In: ADEC, Alaska Department of Environmental Conservation (ed), URL http://dec.alaska.gov, Anchorage, Alaska

Albert D, Shanley C, Baker L (2010) A preliminary classification of bays and estuaries in Southeast Alaska: A hierarchical framework and exploratory analysis. Coastal Ecological Systems in Southeast Alaska. The Nature Conservancy (http://conserveonline.org/library/a-preliminary-classification-of-bays-and-estuaries/view.html), Juneau, AK

Anderson TJ, Syms C, Roberts DA, Howard DF (2009) Multi-scale fish-habitat associations and the use of habitat surrogates to predict the organization and abundance of deep-water fish assemblages. Journal of Experimental Marine Biology and Ecology 379:34-42

Arimitsu ML, Piatt JF, Romano MD, Douglas DD (2004) Distribution of forage fishes in relation to oceanography of Glacier Bay National Park. In: Piatt JF, Gende SM (eds) U.S. Geological Survey Scientific Investigations Report, 1007-5047. Proc Fourth Glacier Bay Science Symposium, October 26-28, 2004. U.S. Geological Survey

Barrantes G, Sandoval L (2009) Conceptual and statistical problems associated with the use of diversity indices in ecology. International Journal of Tropical Biology 57:451-460

Baselga A, Araújo FG (2010) Do community-level models describe community variation effectively. Journal of Biogeography 37:1842-1850

Boulangeat I, Gravel D, Thuiller W (2012) Accounting for dispersal and biotic interactions to disentangle drivers of species distributions and their abundances. Ecology Letters 15:584-593

Breiman L (2001) Random Forests. Machine Learning 45:5-32

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees, Vol The Wadsworth Statistics/Probability Series. Wadsworth International Group, Belmont

Cao Y, Williams DD, Williams NE (1998) How important are rare species in aquatic community ecology and bioassessment? Limnology and Oceanography 43:1403-1409

Cao Y, Williams WP, Bark AW (1997) Similarity measure bias in river benthic Aufwuchs community analysis. Water Environment Federation 69:95-106

Clarke KR, Chapman MG, Somerfield PJ, Needham HR (2007) Dispersion-based weighting of species counts in assemblage analyses. Marine Ecology Progress Series 320:11-17

Clarke KR, Somerfield PJ, Chapman MG (2006) On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray-Curtis coefficient for denuded assemblages. Journal of Experimental Marine Biology and Ecology 330:55-80

Clarke KR, Warwick RM (2001) Change in marine communities: An approach to statistical analysis and interpretation. PRIMER-E, Plymouth

Cunningham RB, Lindenmayer DB (2005) Modeling count data of rare species: some statistical issues. Ecology 86:1135-1142

De'ath G (2002) Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology 83:1105-1117

Digby MJ, Saenger P, Whelan MB, McConchie D, Eyre B, Holmes N, Bucher D (1998) A physical classification of Australian estuaries. Centre for Coastal Management, Southern Cross University,Urban Water Research Association of Australia, (http://auriversinfoorg/library/nrhp/estuary_clasifn/)

Drew CA, Wiersma Y, Huettmann F (2011) Predictive Modeling in Landscape Ecology, Vol 1. Springer, New York, NY

Dunson W, Travis J (1991) Role of abiotic factors in community organization. The American Naturalist 138:1067-1091

Dustan P, Pinckney JL, Jr. (1989) Tidally induced estuarine phytoplankton patchiness. Limnology and Oceanography 34:410-419

78

Edgar G, Barrett N, Graddon D (1999) A classification of Tasmanian estuaries and assessment of their conservation significance using ecological and physical attributes, population, and land use. Marine Research Laboratories - Tasmanian Aquaculture and Fisheries Institute, University of Tasmania, (http://eprints.utas.edu.au/1718/)

Eggleston DB, Armstrong DA, Elis WE, Patton WS (1998) Estuarine fronts as conduits for larval transport: hydrodynamics and spatial distribution of Dungeness crab larvae. Marine Ecology Progress Series 164:73-82

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick J, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Jacob MO, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmerman NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129-151

Elith J, Leathwick J (2009) Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology and Systematics 40:677-697

Ellingsen K, E., Hewitt JE, Thrush S (2007) Rare species, habitat diversity and functional redundancy in marine benthos. Journal of Sea Research 58:291-301

ESRI, Environmental Systems Research Institute (2011) ArcGIS Desktop: Release 10. Redlands, CA

Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. Journal of Applied Ecology 43:393-404

Field JG, Clarke KR, Warwick RM (1982) A practical strategy for analysing multispecies distribution patterns. Marine Ecology Progress Series 8:37-52

Franks PJS, Chen C (1996) Plankton production in tidal fronts: A model of Georges Bank in summer. Journal of Marine Research 54:631-651

Froese R, Pauley D (2012) Fishbase. World Wide Web electronic publication, http://wwwfishbaseorg, version 10/2012

Genuer R, Poggi J-M, Tuleau-Malot C (2010) Variable selection using random forests. Pattern Recognition Letters 31:2225-2236

Gray JS, Bjørgesæter A, Ugland KI (2005) The impact of rare species on natural assemblages. Journal of Animal Ecology 74:1131-1139

Gray JS, Bjørgesæter A, Ugland KI (2006) On plotting species abundance distributions. Journal of Animal Ecology 75:752-756

Gutiérrez-Estrada JC, Vasconcelos R, Costa MJ (2008) Estimating fish community diversity from environmental features in the Tagus estuary (Portugal): Multiple linear regression and artificial neural network approaches. Journal of Applied Ichthyology 24:150-162

Harney J, Morris M, Harper J (2008) Shorezone coastal habitat mapping protocol for the Gulf of Alaska. Coastal & Ocean Resources, Inc (http://wwwfakrnoaagov/shorezone)

Hinchley EK, Nicholson MC, Zajac RN, Irlandi EA (2008) Preface: Marine and coastal applications in landscape ecology. Landscape Ecology 23:1-5

Holsman K, Armstrong DA, Beauchamp DA, Reusink JL (2003) The necessity for intertidal foraging by estuarine populations of subadult Dungeness crab, *Cancer Magister*: Evidence from a bioenergetics model. Estuaries 26:1155-1173

Holsman K, McDonald PS, Armstrong DA (2006) Intertidal migration and habitat use by subadult Dungeness crab, *Cancer Magister*, in a NE Pacific estuary. Marine Ecology Progress Series 308:183-195

Homer C, Huang C, Yang L, Wylie B, Coan M (2004) Development of a 2001 national land cover database for the United States. Photogrammetric Engineering and Remote Sensing 70:829-840

Hovel K, Wahle RA (2010) Effects of habitat patchiness on American lobster movement across a gradient of predation risk and shelter competition. Ecology 91:1993-2002

Huettmann F, Diamond AW (2006) Large-scale effects on the spatial distribution of seabirds in the Northwest Atlantic. Landscape Ecology 21:1089-1108

Hunter-Thomson K (2011) Nearshore fish assemblage patterns with respect to landscape-scale habitats in central California. MS, Department of Marine Science, San Jose State University (http://scholarworks.sjsu.edu/etd_theses/4055),

Inazu D, Sato T, Miura S, Ohta Y, Nakamura K, Fujimoto H, Larsen CF, Higuchi T (2009) Accurate ocean tide modeling in Southeast Alaska and large tidal dissipation around Glacier Bay. Journal of Oceanography 65:335-347

Ishwaran H (2007) Variable importance in binary regression trees and forests. Electronic Journal of Statistics 1:591-537

Jelbart JE, Ross PM, Connolly RM (2006) Edge effects and patch size in seagrass landscapes: an experimental test using fish. Marine Ecology Progress Series 319:93-102

Johnson A, Thedinga J, Neff AD, Harris P, Lindeberg MR, Maselko JM, Rice SD (2011) Fish assemblages in nearshore habitats of Prince William Sound, Alaska. Northwest Science 84:266-280

Johnson MP, Costello MJ (2002) Local and external components of the summertime plankton community in Lough Hyne, Ireland a stratified marine inlet. Journal of Plankton Research 24:1305-1315

Johnson SW, Murphy ML, Csepp DJ, Harris P, Thedinga J (2003) A survey of fish assemblages in eelgrass and kelp habitats of Southeast Alaska. US Department of Commerce, NMFS-AFSC-139

Johnson SW, Thedinga J (2006) Fish use and size of eelgrass meadows in Southeastern Alaska: a baseline for long-term assessment of biotic change. Northwest Science 79:141-155

Johnson SW, Thedinga JF, Neff AD, Harris PM, Lindeberg MR (2010) Fish assemblages in nearshore habitats of Prince William Sound, Alaska. Northwest Science 83:266-280

Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analsyis. John Wiley & Sons, Inc., New York, NY

Kingsford MJ, Suthers IM (1994) Dynamic estuarine plumes and fronts: importance to small fish and plankton in coastal waters of NSW, Australia. Continental Shelf Research 14:655-672

Kissling WD, Dormann CF, Groeneveld J, Hickler T, Kühn I, McInerny GJ, Montoya J, Römermann C, Schiffers K, Schurr FM, Singer A, Svenning J-C, Zimmerman NE, O'Hara RB (2011) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. Journal of Biogeography doi: 10.1111/j.1365-2699.2011.02663.x

Knudby A, Brenning A, LeDrew E (2010) New approaches to modelling fish-habitat relationships. Ecological Modelling 221:503-511

Largier JL (1992) Tidal Intrusion Fronts. Estuaries 15:26-39

Largier JL (1993) Estuarine Fronts: How important are they? Estuaries 16:1-11

Laurel BJ, Stoner AW, Ryer C, Abookire AA (2007) Comparative habitat associations in juvenile Pacific cod and other gadids using seines, baited cameras, and laboratory techniques. Journal of Experimental Marine Biology and Ecology 351:42-55

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. Oecologia 129:271-280

Lunetta KL, Hayward LB, Segal J, Van Eerdewegh P (2004) Screening large-scale association study data: exploiting interactions using random forests. BMC Genetics 5:32-45

Magness DR, Huettmann F, Morton JM (2010) Using random forests to provide predicted species distribution maps as a metric for ecological inventory and monitoring programs. In: Cushman S, F. Huettmann (ed) Spatial complexity, informatics and wildlife conservation. Springer, New York

Manly BF, McDonald LL, Thomas DL, McDonald TL, Erickson WP (2002) Resource Selection by Animals: Statistical Design and Analysis for Field Studies. Kluwer, Dordrecht

Martins GM, Thompson RC, Neto AI, Hawkins SJ, Jenkins SR (2010) Exploitation of intertidal grazers as a driver of community divergence. Journal of Applied Ecology 47:1282-1289

Matabos M, Plouviez S, Hourdez S, Desbruyères D, Legendre P, Warén A, Jollivet D, Thiébaut E (2011) Faunal changes and geographic crypticism indicate the occurrence of a biogeographic transition zone along the southern East Pacific Rise. Journal of Biogeography 38:5757-5594

Matthews JB (1981) The seasonal circulation of the Glacier Bay, Alaska fjord system. Estuarine, Coastal and Shelf Science 12:679-700

McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology 82:290-297

Miller K (2013) Chapter 4: Taxonomic and functional β-diversity in Alaskan estuaries. PhD, University of Alaska Fairbanks,

Morin PJ (2011) Community ecology. John Wiley and Sons, Ltd, Chichester, UK

MSAP (2012) Marine species with aquaculture potential, Oregon State University (http://hmsc.oregonstate.edu/projects/msap/index.html).

Mueter FJ, Norcross BL (1999) Linking community structure of small demersal fishes around Kodiak Island, Alaska to environmental variables. Marine Ecology Progress Series 190:37-51

Neal EG, Walter MT, Coffeen C (2002) Linking the Pacific decadal oscillation to seasonal stream discharge patterns in Southeast Alaska. Journal of Hydrology 263:188-197

NOAA (2012) NOAA Tides and Currents. http://tidesandcurrents.noaa.gov

Norcross BL, Blanchard A, Holladay BA (1999) Comparison of models for defining nearshore flatfish nursery areas in Alaskan waters. Fisheries Oceanography 8:50-67

82

Norcross BL, Mueter FJ, Holladay BA (1997) Habitat models for juvenile pleuronectids around Kodiak Island, Alaska. Fishery Bulletin 95:504-520

Olden JD (2003) A species-specific approach to modeling biological communities and its potential for conservation. Conservation Biology 17:854-863

Oppel S, Huettmann F (2010) Using a random forest model and public data to predict the distribution of prey for marine wildlife management. In: Cushman S, Huettmann F (eds) Spatial Complexity, Informatics and Wildlife Conservation. Springer, Tokyo, Japan

Oppel S, Strobl C, Huettmann F (2009) Alternative methods to quantify variable importance in ecology. Department of Statistics, University of Munich, Technical Report Number 65, Munich

Ouellette MH, Legendre P, Borcard D (2012) Cascade multivariate regression trees: a novel approach for modelling nested explanatory variables. Methods in Ecology and Evolution 3:234-244

Park Y-S, Tison J, Lek S, Giraudel J-L, Coste M, Delmas F (2006) Application of a self-organizing map to select representative species in multivariate analysis: A case study determining diatom distribution patterns across France. Ecological Informatics 1:247-257

Peres-Neto PR (2004) Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. Oecologia 140:352-360

Petrie ME, Ryer C (2006) Laboratory and field evidence for structural habitat affinity of young-of-the-year lingcod. Transactions of the American Fisheries Society 135:1622-1630

Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181-199

Sandri M, Zuccolotto P (2006) Variable selection using random forests. In: Zani S, Cerioli A, Riani M, Vechi M (eds) Data analysis, classification and the forward search: Proceedings of the meeting of the classification and data analysis group (CLADAG) of the Italian Statistical Society, Parma, June 6-8, 2005. Spring, Heidelberg

Schoch GC, Dethier MN (1996) Scaling up: the statistical linkage between organismal abundance and geomorphology on rocky intertidal shorelines. Journal of Experimental Marine Biology and Ecology 201:37-72

Segal M, Xiao Y (2011) Multivariate Random Forests. Data Mining and Knowledge Discovery 1:80-87

Sheaves M, Johnston R (2009) Ecological drivers of spatial variability among fish fauna of 21 tropical Australian estuaries. Marine Ecology Progress Series 385:246-260

SNAP (2011). www.snap.uaf.edu

Stojlgren TJ, Ma P, Kumar S, Rocca M, Morisette JT, Jarnevish CS, Benson N (2010) Ensemble habitat mapping of invasive plant species. Risk Analysis 30:224-235

Stone RP, O'Clair CE (2001) Seasonal movements and distribution of Dungeness crabs, *Cancer magister*, in a glacial southeastern Alaska estuary. Marine Ecology Progress Series 214:167-176

Stoner AW, Spencer ML, Ryer C (2007) Flatfish-habitat associations in Alaska nursery grounds: Use of continuous video records for multi-scale spatial analysis. Journal of Sea Research 57:137-150

Svetnik V, Liaw A, Tong C (2002) Variable selection in random forest with application to quantitative structure-activity relationship. In: Intrator N, Masulli F (eds). Proc 7th Course on Ensemble Methods for Learning Machines, September 22-28, 2002. Springer-Verlag

Townsend CR, Dolédec S, Norris R, Peacock K, Arbuckle C (2003) The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction. Freshwater Biology 48:768-785

USDA (2002) High and low tidelines. USDA Forest Service, Tongass National Forest, Southeast Alaska GIS Library, (http://seakgisalaskaedu)

USGS (1995) Alaska hydrologic units. US Geological Survey (http://nhdusgsgov/datahtml) accessed: 2009

USGS (1997) Alaska 300 m digital elevation model.

Wedding LM, Friedlander AM (2008) Determining the influence of seascape structure on coral reef fishes in Hawaii using a geospatial approach. Marine Geodesy 31:246-266

Wedding LM, Lepcxyk CA, Pittman SJ, Friedlander AM, Jorgensen S (2011) Quantifying seascape structure: extending terrestrial spatial pattern metrics to the marine realm. Marine Ecology Progress Series 427:219-232

Wehrly KE, Breck JE, Wang L, Szabo-Kraft L (2012) A Land-based classification of fish assemblages in sampled and unsampled lakes. Transactions of the American Fisheries Society 141:414-425

Weingartner T (2007) Ecosystem structure. In: Spies RB (ed) Long-term Ecological Change in the Northern Gulf of Alaska. Elsevier, Amsterdam, The Netherlands

Weingartner T, Eisner L, Eckert GL, Danielson S (2009) Southeast Alaska: oceanographic habitats and linkages. Journal of Biogeography 36:387-400

Whitlow WL, Grabowski JH (2012) Examining how landscapes influence benthic community assemblages in seagrass and mudflat habitats in southern Maine. Journal of Experimental Marine Biology and Ecology 411:1-6

Wiens JJ (2011) The niche, biogeography and species interactions. Philosophical Transactions of the Royal Society B: Biological Sciences 366:2336-2350

Wisz MS, Potter J, Kissling WD, Pellissier L, Lenoir J, Damgaard CF, Dormann CF, Forchhammer MC, Gryntes JA, Guisan A, Heikkinen RK, Høye TT, Kühn I, Luoto M, Majorano L, Nilsson MC, Normand S, Ockinger E, Schmidt NM, Termansen M, Timmermann A, Wardle DA, Aastrup P, Svenning J-C (2012) The role of biotic interactions in shaping distributions and realized assemblages of species: implications for species distribution modelling. Biological Reviews of the Cambridge Philosophical Society doi: 10.1111/j.1469-185X.2012.00235.x

Chapter 4: Taxonomic and functional diversity of fishes in Alaskan estuaries

ABSTRACT

The change in diversity between communities (β-diversity) has been widely used as a metric for evaluating the biotic and abiotic factors affecting community composition. In particular, understanding the functional composition of ecosystems can provide insight into how gain or loss of species may affect ecosystem processes. I compared functional and taxonomic β-diversity between 49 estuaries in Southeast Alaska to 1) evaluate how estuaries differed in terms of species composition and functions, 2) investigate patterns of species dominance, and 3) and to examine factors affecting estuarine community assembly. Species turnover within the region was high (74%) but functional turnover was much lower (23%). Functional diversity did not have a strong linear relationship to species richness indicating that the number of species in the estuary was not a good indication of how functionally diverse the estuary was. Both biotic and abiotic factors were important for structuring estuarine fish communities. The results of this research provide insight into the vulnerability of these estuarine ecosystems to change, including alterations in species composition as a result of anthropogenic disturbance or exploitation.

INTRODUCTION

A fundamental issue for ecosystem management is the understanding of the relationship between the composition of the biotic community and the functioning of the ecosystem. How do changes in the composition of species affect ecosystem properties? How well can the ecosystem compensate for species' gains or losses? Many community analyses compare the presence or abundance of species, or the change in species composition, between sampling areas without consideration of the functional roles that those species perform (Devictor et al. 2010, Villéger et al. 2012) resulting in an incomplete view of ecosystem structure. The functional characteristics of species influence ecosystem processes and, in concert with information on taxonomic composition, can better explain ecosystem functioning than taxonomic comparisons of species occurrence or abundance alone (Hooper et al. 2005, Wright et al. 2006, Devictor et al. 2010, Mouchet et al. 2010).

Functional groups, or guilds, are assemblages of species that exploit ecological resources in the same way. The taxonomic diversity in a community is a function of hierarchical processes, including evolutionary processes and physiological constraints defining a regional species pool

that is then additionally limited in the local community by the species' dispersal ability and interspecific interactions (Tilman et al. 1997, Guillemot et al. 2011). Functional diversity is entirely independent of taxonomic restrictions and assignment to functional groups or guilds is based on the species' specific resource use or an ecological process they engage in (Simberloff and Dayan 1991). Competitive interactions among species from the same group or guild can increase functional diversity by enhancing resource partitioning or decrease diversity by competitive exclusion (de Bello et al. 2009). The extent to which mechanisms that change species composition, such as disturbance or invasion, affect ecosystem processes depends on the amount of functional redundancy in the system and how species with similar functional attributes differ in their response to change (Tilman et al. 1997, Pratchett et al. 2011).

Ecosystems have high functional redundancy when there are a number of species that perform the same function. Changes in the composition of functionally similar species will have a smaller effect on ecosystem processes than changes in functionally dissimilar species or loss of entire functional groups (Naeem 1998, Folke et al. 2004, Foley et al. 2010, Pratchett et al. 2011). Ecological niche theory states that species are functionally different and coexist as a result of functional specialization. Habitat filtering results in communities of species with similar functional traits adapted to the conditions of the habitat, while species interactions (e.g., competition) result in communities of functionally dissimilar species (Podani 2009, de Bello et al. 2012). These mechanisms often act simultaneously at different spatial scales (Mouchet et al. 2010). Understanding the functional structure of an ecosystem can provide insight into the system's vulnerability to change, including alterations in species composition as a result of anthropogenic disturbance or exploitation.

Estuaries provide a unique opportunity to evaluate the relative contributions of abiotic and biotic filtering. These complex systems are characterized by widely varying hydrological, morphological and chemical conditions, and estuarine fish demonstrate resource partitioning, particularly for habitat (Attrill and Power 2004, Rooper et al. 2006, Eaton 2010). In Southeast Alaska, more than a third of the commercially managed fish species have been documented to occur in estuaries, many of them as juveniles, making these habitats potentially important nursery grounds (Murphy et al. 2000, Johnson et al. 2003). I compared functional and taxonomic β-diversity of Southeast Alaskan estuarine fish to identify the factors affecting each type of diversity and evaluate whether estuaries with dissimilar species are ecologically dissimilar as well.

MATERIALS AND METHODS

<u>Study area and sampling</u>

      This research was conducted in the Alexander Archipelago, a collection of approximately 1,000 mountainous islands in Southeast Alaska from Dixon Entrance at the Canadian border (54º 47'35"N, 130º 38'06"W) to Lance Point in Lynn Canal (58º 44' 141"N, 135º 13' 996"W). The entire archipelago is a temperate rainforest; precipitation varies locally and regionally with a general gradient of lower precipitation in the northwest and higher precipitation in the southeast. Average annual precipitation in the region is in excess of 1000 mm/yr (Neal et al. 2002) with much of the precipitation being released directly into the marine waters via numerous small streams and wetlands. Stream flow is highly seasonal and influenced both by precipitation and by snow and ice melt. The highest stream flows tend to occur in autumn when precipitation rates are high. Flows decrease in winter as a result of freezing, and increase again in the late spring and summer from melting of snow and ice. The flow of freshwater affects not only nearshore estuarine circulation, but is the driver for larger scale oceanographic circulation within Southeast Alaska's interior channels and on the continental shelf (Weingartner et al. 2009). Stream and river temperatures are influenced both by air temperatures and by runoff from glaciers, snowmelt, and precipitation. Estuaries in the study area differ in their hydrological and geomorphological characteristics: most estuaries have a mixture of soft and hard substrate shorelines, but the amount of each type of substrate varies depending on both oceanographic and terrestrial processes.

      Fish were sampled in 49 estuaries between 1998 and 2005 using both trawl and seine gear (Figure 4.1). Sampling was conducted during daylight hours between February and September. An otter trawl (3 m x 1 m, with 6 mm square mesh in the cod end) was deployed with a bridle scope of approximately 20 m and towed at a speed of approximately 3 kn along a depth contour between 5 m and 10 m in each estuary. One tow in each direction was made along the same transect during high and low slack water, and the modal depth of the tow was recorded. The length of the trawl was calculated by marking the beginning and ending coordinates on the GIS and measuring along the appropriate depth contour on the chart. Fish also were sampled with a 37-m long variable mesh beach seine that tapered from 5 m deep at the center to 1 m at the ends. Outer panels were each 10 m with 32 mm stretch mesh, intermediate panels were each 4 m with 6 mm square mesh, and the bunt was 9 m with 2.3 mm square mesh. The net was set as a round haul by fixing one end on the beach, backing the skiff while deploying the net, and

bringing the other end to shore approximately 18 m down shore from the first end. The latitude and longitude for each estuary were recorded. For both sampling methods, captured fish were identified to species and measured in the field to the nearest millimeter for total length.

Trawl and seine gear target different species. As both gear types were used at all the estuaries in my study, I combined the catch from both gears to calculate composite species composition in each estuary. This results in a more representative grouping of resident nearshore species than using data from each gear type separately (Lapointe et al. 2006, Ruetz III et al. 2007). For trawl samples, catch per unit effort (CPUE) was calculated as the number of species caught divided by trawl length and standardized to number of fish per 100 m. For seine samples, CPUE was the number of fish per haul. The relative abundance of each species in an estuary was calculated by dividing the combined CPUE of the species by the combined catch of all species. All fish and invertebrates were collected by NOAA/NMFS under their collection permits.

<u>Diversity calculations</u>

Methods for assigning estuarine species to functional groups are quite varied, and include life history characteristics such as estuarine use or residence (Elliott et al. 2007, Nicolas et al. 2010, Franca et al. 2011), feeding preferences and feeding mode (Elliott et al. 2007, Franco et al. 2008, Cardoso et al. 2011), reproductive mode (Mathieson et al. 2000, Elliott et al. 2007, Fisher et al. 2011), morphology (Bremner et al. 2003, Dimitriadis and Koutsoubas 2011) and behavior (Fisher et al. 2011). For this research, I collated available life history and ecological data for each species into a matrix of functional traits (Table 4.1) describing body size, body shape, trophic level, prey type, habitat use, and spawning season. This information on available life history and ecological traits was compiled from Fishbase (Froese and Pauley 2012), supplemented by the National Marine Fishery Service's Life History Database (NMFS 2012), and primary literature. Data from a taxonomically close species was used when species-specific information (i.e., nest guarding) was not available. These exceptions are noted in Table 4.1. The functional trait information available for less well studied species was primarily for adult fish. Therefore, I included adult information for all fish species even though many fish occurred primarily as juveniles in the data, and other species were represented by a mix of juveniles and adults (Miller 2013a). There was insufficient information on ontogenic changes in functional traits to include life-stage specific data in the analysis.

Functional distance between species was calculated using the Gower coefficient, which is a similarity measure that can handle a mix of quantitative and categorical variables (Romesburg 2004, Botta-Dukát 2005, Pavoine et al. 2009). Principal components analysis (PCA) was used to visualize each species' relative position in functional space and to identify group membership and species dominance within estuaries and clusters. For the PCA, the table of species morphological traits (Table 4.1) was converted to a matrix with binary (0 and 1) indications for whether a species had the trait or not. PCA was implemented using the vegan package in R (Oksanen et al. 2011).

Functional diversity was calculated using Rao's quadratic entropy (Rao 1982, Ricotta and Marignani 2007) because this is the only diversity estimator that combines measures of species functional dissimilarity with species relative abundance (de Bello et al. 2010). This index can be used to compute functional α- and γ-diversity, from which functional β-diversity can be calculated (de Bello et al. 2010, Mouchet et al. 2010). Rao's Q for α-diversity (Q_α) is defined as:

$$Q_\alpha = \sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij} p_i p_j$$

Where d_{ij} is the functional dissimilarity between species i and j, p_i and p_j are the relative abundances of species i and j in an individual estuary, and S is the species richness of the estuary. The functional diversity, Q_α, corresponds to the expected dissimilarity between two randomly chosen individuals in an estuary and is at a maximum when all species in the estuary are maximally dissimilar and equally abundant.

Rao's Q for γ-diversity (Q_γ) is calculated on the regional species pool: the relative abundances of all species in the sampling area. The formula for Q_γ is identical to the formula for Q_α

$$Q_\gamma = \sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij} p_i p_j$$

where d_{ij} is still the functional dissimilarity between species i and j, but p_i and p_j are regional relative abundances of each species, and S is the total number of species in the sampling region (de Bello et al. 2010). Q_γ is the expected dissimilarity between two randomly chosen individuals in the study area.

Entropies, such as Rao or Shannon-Wiener, are measures of the uncertainty associated with determining whether species randomly drawn from a community are the same. These

entropies are not true diversity indices and values cannot be directly comparable between communities unless they are converted to equivalent or effective number of species: the number of equally abundant species needed to produce the diversity value (Jost 2006, 2007, de Bello et al. 2010). This conversion is accomplished for Q_α and Q_γ as follows:

$$Q_{\alpha Eqv} = \frac{1}{1 - Q_\alpha}$$

and

$$Q_{\gamma Eqv} = \frac{1}{1 - Q_\gamma}$$

The proportional functional β-diversity between sites can be expressed as:

$$\beta_{prop} = \frac{Q_{\gamma Eqv} - Q_{\alpha Eqv}}{Q_{\gamma Eqv}}$$

When applied at the regional level, this equation has an upper limit defined by the number of sampling units and takes a value of 1 or 100% when all sampling units are completely distinct. When applied to pair-wise comparisons of composition between estuaries, this equation represents the percent turnover in composition, or degree of change in composition, between two estuaries. It takes a maximum value of 0.5, or 50%, when two sites do not share any species in common (de Bello et al. 2010).

To maintain consistency in diversity measures, I also used Rao's Q to calculate taxonomic diversity. This is done by setting the pairwise functional distances to 1 for all i≠j (maximum distance) and to 0 for all i=j. Constraining the dissimilarity matrix in this way reduces Rao's Q to the Simpson index (Ricotta and Marignani 2007), which measures the probability that two individuals selected at random from a population belong to different species. Values for taxonomic Q_α, Q_β, Q_γ, and β_{prop} are computed using this constrained distance matrix in the same manner as for functional diversity.

Null models

To evaluate whether observed functional and taxonomic β-diversity differed from random expectations, I constructed null models using permutations of species abundances to construct new species assemblages. Null models test the hypothesis that patterns of taxonomic and compositional diversity are non-random resulting from abiotic or biotic factors that influence community assembly (Wilson 1987, Azevedo et al. 2006, de Bello 2012). Null models were constructed by holding species richness constant and randomly sampling species from the species

pool. This approach assumes that each species has an equal opportunity of occurring in all estuaries. The R library picante (Kembel et al. 2010) was used to construct 1000 random matrices. I computed functional and taxonomic β-diversity independently for each of the random species matrices. Random expectation for functional and taxonomic β-diversity was calculated as the mean of diversities for the 1000 matrices for each diversity type. The difference between observed and predicted β-diversity was calculated by subtracting the mean functional and taxonomic β-diversity of the null models from observed functional and taxonomic β-diversity, respectively. This difference was divided by the standard deviation of the null models to obtain the standardized effect size index (SES) (de Bello et al. 2010, Azeria et al. 2011). SES values greater than zero indicate estuaries with diversity more dissimilar than expected by chance (trait divergence), while SES values below zero indicate that estuaries are more similar than expected by chance (trait convergence).

Estuary Clusters

Miller (2013b) used multivariate random forest models (Segal and Xiao 2011) to predict trawl-captured estuarine fish and invertebrate assemblage composition with respect to tidal height, seasonal precipitation, amount of intertidal habitat, estuary depth and slope, and amount of subtidal vegetation. The proximity matrix from this model was clustered using the partition around the mediod (PAM) clustering method to create three clusters of estuaries with similar environmental composition and relative abundance and composition of species. For each of these clusters, I calculated functional β-diversity and used multivariate dispersion (Anderson et al. 2006) to assess whether functional diversity of estuaries within the clusters was different than diversity between the clusters. ANOVA was then used to test if dispersions between clusters were significantly different. Species abundance for a cluster was calculated by summing abundances across all estuaries in the cluster. These abundances were used to calculate functional β-diversity for the cluster in the same manner as these indices were calculated for individual estuaries. Clusters also were used to evaluate differences in species dominance by evaluating the composition and relative abundance of species in each cluster

.

RESULTS

A total of 6,431 fish were captured from all estuaries comprising 57 species from 23 families. Chum salmon (*Oncorhynchus keta*) were the most numerically abundant fish species,

comprising 12% of the total catch across all years, followed closely by pink salmon (*Oncorhynchus gorbuscha*) with 10% of the total catch. Twelve species were captured at more than 40% of the sites: coho salmon (*Oncorhynchus kisutch*), chum salmon, pink salmon, three-spine stickleback (*Gasterosteus acculeatus*), buffalo sculpin (*Enophrys bison*), rock sole (*Lepidopsetta sp.*), starry flounder (*Platichthys stellatus*), crescent gunnel (*Pholis laeta*), kelp greenling (*Hexagrammos decagrammus*), Bay pipefish (*Syngnathus leptorhynchus*), snake prickleback *(Lumpenus sagitta)*, and Pacific staghorn sculpin (*Leptocottus armatus*). Approximately half the species occurred in 10% or fewer of the estuaries.

The PCA of the functional traits identified three broad functional groups (Figure 4.2). The first two principal components explained 44% of the variation. Functional group 1 was characterized by higher trophic level anadromous and pelagic species and included all seven species of salmonids, Pacific cod (*Gadus macrocephalus*), walleye pollock (*Theragra chalcogramma*), and threespine stickleback (Figure 4.3). This group had the lowest number of taxa and contributed 42% of the total species abundance (Table 4.2). Within functional group 1, chum salmon, pink salmon, and three-spine stickleback comprised 66% of the species abundance (28, 24%, and 14%, respectively). Functional group 2 was characterized by benthic marine species that rest on the sea floor and mostly do not guard their nests. This included all six pleuronectids, sturgeon poacher (*Podothecus accipenserinus*), tubenose poacher (*Pallasina barbata)*, buffalo sculpin, lingcod (*Ophiodon elongatus*), and Pacific tomcod (*Microgadus proximus*). This group contributed 21% of the total species abundance. Starry flounder (*Platichthys stellatus*), yellowfin sole (*Limanda aspera*), and rock sole (*Lepidopsetta sp.*) comprised 72% of the total species abundance for functional group 2 (26%, 28% and 18%, respectively). Functional group 3 was characterized by benthopelagic, brackish, and intertidal species, which occur in the water column above the sea floor. Many of these species guard their nests or brood their young. This group contained the majority of the cottids, bay pipefish, crescent gunnel, saddleback gunnel (*Pholis ornata*), tube-snouts (*Aulorhynchus flavidus*), and Arctic shanny (*Stichaeus punctatus*). This group had the highest number of taxa and contributed 37% of the total species abundance. Species dominance, measured by species abundance, in functional group 3 was less defined. Snake prickleback (*Lumpenus sagitta*), Pacific herring, crescent gunnel, and Pacific staghorn sculpin accounted for 58% of the abundance in relatively equal portions within functional group 3 (18%, 15%, 13% and 12%, respectively).

The three clusters of estuaries from the multivariate random forest model (Miller 2013b) were used to assess differences in the functional composition of fishes (Table 4.2). Tidal height and minimum precipitation were the most influential variables in the multivariate random forest models, and the spatial distribution of the clusters reflects these influences (Figure 4.4). Estuaries in Cluster 1 are located in the areas of highest precipitation and moderate tidal heights. This cluster had 18 estuaries with a total of 39 species and 2056 individuals. Forty-six percent of the total catch for this cluster was species from functional group 1, 18% was from functional group 2, and 36% was from functional group 3 (Figure 4.5). Only two species (chum salmon and pink salmon) each comprised more than 10% of the total relative abundance of species in the cluster, but eleven species were captured in 40% or more of the estuaries. Two estuaries from this cluster did not have any species from functional group 2, and one estuary did not have species from functional group 1. The average taxonomic α-diversity (Q_{α_tax}) for this cluster was 15.5, and average functional α-diversity (Q_{α_funct}) was 2.23. Both of these measures are expressed in equivalent numbers of species. Average species turnover (β_{prop_tax}) between estuaries in this cluster was 39%, while average functional turnover (β_{prop_funct}) was 20%.

Estuaries in Cluster 2 are located in the areas with the lowest precipitation in the rain shadow of the coastal mountains. These estuaries also had relatively high tidal heights. This cluster had 20 estuaries with 47 species and 3214 individuals. Forty-three percent of the total catch for this was species from functional group 3, while the remainder was nearly evenly divided between functional groups 1 (29%) and 2 (28%). Pacific herring (*Clupea palasii*) was the only species with more than 10% of the total relative abundance of species in the cluster with over 90% of the overall Pacific herring captured in cluster 2 estuaries. Over 70% of the overall catch of starry flounder, yellowfin sole, and rock sole also occurred in estuaries in cluster 2. The average taxonomic α-diversity (Q_{α_tax}) for this cluster was 17.2, and average functional α-diversity (Q_{α_funct}) was 2.28. Average species turnover (β_{prop_tax}) between estuaries in this cluster was 39% while average functional turnover (β_{prop_funct}) was 21%.

Cluster 3 was the smallest cluster having 10 estuaries with 31 species and 1161 individuals. These estuaries had the lowest tidal heights and were located adjacent to the open ocean on the outer coast of the archipelago. These estuaries also had relatively high precipitation. Species from functional group 1 comprised 72% of the total species abundance in this cluster. Three species (chum salmon, pink salmon, and three-spine stickleback) from functional group 1 each contributed over 15% of the total relative abundance of this cluster. These three species also

were captured in 90% or more of the estuaries in this cluster. A little over half of the three-spine stickleback was captured in cluster 3 estuaries. Species from functional group 3 contributed 21%, and species from functional group 2 contributed only 6% of the relative abundance. All estuaries in this cluster had species from all of the functional groups 2. The average taxonomic α-diversity (Q_{α_tax}) for this cluster was 5.62, and average functional α-diversity (Q_{α_funct}) was 1.76. Average species turnover (β_{prop_tax}) between estuaries in this cluster was 18% while average functional turnover (β_{prop_funct}) was 5%.

Species richness ranged from 5 to 24 species per estuary, with an average of 12 species in each estuary. β_{prop_tax} between all estuaries in the region was 74%, with 83% of the sites having pair-wise species turnover of 25% or higher. Average β_{prop_tax} across all pair-wise samples of estuaries in the region was 38%. β_{prop_tax} between estuary clusters was highest for clusters 2 and 3 (31%) and 1 and 3 (21%) and lowest between clusters 1 and 2 (14%). All measures of functional diversity were much lower than taxonomic diversity. β_{prop_funct} in the region was 22.7% and fewer than 15% of the pair-wise samples had β_{prop_funct} greater than 20%. Average β_{prop_funct} between pair-wise samples was 11%. Approximately half the estuaries had functional turnover of less than 10%. β_{prop_funct} between estuary clusters followed a similar pattern to β_{prop_tax} with the highest functional turnover occurring between clusters 2 and 3 (6%) and clusters 1 and 3 (4%) and the lowest turnover occurring between estuaries in clusters 1 and 2 (2%). The multivariate dispersion and ANOVA analyses found significant differences in functional β-diversity between cluster 3 and the other two clusters, but not between clusters 1 and 2 (Figure 4.6).

Species richness was not strongly correlated to functional α-diversity (R = 0.56): having a higher number of species present in an estuary did not result in an increase in functional attributes within that estuary (Figure 4.7a). Regionally, functional α-diversity was related to taxonomic α-diversity with functional diversity showing a general increase as taxonomic diversity increased (Figure 4.7b), but the relationship was not strongly linear (R=0.77). For estuaries in cluster 1, and to a lesser extent in cluster 2, changes in taxonomic and functional β-diversity between estuary pairs was not strongly linear (R=0.55, and R=0.73, respectively) (Figures 4.8a and 4.8b). This was not the case for estuaries in cluster 3, where changes in taxonomic β-diversity were strongly correlated to changes in functional β-diversity (R=0.94) (Figure 4.8c).

Neither taxonomic nor functional β-diversity were correlated to β-diversity from the null models (R=0.23 and R=0.12, respectively) (Figures 4.8a and 4.8b). Functional β-diversity was strongly correlated (R=0.90) to differences between observed and null model predictions $(\beta_{\_Diff})$,

but the correlation for taxonomic β-diversity was weaker (R=0.63). Observed taxonomic β-diversity was generally lower than predicted by the null models (Figure 4.8c). SES values for taxonomic diversity were below zero indicating estuaries were less dissimilar than predicted by the null model. In contrast, observed functional β-diversity for many estuaries was higher than null model predictions (Figure 4.9). SES values for the estuaries with low observed functional diversity were zero or below, indicating that these estuaries were less functionally diverse than predicted by the null models, while SES values for estuaries with higher observed functional diversity were above zero indicating estuaries that were more functionally diverse than null models predicted.

DISCUSSION

Comparing differences in functional and taxonomic diversity between estuaries can provide insight into patterns of species co-occurrence and the biotic and abiotic processes shaping community composition. Southeast Alaska estuaries differ substantially in their hydrological and geomorphological characteristics (Miller 2013a, b). Thus it is reasonable to expect that this heterogeneity would affect both species composition within estuaries as well as diversity between estuary pairs. I found regional species richness to be relatively high and to vary substantially between estuaries. This high species richness is coupled with high proportional species turnover. In contrast, functional diversity was low in both the region and the estuary clusters.

Functional diversity did not have a strong linear relationship to species richness indicating that the number of species in the estuary was not a good indication of how functionally diverse the estuary is. The positive relationship between functional α-diversity and taxonomic α-diversity indicates that the loss or addition of a species tended to result in a decrease or increase in functional diversity, but the relationship was not strongly linear. In general, functional diversity showed relatively high values at high values of taxonomic diversity, but at intermediate taxonomic diversity values functional diversity could be either high or low. Along with the overall low functional diversity of the region, this suggest that there is functional redundancy in the estuarine fish community such that in some estuaries, gains or losses in species did not have a strong effect on ecosystem processes, while in other estuaries the effects of species changes were stronger.

Cluster 3 estuaries were substantially less diverse both taxonomically and functionally than estuaries in the other two clusters. These estuaries had very low species turnover and were

dominated by anadromous species from functional group 1. The combination of low taxonomic and functional diversity is consistent with communities where environmental constraints select for similar functional attributes. The strong positive correlation between proportional change species diversity and proportional change in functional diversity indicates that estuaries with the greatest difference in taxonomic composition also tended to be the estuaries that were more functionally diverse. In both this analysis and the multivariate random forest models, estuaries in this cluster had the lowest number of fish taxa. In the multivariate random forest models, these estuaries also had the highest abundance of mobile invertebrates. Invertebrates were not included in this functional analysis because the taxonomic resolution of the invertebrate data was not fine enough for attributing functional traits. The dominance of anadromous species, which are present in the estuaries during only a portion of their life, and the low prevalence of other fish species, suggests that other taxa, such as invertebrates, may be performing functional traits in these estuaries that are performed by fish in the other clusters.

For estuaries in clusters 1 and 2, both niche filtering as a result of environmental factors, and similarity limitation through competition appear to be an important structuring mechanism. Species and functional turnover between estuaries in both of these clusters was high, but turnover between the two clusters was relatively low. Unlike estuaries in cluster 3, estuaries in these two clusters demonstrated high functional redundancy. At low levels of both functional and taxonomic diversity, proportional changes in species composition were more likely to result in an equivalent proportional change in functional diversity. However, this relationship did not hold well at higher levels of taxonomic diversity. Large differences in species composition between estuaries did not necessarily equate to large differences in functional diversity. Even where turnover in species between the estuaries was complete and the estuaries shared no species in common, functional turnover could be quite low. This results from a large pool of functionally similar species that occurred in a relatively small portion of the estuaries. Twelve species occurred in 40% or more of the estuaries at relative abundances of 5% or more, but nearly half of all species occurred in fewer than 10% of the estuaries. While abundant species were evenly distributed across the three functional groups (7, 6 and 9 species), the distribution of rare species was more uneven. Twenty percent (10 species) of the rare species were benthopelagic, intertidal and brackish species from functional group 3, 13% (7 species) were benthic marine species from functional group 2 and only 8% (5 species) were anadromous and pelagic species from functional

group 1. Clusters 1 and 2 both had similar functional composition, with 50% or more of the species abundance coming from the functional groups with the largest number of rare species.

Taxonomic diversity was substantially lower than predicted by the null models, indicating that, with respect to species composition, estuaries were more similar than predicted by chance. The null models were constructed to give all species equal consideration in every estuary. The taxonomic model results suggest that, in reality, species distributions may be more constrained and that some species are limited by either environmental (e.g., currents and dispersal opportunities) or biotic (e.g., resource partitioning) factors, or both. Random assignment of species to estuaries in the null models thus resulted in higher potential diversity than actually existed. For many estuaries, the null models predicted higher functional overlap or similarity between estuaries than was observed. This pattern is suggestive of environmental filtering within each estuary that selects for species that are adapted to the environmental conditions there (Hooper et al. 2005, de Bello et al. 2009).

These results provide information about potential gains and losses in functional diversity associated with changes in estuarine fish communities. Although the relationship between functional and taxonomic diversity is positive, it is not strictly linear, and in this research I have shown that for Southeast Alaskan estuaries, higher taxonomic diversity does not equate to higher functional diversity. This finding is important for biodiversity conservation. Some management guides have suggested that maintaining species diversity will adequately protect functional diversity (Foley et al. 2010) , but this research suggests that that is not necessarily the case. The estuaries in cluster three show the greatest potential for loss of functional diversity if species were removed. Since the dominant species in these estuaries are anadromous, they are susceptible to changes in both marine and terrestrial habitats. Because of their higher functional redundancy, estuaries in the other two clusters are less susceptible to ecosystem perturbations from changes in species composition. It should be noted, however, that this analysis does not address the potential role of these estuaries as nursery areas for juvenile fish. Many fish occurred in these estuaries primarily as juveniles or as mixtures of juveniles and adults, but the functional trait data used in this analysis were only for adults because data for juveniles were limited. Juvenile fish may have different habitat and feeding patterns than adults, and incorporating this information could change the functional relationships between species in the data. Research on coastal marine assemblages in California found that even a small increase in functional specificity can have a substantial impact on the sensitivity of the functional diversity analysis (Micheli and Halpern 2005).

Finer resolution trophic data also would be very useful. For all of the species in the data I was able to obtain general information on prey items, but even for well-studied species there is little information on preferred prey or prey sizes and this lack of specificity results in greater functional redundancy. For species like poachers, gunnel, and sculpin, which are not of commercial interest or are not forage for commercial species, all types of functional data are particularly limited. This is unfortunate, as some of these taxa, such as sculpin, prey on the commercial species at certain life-stages (Spies et al. 2012). In combination with length-frequency data more detailed information on predator-prey interactions would be especially valuable in evaluating niche specialization or competition. As fisheries management moves toward ecosystem-based approaches, more research will need to be directed toward obtaining functional trait data for commercial and non-commercial species at various life stages.

The diversity calculation I used for this analysis is susceptible to sampling issues associated with schooling and migratory species. By considering species abundance, Rao's index gives more weight to differences in functional diversity between dominant species than less dominant species (de Bello et al. 2009). That is because a larger number of individuals will contribute more to a function than a smaller number of similar species. Juvenile salmon, shiner perch (*Cymatogaster aggregata*), and Pacific herring were among the most abundant species in my estuaries and all of these species exhibit schooling behavior. Salmon were captured in a large proportion of estuaries, and they are generally well sampled by the gears used. It is therefore unlikely that salmon are artificially skewing the data in this analysis. However, salmon also are seasonal users of estuaries in the study area, with highest abundance occurring in spring during outmigration from streams. Thus, their high abundance during the period when the estuaries were sampled may not reflect their ecological role during other times of the year. Addressing this would require repeated sampling to allow comparison across season. Shiner perch and Pacific herring were captured in 5% or fewer of the estuaries. Because these species are common in the study area, it is likely that this represents a sampling bias rather than a true distribution. Both species exhibit onshore and offshore movement within estuaries (Penttila 2007, Froese and Pauley 2012), possibly related to temperature. Temperature changes could induce these species to move into and out of the capture range of the gear used in this study, thereby skewing the diversity comparisons between estuaries where these species were captured in large numbers and estuaries where they were not. There is no easy way to address this potential bias except with more comprehensive sampling.

Management and conservation of estuarine ecosystems will require understanding of how species composition and species functional roles affect ecosystem processes. These relationships can provide insight into the vulnerability of these ecosystems to anthropogenic or natural change, and can inform management and restoration efforts. My results suggest that both environmental niche filtering and biotic interactions are important in structuring fish assemblages in Southeast Alaska estuaries, and that these factors act within estuary clusters as well as between individual estuaries. This information can be used to develop strategies for habitat protection, resource allocation, or response to perturbations such as the introduction of invasive species, ocean acidification, sea level rise, and climate change.
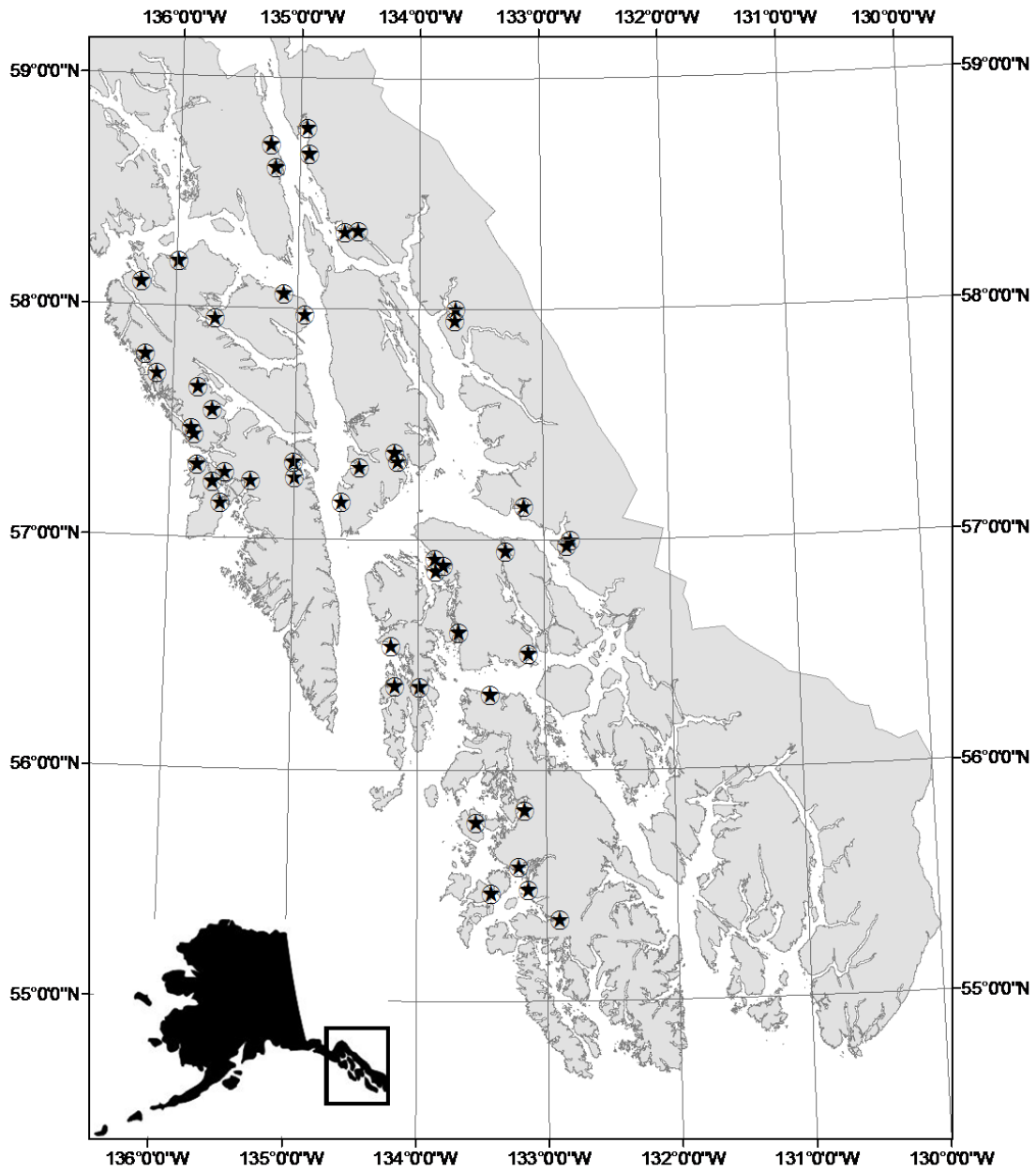
Figure 4.1: Estuaries sampled between 1998 and 2005 in Southeast Alaska.
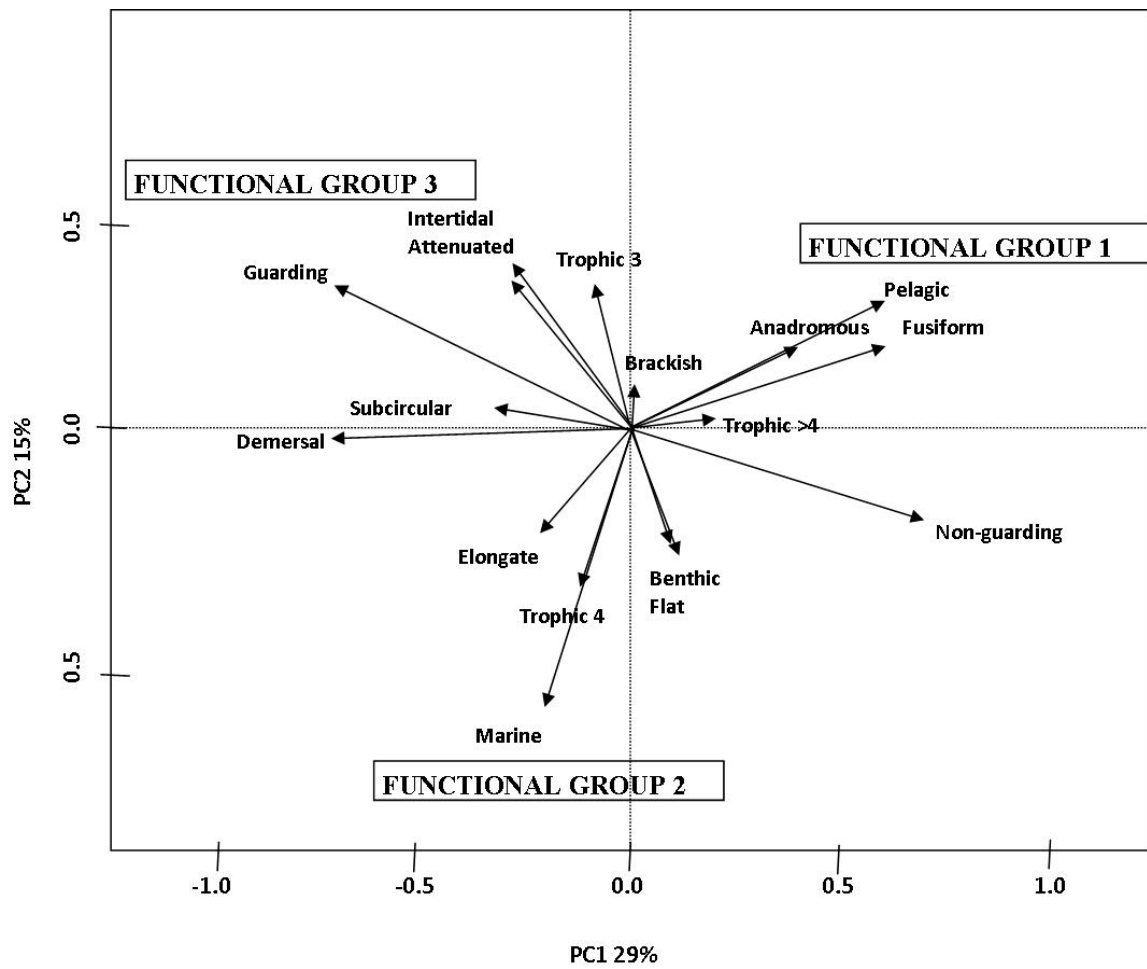
Figure 4.2: Principal component analysis of species functional traits for estuarine fish showing variation among species with vectors of traits. The first two principal component axes account for 44% of the variation among species.
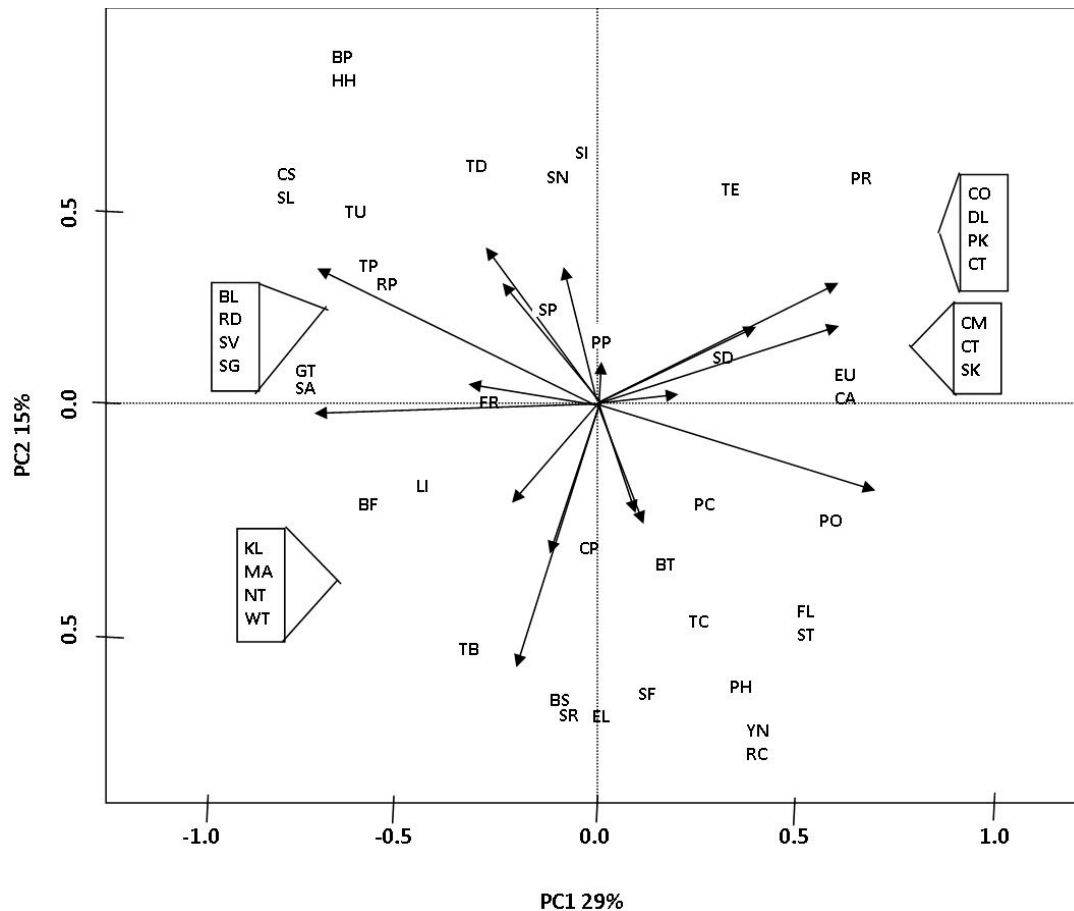
Figure 4.3: Variation in sign and strength of the life history traits. Codes for species are: BF=Buffalo sculpin, BL=Black Irish lord, BP=Black prickleback, BT=Butter sole, CA=Capelin, CM=Chum salmon, CO=Coho salmon, CP=Canary rockfish, CS=Crescent gunnel, CT=Cutthroat trout, DL=Dolly varden, EL=English sole, EU=Eulachon, FT=Flathead sole, FR=Frog sculpin, GT=Great sculpin, HH=High cockscomb, KL=Kelp greenling, LI=Lingcod, MA=Masked greenling, NT=Northern sculpin, PC=Pacific cod, PH=Pacific halibut, PK=Pink salmon, PO=Walleye Pollock, PP=Bay pipefish, PR=Pacific herring, RC=Rock sole, RD=Red Irish lord, RC=Rock greenling, RP=Rock prickleback, SA=Sailfin sculpin, SD=Pacific sand land, SF=Pacific sand fish, SG=Staghorn sculpin, SH=Rainbow trout, SH=Shiner perch, SK=Sockeye salmon, SL=Saddleback gunnel, SN=Arctic shanny, SP=Snake prickleback, SR=Sturgeon poacher, SS=Spinynose sculpin, ST=Starry flounder, SV=Silverspotted sculpin, TB=Tubenose poacher, TC=Pacific tom cod, TD=Tadpole sculpin, TE=Three-spine stickleback, TP=Tidepool sculpin, TU=Tube-snout, WT=Whitespotted greening, YN=Yellowfin sole.

Figure 4.4: Spatial distribution of estuaries clustered using the multivariate random forest model (Miller 2013b) and PAM clustering. Cluster 1=● ,Cluster 2 = ●, and Cluster 3 = ●
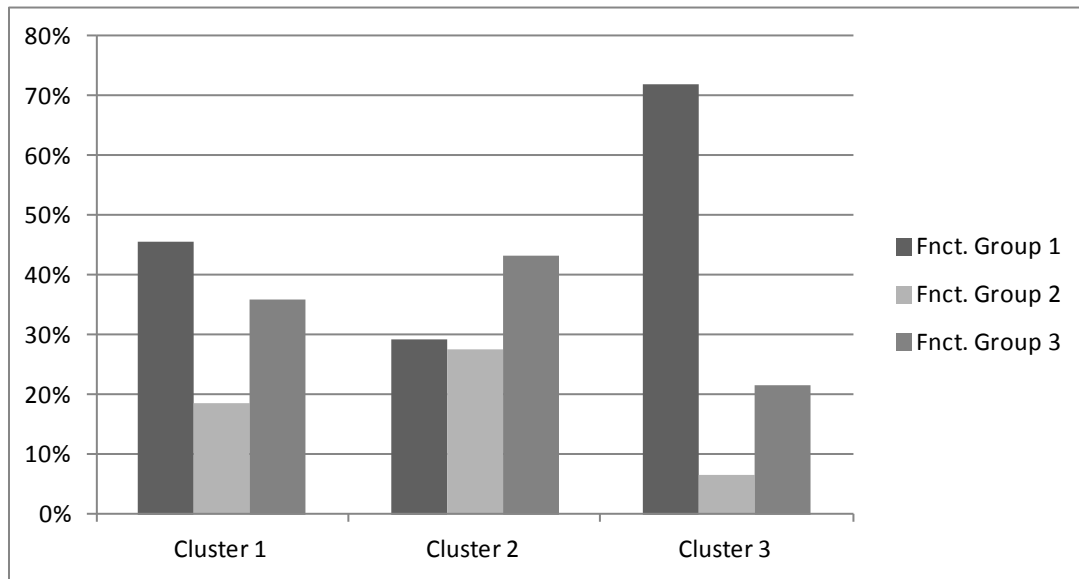
Figure 4.5: Relative of abundance of species from each functional group in the random forest clusters (Miller 2013b).
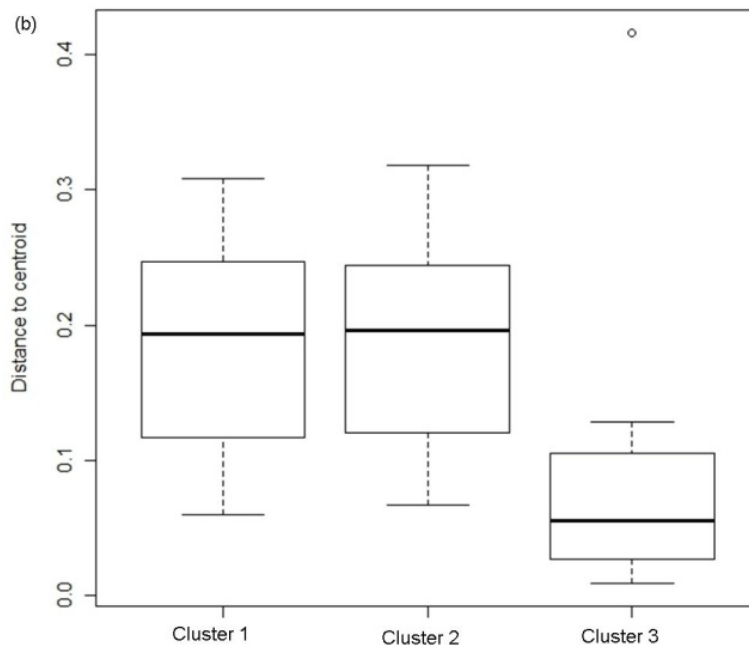
Figure 4.6: Boxplot of ANOVA analysis of differences in multivariate dispersion of functional β-diversity for estuaries in the random forest clusters (Miller 2013b). Lines mark median values while circles show extreme values. A significant difference was found only between cluster 3 and the other two clusters.
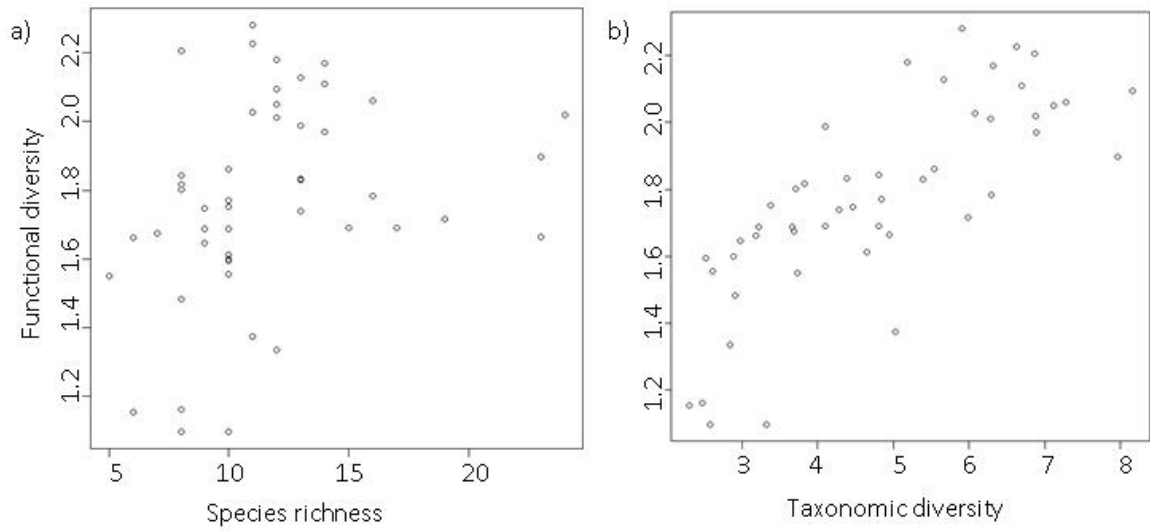
Figure 4.7: Relationship between species richness and functional β-diversity (a), and between taxonomic β-diversity and functional β-diversity (b).
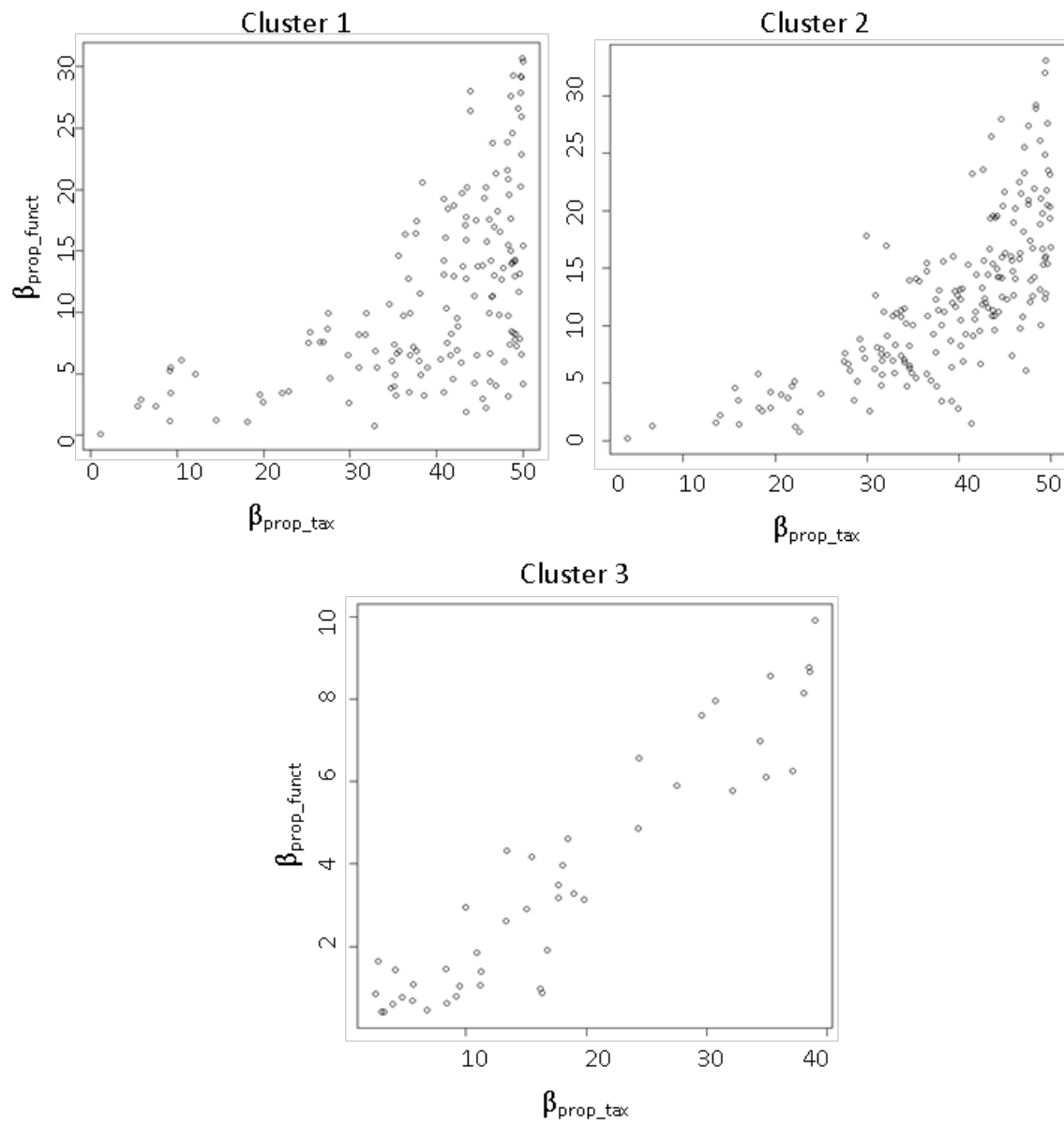
Figure 4.8: Relationship between changes in species turnover (β_{prop_tax}) and functional turnover (β_{prop_funct}) between estuaries in the three clusters
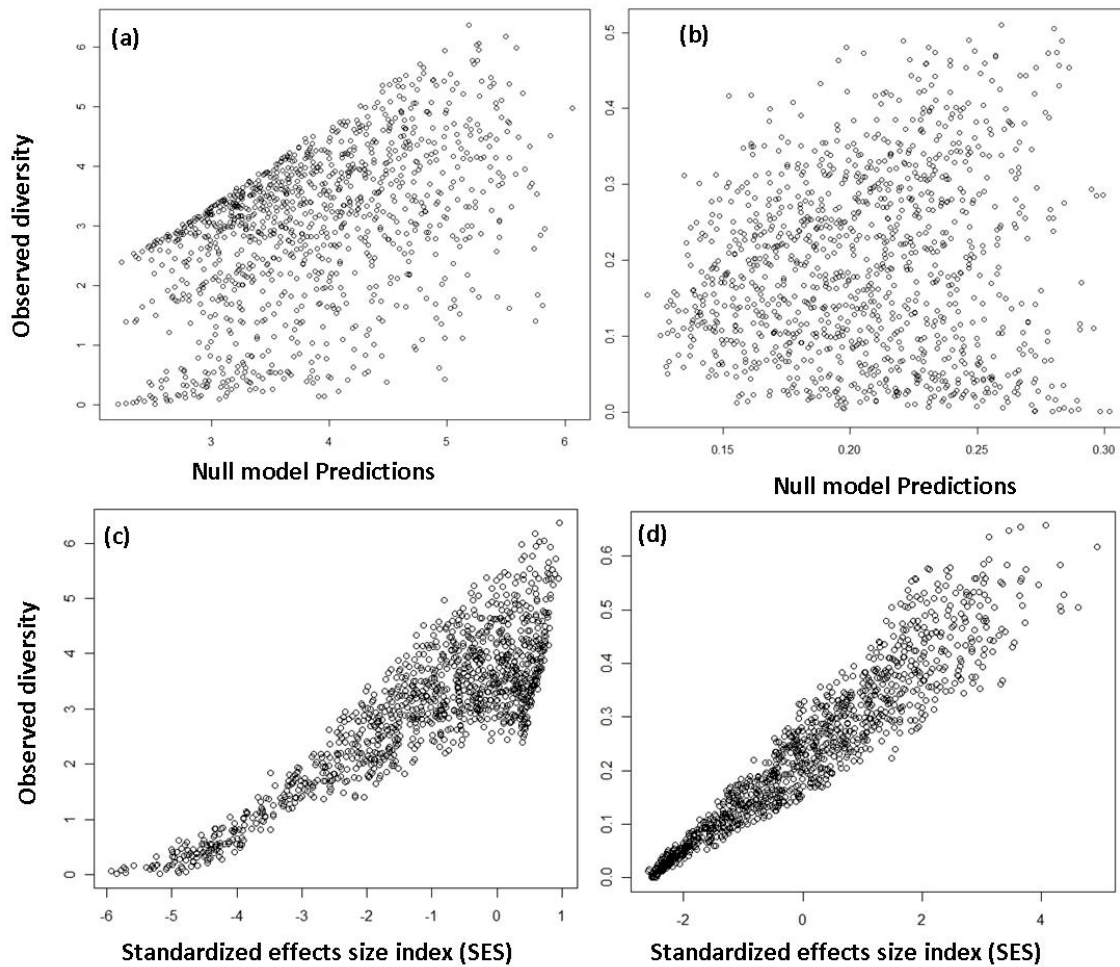
Figure 4.9: Relationship between observed β-diversity and β-diversity predicted by the null models. Taxonomic (a) and functional (b) β-diversity showed low correlation to null model predictions. Taxonomic (c) and functional (d) β-diversity plotted against SES values

Table 4.1: List of fish species and functional traits. Shape codes: A=attenuated, C=laterally compressed, E=elongate, FL=flat, FU=fusiform. Nest guarding codes: N=no guarding, Y=guarding, B=brooding. Spawning season codes: S=spring, SU=summer, F=fall, W=winter. Primary food codes: A=amphipods, AN=annelids, C=cephalopods, D=decapods, F=fish, P=polychaetes, Z=zooplankton. Habitat codes: A=anadromous, BR=brackish, I=intertidal, M=marine. Location codes: B=benthic, D=demersal, P=pelagic. References: 1) Froese & Pauley 2012, 2) NMFS 2012, 3) MSAP 2012, 4) similar species, 5) Hughes 1985.

Family Species	Code	Max. Length (mm)	Shape	Nest guard	Spawn season	Trophic level	Primary food	Habitat	Location
Clupeidae									
Clupea pallasii (Pacific herring) 1, 2	PR	42	FU	N	S	2.3	Z	BR	P
Osmeridae									
Mallotus villosus (capelin) 1,2	CP	20	FU	N	F	3.15	Z	BR	P
Thaleichthys pacificus (eulachon) 1,2	EU	34	FU	N	S	3.2	Z	BR	P
Salmonidae									
Salvelinus malma (Dolly Varden) 1	DL	127	FU	N	F	4.2	F	A	P
Oncorhynchus clarkii (cutthroat trout) 1	CT	99	FU	N	S	3.2	F	A	P
Oncorhynchus mykiss (rainbow trout) 1	SH	120	FU	N	W	4.4	C	A	P
Oncorhynchus gorbuscha (pink salmon) 1	PK	76	FU	N	F	4.2	F	A	P
Oncorhynchus kisutch (coho salmon) 1	CO	108	FU	N	F	4.2	F	A	P
Oncorhynchus keta (chum salmon) 1	CM	100	FU	N	F	3.5	F	A	P
Oncorhynchus nerka (sockeye salmon) 1	SK	84	FU	N	S	3.7	F	A	P
Gadidae									
Theragra chalcogramma (walleye pollock) 1,2	PO	105	FU	N	S	3.5	Z	M	P
Microgadus proximus (Pacific tomcod) 1,2	TC	31	FU	N	S	3.6	A	M	D
Gadus macrocephalus (Pacific cod) 1,2	PC	147	FU	N	W	4	Z	M	D
Aulorhynchidae									
Aulorhynchus flavidus (tube-snout) 1,3	TU	18	A	G	S	3.4	Z	BR	D
Gasterosteidae									
Gasterosteus aculeatus (threespine stickleack) 1,3	TE	11	FU	G	SU	3.5	AN	A	P
Sygnathidae									
Syngnathus leptorhynchus (Bay pipefish) 1	PP	33	A	B	S	3.2	A	I	B
Scorpaenidae									
Sebastes caurinus (canary rockfish) 1,2,3	CP	58	A	N	S	4.1	F	M	D

Table 4.1 (cont.)

Family Species	Code	Max. Length (mm)	Shape	Nest guard	Spawn season	Trophic level	Primary food	Habitat	Location
Hexagrammidae									
Ophiodon elongatus (lingcod) 1,2	LI	127	E	G	W	4.3	D	M	D
Hexagrammos stelleri (whitespotted greenling) 1,2	WT	32	E	G	S	3.3	Z	M	D
H.octogrammus (masked greenling) 1,2	MA	42	E	G	W	3.5	D	M	D
H. decagrammus (kelp greenling) 1,2	KL	52	E	G	F	3.6	D	M	D
H. lagocephalus (rock greenling) 1,2	RG	61	E	N	W	3.9	A	M	D
Cottidae									
Hemilepidotus spinosus (brown irish lord) 1,2,3	BL	29	S	G	W	3.5	A	BR	D
Hemilepidotus hemilepidotus (red irish lord) 1,2,3	RD	51	S	G	S	3.5	D	BR	D
Icelinus borealis (northern sculpin) 1,2,4	NT	10	E	G	S	3.6	D	M	D
Radulinus taylori (spinynose sculpin) 4	SS	8	E	N	S	3.3	A	M	D
Leptocottus armatus (Pacific staghorn sculpin) 1,2	SG	46	S	G	W	3.5	P	BR	D
Enophrys bison (buffalo sculpin) 1,2	BF	37	S	G	S	3.3	I	M	D
Myoxocephalus stelleri (frog sclupin) 1, 4	FR	60	S	N	U	3.9	F	BR	D
M. polyacanthocephalus (great sculpin) 1,2	GT	72	S	G	W	4.1	D	M	D
Hemitripteridae									
Blepsias cirrhosus (silverspotted sculpin) 1,4	SV	20	S	G	SU	3.3	A	BR	D
Nautichthys oculofasciatus (sailfin sculpin) 1,4	SA	20	S	G	W	4	D	M	D
Psychrolutidae									
Psychrolutes paradoxus (tadpole sculpin) 1,2	TDPS	7	S	G	S	3.1	D	I	B
Oligocottus maculosus (tidepool sculpin) 1,3	TP	9	S	G	S	3.4	P	I	D
Agonidae									
Pallasina barbata (tubenose poacher) 1,2, 4	TB	19	E	N	SU	3.2	Z	M	D
Podothecus accipenserinus (sturgeon poacher) 1,2	SR	28	E	N	U	3.4	A	M	B
Embiotocidae									
Cymatogaster aggregata (shiner perch) 1, 3	SI	20	C	B	SU	3	I	BR	P

Table 4.1 (cont.)

Family Species	Code	Max Length (mm)	Shape	Nest guard	Spawn season	Trophic level	Primary food	Habitat	Location
Stichaediae									
Lumpenus sagitta (snake prickleback) 1,2	SP	51	E	N	F	3.1	Z	I	D
Stichaeus punctatus (Arctic shanny) 1,2	SN	22	A	N	S	3.1	P	I	D
Anoplarchus purpurescens (high cockscomb) 1,3	HH	20	A	G	W	2.8	P	I	D
Xiphister mucosus (rock prickleback) 1,3, 4	RP	58	A	G	F	2.2	AL	M	D
X.(black prickleback) 1,3	BP	31	A	G	S	2.41	Z	I	D
Pholidae									
Pholis laeta (crescent gunnel) 1,3,5	CS	25	A	G	W	3.3	I	I	D
Pholis ornata (saddleback gunnel) 5, 4	SA	30	A	G	W	3.5	I	I	D
Trichodintidae									
Trichodon trichodon (Pacific sandfish)1,2,3	SF	31	S	N	W	3.7	F	M	B
Ammodytidae									
A.s hexapterus (Pacific sand lance) 1,2,3	SD	22	A	N	F	3.1	Z	BR	B
Pleuronectidae									
Hippoglossus stenolepis (Pacific halibut) 1,2	PH	258	FL	N	W	4.1	D	M	B
Hippoglossoides elassodon (flathead sole) 1, 2	FL	74	FL	N	S	3.6	P	BR	B
Platichthys stellatus (starry flounder) 1,2	ST	91	FL	N	S	3.3	P	BR	B
Lepidopsetta sp. (rock sole) 1,2	RC	65	FL	N	S	3.2	AN	M	B
Isopsetta isolepis (butter sole) 1,3	BY	55	FL	N	S	3.6	P	BR	B
Limanda aspera (yellowfin sole) 1,2	YN	63	FL	N	SU	3.2	A	M	B
Parophrys vetulus (english sole) 1,2,3	EL	49	FL	N	W	3.4	P	M	B

Table 4.2: Species arranged by functional group and ordered by relative abundance in the data set. The number of sites within each cluster where the species were captured and the relative abundance within each cluster are shown.

FUNCTIONAL GROUP ONE	% sites			% Rel. abundance				
Species	**In data**	**Cluster 1**	**Cluster 2**	**Cluster 3**	**In data**	**Functional Group**	**Cluster 1**	**Cluster 2**	**Cluster 3**
chum salmon (*Oncorhynchus keta*)	62.5	56.0	50.0	100.0	11.9	28.1	11.4	6.8	27.0
pink salmon (*Oncorhynchus gorbuscha*)	60.4	44.4	50.0	100.0	10.1	24.0	11.0	4.6	23.9
three-spine stickleback (*Gasterosteus aculeatus*)	47.9	16.7	55.0	90.0	6.0	14.2	2.0	4.6	16.8
coho salmon (*Oncorhynchus kisutch*)	47.9	33.0	80.0	10.0	5.1	12.0	7.4	5.4	0.1
Pacific sand lance (*Ammodytes hexapterus*)	33.3	44.4	35.0	10.0	4.0	9.4	7.0	3.5	0.1
Dolly Varden (*Salvelinus malma*)	31.3	28.0	50.0	0.0	2.8	6.5	5.4	2.1	0.0
sockey salmon (*Oncorhynchus nerka*)	22.9	6.0	35.0	30.0	1.5	3.6	0.4	1.4	3.8
Pacific cod (*Gadus macrocephalus*)	18.8	33.3	15.0	0.0	0.4	0.8	0.8	0.2	0.0
capelin (*Mallotus villosus*)	6.3	0.0	15.0	0.0	0.2	0.6	0.0	0.0	0.0
walleye pollock (*Theragra chalcogramma*)	8.3	11.1	5.0	10.0	0.1	0.3	0.2	0.1	0.2
cutthroat trout (*Oncorhynchus clarkii*)	6.3	0.0	15.0	0.0	0.1	0.3	0.0	0.2	0.0
rainbow trout (*Oncorhynchus mykiss*)	6.3	0.0	15.0	0.0	0.1	0.1	0.0	0.1	0.0
eulachon (*Thaleichthys pacificus*)	6.3	0.0	5.0	20.0	0.0	0.1	0.0	0.0	0.2
TOTAL					42.0				

FUNCTIONAL GROUP TWO	% sites			% Relative abundance				
Species	**In data**	**Cluster 1**	**Cluster 2**	**Cluster 3**	**In data**	**Functional Group**	**Cluster 1**	**Cluster 2**	**Cluster 3**
yellowfin sole (*Limanda aspera*)	31.3	27.8	45.0	10.0	5.8	27.9	5.1	8.3	0.1
starry flounder (*Platichthys stellatus*)	62.5	44.4	75.0	70.0	5.4	26.1	1.6	8.7	3.1
rock sole (*Lepidopsetta* sp.)	68.8	50.0	85.0	70.0	3.8	18.1	1.9	5.7	1.8
buffalo sculpin (*Enophrys bison*)	52.1	78.0	30.0	50.0	1.9	9.0	3.8	1.1	0.4
tubenose poacher (*Pallasina barbata*)	12.5	5.6	20.0	10.0	1.2	6.0	0.2	2.3	0.1
rock greenling (*Hexagrammos lagocephalus*)	10.4	11.1	15.0	0.0	0.7	3.6	2.2	0.1	0.0
sturgeon poacher (*Podothecus accipenserinus*)	20.8	22.2	30.0	0.0	0.5	2.2	0.7	0.5	0.0
flathead sole (*Hippoglossoides elassodon*)	12.5	11.1	10.0	20.0	0.4	1.7	0.6	0.2	0.3

Table 4.2 (cont.)

FUNCTIONAL GROUP TWO	% sites				% Relative abundance				
Species	In data	Cluster 1	Cluster 2	Cluster 3	In data	Functional Group	Cluster 1	Cluster 2	Cluster 3
Pacific halibut (*Hippoglossus stenolepis*)	4.2	5.6	5.0	0.0	0.3	1.6	1.0	0.0	0.0
Pacific tomcod (*Gadus macrocephalus*)	2.1	0.0	5.0	0.0	0.2	1.2	0.0	0.0	0.0
butter sole (*Isopsetta isolepis*)	10.4	22.2	5.0	0.0	0.2	0.8	0.5	0.0	0.0
lingcod (*Ophiodon elongatus*)	8.3	22.2	0.0	0.0	0.1	0.7	0.4	0.0	0.0
spinynose sculpin (*Radulinus taylori*)	6.3	0.0	0.0	30.0	0.1	0.5	0.0	0.0	0.6
canary rockfish (*Sebastes caurinus*)	6.3	11.1	0.0	10.0	0.1	0.4	0.2	0.0	0.1
english sole (*Parophrys vetulus*)	2.1	5.6	0.0	0.0	0.0	0.1	0.0	0.0	0.0
Pacific sandfish (*Trichodon trichodon*)	2.1	0.0	5.0	0.0	0.0	0.1	0.0	0.0	0.0
TOTAL					21.0				
FUNCTIONAL GROUP THREE	% sites				% Relative abundance				
Species	In data	Cluster 1	Cluster 2	Cluster 3	In data	Functional Group	Cluster 1	Cluster 2	Cluster 3
snake prickleback (*Lumpenus sagitta*)	45.8	38.9	75.0	0	6.5	17.6	7.2	8.4	0
Pacific herring (*Clupea pallasii*)	25.0	22.2	35.0	10.0	5.5	15.0	1.6	10.0	0.1
crescent gunnel (*Pholis laeta*)	68.8	72.0	75.0	50.0	5.0	13.4	8.3	4.4	0.8
Bay pipefish (*Syngnathus leptorhynchus*)	45.8	44.4	25.0	90.0	4.7	12.6	6.6	0.7	12.1
Pacific staghorn sculpin (*Leptocottus armatus*)	70.8	77.8	70.0	60.0	4.4	11.9	4.2	5.5	1.6
tube-snout (*Aulorhynchus flavidus*)	27.1	11.1	35.0	40.0	3.0	8.0	0.2	4.4	3.9
kelp greenling (*Hexagrammos decagrammus*)	41.7	55.6	40.0	20.0	2.1	5.6	3.6	1.8	0.2
rock prickleback (*Xiphister mucosus*)	25.0	27.8	35.0	0.0	1.7	4.7	2.4	1.9	0.0
shiner perch (*Cymatogaster aggregata*)	16.7	11.1	30.0	0.0	1.6	4.5	0.1	3.2	0.0
silverspotted sculpin (*Blepsias cirrhosus*)	25.0	27.8	25.0	10.0	0.9	2.5	0.8	1.3	0.1
great sculpin (*M. polyacanthocephalus*)	22.9	17.0	35.0	10.0	0.3	0.8	0.4	0.3	0.1
northern sculpin (*Icelinus borealis*)	14.6	11.1	15.0	20.0	0.2	0.6	0.1	0.1	0.7
tidepool sculpin (*oligocottus maculosus*)	4.2	0.0	5.0	10.0	0.2	0.5	0.0	0.0	0.9
whitespotted greenling (*Hexagrammos stelleri*)	10.4	0.0	10.0	30.0	0.1	0.3	0.0	0.1	0.4
high cockscomb (*Anoplarchus purpurescens*)	10.4	5.6	5.0	30.0	0.1	0.3	0.0	0.1	0.3

Table 4.2 (cont)

FUNCTIONAL GROUP THREE	% sites				% Relative abundance				
Species	In data	Cluster 1	Cluster 2	Cluster 3	In data	Functional Group	Cluster 1	Cluster 2	Cluster 3
frog sculpin (*Myoxocephalus stelleri*)	8.3	0.0	20.0	0.0	0.1	0.3	0.0	0.2	0.0
brown irish lord (*Hemilepidotus spinosus*)	4.2	0.0	10.0	0.0	0.1	0.2	0.0	0.2	0.0
sailfin sculpin (*Nautichthys oculofasciatus*)	4.2	5.6	5.0	0.0	0.1	0.2	0.1	0.1	0.0
masked greenling (*Hexagrammos octogrammus*)	4.2	5.6	5.0	0.0	0.1	0.2	0.1	0.1	0.0
saddleback gunnel (Pholis ornata)	4.2	5.6	0.0	10.0	0.1	0.2	0.1	0.0	0.1
tadpole sculpin (*Psychrolutes paradoxus*)	8.3	11.1	5.0	0.0	0.1	0.2	0.1	0.0	0.0
red irish lord (*Hemilepidotus hemilepidotus*)	4.2	0.0	5.0	10.0	0.0	0.1	0.0	0.1	0.1
Arctic shanny (*Stichaeus punctatus*)	4.2	0.0	10.0	0.0	0.0	0.1	0.0	0.1	0.0
black prickleback (*Xyphister atropurpureus*)	2.1	0.0	0.0	10.0	0.0	0.1	0.0	0.0	0.2
TOTAL					**37.0**				

REFERENCES

Anderson, M. J., K. E. Ellingsen, and B. H. MxArdle. 2006. Multivariate dispersion as a measure of beta diversity. Ecology Letters **9**:683-693.

Attrill, M. J. and M. Power. 2004. Partitioning of temperature resources amongst an estuarine fish assemblage. Estuarine, Coastal and Shelf Science **61**:725-738.

Azeria, E., J. Izbarzabal, J. Boucher, and C. Hébert. 2011. Towards a better understanding of beta diversity: Deconstructing composition patterns of saproxylic beetles breading in recently burnt boreal forest. Page 364 *in* I. Y. Pavlinov, editor. Research in Biodiversity - Models and Applications InTech (http://www.intechopen.com/books/research-in-biodiversity-models-and-applications).

Azevedo, M. C. C. d., F. G. Araújo, A. L. M. Pessanha, and M. d. A. Silva. 2006. Co-occurrence of demersal fishes in a tropical bay in southeastern Brazil: A null model analysis. Estuarine, Coastal and Shelf Science **66**:315-322.

Botta-Dukát, Z. 2005. Rao's quadratic entropy as a measure of functional diversity based on muliple traits. Journal of Vegetation Science **16**:533-540.

Bremner, J., S. I. Rogers, and C. L. J. Frid. 2003. Assessing functional diversity in marine benthic ecosystems: a comparison of approaches. Marine Ecology Progress Series **254**:11-25.

Cardoso, I., S. França, M. P. Pais, S. Henriques, L. Cancela da Fonseca, and H. N. Cabral. 2011. Fish assemblages of small estuaries of the Portuguese coast: A functional approach. Estuarine, Coastal and Shelf Science **93**:40-46.

de Bello, F. 2012. The quest for trait convergence and divergence in community assembly: are null-models the magic wand? Global Ecology and Biogeography **21**:312-317.

de Bello, F., S. Lavergne, C. Meynard, J. Lepš, and W. Thuiller. 2010. The partioning of diversity: showing Theseus a way out of the labyrinth. Journal of Vegetation Science **21**:992-1000.

de Bello, F., J. N. Price, T. Münkemüller, J. Liira, M. Zobel, W. Thuiller, P. Gerhold, L. Götzenberger, S. Lavergne, J. Lepš, K. Zobel, and M. Pärtel. 2012. Functional species pool framework to test for biotic effects on community assembly. Ecology **10**:2263-2273.

de Bello, F., T. Wilfried, J. Lepš, P. Choler, J.-C. Clément, P. Macek, M.-T. Sebastiá, and S. Lavorel. 2009. Partitioning of functional diversity reveals the scale and extent of trait convergence and divergence. Journal of Vegetation Science **20**:475-486.

Devictor, V., D. Mouillot, C. Meynard, F. Jiguet, W. Thuiller, and N. Mouquet. 2010. Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. Ecology Letters **13**:1030-1040.

Dimitriadis, C. and D. Koutsoubas. 2011. Functional diversity and species turnover of benthic invertebrates along a local environmental gradient induced by an aquaculture unit: the contribution of species dispersal ability and rarity. Hydrobiologia **670**:307-315.

Eaton, C. D. 2010. Resource paritioning, habitat connectivity, and resulting foraging variation among salmonids in the estuarine habitat mosaic. University of Washington, Seattle, WA.

Elliott, M., A. K. Whitfield, I. C. Potter, S. J. M. Blaber, D. P. Cyrus, F. G. Nordlie, and T. D. Harrison. 2007. The guild approach to categorizing estuarine fish assemblages: a global view. Fish and Fisheries **8**:241-268.

Fisher, J. A. D., K. T. Frank, V. E. Kostylev, N. L. Shackell, T. Horsman, and C. G. Hannah. 2011. Evaluating a habitat template model's predictions of marine fish diversity on the Scotian Shelf and Bay of Fundy, Northwest Atlantic. ICES Journal of Marine Science **68**:2096-2105.

Foley, M. M., B. S. Halpern, F. Micheli, M. H. Armsby, M. R. Caldwell, C. M. Crain, E. Prahler, N. Rohr, D. Sivas, M. W. Beck, M. H. Carr, L. B. Crowder, J. Emmett Duffy, S. D. Hacker, K. L. McLeod, S. R. Palumbi, C. H. Peterson, H. M. Regan, M. H. Ruckelshaus, P. A. Sandifer, and R. S. Steneck. 2010. Guiding ecological principles for marine spatial planning. Marine Policy **34**:955-966.

Folke, C., S. Carpenter, B. Walker, M. Scheffer, T. Elmqvist, L. Gunderson, and C. S. Holling. 2004. Regime shifts, resilience, and biodiversity in ecosystem management. Annual Review of Ecology, Evolution, and Systematics **35**:557-581.

Franca, S., M. J. Costa, and H. N. Cabral. 2011. Inter- and intra-estuarine fish assemblage variability patterns along the Portuguese coast. Estuarine, Coastal and Shelf Science **91**:262-271.

Franco, A., M. Elliott, P. Franzoi, and P. Torricelli. 2008. Life strategies of fishes in European estuaries: the functional guild approach. Marine Ecology Progress Series **354**:219-228.

Froese, R. and D. Pauley. 2012. Fishbase. World Wide Web electronic publication, http://www.fishbase.org, version 10/2012.

Guillemot, N., M. Kulbicki, P. Chabanet, and L. Vigliola. 2011. Functional redundancy patterns reveal non-random assembly rules in a species-rich marine assembly. PlosOne **6**: e26735.

Hooper, D. U., F. S. Chapin, J. J. Ewel, A. Hector, P. Inchausti, S. Lavorel, J. H. Lawton, D. M. Lodge, M. Loreau, S. Naeem, B. Schmid, H. Setälä, A. J. Symstad, J. Vandermeer, and D. A. Wardle. 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecological Monographs **75**:3-35.

Johnson, S. W., M. L. Murphy, and D. J. Csepp. 2003. Distribution, habitat and behavior of rockfishes, *Sebastes* spp., in nearshore waters of southeastern Alaska: observations from a remotely operated vehicle. Environmental Biology of Fishes **66**:259-270.

Jost, L. 2006. Entropy and diversity. Oikos **113**:363-375.

Jost, L. 2007. Partitioning diversity into independent alpha and beta components. Ecology **88**:2427-2439.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, D. D. Morlon, S. P. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: Tools for integrating phylogenies and ecology. Bioinformatics **26**:1463-1464.

Lapointe, N. W. R., L. D. Corkum, and N. E. Mandrak. 2006. A comparision of methods for sampling fish diversity in shallow offshore waters of large rivers. North American Journal of Fisheries Management **26**:503-513.

Mathieson, S., A. Cattrijsse, M. J. Costa, P. Drake, M. Elliott, J. Gardner, and J. Marchand. 2000. Fish assemblages of European tidal marshes: a comparision based on species, families and functional guilds. Marine Ecology Progress Series **204**:225-242.

Micheli, F. and B. S. Halpern. 2005. Low functional redundancy in coastal marine assemblages. Ecology Letters **8**:391-400.

Miller, K. 2013a. Chapter 2: Spatial multispecies models for predicting the occurrence of estuarine-associated fish and invertebrates. University of Alaska Fairbanks.

Miller, K. 2013b. Chapter 3: Multivariate random forest models of estuarine-associated fish and invertebrate communities. Dissertation. University of Alaska Fairbanks.

Mouchet, M. A., S. Villéger, N. W. H. Mason, and D. Mouillot. 2010. Functonal diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. Functional Ecology **24**:867-876.

Murphy, M. L., S. W. Johnson, and D. J. Csepp. 2000. A comparison of fish assemblages in eelgrass and adjacent subtidal habitats near Craig, Alaska. Alaska Fishery Bulletin **7**:11-21.

Naeem, S. 1998. Species redundancy and ecosystem reliability. Conservation Biology **12**:39-45.

Neal, E. G., M. T. Walter, and C. Coffeen. 2002. Linking the Pacific decadal oscillation to seasonal stream discharge patterns in Southeast Alaska. Journal of Hydrology **263**:188-197.

Nicolas, D., J. Lobry, O. Le Pape, and P. Boët. 2010. Functional diversity in European estuaires: relating the composition of fish assemblages to the abiotic environment. Estuarine, Coastal and Shelf Science **88**:329-338.

NMFS. 2012. Life History Database, Resource Ecology and Fisheries Management Division, http://access.afsc.noaa.gov/reem/LHWeb/Index.cfm.*in* h. a. a. n. g. r. L. I. c. Resource Ecology and Fisheries Management Division, editor.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2011. vegan: Community ecology package. R package version 1.0-2 http://CRAN.R-project.org/package=vegan.

Pavoine, S., J. Vallet, A.-B. Dufour, S. Gachet, and H. Daniel. 2009. On the challenge of treating various types of variables: application for improving the measure of functional diversity. Oikos **118**:391-402.

Penttila, D. 2007. Marine forage fishes in Puget Sound. Washington Department of Fish and Wildlife, www.pugetsoundnearshore.org/technical_papers/marine_fish.pdf.

Podani, J. 2009. Convex hulls, habitat filtering and functional diversity: mathematical elegance versus ecological interpretability. Community Ecology **10**:244-250.

Pratchett, M. S., A. S. Hoey, S. K. Wilson, V. Messmer, and N. A. J. Graham. 2011. Changes in biodiversity and functioning of reef fish assemblages following coral bleaching and coral loss. Diversity **3**:424-452.

Rao, C. R. 1982. Diversity and dissimilarity coefficients: a unified approach. Theoretical Population Biology **21**:24-43.

Ricotta, C. and M. Marignani. 2007. Computing β-diversity with Rao's quadratic entropy: a change of perspective. Diversity and Distributions **13**:237-241.

Romesburg, H. C. 2004. Cluster Analysis for Researchers. Lulu Press, North Carolina.

Rooper, C. N., D. R. Gunderson, and D. A. Armstrong. 2006. Evidence for resource partitioning and competition in nursery estuaries by juvenile flatfish in Oregon and Washington. Fishery Bulletin **104**:616-622.

Ruetz III, C. R., D. G. Uzarski, D. M. Krueger, and E. S. Rutherford. 2007. Sampling a littoral fish assemblage: comparison of small-mesh fyke netting and boat electrofishing. Peer-reviewed scientific publications, Paper 1, http://scholarworks.gvsu.edu/peerscipub/1.

Segal, M. and Y. Xiao. 2011. Multivariate Random Forests. Data Mining and Knowledge Discovery **1**:80-87.

Simberloff, D. and T. Dayan. 1991. The guild concept and structure of ecological communities. Annual Reivew of Ecology and Systematics **22**:115-143.

Spies, I., O. A. Ormseth, and T. T. TenBrink. 2012. Chapter 19: Bering Sea and Aleutian Island sculpins. North Pacific Regional Fishery Management Council, Anchorage, AK.

Tilman, D., J. Knops, D. Wedin, P. Reich, M. Ritchie, and E. Siemann. 1997. The influence of functional diversity and composition on ecosystem processes. Science **277**:1300-1302.

Villéger, S., J. R. Miranda, D. F. Hernandez, and D. Mouillot. 2012. Low functional β-diversity despite high taxonomic β-diversity among tropical estuarine fish communites. PlosOne **7**:e40679. .

Weingartner, T., L. Eisner, G. L. Eckert, and S. Danielson. 2009. Southeast Alaska: oceanographic habitats and linkages. Journal of Biogeography **36**:387-400.

Wilson, J. B. 1987. Methods for detecting non-randomness in species co-occurrences: a contribution. Oecologia **73**:579-582.

Wright, J. P., S. Naeem, A. Hector, C. Lehman, P. B. Reich, B. Schmid, and D. Tilman. 2006. Conventional functional classification schemes underestimate the relationship with ecosystem functioning. Ecology Letters **9**:111-120.

Chapter 5: Summary, Conclusions and Recommendations

SUMMARY

- Accurate predictive models of species occurrence can be developed for estuarine-associated fish and invertebrates using regional-scale spatial datasets as surrogates for environmental data collected in the field. Models confirmed species-habitat relationships documented in previous research and highlighted new relationships that could further refine understanding of factors affecting the spatial distribution of species.

- The random forest algorithm produced highly accurate species occurrence predictions when validated against independent data. Sixty-six percent of the fish and 72% of the invertebrate models had predictive accuracies greater that 70%.

- Individual species models plotted in GIS can provide critical information for managers and researchers on spatial distributions and predicted species occurrence in relation to anthropogenic or other habitat changes. Because many fish species occur in estuaries as juveniles, these models can help identify nearshore areas important to fish early life stages.

- Species composition within estuaries showed strong associations with regional patterns of tidal height and precipitation and with local availability of intertidal habitat, kelp, and eelgrass. Multispecies models can provide insight into species co-occurrence patterns and the effect of co-occurring species on abundances of individual species.

- Methods, such as multivariate random forests, that can model multivariate responses against multivariate predictors retain information on individual species responses to both biotic and abiotic factors, and therefore provide information necessary for ecosystem-level management

- Both environmental niche filtering and biotic interactions are important structuring mechanisms for estuarine communities. Functional diversity in the estuaries was not correlated to the number of species present, but showed a slight positive relationship to change in species composition. Rare species that occurred in fewer than 10% of the sampled estuaries were primarily responsible for the high taxonomic diversity in the region, and the relatively high functional redundancy among these species contributed to the relatively low functional diversity in the region

- Taxonomic diversity was substantially lower than expected by chance, a result consistent with the large number of rare species in the data and environmental filtering limiting dispersal of these species across estuaries. In contrast, functional diversity was higher than

expected by chance. This result is also consistent with environmental filtering within estuaries selecting for species adapted to the environmental conditions of that estuary

CONCLUSIONS AND RECOMMENDATIONS

The spatial distribution of estuarine species can be quantifiably related to differences in the quality and composition of habitat at the landscape scale. This finding is of substantial value for improving species management, understanding ecosystem processes, and identifying areas for conservation. For areas like Southeast Alaska where comprehensive data collection is hampered by the area's large size and challenging climate, model-based estimates of species occurrence and community composition can provide information that can guide research and inform management decisions.

A potential application of this research is to provide a stronger scientific and analytical basis for evaluating how environmental changes may affect species occurrence and distribution. The National Marine Fisheries Service (NMFS) has been tasked under the Magnuson-Stevens Fisheries Management Act with evaluating how fishing and non-fishing activities may adversely affect essential fish habitat. For nearshore development in Alaska, this determination is made with very little scientific information on species distributions or habitat use because for much of coastal Alaska the relative abundance of species within different habitats is virtually unknown. This research has identified species-habitat relationships for a number of species of concern to NMFS. The models constructed in Chapter 1 provide a scientific and quantifiable basis for evaluating scenarios of habitat loss that may occur from nearshore development and determining how this loss may affect species occurrence and distribution. These models can also assist in identifying areas of importance to species of commercial interest which can be used to prioritize areas for conservation.

As fishery management methods change to focus more on ecosystem processes than individual species, modeling will be a critical part of the approach used to understand the composition of estuarine communities in relation to the biotic and abiotic environment. Models that evaluate species-habitat relationships at the community level have the advantage of being able to evaluate species interactions as they affect community composition in a way that individual models cannot. The use of co-occurring species to understand patterns in target species is not new, but it is only recently that accurate multivariate methods have been developed that can incorporate not only several species as the response, but also a large number of predictor

variables. This provides the opportunity to explore new relationships between individual species and between species and their environment that are not possible with standard statistical methods. As shown in Chapter 2, community composition in Southeast Alaska estuaries is strongly influenced by environmental gradients in precipitation, tidal energy, and intertidal habitat availability. These models can be used to justify additional funding to further define the environmental gradients along which species are distributed.

Ecosystem processes are affected by the functional characteristics of species, which is only weakly related to a species' taxonomic identity. Yet most ecosystem management approaches still focus on the presence or abundance of species without evaluating either the functional role those species play in the ecosystem or the functional similarity between areas with dissimilar species. These approaches assume functional diversity will be maintained if species diversity is maintained (Foley et al. 2010). In Chapter 3, I demonstrated that functional diversity of fishes in Southeast Alaska estuaries was not correlated to species richness and only moderately related to taxonomic diversity. Most of the estuaries have a relatively high level of functional similarity despite having high taxonomic diversity as a result of having a large pool of rare species that occur in only a few estuaries. However, a portion of the estuaries are dominated by a small number of anadromous species and have little functional redundancy. These estuaries are most vulnerable to changes in ecosystem processes resulting from anthropogenic or natural changes in species composition. The results of this research have direct application to decisions on habitat management and conservation.

Although this research makes major strides in understanding estuarine ecosystem processes and developing methods for evaluating large complex data, it is but a start. Habitat, in the many ways it is defined, is well recognized within the fisheries and marine research communities as a critical factor that affects species abundance and distribution (Foley et al. 2010, Yoklavich et al. 2010), yet it is also peripheral to most fisheries management decisions and management-related research (Armstrong & Falk-Petersen 2008). As an example, consider that the standard stock-recruit model used for assessing fish abundance for setting allowable fishing levels has no habitat variable: the influence of habitat is assumed through spawner and recruit survival and the carrying capacity. There have been several approaches to incorporating habitat more explicitly into these models (Scheuerell et al. 2006, Stier & Osenberg 2010, Mizerek et al. 2011, Su et al. 2012), but methods incorporating habitat are still not widely used. Ecosystem fisheries management demands an understanding of the ecological linkages between species and

their habitats (Link 2000). The models developed in this research could be used to develop a stronger understanding of the relationship among species, habitats and co-occurring species that could inform fisheries stock assessments and ecosystem management activities. Especially for juveniles, models that can estimate species abundance in relation to biotic and abiotic factors could be incorporated into stock assessments (Yoklavich et al. 2010). This research is one example of how these models can be developed. A future challenge will be to bring these modeling methods to bear on the extensive fisheries and habitat data available and assess how they enhance fisheries stock assessment and ecosystem management.

REFERENCES

Armstrong CW, Falk-Petersen J (2008) Habitat-fisheries interactions: a missing link? ICES Journal of Marine Science 65:817-821

Foley MM, Halpern BS, Micheli F, Armsby MH, Caldwell MR, Crain CM, Prahler E, Rohr N, Sivas D, Beck MW, Carr MH, Crowder LB, Emmett Duffy J, Hacker SD, McLeod KL, Palumbi SR, Peterson CH, Regan HM, Ruckelshaus MH, Sandifer PA, Steneck RS (2010) Guiding ecological principles for marine spatial planning. Marine Policy 34:955-966

Link JS (2000) Fisheries management in an ecosystem context: what does this mean, what do we want, and can we do it? In: Mace PM (ed). Proc Incorporating Ecosystem Considerations into Stock Assessments and Management Advice. National Marine Fisheries Service

Mizerek T, Regan HM, Hovel KA (2011) Seagrass habitat loss and fragmentation influence management strategies for a blue crab Callinectes sapidus fishery. Marine Ecology Progress Series 427:247-257

Scheuerell MD, Hilborn R, Ruckelshaus MH, Bartz KK, Lagueux KM, Haas AD, Rawson K (2006) The Shiraz model: a tool for incorporating anthropogenic effects and fish–habitat relationships in conservation planning. Canadian Journal of Fisheries and Aquatic Sciences 63:1596-1607

Stier AC, Osenberg CW (2010) Propagule redirection: habitat availability reduces colonization and increases recruitment in reef fishes. Ecology 91:2826-2832

Su N-J, Sun C-L, Punt AE, Yeh S-Z, DiNardo G (2012) Incorporating habitat preference into the stock assessment and management of blue marlin (*Makaira nigricans*) in the Pacific Ocean. Marine and Freshwater Research 63:565-575

Yoklavich MM, Blackhard K, Brown SK, Greene C, Minello T, Noji T, Parke M, Parrish F, Smith K, Stone RP, Wakefield WW (2010) Marine fisheries habitat assessment improvement plan. In: U. S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service (ed), Silver Spring, MD