Spatial Dynamics Modeling for Data-poor Species Using Examples of Longline Seabird Bycatch And Endangered White Abalone

Yan Li

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Fish and Wildlife Conservation

Yan Jiao (Chair)
Feng Guo
Carola Haas
Jess Jones
Laura Rogers-Bennett

May 5, 2014
Blacksburg, Virginia

Keywords: spatial nonstationarity, spatial filter, risk assessment, data-poor species, zero-inflated data, demographic model, delta model

Spatial Dynamics Modeling for Data-poor Species Using Examples of Longline Seabird Bycatch And Endangered White Abalone

Yan Li

ABSTRACT

Spatial analysis of species for which there is limited quantity of data, termed as the data-poor species, has been challenging due to limited information, especially lack of spatially explicit information. However, these species are frequently of high ecological, conservation and management interest. In this study, I used two empirical examples to demonstrate spatial analysis for two kinds of data-poor species. One example was seabird bycatch from the U.S. Atlantic pelagic longline fishery, which focused on rare events/species for which data are generally characterized by a high percentage of zero observations. The other example was endangered white abalone off the California coast, which focused on endangered species whose data are very limited. With the seabird bycatch example, I adopted a spatial filtering technique to incorporate spatial patterns and to improve model performance. The model modified with spatial filters showed superior performance over other candidate models. I also applied the geographically weighted approach to explore spatial nonstationarity in seabird bycatch, i.e., spatial variation in the parameters that describe relationships between biological processes and environmental factors. Estimates of parameters exhibited high spatial variation. With the white abalone example, I demonstrated the spatially explicit hierarchical demographic model and conducted a risk assessment to evaluate the efficacy of hypothetical restoration strategies. The model allowed for the Allee effect (i.e., density-dependent fertilization success) by using spatial explicit density estimates. Restoration efforts directed at larger-size individuals may be more effective in increasing population density than efforts focusing on juveniles. I also explored the spatial

nonstationarity in white abalone catch data. I estimated the spatially explicit decline rate and linked the decline rate to environmental factors including water depth, distance to California coast, distance to land, sea surface temperature and chlorophyll concentration. The decline rate showed spatial variation. I did not detect any significant associations between decline rate and these five environmental factors. Through such a study, I am hoping to provide insights on applying or adapting existing methods to model spatial dynamics of data-poor species, and on utilizing information from such analyses to aid in their conservation and management.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my Ph.D. advisor Dr. Jiao for offering me such a great opportunity to be her student and to work with her. I also took a M.S. degree with her. Over the two years of M.S. study and four years of Ph.D. study, Dr. Jiao has guided me on conducting quantitative analysis and developing statistical and simulation models that would benefit the field of ecology and fisheries. Most importantly, Dr. Jiao has guided me on finding my research interest and building my career goal. Before working with Dr. Jiao, my research in China focused on fresh water ecology that was heavily field-experiment oriented, and I did not have a strong background in quantitative analysis. Once I joined Dr. Jiao's lab and started working with her, I was amazed by how the world that I observe can be described quantitatively, and how part of the world that I will never have a chance to observe can be inferred with available data and knowledge. Then I had my career goal clear, to be a scientist in the interdisciplinary field of population dynamics, ecological modeling, fisheries management and conservation. Dr. Jiao offered me as many opportunities as she could to train me and strengthen my skills that I will need to pursue my career goal, e.g., attending conferences and presenting research projects, writing proposals, publishing manuscripts, supporting me to obtain the M.S. degree in Statistics at Virginia Tech, collaborating with senior scientists. I feel so lucky to have had the opportunity to work with her.

I am highly indebted to my advisory committee for their great support and valuable advice over the past four years. Also as my M.S. advisor in Statistics at Virginia Tech, Dr. Guo offered me valuable advice on the modeling techniques that I used throughout the program, and had guided me on understanding theories of statistical analysis and applying statistics in solving real problems such as the problems that I encountered in my Ph.D. program. Dr. Haas introduced

me to many ecological and biological concepts related to my research, and offered me lots of great advice on my graduate program, job application and building career path. I was amazed by the great detail that Dr. Jones revised my dissertation chapters, from paragraphs to sentences, and to small typos and even a symbol in the text. Dr. Jones also brought me great perspectives regarding my chapters, especially the chapter for white abalone risk assessment. Dr. Rogers-Bennett offered me advice and support in lots of ways, e.g., providing data information (for white abalone), providing advice on developing demographic models (for white abalone), reviewing manuscripts, suggesting target journals for manuscript submission.

I extend my acknowledgement to all members in my lab, Dr. Hao Yu (graduated in 2010), Dr. Robert Leaf (graduated in 2010), Qing He (graduated in 2010), Joshua Hatch (graduated in 2011), Matt Vincent (graduated in 2013), Dan Hua, Man Tang and Yafei Zhang. Their help and support are extremely important and precious to me.

Lastly, I give my deep gratitude to my family. My parents and my parents-in-law gave me great support to pursue my career in scientific research. My husband Laurence Lin has been a great husband and the greatest daddy ever. My daughter Mavis just joined our family six months ago, and she is the sweetest gift in my life. Laurence offered Mavis and me all his deep-heart, care and love, and plus he always helped me out with numerical and computing problems that I encountered in my project.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Data-poor species

Analysis of species for which there is a limited quantity of data, termed as the data-poor species, has been recognized to be very challenging due to the high uncertainty involved in the analysis. Spatial analysis that requires additional information on spatial distribution makes the analysis for data-poor species even more complicated. However, these data-poor species are oftentimes of high interest in ecology, conservation and management, e.g., threatened or endangered species and bycatch species in fisheries. Methodologies that can fully utilize whatever information is available and can improve reliability of the analysis for data-poor species are thus urgently needed. In this study, I focused on two kinds of data-poor species. The first kind was rare events/species for which data are generally characterized by a high percentage of zero observations (termed as the zero-inflated data) and a large dataset if data are collected through a long-term survey. The second kind of data-poor species in this study was endangered species whose data are very limited, especially the biological and ecological information, and whose current abundance in the wild is too low to support further data collection. In this study, I demonstrate the methodologies using two examples, i.e., seabird bycatch from the U.S. Atlantic pelagic longline fishery and endangered white abalone (*Haliotis sorenseni*) off California coast. I used the seabird bycatch example to demonstrate the methodology focusing on spatial analysis of rare events/species with highly zero-inflated data. The white abalone example was employed to illustrate the methodology focusing on spatial analysis on endangered species with limited data. Both examples were of high conservation and management interest, and previous studies have

suggested that spatial patterns may exist in both examples (Hobday and Tegner, 2000; Karpov et al., 2000; Butler et al., 2006; Li and Jiao, 2011; Li and Jiao, 2012).

Incidental mortality, known as seabird bycatch (Brothers et al., 1999; Tasker et al., 2000; Belda and Sanchez, 2001; Furness, 2003), has been believed to partially contribute to population declines of several albatross species (in the family Diomedeidae) and other species in the procellariiform order (Croxall et al., 1982; Brothers et al., 1999; Tasker et al., 2000; Belda and Sanchez, 2001; Furness, 2003). Adult seabirds get hooked or entangled when attempting to eat the bait on hooks and are drowned as hooks sink, which may also cause mortality of chicks if one or both parents are killed (Brothers et al., 1999; Gilman, 2001). Biological characteristics of seabirds such as late maturity, low reproductive rate and maternal care for chicks, make their populations vulnerable to the incidental mortality from all fisheries. The U.S Atlantic pelagic longline fishery primarily targets tunas (*Thunnus*) and swordfish (*Xiphias gladius*), and secondarily targets dolphinfish (*Coryphaena hippurus*) and sharks (Selachimorpha). This fishery has been monitored by the National Marine Fisheries Service Southeast Fisheries Science Center Pelagic Observer Program (POP) since 1992. In the POP, randomly selected fishing trips are attended by a trained observer, and detailed information on gear specification, fishing effort, fishing tactics, target catch, bycatch and environmental conditions is recorded. The seabird bycatch data that I used in this study were provided by the POP program. The POP data contained greater than 99% zero observations and served as a typical example of rare events/rare species with highly zero-inflated data.

Five abalone species, including white abalone, historically supported a valuable commercial and recreational fishery in California (Haaker, 1994; Hobday and Tegner, 2000). The white abalone fishery began in 1968, peaked in 1972 and collapsed by 1978, driven by

intensive exploitation and potentially by other factors such as environmental changes, diseases, hybridization, accidental mortality, and predation and competition (Lundy, 1997; Hobday and Tegner, 2000). A decline of greater than 99% was estimated, and recruitment has not been observed in the wild for decades. Remnant populations may suffer the Allee effect (Allee, 1931) where fertilization fails due to low density of adult abalones (Cox, 1960; Hobday and Tegner, 2000; Stierhoff et al., 2012). By 1997, the State of California closed all commercial and recreational abalone fisheries throughout the state except the recreational fishery for red abalone (*Haliotis rufescens*) located north of San Francisco (Altstatt et al., 1996; Hobday and Tegner, 2000). In 2001, white abalone was listed as an endangered species under the Endangered Species Act (Haaker, 1998; Hobday and Tegner, 2000). As the deepest-living and the latest harvested species among all abalone species along the west coast, the white abalone has been poorly studied (Hobday and Tegner, 2000). Limited data on its biology and ecology make restoration of the species even more challenging.

1.2 Spatial modeling for data-poor species

Spatial models and non-spatial models differ substantially in the assumptions about relationships among observations. Spatial models assume observations to spatially depend on each other, whereas non-spatial models assume that observations are independent of each other (Cressie, 1993). Spatial analysis has been applied to improve stock assessments for spatially aggregated species, especially for invertebrate species that have limited movement (Conan, 1985; Gonzalez-Gurriaran et al., 1993). With observations being spatially dependent of each other, the commonly used assumption of independence will not be appropriate any more. Spatial analysis also helps detect spatial patterns of a species which can further benefit its management

and restoration (Vignaux, 1996; Walter et al., 2007). For example, hotspots of bycatch events identified through a spatial analysis will provide implications in the development of bycatch reduction strategies. Spatial variation in key parameters (e.g., the parameter describing relationships between biological processes and environmental factors, and the biological parameter such as growth rate, survival and reproduction), also known as the spatial nonstationarity, will help identify driving factors and develop spatially explicit management plans.

Spatial analysis of data-poor species is confronted with many challenges. First, excess zeros in the data of rare events/species invalidate the normality assumption that we commonly use in ecological data analyses and may cause computational problems (Maunder and Punt, 2004; Cunningham and Lindenmayer, 2005). For example, the longline seabird bycatch data that I used in this study contained greater than 99% zero observations. The simplest ways to deal with zeros is to ignore them, or to aggregate multiple records (Maunder and Punt, 2004; Winter et al., 2011). However, either ignorance of a considerable proportion of zeros or aggregation of multiple records will result in a loss of information that may reflect the spatial or temporal distribution characteristics of the species. Second, a large dataset makes computation even more difficult if the traditional methods for spatial analysis are employed which assume a variance-covariance matrix to describe spatial dependence between observations (Cressie, 1993). For example, there were a total of 12,894 longline sets analyzed for the seabird bycatch data in this study, and the traditional method may not be applicable for such a large dataset because inverting such a large variance-covariance matrix when calculating its normal likelihood requires special numerical techniques. Third, lack of ecological and biological information for endangered species already makes their analysis difficult and highly uncertain. In terms of spatial analysis

that requires spatially explicit ecological and biological information, the situation is even more difficult. However, analysis with whatever data are available will be important to the management and restoration of these endangered species.

There are many methodologies available to investigate spatial patterns, to deal with zero-inflated data or to model species with limited information. For example, to model spatial patterns, the semivariogram model, the spatial filtering technique (e.g., Griffith and Peres-Neto, 2006), the geographically weighted approach (e.g., Brunsdon et al., 1996), the simultaneously autoregressive model and the conditional autoregressive model (Cressie, 1993) have long been used. To deal with zero-inflated data, the generalized linear and additive models with an additive constant (e.g., Ortiz et al., 2000), the delta model (e.g., Lo et al., 1992; Stefansson, 1996), and the zero-inflated model (e.g., Minami et al., 2007) are commonly applied in ecological studies; the Tweedie distribution model (e.g., Shono, 2008), and the delta model combined with the AdaBoost model (e.g., Li et al., 2011) are recently applied. To incorporate high uncertainty due to data limitation, hierarchical models (e.g., Jiao et al., 2011), Bayesian models (e.g., Jiao et al., 2011) and demographic models (e.g., Rogers-Bennett and Leaf, 2006) are oftentimes options to consider. I would like to emphasize that with such a large pool of methodologies being available, I only explored or adopted four of them in this study, namely the spatial filtering technique, the geographically weighted approach, the hierarchical demographic model and the spatial nonstationary analysis. I decided to focus on these four methodologies for three reasons. First, they are relatively new and have a lot of potential applications in ecological studies that are worthy a further exploration. Second, they can potentially be modified or adopted into existing modeling approaches and thus become appropriate for data-poor species. For example, I modified and extended the spatial filtering technique and the geographically weighted approach

so that they can be appropriately applied for zero-inflated data. Third, they are flexible tools to explore spatial patterns. By adjusting the assumptions on patterns and strength of spatial dependence between observations, spatial patterns that appropriately describe the data and fit the research interest can be explored.

1.3 Implications in conservation and management

In the traditional statistical analysis, we usually assume observations to be independent. However, with spatial dependence existing in the data, this assumption becomes violated. By incorporating spatial patterns in the model, bias caused by violation of this assumption may be reduced, and estimation and prediction (e.g., number of seabird bycatch) from the model may be improved. Additionally, such analysis may help identify spatial distribution of species and hotspots on which management and conservation effort could focus. For example, hotspots of seabird bycatch can be identified through such spatial analysis, and the bycatch reduction measures focusing on these hotspots may be more efficient; if population dynamics of white abalone show spatial variation, restoration may need to consider both global and local scale.

1.4 Objectives

Given the challenges and importance of spatial analysis for data-poor species, the overall goal of this study was to develop alternative modeling approaches to explore spatial dynamics of data-poor species and demonstrate these approaches using empirical examples of seabird bycatch and white abalone. Specifically, in Chapters 2 and 3, I focused on rare events/species with highly zero-inflated data using the example of seabird bycatch. In Chapter 2, I aimed to adopt a relatively new technique, the eigenfunction-based spatial filtering, into an existing modeling

approach, the delta model, to incorporate spatial dependence in analysis of a large dataset. In Chapter 3, I aimed to extend a relatively new approach, the geographically weighted regression, by incorporating statistical distributions other than the normal distribution and to develop the geographically weighted generalized linear model to explore spatial nonstationarity in seabird bycatch data. In Chapters 4 and 5, I focused on endangered species with limited data using the example of white abalone. The purposes of Chapter 4 were to develop the spatially explicit demographic model and to conduct risk assessment for white abalone where the efficacy of hypothetical restoration strategies were evaluated. The purposes of Chapter 5 were to explore the spatial nonstationarity in white abalone commercial catch data and to link decline of catch and environmental factors. Through such a study, I am hoping to provide insights on applying or adapting existing methods for modeling spatial dynamics of data-poor species, and on utilizing information from such analyses to aid in conservation and management of these species.

1.5 References

Allee, W.C., 1931. Animal aggregations: a study in general sociology. University of Chicago Press, Chicago.

Altstatt, J., Ambrose, R., Engel, J., Haaker, P., Lafferty, K., Raimondi, P., 1996. Recent declines of black abalone *Haliotis cracherodii* on the mainland coast of central California. MEPS 142, 185-192.

Belda, E., Sanchez, A., 2001. Seabird mortality on longline fisheries in the western Mediterranean: factors affecting bycatch and proposed mitigating measures. Biol Conserv 98, 357-363.

Brothers, N., Cooper, J., Løkkeborg, S., 1999. The incidental catch of seabirds by longline fisheries: worldwide review and technical guidelines for mitigation. FAO Fisheries Circular No. 937, Rome.

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical analysis 28, 281-298.

Butler, J., Neuman, M., Pinkard, D., Kvitek, R., Cochrane, G., 2006. The use of multibeam sonar mapping techniques to refine population estimates of the endangered white abalone (*Haliotis sorenseni*). Fish Bull 104, 521-532.

Conan, G.Y., 1985. Assessment of shellfish stocks by geostatistical techniques. International Council for the Exploration of the Sea Committee Meeting paper (Shellfish Committee) K 30, 1-24.

Cox, K.W., 1960. Review of the abalone of California. California Fish and Game Bulletin 46, 381-406.

Cressie, N., 1993. Statistics for spatial data. Wiley, New York.

Croxall, J., Evans, P., Schreiber, R., 1982. Status and conservation of the world's seabirds. ICBP Technical Publication 2, 1-4.

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

Furness, R., 2003. Impacts of fisheries on seabird communities. Scientia Marina 67, 33-45.

Gilman, E., 2001. Integrated management to address the incidental mortality of seabirds in longline fisheries. Aquat Conserv: Mar Freshwat Ecosyst 11, 391-414.

Gonzalez-Gurriaran, E., Freire, J., Fernandez, L., 1993. Geostatistical analysis of spatial distribution of *Liocarcinus depurator*, *Macropipus tuberculatus* and *Polybius henslowii*

(Crustacea: Brachyura) over the Galician continental shelf (NW Spain). Mar Biol 115, 453-461.

Griffith, D.A., Peres-Neto, P.R., 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. Ecology 87, 2603-2613.

Haaker, P.L., 1994. Assessment of abalone resources at the Channel Islands. in: Halvorson, W.L., Maender, G.J., eds. The Fourth California Islands Symposium: Update on the Status of Resources. Santa Barbara, CA: Santa Barbara Museum of Natural History;

Haaker, P.L., 1998. White abalone-off the deep end forever? Outdoor California January-February, 17-20.

Hobday, A.J., Tegner, M.J., 2000. Status review of white abalone (*Haliotis sorenseni*) throughout its range in California and Mexico. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southwest Region Office, Long Beach, California.

Jiao, Y., Cortes, E., Andrews, K., Guo, F., 2011. Poor-data and data-poor species stock assessment using a Bayesian hierarchical approach. Ecol Appl 21, 2691-2708.

Karpov, K.A., Haaker, P.L., Taniguchi, I.K., Rogers-Bennett, L., 2000. Serial depletion and the collapse of the California abalone (*Haliotis spp.*) fishery. In: Campbell, A. (Ed), Workshop on Rebuilding Abalone Stocks in British Columbia. NRC Research Press, Ottawa, Ontario, Canada.

Li, Y., Jiao, Y., 2011. Estimated seabird bycatch in the U.S. Atlantic pelagic longline fishery during 1992-2010 based on observer and logbook data. Report to NOAA Fisheries Service, Southeast Fisheries Science Center, Miami, Florida. Virginia Polytechnic

Institute and State University, Department of Fish and Wildlife Conservation, Blacksburg, Virginia.

Li, Y., Jiao, Y., 2012. Estimated seabird bycatch in the U.S. Atlantic pelagic longline fishery during 1992-2011 based on observer and logbook data. Report to NOAA Fisheries Service, Southeast Fisheries Science Center, Miami, Florida. Virginia Polytechnic Institute and State University, Department of Fish and Wildlife Conservation, Blacksburg, Virginia.

Li, Y., Jiao, Y., He, Q., 2011. Decreasing uncertainty in catch rate analyses using Delta-AdaBoost: An alternative approach in catch and bycatch analyses with high percentage of zeros. Fish Res 107, 261-271.

Lo, N., Jacobson, L., Squire, J., 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can J Fish Aquat Sci 49, 2515-2526.

Lundy, A.L., 1997. The California abalone industry, a pictorial history. Best Publishing Company, Flagstaff, Arizona.

Maunder, M., Punt, A., 2004. Standardizing catch and effort data: a review of recent approaches. Fish Res 70, 141-159.

Minami, M., Lennert-Cody, C., Gao, W., Rom·n-Verdesoto, M., 2007. Modeling shark bycatch: the zero-inflated negative binomial regression model with smoothing. Fish Res 84, 210-221.

Ortiz, M., Legault, C., Ehrhardt, N., 2000. An alternative method for estimating bycatch from the US shrimp trawl fishery in the Gulf of Mexico, 1972-1995. Fish Bull 98, 583-599.

Rogers-Bennett, L., Leaf, R.T., 2006. Elasticity analyses of size-based red and white abalone matrix models: management and conservation. Ecol Appl 16, 213-224.

Shono, H., 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fish Res 93, 154-162.

Stefansson, G., 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES J Mar Sci 53, 577-581.

Stierhoff, K.L., Neuman, M., Butler, J.L., 2012. On the road to extinction? Population declines of the endangered white abalone, *Haliotis sorenseni*. Biol Conserv 152, 46-52.

Tasker, M., Camphuysen, C., Cooper, J., Garthe, S., Montevecchi, W., Blaber, S., 2000. The impacts of fishing on marine birds. ICES Journal of Marine Science: Journal du Conseil 57, 531.

Vignaux, M., 1996. Analysis of spatial structure in fish distribution using commercial catch and effort data from the New Zealand hoki fishery. Can J Fish Aquat Sci 53, 963-973.

Walter, J.F., Christman, M.C., Hoenig, J.M., Mann, R., 2007. Combining data from multiple years or areas to improve variogram estimation. Environmetrics 18, 583-598.

Winter, A., Jiao, Y., Browder, J.A., 2011. Modeling low rates of seabird bycatch in the U.S. longline fishery. Waterbirds 34, 289-303.

Chapter 2

Modeling Spatial Patterns of Rare Species Using Eigenfunction-based Spatial Filters: An Example of Modified Delta Model for Zero-inflated Longline Seabird Bycatch Data

2.1 Abstract

Data of rare species usually contain a high percentage of zero observations due to their low abundance. Such data are generally referred to as zero-inflated data. Modeling spatial patterns in such data has been challenging, especially when large datasets are involved and intensive computing are required. The eigenfunction-based spatial filtering provides a flexible tool that allows the existing modeling approaches that can handle zero-inflated data, such as the delta model, to be applied in the presence of spatial dependence. With a real dataset, the longline seabird bycatch data, the present study demonstrated a modification of the delta model with spatial filters to investigate spatial patterns in zero-inflated data for rare species. I explored a total of 108 spatial weighting matrices, and modified the delta model by incorporating the spatial filters generated from the best spatial weighting matrix. I applied a five-fold cross-validation to compare performance of the modified delta model with the three other candidate models based on the mean absolute error and the mean bias. The three candidate models included the baseline model without spatial dependence considered, the trend-surface generalized additive model and the random areal effect model. The best matrix was constructed with the Gabriel graph connectivity matrix and the exponential weighting function with parameter $m = 0.5$. The delta model modified with spatial filters showed superior performance over the other three candidate

models in the seabird bycatch example. This study provides an alternative to incorporate spatial dependence in the existing approaches for modeling spatial patterns in zero-inflated data for rare species.

2.2 Introduction

Data of rare species are usually characterized by a high percentage of zero observations due to their low abundance. Such data are generally referred to as zero-inflated data. The excess zeros may invalidate the normality assumption that we commonly use in ecological data analyses and may cause computational problems (Cunningham and Lindenmayer, 2005). Ignorance of a considerable proportion of zeros or merging of multiple records would likely result in a loss of information that reflects the spatial or temporal distribution characteristics of the species; ignorance of the spatial patterns that are likely to exist may also cause bias in estimation because the assumption of independence among observations is violated. Therefore, analysis of such data requires specialized statistical techniques, and it is quite challenging to incorporate spatial patterns in the analysis. Because rare species are frequently of ecological, conservation or management interest, analyses of such data may provide valuable information for management and conservation of rare species, e.g., to help identify hotspots of bycatch events or provide guidance for habitat restoration.

Several modeling approaches have been developed to deal with zero-inflated data. In these approaches, the non-zero data and zero data are handled either in two separate sub-models (e.g., the delta model or hurdle model [Aitchison, 1955; Aitchison and Brown, 1957; Pennington, 1983; Lo et al., 1992; Fletcher et al., 2005] and the zero-inflated model [Mullahy, 1986; Welsh et al., 1996; Hall, 2000]), or simultaneously in a single model (e.g., the Tweedie distribution model

[Tweedie, 1984] and the classification tree model [Breiman et al., 1984; Li and Jiao, 2011b]). Given these existing modeling approaches, the question then is how to incorporate spatial dependence in these approaches.

An early example of incorporating spatial patterns in the traditional regression model is the trend-surface analysis. The trend-surface models describe spatial patterns using polynomial terms of geographic coordinates (Legendre, 1990). This method is devised to capture broad-scale spatial patterns with simple shapes like planes, saddles or parabolas, but insufficient to capture fine-scale structures that may require too many parameters to estimate (Borcard and Legendre, 2002). Another major problem with the trend-surface analysis is the correlation inherent in those polynomial terms, which may cause bias in parameter estimation and statistical testing; although this problem can be minimized by orthogonalizing the polynomial terms, difficulties may arise for interpretation and prediction.

Recently applied in ecological studies, the eigenfunction-based spatial filtering provides an alternative. The principal coordinates of neighbor matrices (PCNM) proposed by Borcard and Legendre (2002) and modified by Dray et al. (2006) is one kind of the commonly used eigenfunction-based spatial filtering methods. To apply the eigenfunction-based spatial filtering, a spatial weighting matrix is constructed based on the relationship between each pair of locations: whether they are neighbors and how strong their dependence is given that they are neighbors. The spatial dependence between two locations is assumed to be a function of their Euclidean distance. The eigenvalues and their associated eigenvectors of a centered spatial weighting matrix constitute the eigenfunction. These eigenvectors (called the spatial filters hereafter) are believed to describe all the possible spatial patterns among sampling locations, where the spatial filters associated with larger eigenvalues represent broader-scale variation and

those associated with smaller eigenvalues represent finer-scale variation. The spatial filters can further be used as explanatory variables in the traditional regression model. The eigenfunction-based spatial filtering has been used with success in several ecological applications (Borcard and Legendre, 2002; Diniz-Filho and Bini, 2005; Dray et al., 2006; Griffith and Peres-Neto, 2006). Its advantages may include the flexibility in conjunction with the regression model framework, less demand for computation for large datasets and the capability of capturing both large- and fine-scale variations. Therefore, I was motivated to modify existing modeling approaches that handle zero-inflated data with eigenfunction-based spatial filters, with the hope that this modification may provide an alternative to incorporate spatial dependence in the data analysis for rare species.

In the present study, I demonstrated such a modification of the delta model. The delta model, also called the hurdle model, was developed on the basis of the delta distribution proposed by Aitchison (1955) and Aitchison and Brown (1957) to deal with excess zeros in data analyses. The delta model and zero-inflated model are similar in that they both model zero and non-zero data in two separate sub-models. Their difference lies in the sub-model that handles zero data (the probability sub-model). The probability sub-model in the delta model gives the probability of obtaining non-zero data while the one in the zero-inflated model gives the probability of obtaining zeros. Although in this study, I adopted the delta model as an example to demonstrate the modification with spatial filters for zero-inflated data, the methodology can also be applied to other modeling approaches such as the zero-inflated model.

The overall goal of this study was to demonstrate the modification of existing models with eigenfunction-based spatial filters to explore spatial patterns in zero-inflated data for rare species using the example of modified delta model fitted to a real dataset, the longline seabird

bycatch observer data. Specifically, I aimed to (1) extract spatial filters from appropriate spatial weighting matrices; (2) modify the delta model by incorporating these spatial filters; (3) compare the performance of the modified delta model with other commonly used modeling approaches including the baseline model without spatial dependence considered, the trend-surface generalized additive model and the random areal effect model.

2.3 Methods

Delta model

I applied the delta model as an example to deal with zero-inflated data for rare species. The delta model consists of two sub-models, one sub-model (positive sub-model) to analyze positive data, and the other one (probability sub-model) to estimate the probability of obtaining positive data. The product of the estimates from these two sub-models gives the final estimates:

$$\hat{c} = \hat{d} \times \hat{p}\,, \qquad (1)$$

where $\hat{d}$ is the estimated value when only the positive data are analyzed, $\hat{p}$ is the estimated probability of obtaining positive data, and $\hat{c}$ is the final estimates from the delta model. For the positive sub-model, a generalized linear model was applied where I assumed a certain statistical distribution (e.g., a normal distribution) and a link function (e.g., an identity link) for the positive data:

$$g(\hat{d}) = \beta_0 + \sum \beta_q X_q\,, \qquad (2)$$

where β_0 is the intercept; β_q is the coefficient for the qth explanatory variable X_q, and $g(.)$ is the link function. In the probability sub-model, the data were converted into the presence/absence data (y) that take a value of one for those positive observations and a value of zero otherwise. A

generalized linear model with an assumption of a binomial distribution and a logit link function was used as the probability sub-model:

$$\ln(\frac{p}{1-p}) = \alpha_0 + \sum \alpha_q X_q , \quad (3)$$

where α_0 is the intercept; α_q is the coefficient for the qth explanatory variable X_q.

The explanatory variables were selected through a stepwise approach based on the Akaike Information Criterion (AIC) and a chi-square test (Akaike, 1974; Burnham and Anderson, 2002). Model development started with a model only including an intercept. At each step of the stepwise selection, the variable that reduced the AIC value most or showed the most significant effects on the response variable was selected into the model. I repeated this step until no substantial improvement was obtained from including an additional variable. Two-way interactions were not included in the models because they were either insignificant or correlated with main factors.

Eigenfunction-based spatial filters and the modified delta model

The spatial filters were obtained by calculating the eigenvectors and eigenvalues of the $n \times n$ spatial weighting matrix $W = [w_{ij}]$ after centering, where i and j index the ith and jth locations and n is the total number of locations of the observations (Dray et al., 2006; Griffith and Peres-Neto, 2006). The spatial weighting matrix can be seen as the Hadamard product (element-wise product) of a connectivity matrix $B=[b_{ij}]$ by a weighting function matrix $A=[a_{ij}]$, i.e., $[w_{ij}]=[b_{ij}\, a_{ij}]$ (Dray et al., 2006). Elements of the connectivity matrix B take a value of one for two locations that are neighbors (i.e., connected) and zero otherwise. I constructed five connectivity matrices for comparison, including the distance-based neighborhood based on

minimum spanning tree (MST), one based on semivariogram range (SVG), the Delaunay triangulation (DEL), the Gabriel graph (GAB) and the relative neighborhood graph (REL). The last three were based on topology. The MST defined two locations to be neighbors if $d_{ij} \leq 6.19$ (10^2 km, estimated from the POP data), where d_{ij} is their Euclidian distance. Similarly, the SVG defined two locations to be neighbors if $d_{ij} \leq 20$ (10^2 km, estimated by fitting the semivariogram to the POP data, Cressie, 1993; Appendix A). In the DEL, I first defined a region $T_i = \{s : d_{is} \leq d_{js} \ \ for \ \ any \ \ j \neq i\}$ in which the distance from any point s to location i was no larger than the distance to any other locations; two locations are considered to be neighbors if T_i and T_j share a side. The GAB defined two locations to be neighbors if $d_{ij} \leq \min\left(\sqrt{d_{ik}^2 + d_{jk}^2}\right)$ for $k = 1,\ldots,$ n and $k \neq i, j$. The REL defined two locations to be neighbors if $d_{ij} \leq \max\left(d_{ik}, d_{jk}\right)$ for any $k =$ $1,\ldots, n$ and $k \neq i, j$. Details on constructing these five connectivity matrices can be found in Toussaint (1980), Lee and Schachter (1980), and Jaromczyk and Toussaint (1992).

The weighting function matrix A measures the intensity and patterns of the spatial dependence between two locations, and was obtained by applying the weighting function to their Euclidean distance. Dray et al. (2006) introduced the notion of geographic similarity in A and assumed that $0 \leq a_{ij} \leq 1$, where a value of zero measured independence and a value of one measured extreme dependence. Following this assumption and modified from the spatial correlation functions that were widely used in literatures (Weber and Talkner, 1993; Schlather, 1999; Dray et al., 2006), I constructed six types of weighting functions, including the equal weights function (F1), the concave-down decay function (F2), the inverse decay function (F3), the Gaussian decay function (F4), the exponential decay function (F5), and the spherical decay function (F6):

$$\text{F1: } a_{ij} = 1, \tag{4}$$

$$\text{F2: } a_{ij} = 1 - \left(\frac{d_{ij}}{\max(d_{ij})} \right)^{m},$$

$$\text{F3: } a_{ij} = \frac{1}{1 + d_{ij}^{m}},$$

$$\text{F4: } a_{ij} = \exp(-(md_{ij})^{2}),$$

$$\text{F5: } a_{ij} = \exp(-md_{ij}),$$

$$\text{F6: } a_{ij} = 1 - \frac{3}{2}(md_{ij}) + \frac{1}{2}(md_{ij})^{3},$$

where m is a shape parameter that controls how fast the dependence between two locations decays as their distance increases. I evaluated four different m values for each of the weighting functions F2-F6, i.e., $m = \{0.5, 1, 2, 3\}$; when m = 1, the function F2 represented a linear decay pattern. For weighting functions F4-F6, I additionally tested m = 1/20 where the value 20 (10^2 km) was the estimated asymptotic range by fitting a semivariogram model to the POP data (Appendix A), and the resulting weighting functions corresponded to the canonical spatial correlation functions (Cressie, 1993).

The spatial weighting matrix W was centered as follows (Borcard and Legendre, 2002; Dray et al., 2006; Griffith and Peres-Neto, 2006):

$$W_c = (I - \frac{J}{n})W(I - \frac{J}{n}) \tag{5}$$

where W_c is the centered spatial weighting matrix, I is a $n \times n$ identity matrix and J is a $n \times n$ matrix of one's. The eigenvectors and their corresponding eigenvalues of W_c constituted the eigenfunction, and those eigenvectors (i.e., spatial filters) associated with positive eigenvalues

were used as explanatory variables for model development (Dray et al., 2006; Griffith and Peres-Neto, 2006).

I obtained a large number of spatial filters associated with positive eigenvalues (Appendix A) because of the large dataset, I restricted the variable selection to those having $MC/MC_{\max} \geq 0.25$ for model development (Griffith, 2004), where MC refers to the Moran's I coefficient (Moran, 1950; Dray et al., 2006):

$$MC(v) = \frac{n}{1^t W 1} \frac{v^t (I - J/n) W (I - J/n) v}{v^t (I - J/n) v}, \qquad (6)$$

where $\mathbf{1}$ is a $n \times 1$ vector of one's, t refers to the transpose of a matrix or a vector, and v is a $n \times 1$ vector for which the Moran's I coefficient is calculated. For the positive sub-model, I had to further restrict the variable selection to those top 20 with the largest values of $MC/MC_{\max}$ due to few positive observations. I started with the baseline model (BL), and selected the eigenvectors through the forward selection procedure as I described in the section of delta model.

I explored a total of 108 spatial weighting matrices obtained by combinations of different connectivity matrices B and spatial weighting function matrices A. I modified the BL model by incorporating spatial filters from each of these spatial weighting matrices. Additionally, I evaluated the PCNM method by Dray et al. (2006). The PCNM uses a weighting matrix similar to the one obtained by the combination of MST connectivity matrix and F2 weighting function with $m = 2$, except that: (1) each location does not connect to itself (i.e., each location has no influence on itself) and thus the connectivity matrix B has values of zero on its diagonal; (2) max (d_{ij}) in F2 is replaced with a value that is four times the threshold; I used 6.19 (10^2 km) as the threshold in this study based on the POP data.

I compared the spatial weighting matrices based on performance of the correspondingly modified delta models. I used AIC as the main criterion for selecting the best spatial weighting matrix (matrices), and the one (s) associated with the smallest AIC value (s) would be considered to yield the best model goodness-of-fit and would be used as candidate models to compare with other commonly-used spatial modeling approaches in the next section (the section of model comparison). Because performance of the delta model is determined by both sub-models and it is possible that the two sub-models favor different spatial weighting matrices, I selected the best spatial weighting matrices favored by each sub-model, and then consistently applied each matrix to both sub-models for convenience of interpretation. Other than the AIC values, I also tracked the MC values of the response residuals ($y - \hat{p}$) for both sub-models (Griffith, 2004) and the area under the receiver operating characteristic (ROC) curve (AUC) for the probability sub-model. The ROC curve plots the true positive rate $\Pr(\hat{y} = 1 \mid y = 1)$ against the false positive rate $\Pr(\hat{y} = 1 \mid y = 0)$, and has long been used to evaluate the prediction accuracy of a classification algorithm (Bradley, 1997). Here the probability sub-model can be treated as a classification algorithm and a larger AUC value suggests higher discrimination accuracy of the model. I reported the residual MC values for both the BL and the SF models for the purpose of comparing the MC values between these two models and demonstrating how efficiently the SF model reduces spatial autocorrelation in residuals by including spatial filters.

Model comparison

I compared performance of the delta models (SF) modified with selected spatial weighting matrices with other three candidate delta models, including the baseline model (BL) that did not account for spatial patterns, the trend-surface generalized additive model (TS) using

a smooth function of latitude and longitude to describe the large scale spatial patterns (Hastie and Tibshirani, 1990; Dormann et al., 2007) and the random areal effect model (RE) that assumed that areas had a random effect on the response variable and this random effect followed a normal distribution with a zero mean (Bolker et al., 2009; Li and Jiao, 2012). I applied the five-fold cross-validation to evaluate the performance of candidate models (Hastie et al., 2001; Damalas et al., 2007; Shono, 2008). In the five-fold cross-validation, I randomly divided the whole dataset into five sub-datasets with roughly equal size. Each subset was used as test data to predict from the model, and the other four subsets were combined as training data to fit the model. The comparison criterion was calculated for each set of training and test. I used the mean absolute error (MAE) and the mean bias as criteria. The model providing a lower MAE value and a mean bias closer to zero would be judged to have better performance. The MAE and mean bias were calculated as follows respectively:

$$MAE = \frac{1}{n}\sum\left|\hat{c}_i - c_i\right|, \tag{7}$$

$$Mean\ bias = \frac{1}{n}\sum\left(\hat{c}_i - c_i\right). \tag{8}$$

Seabird bycatch example

In this study, I used the seabird bycatch data provided by National Marine Fisheries Service Southeast Fisheries Science Center Pelagic Observer Program (POP). The POP program was initiated for monitoring the United States Atlantic pelagic longline fishery. In the POP program, randomly selected fishing trips are attended by an observer, and the detailed information on gear specification, fishing effort, fishing tactics, target catch, bycatch and environmental conditions is recorded. I analyzed 12,894 longline sets of the POP data during

1992 to 2011, of which only 73 sets caught 140 seabirds; therefore, approximately 99.4% zero observations occurred. The POP data covered 11 fishing areas (Fig. 1, Lee and Brown, 1998). In the following analyses, I projected the geographic coordinates recorded in the POP data (i.e., degree north and degree west) to the Universal Transverse Mercator (UTM) system zone 15 based on the NAD83 geodetic datum. The explanatory variables analyzed are listed in Table 1.

I decided to use the POP seabird bycatch data as an example for three reasons: (1) this dataset represents a typical structure of survey data for rare species with an extremely high percentage of zero observations; (2) seabird bycatch in the POP showed spatial patterns, i.e., most seabird bycatch occurred along the east coast of the United States (Fig. 1), which needs to be considered to increase the accuracy of assessment; (3) the large dataset and the extreme low capture rate have posed challenges in modeling spatial patterns.

With the seabird bycatch data, $\hat{d}$ is the estimated bycatch when only the longline sets with at least one seabird caught were analyzed, $\hat{p}$ is the estimated probability of catching seabirds, and $\hat{c}$ is the estimated number of seabird bycatch on a longline set. In the positive sub-model, I assumed the log-transformed data to follow a normal distribution. I applied log-transformation because it has been widely used in fisheries catch/bycatch analysis (e.g., Stefansson, 1996; Ortiz et al., 2000). I used normal distribution instead of Poisson or negative binomial distributions for three reasons. First, the bycatch data were no longer count data after log-transformation. Second, the bycatch data can be treated as bycatch rate (number of seabird bycatch per longline set) as in Winter et al. (2011) where a normal distribution was also applied after log-transformation. Third, the probability density function of continuous data has been assumed on log-transformed count data in other literatures (Pennington, 1983; Stefansson, 1996; Li et al., 2011).

Fitting the baseline model (BL) to the POP data, the positive sub-model selected soak duration in the model and the probability sub-model included latitude, longitude, season and target species. The other candidate models fitted to the POP data including the one modified with spatial filters (SF) are shown in Table 2. For the RE model, I did not include an areal random effect in the positive catch sub-model and used the same sub-model as the BL model because the POP data obtained few captures of seabirds and all captures occurred in five of the 11 fishing areas (Fig. 1).

2.4 Results

I present in Table 3 the results of the best two spatial weighting matrices from each type of connectivity matrix, one favored by the positive sub-model (i.e., yielding the smallest AIC in the positive sub-model) and the other one favored by the probability sub-model (i.e., yielding the smallest AIC in the probability sub-model). The full list of results are summarized in Appendix A. Compared with the BL model, including spatial filters greatly decreased the AIC values for the probability sub-model but not for the positive sub-model (Tables 2 and 3). Among all the 108 spatial weighting matrices and the PCNM method, two matrices performed equally best in the positive sub-model, yielding the smallest AIC value (139.5); one was constructed with the MST connectivity matrix and the other one with the SVG connectivity matrix, and both were weighted by the exponential function (F5) with $m = 2$. However, the differences in AIC values between candidate positive sub-models were not obvious because a difference of four in AIC values will be considered to be statistically significant. The best spatial weighting matrix selected by the probability sub-model, corresponding to the lowest value of AIC (646.9), combined the GAB connectivity matrix and the exponential weighting function (F5) with $m = 0.5$. The PCNM

method performed better than the BL model, but many of the other spatial weighting matrices produced a better fit. Further comparison between these three spatial weighting matrices through the five-fold cross validation determined the one combining the GAB and F5 with $m = 0.5$ to be the best for the overall delta model (Table 2), yielding the smallest MAE value for training data on average (0.0172 on average), competing MAE value for test data (0.0192 on average) and the smallest mean bias (-0.0012 for training data and -0.0009 for test data on average).

The best delta model that I selected above (SF-GAB), i.e., the one modified with the spatial weighing matrix combining the GAB and F5 with $m = 0.5$, selected two spatial filters in the positive sub-model and 31 spatial filters in the probability sub-model (Tables 2 and 3). Both spatial filters in the positive sub-model were associated with relatively large eigenvalues (the 11th largest value of 8.6 and the 17th largest value of 7.7) representing broader-scale spatial patterns. The SF-GAB model reduced the MC value in the positive data (0.793) by 3.2% while the BL model reduced the observation MC value by only 1.6%. Of the 31 spatial filters in the probability sub-model, four were associated with relatively large eigenvalues (7.1-13.6) representing broader-scale spatial patterns and the rest associated with small eigenvalues (4.9-6.6) representing finer-scale spatial patterns. Compared to the MC value in the presence/absence data (0.244), the SF-GAB model had the MC value of its response residuals drop by 45.9% while the BL model had it drop by only 2.8%. The AUC value produced by the SF-GAB model was relatively high (0.9401). Spatial patterns in the positive catch depicted by the two spatial filters in the SF-GAB model are shown in Fig. 2, and examples of the spatial patterns in the probability of catching seabirds depicted by spatial filters are shown in Fig. 3.

Results of model comparison between the SF-GAB model and the other three commonly used models are summarized in Tables 4 and 5. Of the four candidate models, the SF-GAB

produced the lowest MAE value for training data; although its MAE value for test data was larger than the BL and the RE models, the differences between candidate models was slight (Table 4). The SF-GAB had the smallest mean bias for both training and test data (Table 5). Spatial patterns of the estimated positive catch, the estimated probability of catching seabirds and the estimated seabird bycatch from these four candidate models are shown in Figs. 4, 5 and 6. The SF-GAB model captured more detailed spatial patterns than the other three candidate models. For the estimated positive catch, the SF-GAB model and the BL model produced similar spatial patterns, i.e., higher positive catch along east coast of the United States and in the Gulf of Mexico (Figs. 4a and 4b), with the SF-GAB model providing more local structures especially in the Sargasso region and the north central Atlantic (Fig. 4b); however, the TS model showed strength in capturing the larger-scale trend with higher positive catch along east coast and lower in the Gulf of Mexico and the northeast district (Fig. 4c). For the probability of catching seabirds, the four candidate models gave similar spatial patterns, i.e., higher probability along northeast coast of the United States and in the northeast district, and more small-scale patterns were captured by the SF-GAB model especially for the northeast coast (Fig. 5b). When combining the estimates from both sub-models, all four candidate models described a similar pattern with higher seabird bycatch along northeast coast of the United States and in the northeast district; the BL model also highlighted the Gulf of Mexico with higher seabird bycatch (Fig. 6a). The estimated seabird bycatch by the SF-GAB model along the northeast coast was lower than the other three candidate models; the estimates in the southwest corner of the northeast district was lower than the rest of this district, which added a local structure to the overall trend (Fig. 6b).

2.5 Discussion

Compared with the BL model that considered no spatial dependence, the delta model showed a great improvement in goodness-of-fit after incorporating the spatial filters in the probability sub-model but little improvement was achieved for the positive sub-model. For example in the SF-GAB model, the spatial filters led to a decrease of 147 in AIC value for the probability sub-model but only a decrease of one for the positive sub-model (Table 2). This result suggested the existence of spatial dependence in the probability of catching seabirds but little large-scale spatial patterns in the positive catch data, which agreed with the result from semivariograms (Appendix A). However, this result does not provide evidence for the absence of local spatial structure in the positive catch data because I restricted our selection to those spatial filters associated with large eigenvalues that described larger-scale spatial patterns due to few positive observations. If I also investigate those spatial filters associated with small eigenvalues that describe smaller-scale spatial patterns, I may find substantial improvement in goodness-of-fit for the modified positive sub-model if local variation did play an important role. I ignored those spatial filters with small eigenvalues in this study due to intensive computing. For example in the SF-GAB model, I would have more than 7,702 spatial filters to select from if I consider all spatial filters with positive eigenvalues (Appendix A). I did find support for the existence and importance of local spatial variation in the probability of catching seabirds. Of the 31 spatial filters selected in the probability sub-model, 27 were associated with small eigenvalues. The importance of spatial impacts on seabird bycatch in longline fisheries has been well documented (Klaer and Polacheck, 1998; Brothers et al., 1999; Jimenez et al., 2010; Li and Jiao, 2011a).

Besides the improvement in model performance due to incorporation of spatial dependence, another significant advantage of the spatial filtering is its flexibility. First, all spatial

filters are mutually independent; they can capture both broader- and finer-scale spatial patterns. Second, spatial filters can be incorporated as explanatory variables in the commonly used statistical models such as the generalized linear models. Third, the spatial weighting matrix can be constructed in various ways to meet the assumptions of spatial patterns in the data, and does not require the positive-definiteness that is generally required by the spatial variance-covariance matrix. In this study, I constructed a total of 109 spatial weighting matrices that combined different neighbour structures and different patterns of how spatial dependence decays with separation distance (Table 3), and some of these weighting matrices utilized information from semivariograms (i.e., the range estimate) that would also be used when constructing a spatial variance-covariance matrix. A major drawback of the eigenfunction-based spatial filtering is that centering of the spatial weighting matrix may require intensive computing for large datasets such as the POP data. Fortunately, some numerical techniques can relieve this problem, for example, I reduced those calculations on matrices to the calculations on vectors in this study. Additionally, Griffith (2000) provided an analytical formulation for eigenvalues and eigenvectors for regular square tessellations, which would be useful for large raster datasets.

The choice of a spatial weighting matrix has proven critical for data with irregular distribution of locations (Tiefelsdorf et al., 1999; Dray et al., 2006), which is typical for the data of rare species. For data with irregular distributions, the spatial weighting matrix greatly influence the number of positive eigenvalues and the spatial patterns depicted by their associated spatial filters. Choice of the best spatial weighting matrix may depend on several factors including selection criterion, the number of alternative matrices compared and the data structure. However, no matter what weighting matrix that I choose, the overall assumption for constructing these weighting matrices is that those observations nearer to a location have greater influence

than those farther away; their major difference lies in how such influence decays as distance increases. Having data to select the best spatial weighting matrix based on certain criteria such as the AIC provides a useful tool to determine the best spatial weighting matrix for generating spatial filters. Additionally, the ecology behind the choice of a specific weighting matrix is also an important criterion. For example, the best spatial weighting matrix for the seabird bycatch data used the GAB connectivity matrix. The GAB matrix connected individual locations into five clusters (Fig. 7), namely the northeast corner of my study region, the U.S. Atlantic coast from gulf of main to Florida, the Gulf of Mexico, the north central Atlantic and the southeast corner of our study region. I would hypothesize that locations within each of the clusters were highly connected due to its unique oceanographical features, and locations across different clusters were less connected due to great difference in oceanography between clusters. For example, the U.S. Atlantic coast covers large latitudinal range from subtropical to temperate, which results in great latitudinal variation in climate and weather-related physical processes (Atkinson and Menzel, 1985). In contrast, the other four clusters cover either temperate (the northeast corner), subtropical (the Gulf of Mexico and the north central Atlantic) or tropical zone (the southeast corner), which provides a relatively less variable climate. The northeast and southeast corners of my study region are away from the U.S. Atlantic coast towards open sea, which results in deeper waters, lower primary productivity and less dynamic surface current than the coast. The Gulf of Mexico is shallow waters largely surrounded by continent, and hosts the Gulf Stream that greatly influences the marine ecosystem in the gulf and south Atlantic bight (Weaver, 1955; Atkinson and Menzel, 1985).

The spatial filters selected in the model may represent certain environmental factors that highly vary over space but were not included in the model. For example, spatial patterns of the

spatial filters in Figs. 2 and 3 showed a general trend that values along coast were different from those towards open sea. Distinct oceanographical and biological features near shore versus in open sea, such as water depth, ocean current and primary productivity, may contribute to the formation of these patterns. In general, there are deeper waters and lower primary productivity, and less variable ocean surface current when moving away from coast.

In the spatial filtering method, the purpose of centering is to guarantee the extracted spatial filters to be mutually orthogonal and independent (Griffith, 2000). Tiefelsdorf & Griffith (2007) mentioned that different ways to center the spatial weighting matrix can link the eigenfuncion-based filtering approach to several parametric spatial autoregressive regression models. According to Tiefelsdorf & Griffith (2007), compared to centering on residuals, the Eq (5) that I used to center the spatial weighting matrix may give spatial filters that are correlated with explanatory variables. To reduce such a problem, I applied stepwise selection approach to select significant spatial filters into the model. Those spatial filters highly correlated with explanatory variables will not be selected due to its little contribution to model fitting.

As applied in most ecological studies, I only focused on the spatial filters associated with positive eigenvalues for two reasons. First, the seabird bycatch in the POP data more likely showed positive correlation among locations rather than a negative one given that the semivariograms showed a decline in dependence as separation distance increased (Appendix A). Second, the large dataset led to a large number of spatial filters generated from a spatial weighting matrix (Appendix A); however, the few captures of seabirds limited the number of spatial filters that can be incorporated in the model, and thus I had decided to constrain our interest to those associated with positive eigenvalues. The spatial filters with negative

eigenvalues also contain information about spatial patterns, and are available to use when negative spatial correlation appears in data (Dray et al., 2006).

Performance of the delta model is influenced by both sub-models; this feature may introduce great complexity in applying spatial filtering technique to zero-inflated data. Because the observed locations associated with positive catch data form a subset of all locations in the study area, there may be two ways to apply spatial filtering in the delta model. In the first method, I construct a spatial weighting matrix for the positive sub-model and the probability sub-model separately. More detailed, I constructed one spatial weighting matrix using all locations and only used this matrix for the probability sub-model; another spatial weighting matrix can be constructed using those locations associated with positive data and then extrapolated to all the other locations in the whole study area, i.e., expand the spatial patterns built on the positive catch locations to the whole study area. In the second method, I constructed one spatial weighting matrix using all locations and applied this matrix to both sub-models by assuming spatial patterns in a sub-region integrates with those in the whole study area. Both methods may suffer the problem that the orthogonality among spatial filters is not guaranteed. In this study, as a pioneer trial, I applied the second method because it is easier to practice and interpret, and can avoid introducing more uncertainty that occurs in extrapolation in the first method. However, I would like to bring the following three questions to an open discussion, and I am hoping that better alternatives can be explored through such discussion. First of all, can spatial filters that describe spatial patterns in the whole area also represent spatial patterns in the sub-region? The spatial weighting matrix where spatial filters are extracted uses only the information on relative location. Intuitively, if spatial filters describe the spatial patterns over the whole study area, it should already cover the spatial patterns in the sub-region because the relative location of

observations within the sub-region will remain the same as in the whole area. Thus, I think that it is intuitively acceptable to define the spatial filters in the whole area and apply it to the sub-region. Second, how important is it to keep spatial filters strictly orthogonal? In my opinion, orthogonality among spatial filters is one of the nice features inherent in the eigenfuncion-based spatial filtering technique because this feature meets the assumption in linear regression that all explanatory variables are independent. This assumption is to avoid the multicolinearity problem. However in reality, the assumption of independence among explanatory variables may be an exception rather than the rule. There are several methods to avoid the multicolinearity problem in linear regression, and one of them is to select the explanatory variables through a stepwise procedure as I applied in this study. In the stepwise selection procedure, those variables highly correlated with the ones that are already included in the model will not be selected anymore due to their redundant information. In my case of spatial filters, given the answer to the first question being positive, spatial filters are treated as explanatory variables and the stepwise selection procedure should work to avoid multicolinearity. Third, are there better alternatives to construct spatial filters for sub-region? The eigenfunction-based spatial filtering technique has been recently applied to ecological studies. To my best knowledge, the question of how to apply it to zero-inflated data still remains open. A simulation study may help address these questions in future studies.

Regardless of how to construct the spatial weighting matrix for both sub-models, applying spatial filters to the delta model may suffer another problem when selecting the best spatial weighting matrix because the two sub-models may favor different matrices that represent different spatial patterns, which is also a drawback inherent in the delta model. To handle this problem in the present study, I practiced a two-step selection. In the first step, I selected those

spatial weighting matrices that were favored by each sub-model and consistently applied each matrix to both sub-models; in the second step, I determined the best one among all matrices that were selected in the first step through the five-fold cross-validation. Another alternative way to select the best spatial weighting matrix in the delta model could be based on the sum of AICs from the two sub-models, i.e., the best weighting matrix is the one with the lowest sum of two AICs.

Other than using information on relative locations to generate spatial filters, I also included latitude and longitude as environmental factors in the probability sub-model. There are two reasons for this practice. First, latitude and longitude were selected by the model itself due to their significance in explaining the variation in the POP data. Furthermore, there may be information imbedded in these two variables that are related to other influential factors such as bait and ocean currents, which needs to be further explored. The spatial filters generated based on relative locations contributed to explaining part of the variation in the POP data that cannot be explained by those environmental factors. Second, I conducted the same analysis when excluding latitude and longitude in candidate models, and obtained the same conclusions but larger AIC values than the corresponding models with these two variables included (results are not presented here).

In this study, I did not consider the spatial patterns in explanatory variables. Incorporation of these spatial patterns will provide a direction for future improvement of the current study. Centering the spatial weighting matrix in a different way, e.g., centering on model residuals (Tiefelsdorf and Griffith, 2007), may help achieve such an attempt.

In summary, using the POP seabird bycatch data as an example, the delta model modified with the eigenfunction-based spatial filters showed superior performance over the model without

spatial dependence considered, or those alternatives with trend surface or with random areal effects. The eigenfuncion-based spatial filtering provides a flexible tool to incorporate spatial dependence in the existing modeling approaches that handle zero-inflated data for rare species. Although I only illustrated a modification of the delta model, such an application can also be extended to other modeling approaches such as the zero-inflated model. My conclusions may be limited by selection criterion, candidate models compared and data structure (especially the percentage of zero observations in the data). I am hoping to provide important insights in alternative ways of modeling spatial patterns in zero-inflated data for rare species.

2.6 Acknowledgements

This research was supported by the USDA Cooperative State Research, Education and Extension Service, Hatch project #0210510 to Dr. Y. Jiao, and a grant for Modeling Pelagic Longline Seabird Bycatch awarded to Y. Jiao by the National Oceanic and Atmospheric Administration (NOAA), National Marine Fisheries Service (NMFS) Southeast Fisheries Science Center (SEFSC), as part of the NOAA Fisheries National Seabird Program.

2.7 References

Aitchison, J., 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. Journal of the American Statistical Association 50, 901-908.

Aitchison, J., Brown, J., 1957. The lognormal distribution. Cambridge University Press, Cambridge, UK.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-723.

Atkinson, L.P., Menzel, D.W., 1985. Introduction: Oceanography of the southeast United States continental shelf. In: Atkinson, L.P., Menzel, D.W., Bush, K.A. (Eds), Oceanography of the southeastern US continental shelf. American Geophysical Union, Washington, D.C.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24, 127-135.

Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol Model 153, 51-68.

Bradley, A., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition 30, 1145-1159.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees. Wadsworth Publishing Company, Belmont, CA.

Brothers, N., Gales, R., Reid, T., 1999. The influence of environmental variables and mitigation measures on seabird catch rates in the Japanese tuna longline fishery within the Australian Fishing Zone, 1991-1995. Biol Conserv 88, 85-101.

Burnham, K., Anderson, D., 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

Cressie, N., 1993. Statistics for spatial data. Wiley, New York.

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

Damalas, D., Megalofonou, P., Apostolopoulou, M., 2007. Environmental, spatial, temporal and operational effects on swordfish (*Xiphias gladius*) catch rates of eastern Mediterranean Sea longline fisheries. Fish Res 84, 233-246.

Diniz-Filho, J.A.F., Bini, L.M., 2005. Modelling geographical patterns in species richness using eigenvector-based spatial filters. Global Ecol Biogeogr 14, 177-185.

Dormann, C.F., McPherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemuller, R., Peres-Neto, P.R., Reineking, B., Schroder, B., Schurr, F.M., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609-628.

Dray, S., Legendre, P., Peres-Neto, P.R., 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model 196, 483-493.

Fletcher, D., MacKenzie, D., Villouta, E., 2005. Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. Environ Ecol Stat 12, 45-54.

Griffith, D.A., 2000. Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. Linear Algebra and its Applications 321, 95-112.

Griffith, D.A., 2004. A spatial filtering specification for the autologistic model. Environment and Planning A 36, 1791-1812.

Griffith, D.A., Peres-Neto, P.R., 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. Ecology 87, 2603-2613.

Hall, D.B., 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56, 1030-1039.

Hastie, T., Tibshirani, R., 1990. Generalized additive models. Chapman and Hall, London, UK.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning: data mining, inference and prediction. Springer, New York.

Jaromczyk, J.W., Toussaint, G.T., 1992. Relative neighborhood graphs and their relatives. Proceedings of the IEEE 80, 1502-1517.

Jimenez, S., Abreu, M., Pons, M., Ortiz, M., Domingo, A., 2010. Assessing the impact of the pelagic longline fishery on albatrosses and petrels in the southwest Atlantic. Aquat Living Resour 23, 49-64.

Klaer, N., Polacheck, T., 1998. The influence of environmental factors and mitigation measures on by-catch rates of seabirds by Japanese longline fishing vessels in the Australian region. Emu 98, 305-316.

Lee, D.T., Schachter, B.J., 1980. Two algorithms for constructing a Delaunay triangulation. International Journal of Parallel Programming 9, 219-242.

Lee, D.W., Brown, C.J., 1998. Southeast Fisheries Science Center pelagic observer program data summary for 1992-1996. NOAA Technical Memorandum NMFS-SEFSC-408. National Marine Fisheries Service Southeast Fisheries Science Center, Miami, FL.

Legendre, P., 1990. Quantitative methods and biogeographic analysis. In: Grarbary, D.J., South, R.G. (Eds), Evolutionary biogeography of the marine algae of the North Atlantic. Springer-Verlag, Berlin.

Li, Y., Jiao, Y., 2011a. Estimated seabird bycatch in the U.S. Atlantic pelagic longline fishery during 1992-2010 based on observer and logbook data. Report to NOAA Fisheries Service, Southeast Fisheries Science Center, Miami, Florida. Virginia Polytechnic

Institute and State University, Department of Fish and Wildlife Conservation, Blacksburg, Virginia.

Li, Y., Jiao, Y., 2011b. Influences of gillnet fishing on lake sturgeon bycatch in Lake Erie and implications for conservation. Endangered Species Research 13, 253-261.

Li, Y., Jiao, Y., 2012. Estimated seabird bycatch in the U.S. Atlantic pelagic longline fishery during 1992-2011 based on observer and logbook data. Report to NOAA Fisheries Service, Southeast Fisheries Science Center, Miami, Florida. Virginia Polytechnic Institute and State University, Department of Fish and Wildlife Conservation, Blacksburg, Virginia.

Li, Y., Jiao, Y., Reid, K., 2011. Assessment of landed and non-landed by-catch of walleye, yellow perch and white perch from the commercial gillnet fisheries of Lake Erie, 1994–2007. J Great Lakes Res 37, 325-334.

Lo, N., Jacobson, L., Squire, J., 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can J Fish Aquat Sci 49, 2515-2526.

Moran, P.A., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17-23.

Mullahy, J., 1986. Specification and testing of some modified count data models. Journal of Econometrics 33, 341-365.

Ortiz, M., Legault, C., Ehrhardt, N., 2000. An alternative method for estimating bycatch from the US shrimp trawl fishery in the Gulf of Mexico, 1972-1995. Fish Bull 98, 583-599.

Pennington, M., 1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics 39, 281-286.

Schlather, M., 1999. An introduction to positive definite functions and to unconditional simulation of random fields. Department of Mathematics and Statistics, Lancaster University, UK.

Shono, H., 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fish Res 93, 154-162.

Stefansson, G., 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES J Mar Sci 53, 577-581.

Tiefelsdorf, M., Griffith, D.A., 2007. Semiparametric filtering of spatial autocorrelation: the eigenvector approach. Environment and Planning A 39, 1193.

Tiefelsdorf, M., Griffith, D.A., Boots, B., 1999. A variance-stabilizing coding scheme for spatial link matrices. Environment and Planning A 31, 165-180.

Toussaint, G.T., 1980. The relative neighbourhood graph of a finite planar set. Pattern recognition 12, 261-268.

Tweedie, M., 1984. An index which distinguishes between some important exponential families. In: Ghosh, J., Roy, J. (Eds), Statistics: Applications and New Directions Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. Indian Statistical Institute, Calcutta.

Weaver, P., 1955. Gulf of Mexico. Geological Society of America Special Papers 62, 269-278.

Weber, R.O., Talkner, P., 1993. Some remarks on spatial correlation function models. Monthly Weather Review USA 121, 2611-2611.

Welsh, A., Cunningham, R.B., Donnelly, C., Lindenmayer, D.B., 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecol Model 88, 297-308.

Winter, A., Jiao, Y., Browder, J.A., 2011. Modeling low rates of seabird bycatch in the U.S. longline fishery. Waterbirds 34, 289-303.

Table 1

List of explanatory variables that were analyzed.

Variables	Notations	Type	Categories/mean	Details
Area	AREA	Categorical	11	See Fig. 1 for details about the 11 fishing areas
Season	SEA	Categorical	4	Spring (Mar-May), Summer (Jun-Aug), Fall (Sep-Nov), Winter (Dec-Feb)
Target species	TAR	Categorical	5	Mixed, Swordfish, Tuna, Shark, Dolphinfish
Latitude (10^2 km)	LAT	Continuous	35.32	Beginning of a set
Longitude (10^2 km)	LON	Continuous	17.85	Beginning of a set
Soak duration (h)	SOAK	Continuous	8.7	Time period between time of setting and time of hauling

Table 2

Candidate models for estimating seabirds bycatch. BL-the baseline model without spatial dependence considered; SF-the delta model modified with spatial filters from spatial weighting matrices (a total of three spatial weighting matrices were selected according to Table 3, one with MST connectivity matrix and weighting function F5 at $m = 2$, one with SVG connectivity matrix and weighting function F5 at $m = 2$, and one with GAB connectivity matrix and weighting function F5 at $m = 0.5$); TS-the trend-surface generalized additive model; RE-the random areal effect model. See Table 1 for the abbreviation of explanatory variables.

Candidate models	Positive sub-model		Probability sub-model	
	Model structure	AIC	Model structure	AIC
BL	Intercept + SOAK	142.342	Intercept + LAT + LON + SEA + TAR	793.551
SF-MST	Intercept + SOAK + 1 spatial filter	139.478	Intercept + LAT + LON + SEA + TAR + 8 spatial filters	752.745
SF-SVG	Intercept + SOAK + 1 spatial filter	139.478	Intercept + LAT + LON + SEA + TAR + 8 spatial filters	752.745
SF-GAB	Intercept + SOAK + 2 spatial filters	141.392	Intercept + LAT + LON + SEA + TAR + 31 spatial filters	646.898
TS	Intercept + SOAK + f^{a}(LAT) + f(LON)	152.623	Intercept + f(LAT) + f(LON) + SEA + TAR	748.011
RE	Intercept + SOAK	142.342	Intercept + LAT + LON + SEA + TAR + h^{b}	763.911

[a]: The $f(.)$ refers to a smooth function in the generalized additive model.

[b]: The h refers to the areal random effect following a normal distribution with a zero mean.

Table 3

Performace of the 108 spatial weighting matrices and the PCNM method for the POP data. Reported here are the results of the best two spatial weighting matrices from each type of the connectivity matrix, one favored by the positive sub-model (i.e., yielding the smallest AIC in the positive sub-model) and the other one favored by the probability sub-model (i.e., yielding the smallest AIC in the probability sub-model). The best spatial weighting matrics selected are bolded. The full list of results are shown in Appendix A.

				Positive sub-model					Probability sub-model						Sub-model
B[a]	A[b]	m	# pos[c]	# sel[d]	AIC	Obs MC[e]	BL MC[f]	SF MC[g]	# sel	AIC	Obs MC	BL MC	SF MC	AUC	that favors this matrix
MST	**F5**	**2**	**11909**	**1**	**139.478**	**0.141**	**0.158**	**0.118**	**8**	**752.745**	**0.020**	**0.010**	**0.004**	**0.898**	**Positive**
	F6	1	8948	1	141.650	-0.009	-0.006	-0.014	2	737.139	0.012	0.004	-0.001	0.901	Probability
SVG	**F5**	**2**	**11909**	**1**	**139.478**	**0.141**	**0.158**	**0.118**	**8**	**752.745**	**0.020**	**0.010**	**0.004**	**0.898**	**Positive**
	F3	1	10390	1	142.309	0.035	0.039	0.037	2	741.521	0.008	0.003	≈ 0	0.899	Probability
DEL	F2	0.5	7119	1	140.310	0.819	0.804	0.818	38	659.871	0.175	0.170	0.109	0.963	Positive
	F5	0.5	7203	0	142.342	0.819	0.804	0.804	40	655.171	0.180	0.175	0.104	0.962	Probability
GAB	**F5**	**0.5**	**7702**	**2**	**141.392**	**0.793**	**0.780**	**0.767**	**31**	**646.898**	**0.244**	**0.237**	**0.132**	**0.940**	**Both**
REL	F3	0.5	8466	2	140.020	0.792	0.787	0.781	10	768.321	0.312	0.304	0.279	0.880	Positive

	F4	3	8252	1	141.488	0.774	0.768	0.764	11	763.585	0.292	0.283	0.258	0.883	Probability
PCNM	-	-	2271	0	142.342	-0.015	-0.014	-0.014	3	748.497	0.010	0.002	≈ 0	0.894	-

[a]: The *B* refers to the connectivity matrix.

[b]: The *A* refers to the spatial weighting function matrix.

[c]: the number of positive spatial filters generated from a spatial weighting matrix.

[d]: the number of spatial filters selected into the sub-model.

[e]: the Moran's I coefficient of the observations.

[f]: the Moran's I coefficient of the response residuals for the baseline model without spatial dependence considered.

[g]: the Moran's I coefficient of the response residuals for the model modified with spatial filters.

Table 4

Mean absolute error (MAE) through five-fold cross-validation for the POP data. See Table 2 for the explanation of candidate models.

Model	Training						Test					
	1	2	3	4	5	Average	1	2	3	4	5	Average
BL	0.0181	0.0200	0.0187	0.0197	0.0189	0.0191	0.0205	0.0171	0.0198	0.0188	0.0194	0.0191
SF-MST	0.0182	0.0198	0.0190	0.0196	0.0188	0.0191	0.0206	0.0170	0.0203	0.0185	0.0196	0.0192
SF-SVG	0.0182	0.0198	0.0190	0.0196	0.0188	0.0191	0.0206	0.0170	0.0203	0.0185	0.0196	0.0192
SF-GAB	0.0161	0.0181	0.0178	0.0174	0.0168	0.0172	0.0187	0.0167	0.0207	0.0208	0.0191	0.0192
TS	0.0184	0.0199	0.0188	0.0197	0.0187	0.0191	0.0208	0.0170	0.0200	0.0187	0.0193	0.0192
RE	0.0179	0.0198	0.0186	0.0195	0.0187	0.0189	0.0203	0.0169	0.0197	0.0187	0.0194	0.0190

Table 5

Mean bias through five-fold cross-validation for the POP data. See Table 2 for the explanation of candidate models.

Model	Training						Test					
	1	2	3	4	5	Average	1	2	3	4	5	Average
BL	-0.0024	-0.0024	-0.0023	-0.0021	-0.0025	-0.0023	-0.0048	-0.0005	-0.0032	-0.0011	-0.0020	-0.0023
SF-MST	-0.0021	-0.0023	-0.0019	-0.0019	-0.0023	-0.0021	-0.0044	-0.0006	-0.0024	-0.0014	-0.0017	-0.0021
SF-SVG	-0.0021	-0.0023	-0.0019	-0.0019	-0.0023	-0.0021	-0.0044	-0.0006	-0.0024	-0.0014	-0.0017	-0.0021
SF-GAB	-0.0021	-0.0009	-0.0007	-0.0011	-0.0013	-0.0012	-0.0037	0.0000	-0.0004	0.0012	-0.0017	-0.0009
TS	-0.0018	-0.0023	-0.0021	-0.0019	-0.0025	-0.0021	-0.0043	-0.0004	-0.0028	-0.0011	-0.0018	-0.0021
RE	-0.0024	-0.0024	-0.0023	-0.0021	-0.0026	-0.0024	-0.0049	-0.0006	-0.0032	-0.0012	-0.0018	-0.0023

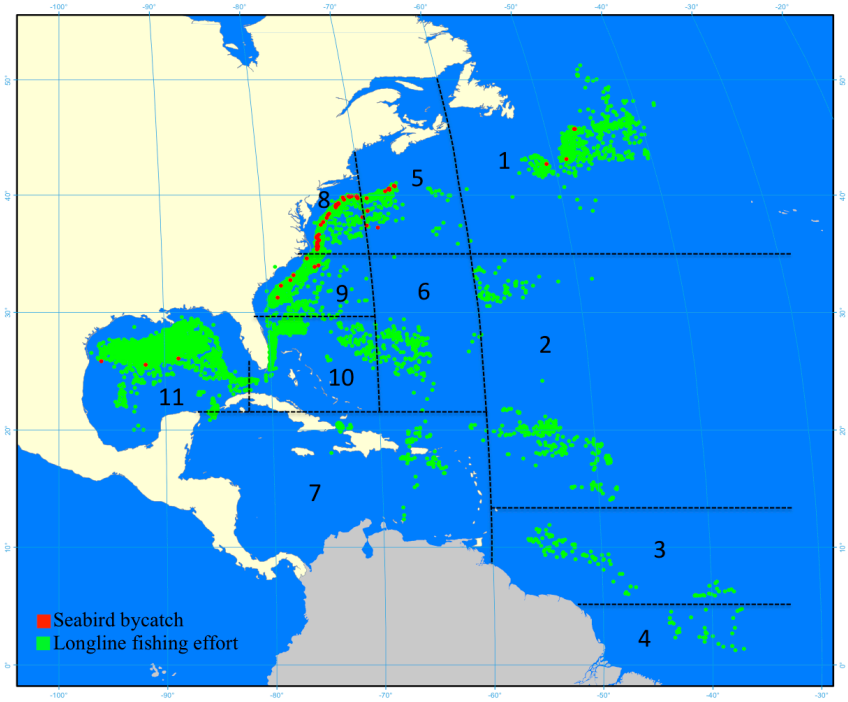

Figure 1. Spatial distribution of the observed longline sets and those with seabirds caught. 1-Northeast district, 2-North Central Atlantic, 3-Tuna North, 4-Tuna South, 5-Northeast coast, 6-Sargasso region, 7-Caribbean region, 8-Middle Atlantic bight, 9-South Atlantic bight, 10-Florida east coast, 11-Gulf of Mexico.

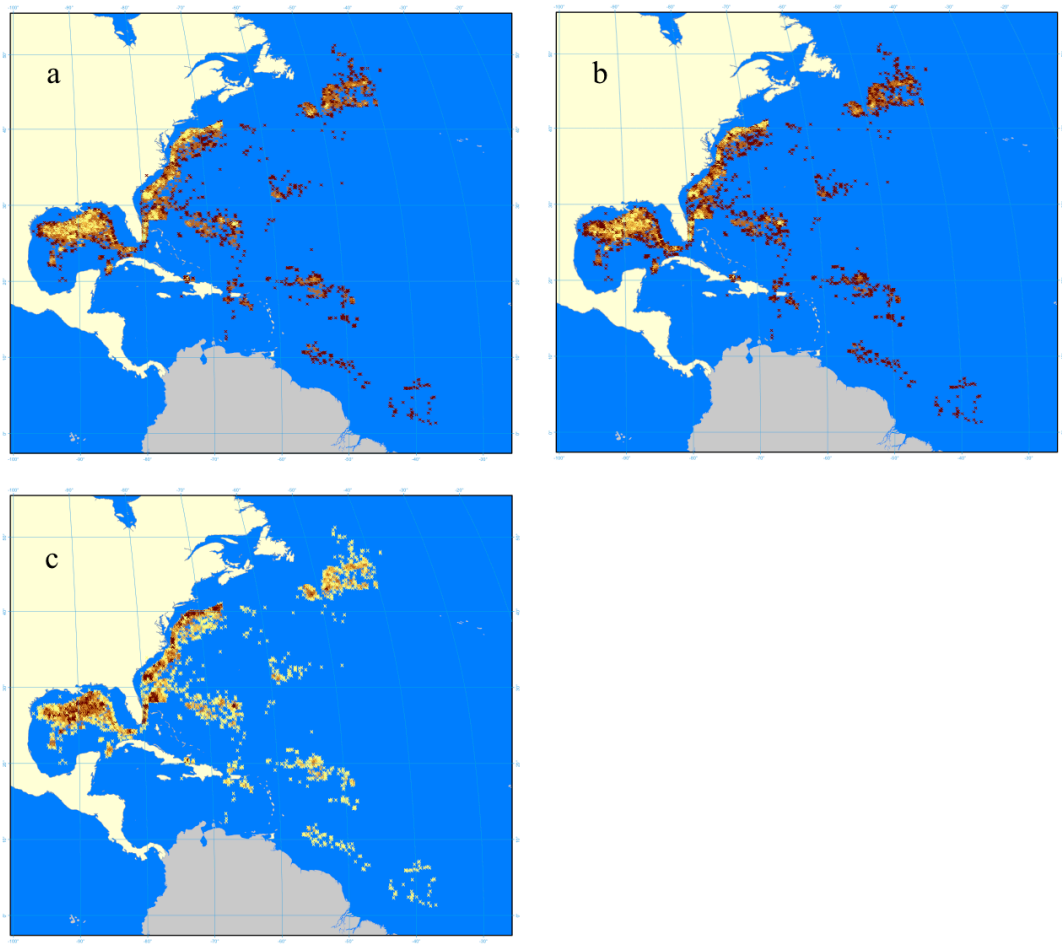

Figure 2. Spatial patterns described by the two eigenvectors in SF-GAB model for estimating the positive catch of seabirds, i.e., (a) the one associated with the 11th largest eigenvalue, (b) the one associated with the 17th largest eigenvalue, and (c) the linear combination of these two eigenvectors. Colors are based on the values of the eigenvector, for example, darker colors for larger values and lighter colors for smaller values. See Table 2 for the explanation of SF-GAB model.

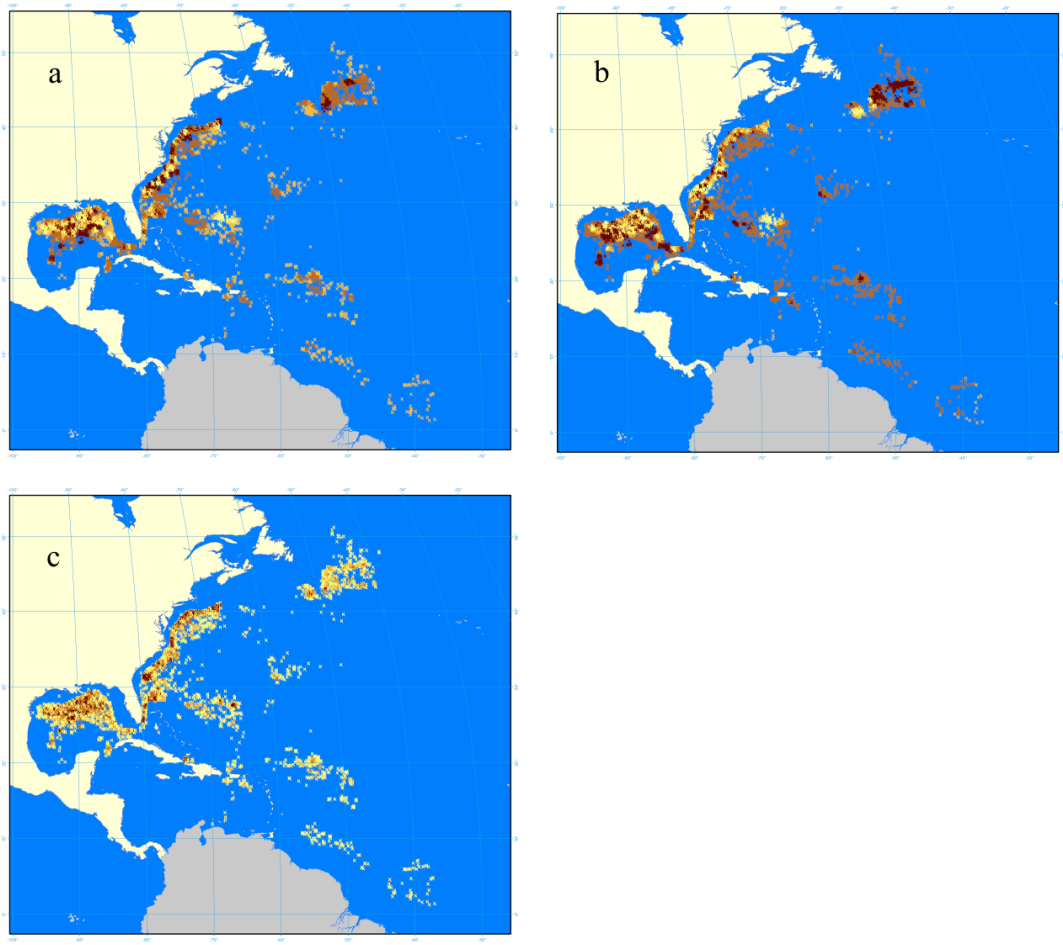

Figure 3. Spatial patterns described by the eigenvectors in SF-GAB model for estimating the probability of catching seabirds. Here are shown two eigenvectors as examples, (a) the one associated with the second largest eigenvalue describing broader-scale spatial variation, (b) the one associated with the 498th largest eigenvalue describing finer-scale spatial variation, and (c) the linear combination of the 31 eigenvectors in SF-GAB model. Colors are based on the values of the eigenvector, for example, darker colors for larger values and lighter colors for smaller values. See Table 2 for the explanation of SF-GAB model.

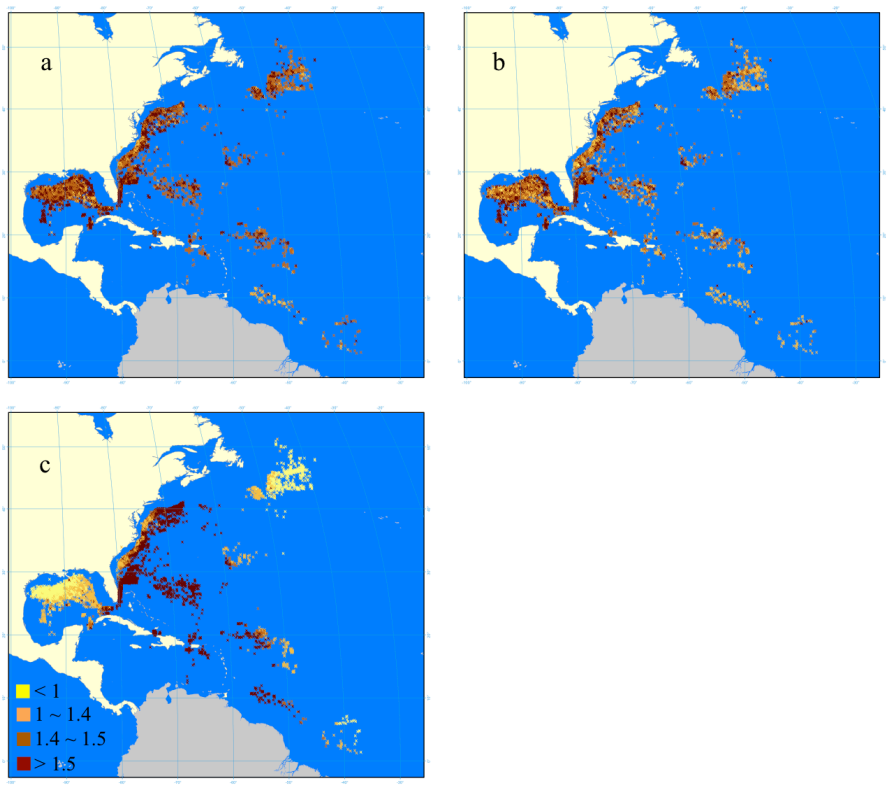

Figure 4. Spatial patterns of the positive catch of seabirds (in number) estimated by candidate models, i.e., (a) BL, (b) SF-GAB, and (c) TS. The estimates were medians from 100 iterations of bootstrapping. See Table 2 for the explanation of candidate models.

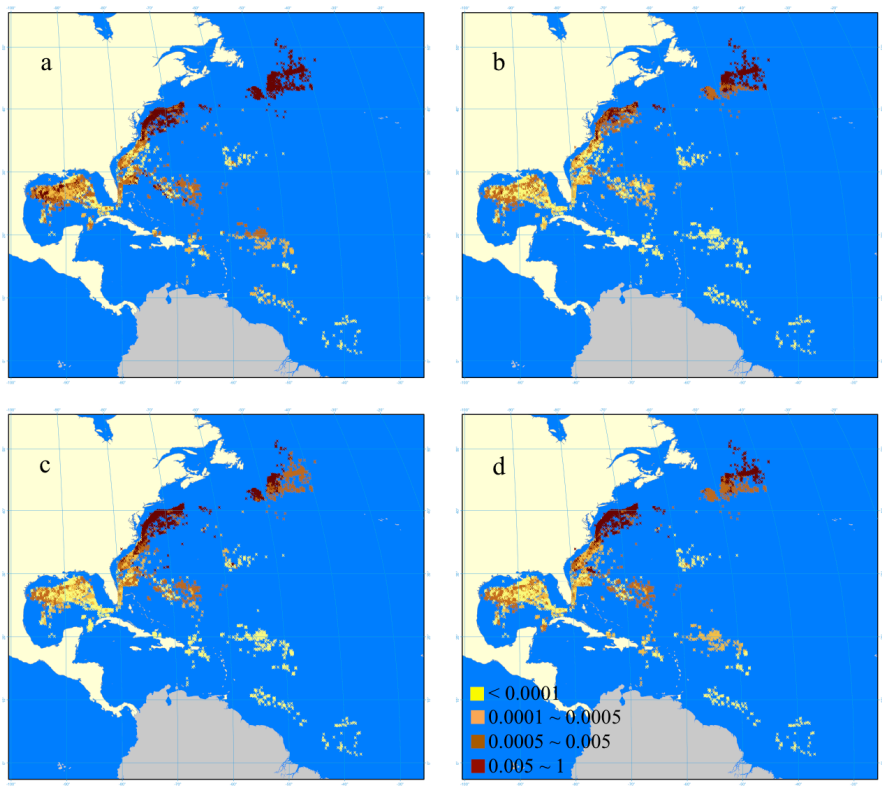

Figure 5. Spatial patterns of the probability of catching seabirds estimated by candidate models, i.e., (a) BL, (b) SF-GAB, (c) TS and (d) RE. The estimates were medians from 100 iterations of bootstrapping. See Table 2 for the explanation of candidate models.

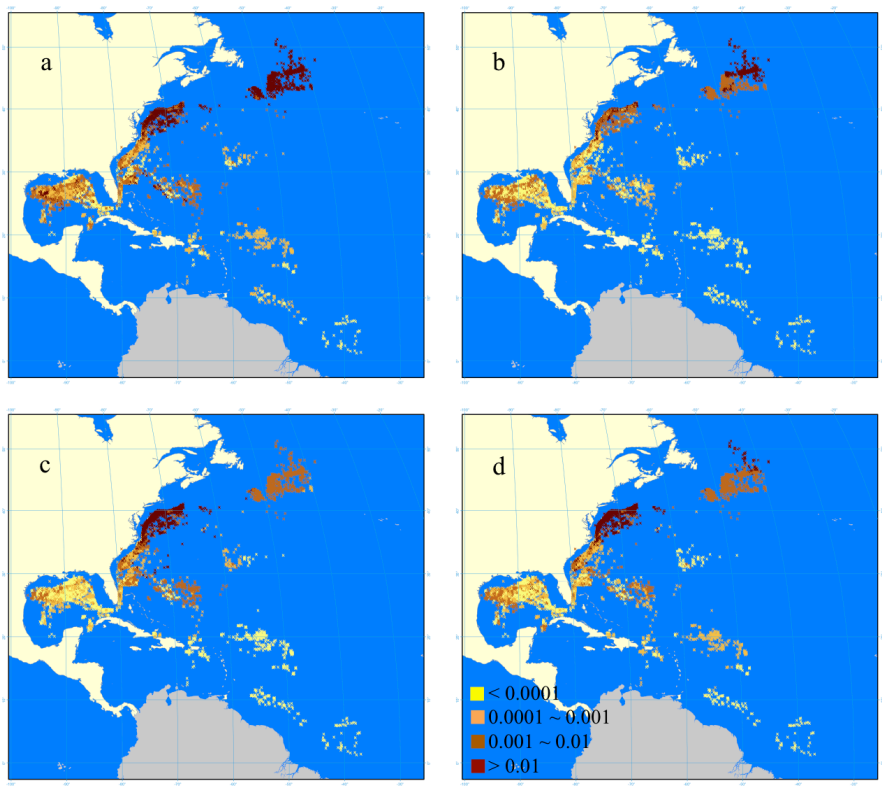

Figure 6. Spatial patterns of the seabird bycatch (in number) estimated by candidate models, i.e., (a) BL, (b) SF-GAB, (c) TS and (d) RE. The estimates were medians from 100 iterations of bootstrapping. See Table 2 for the explanation of candidate models.

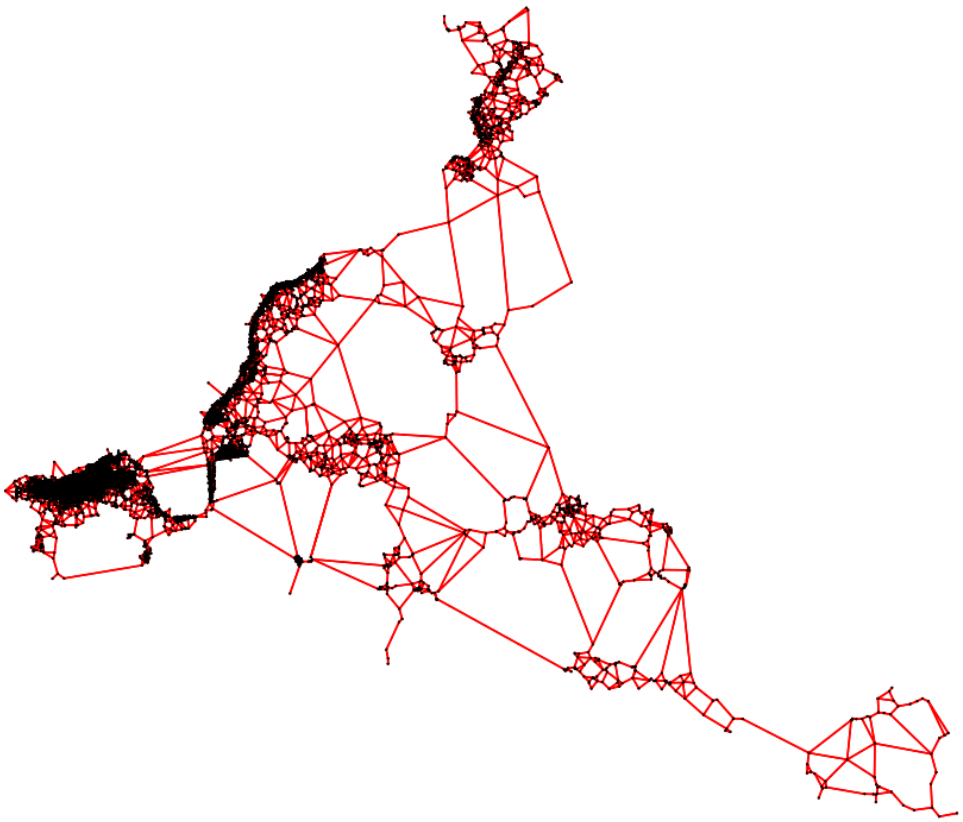

Figure 7. Connectivity matrix based on the Gabriel graph (GAB) that was used in the best spatial weighting matrix for seabird bycatch data.

Chapter 3

Exploring Spatial Nonstationarity Using Geographically Weighted Generalized Linear Model: An Example of Longline Seabird Bycatch

3.1 Abstract

Spatial nonstationarity describes the spatially varying relationships between biological and environmental processes. Geographically weighted regression (GWR) is a relatively new technique to explore spatial nonstationarity in ecological data, which allows parameters to vary across the study region. In the present study, I extended GWR to a geographically weighted generalized linear model (GW-GLM) by incorporating other statistical distributions in exponential family besides normal distribution. I demonstrated the application of GW-GLM with an empirical example of longline seabird bycatch. Due to high percentage of zero observations in the seabird bycatch data, I analyzed the positive catch rates and the probability of catching a seabird separately. I developed a GW-GLM with an assumption of normal distribution and identity link for analyzing the logarithm of searbird catch rates, which is equivalent to a GWR. I developed a GW-GLM with an assumption of binomial distribution and logit link for estimating the probability of catching a seabird. Estimates of parameters exhibited great spatial variation. I compared model performance of GW-GLM with a global generalized linear model (GLM) and a spatial expansion (SE) model that is an early technique to explore spatial nonstationarity by modeling each of the parameters with a function of locations. The GW-GLM performed the best. I conducted a permutation test for spatial nonstationarity. Target catch rate showed significant spatial variation when analyzing positive catch rates (p-value = 0.01). Relationships between the

probability of catching a seabird and environmental factors including intercept (p-value = 0.01), water depth (p-value << 0.0001) and water temperature (p-value = 0.01) varied significantly across the study region.

Keywords: Geographically weighted regression; geographically weighted generalized linear model; spatial expansion; spatial nonstationarity; spatial variation

3.2 Introduction

Understanding key relationships between biological processes and environmental factors is important in natural resources management and conservation, e.g., catch rate standardization using relationships between catch rate and environmental factors to estimate annual patterns of catch rate for fish stock assessment in fisheries (e.g., Stefansson, 1996; Maunder and Punt, 2004; Damalas et al., 2007); bycatch assessment using these relationships to predict hotspots for occurrence of a bycatch event (e.g., Murray, 2004; Li et al., 2011; Winter et al., 2011). A "global" model is commonly applied to examine the relationships between biological processes and environmental factors, which assumes these relationships to be stationary over the entire study area, i.e., constant and independent of location and direction. Given the dynamic spatial interactions between biological and environmental factors in natural ecosystems, the assumption of spatial stationarity may be too ideal to be realistic. Instead, spatial nonstationarity, by its narrow definition, describes the variation in relationships between biological and environmental factors over space (Brunsdon et al., 1996), and is more appropriate to capture the spatial distribution of species based on environmental factors.

One of the early techniques to explore spatial nonstationarity is the spatial expansion (SE) model (Casetti, 1972; Casetti and Jones III, 1992). In the SE model, each of the parameters

is itself a function of location. The nature of the function must be predetermined (e.g., linear or polynomial) based on prior knowledge. However, model results may be sensitive to the specification of expansion function (Charlton et al., 2009). Additionally, for the polynomial expansion function, computation and interpretation may become difficult when higher order polynomials are used, especially in the case of limited data.

A relatively new technique to address spatial nonstationarity in ecological data is the geographically weighted regression (GWR, Brunsdon et al., 1996; Fotheringham et al., 1998). In GWR, the neighboring observations are weighted based on the strength of the spatial dependence and thus drive parameters to vary across the study area. The GWR has several advantages in exploration of spatial nonstationarity. First, similar to other alternatives, GWR yields a set of estimates for each of the parameters that vary over space. The set of estimates can be mapped over the entire study area, which provides a visualization of the spatially varying relationships between biological and environmental variables. Second, GWR extends the traditional regression approach with spatially varying parameters. It can be fitted using the weighted least square (WLS) method, which is easier to operate compared to other alternative estimators such as the maximum likelihood method. Third, in GWR, spatial relationships among observations can be investigated by testing different spatial weighting matrices that combine information on neighbor structure and spatial dependence. However, GWR, by its original formulation (Brunsdon et al., 1996; Fotheringham et al., 1998), is limited to the normal assumption. In this study, I extended GWR to the geographically weighted generalized linear model (GW-GLM) by applying other distributions from the exponential family (i.e., the binomial and Poisson distributions) besides the normal distribution. The GW-GLM can be fitted using the iteratively reweighted least square method (IRLS, McCullagh and Nelder, 1989). The GW-GLM is also an extension of global

generalized linear model (GLM) by applying spatial weights in the IRLS procedure to consider spatial locations in parameter estimation. The GWR is a special case of GW-GLM in which a normal distribution and an identity link function are applied, and spatial weights are directly applied to observations.

In this study, I demonstrated the application of GW-GLM using an empirical example of longline seabird bycatch. The incidental mortality of seabirds (termed "bycatch") in longline fisheries has raised a global concern among marine conservationists because it has threatened several albatross species (Diomedeidae) and other species in the order of Procellariiformes (Brothers et al., 1999a; Tasker et al., 2000; Belda and Sanchez, 2001; Furness, 2003). Occurrence of seabird bycatch has been related to environmental variables such as wind speed (Klaer and Polacheck, 1998; Brothers et al., 1999b). Given the highly mobile nature of seabirds and the large heterogeneous habitat covered by longline fishing, occurrence of seabird bycatch is likely to be spatially nonstationary. Therefore, conservation and management of seabird bycatch will benefit from a more complete understanding of the relationships (e.g., the spatial nonstationarity) between population dynamics and environmental factors.

The purpose of this study was to extend the geographically weighted regression (GWR) model to the geographically weighted generalized linear model (GW-GLM) and to demonstrate the application of GW-GLM to explore spatial nonstationarity through an empirical example of longline seabird bycatch. I addressed two major questions in the study of spatial nonstationarity. First, is GW-GLM better than the global GLM and the spatial expansion (SE) model? Second, are relationships between biological processes and environmental factors in GW-GLM significantly nonstationary?

3.3 Methods

Geographically weighted generalized linear model (GW-GLM)

The GW-GLM model is an extension of the global generalized linear model (GLM) by considering locations in parameter estimation and thus allowing parameters to vary locally rather than being constant over space. A global GLM takes the form:

$$g\left(E\left[y_i\right]\right) = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} \,, \tag{1}$$

where $i = 1, 2,\ldots, n$ denotes the ith location, y is the observation, $E[y]$ is the expectation of y, β_0 is the intercept, β_k is parameter for the kth explanatory variable x_k and k = 1, 2,…, p, and $g(.)$ is the link function that describes relationship between $E[y]$ and explanatory variables. With the iterative reweighted least squares (IRLS) method, the parameters in the global GLM can be estimated through an iterative algorithm in which the weighted least squares method (WLS) is applied to the adjusted response variable $Z = [z_1, z_2,..., z_n]^T$ at each iteration step (McCullagh and Nelder, 1989):

$$\hat{\beta}^{new} = \left(X^T W^{old} X\right)^{-1} X^T W^{old} Z \,, \tag{2}$$

where T denotes the transpose of a matrix, $\hat{\beta}^{new} = \left[\hat{\beta}_0^{new}, \hat{\beta}_1^{new}, ..., \hat{\beta}_p^{new}\right]^T$ is a vector of estimated parameters that are updated at each iteration step, X is the design matrix defined as:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}. \tag{3}$$

The adjusted response variable z_i is definded as:

$$z_i = x_i^T \hat{\beta}^{old} + \left(y_i - E\left[y_i\right]\right)\left(G'\left(x_i^T \hat{\beta}^{old}\right)\right)^{-1}, \qquad (4)$$

where G(.) is the inverse function of g(.) and G'(.) denotes the first derivative of function G(.) with respect to parameters β. The $n \times n$ weighting matrix $W = [w_i]$ can be written as:

$$W = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_n \end{bmatrix}, \qquad (5)$$

and w_i is updated accordingly at eatch iteration step:

$$w_i^{new} = \frac{G'\left(x_i^T \hat{\beta}^{new}\right)^2}{V\left(E\left[y_i\right]\right)}, \qquad (6)$$

where V(.) is the score function that is defined as the first derivative of the log-likelihood with respect to parameters β. The whole algorithm progresses until parameter estimates converge. In the global GLM, a unity weight is assigned to each observation as the prior weights. By controlling prior weights, parameter estimation can be adjusted. The linear regression is a special case of GLM in which $G'(.) \equiv 1$ and $E[y_i]=x_i^T\beta$, and thus prior weights can be applied directly on observations and no iteration is necessary.

As an extension of the global GLM with locally varying parameters, the geographically weighted generalized linear model (GW-GLM) can be written as:

$$g\left(E\left[y_i\right]\right) = \beta_0\left(u_i, v_i\right) + \sum_{k=1}^{p} \beta_k\left(u_i, v_i\right) x_{ik}, \qquad (7)$$

where (u_i, v_i) denotes the coordinates of the ith location (longitude, latitude), and $\beta_0 (u_i, v_i)$ and $\beta_k (u_i, v_i)$ are parameters for the ith location. Same as most of the spatial analyses, GW-GLM

implicitly assumes that observations nearer to location *i* have more influence on the parameter estimation for this location than observations farther away from *i* (Cressie, 1993; Fotheringham et al., 1998). Thus, in a GW-GLM, parameters for the *i*th location are estimated using its neighboring observations through the IRLS method as described above except that prior weights are assigned to its neighboring observations according to their influence on the *i*th location. I denote the prior weight for the *j*th neighboring observation according to its influence on location *i* by w_{ij} and $j = 1, 2, \ldots, n$.

Same as the geographically weighted regression (GWR), in GW-GLM, the prior weight w_{ij} is assumed to be a function of distance between locations *i* and *j* (d_{ij}) and the weight decreases as the distance increases. This weighting function can take various forms (Brunsdon et al., 1996; Fotheringham et al., 1998). In this study, I applied the Gaussian weighting function that has been widely used in spatial analysis due to its continuity:

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{2h^2}\right), \tag{8}$$

where the pamameter *h* has great impact on model outcomes, i.e., smaller *h* leads to highly localized parameter estimates while larger *h* drives parameter estimates towards those from the global model (Brunsdon et al., 1996; Fotheringham et al., 1998). The optimal value of *h* can be determined by optimizing the selection criterion, e.g., minimizing the leave-one-out sum of squares (also termed the cross-validation score, Brunsdon et al., 1996; Fotheringham et al., 1998), or minimizing the Akaike Informaiton Criterion (AIC, Akaike, 1974; Fotheringham et al., 2002). The cross-validation score may not be the best choice for certain types of data such as the binary data that take values of either zero or one. The AIC works well for GW-GLM with an assumption of normal distribution and an identity link (i.e., GWR), but determination of the

effective number of parameters may be problematic for GW-GLM with other distributions. In the seabird bycatch example, to determine the optimal value of *h*, I minimized the AIC for the normal data (positive catch rates of seabird) and maximized the area under the receiver operating characteristic (ROC) curve (AUC) for the binary data (capture or no capture of a seabird). The ROC curve plots the true positive rate $\Pr(\hat{y} = 1 \mid y = 1)$ against the false positive rate $\Pr(\hat{y} = 1 \mid y = 0)$, and has long been used to evaluate the prediction accuracy of a classification algorithm (Bradley, 1997). Here modeling the binary data can be treated as a classification algorithm and a larger AUC value suggests higher discrimination accuracy of the model. In the determination of *h*, a very small value of *h* will drive the weights of all neighboring observations to be negligible except for *i*th location itself, which suggests that the fitted value for *i*th location will tend to be the observed value (Brunsdon et al., 1996; Fotheringham et al., 1998) and an AUC value close to one will be yielded. Thus, in the calculation of AUC, I estimated the probability for *i*th location using data excluding the observation at *i*th location (i.e., leave one out) to avoid this problem.

A permutation test was conducted to test for the significance of spatial nonstationarity (Brunsdon et al., 1996; Fotheringham et al., 1998). Under the null hypothesis of constant parameters over space, any permutaion of (u_i, v_i) pairs amongst the geographical sampling points *i* are equally likely to occur. Same as in GWR, in GW-GLM, there are *n* estimates for each of the parameters. The testing statistic is given by an estimate of variability in these *n* parameter estimates, i.e, its standard deviation.

Spatial expansion (SE) model

The original spatial expansion (SE) model is also an extension of the traditional regression model (Casetti, 1972; Casetti and Jones III, 1992). The SE model expands parameters to be a function of locations (u_i, v_i) so that parameter estimates vary over space and the spatial nonstationarity is explored. In this study, I employed a first order linear polynomial function as the expansion function due to limited data in the seabird bycatch example:

$$\begin{aligned} \beta_0(u_i, v_i) &= a_0 + a_1 u_i + a_2 v_i, \\ \beta_k(u_i, v_i) &= b_{0k} + b_{1k} u_i + b_{2k} v_i, \end{aligned} \tag{9}$$

where (a_0, a_1, a_2) and (b_{k0}, b_{k1}, b_{k2}) are parameters in the expansion function to determine β_0 and β_k for the ith location, respectively. Applying the spatial expansion technique to the global GLM and modifying the global model with such a linear expansion function, the SE model can be written as:

$$g(E(y_i)) = a_0 + a_1 u_i + a_2 v_i + \sum_{k=1}^{p} (b_{0k} + b_{1k} u_i + b_{2k} v_i) x_{ik}, \tag{10}$$

The SE model can also be easily fitted using the IRLS method with unity prior weights.

Model comparison

In this study, I compared the performance of GW-GLM with the global model and SE model. Again in the seabird bycatch example, I used the AIC and AUC as criterion for normal data and binary data respectively. The model providing a lower AIC value or a higher AUC value would be judged to have better performance.

Examples of longline seabird bycatch

In this study, I demonstrated the application of GW-GLM model to investigate spatial nonstationarity through an empirical example of longline seabird bycatch. The seabird bycatch data were provided by the National Marine Fisheries Service Southeast Fisheries Science Center Pelagic Observer Program (POP). The POP program was initiated for monitoring the United States Atlantic pelagic longline fishery. In the POP program, randomly selected fishing trips are attended by an observer, and the detailed information on gear specification, fishing effort, fishing tactics, target catch, bycatch and environmental conditions is recorded. The environmental variables that I analyzed included water depth (fathom), water temperature (Fahrenheit), target species catch rate (number per 1000 hooks), wind speed (knot), wind direction (degree) and wave height (feet). The POP data covered 11 fishing areas (Lee and Brown, 1998), and I constrained our analysis within the three areas with highest seabird bycatch, i.e., northeast coast, middle Atlantic bight and south Atlantic bight (Fig. 1). I analyzed 3,751 longline sets of the POP data during 1992 to 2012, of which only 58 sets caught 120 seabirds; therefore, approximately 98.5% zero observations occurred. I projected the geographic coordinates in the data (i.e., degree north and degree west) to the Universal Transverse Mercator (UTM) system zone 17 based on the NAD83 geodetic datum.

Due to the high percentage of zero observations in the POP data, I analyzed the positive catch rate (number per 1000 hooks) and the probability of capturing a seabird (p) separately, which is the same idea as the delta model or hurdle model that handles data with excess zeros (Aitchison, 1955; Lo et al., 1992; Fletcher et al., 2005). To model the probability of catching a seabird, I converted the original data into binary data, i.e., zero or one in which the value of zero represented the event of no seabird captured and the value of one represented the event of at least one seabird captured from a longline set. The global GLM for positive catch rate assumed

logarithm of the catch rate to follow a normal distribution with an identity link, which is equivalent to a GWR model:

ln(seabird catch rate) = intercept + water depth + water temperature + target species catch rate + wind speed + wind direction + wave height,

The global GLM for the probability of catching a seabird assumed catch event to follow a binomial distribution with a logit link:

$$\ln\left(\frac{p}{1-p}\right) = \text{intercept} + \text{water depth} + \text{water temperature} + \text{target species catch rate} + \text{wind speed} + \text{wind direction} + \text{wave height}.$$

To determine the value of *h* and to compare candidate models, I applied the AIC for the positive catch rate models and the leave-one-out AUC for the probability models. With the AIC criterion, I determined *h* = 7.24 (100 km) for the positive catch rate GW-GLM (AIC = 120.71), and with the AUC criterion, I determined *h* = 1.01 (100 km) for the probability GW-GLM (leave-one-out AUC = 0.693). This study was programed using R package 'spgwr'.

3.4 Results

Parameter estimates from the three candidate models are summarized in Tables 1 and 2, and the spatially varying parameter estimates from GW-GLM and the SE model are mapped in Figs. 2-5. These spatial patterns of parameters appeared to characterize the three fishing zones defined in the U.S. Atlantic pelagic longline fishery (Fig. 1). Estimates of parameters in the positive catch rate model showed great spatial variation (Figs. 2 and 3). Derived from both GW-GLM and the SE model, estimates of intercept and wave height were positive and appeared to become higher from north towards south of the study region. Estimates of target catch rate were

mostly negative and became more negative towards the north. In GW-GLM, estimates of water depth, water temperature and wind direction were all negative and exhibited a trend of higher values in terms of magnitude running from north towards south except for water depth which had a saddle in the middle (Fig. 2). In the SE model, estimates of water depth, water temperature and wind direction were mostly negative, with values being more negative from the middle of the study region spreading towards north and south (Fig. 3). Maps of wind speed derived from the GW-GLM and SE model both exhibited a trend of becoming more positive towards north but different in estimated values (Figs. 2 and 3). Estimates of wind speed from GW-GLM were dominated by negative values while those from the SE model were mostly positive. Overall, estimates from the SE model were higher in magnitude than those from GW-GLM when analyzing positive catch rates.

Estimates of parameters in the probability model also varied across the study area (Figs. 4 and 5). Estimates of intercept from both GW-GLM and the SE model were more positive in the middle and the south-most tip, and became more negative towards north and south. Estimates of target catch rate and wind direction from both models tended to become more positive towards north of the study region. Estimates of water depth and wind speed from the SE model showed a reverse trend of being more positive towards south (Fig. 5). Compared to the SE model, GW-GLM produced quite different maps for water depth and wind speed (Fig. 4). Estimates of water depth from GW-GLM had positive values concentrated in the middle and became more negative towards north and south. Estimates of wind speed from GW-GLM depicted a complicated trend in which more negative values were found in the middle, north-most and south-most tips, and more positive values lied in between. Estimates of water temperature and wave height from both GW-GLM and the SE model were dominated by positive values, and became larger in the north

and south and smaller in the middle, suggesting that the probability of catching a seabird was less sensitive to water temperature and wave height in the middle of study region. Overall, maps of parameter estimates derived from GW-GLM showed greater details than those from SE model. By contrast to the estimates in the positive catch rate model, estimates from SE model had a smaller magnitude than those from GW-GLM when estimating the probability of catching a seabird.

When analyzing positive catch rates, only target catch rate exhibited significant spatial variation (p-value = 0.01) at a significance level of 0.05 (Table 1). Relationships between the probability of catching a seabird and environmental factors including intercept (p-value = 0.01), water depth (p-value << 0.0001) and water temperature (p-value = 0.01) varied significantly across the study region (Table 2).

Among the three candidate models (Tables 1 and 2), GW-GLM performed the best when analyzing positive catch rates and when estimating the probability of catching a seabird, producing the smallest AIC value (120.71) and the highest AUC value (0.84) respectively. For analyzing positive catch rates, the global model outperformed the SE model with an AIC value 5.5% lower than the one in the SE model. However, for the probability of catching a seabird, the SE model showed better performance than the global model, yielding an AUC value 6% higher than the one in global model.

3.5 Discussion

The geographically weighted approach (e.g, GWR and GW-GLM) is a flexible tool to explore spatial nonstationarity. The spatial weighting matrix contains all information about spatial distribution of observations, and the weighting matrix can take various forms by assuming

different neighbor structure and distance-decay functions. Whatever a specific weighting matrix is employed, the essential idea of the geographically weighted approach is that for a location *i*, the observations closer to *i* have more influence on the estimation of parameters at location *i* than those observations farther away (Fotheringham et al., 1998). By controlling the weighting matrix, several other techniques can be generalized into the geographically weighted approach. For example, a weighting matrix with a weight of unity for all locations results in a global model, and a weighting matrix that assigns a weight of unity if locations *i* and *j* lie within a square and zero otherwise leads to the moving window method (Brunsdon et al., 1996; Fotheringham et al., 1997).

Besides the SE model, there are other techniques to explore spatial nonstationarity in the early development of the methodology. For example, the spatially adaptive filtering (Foster and Gorr, 1986) iteratively adjusts parameters based on neighboring observations. The random coefficient model (Rao, 1965; Swamy, 1971) and the multilevel modeling approach (Goldstein, 1987) allows parameters to randomly vary across the study area. The random coefficient model assumes parameters to follow a certain statistical distribution. The multilevel modeling approach uses hierarchies to represent variations between spatial regions and combines models for individual regions within an overall model. In this study, I decided to use the SE model as one of the candidate models for comparison because the SE model can be easily fitted using the same fitting method as GW-GLM and global GLM, i.e., the Iterative Reweighted Least Squares (IRLS), which may reduce the bias in comparison due to different fitting methods.

Although various discrete weighting functions are also available, they should be used with great caution. First, discrete functions may cause a sudden change in parameter estimates, which may barely occur given the interactions among neighboring locations. Second, with

discrete weighting functions, localized parameter estimates may become sensitive to the observations at the cutting edge, i.e., parameter estimates may change dramatically when one observation is determined to include as a neighbor or not (Brunsdon et al., 1996). Third, use of discrete weighting functions may introduce the edge effect in which localized parameter estimates near edges suffer large variance when too few observations are determined to be neighbors (Brunsdon et al., 1996).

In this study, I did not include categorical variables as environmental factors when exploring spatial nonstationarity. Same as in GWR, categorical variables could be redundant locally when categories are spatially clustered or some categories are missing in some areas. Zhang et al. (2011) suggested to avoid using categorical variables that have values spatially clustered or to combine and reduce the number of categories, or to combine with other variables. Additionally, interpretation and mapping of the localized estimates of a categorical variable is also problematic.

Results showed that GW-GLM and the SE model produced parameter estimates that were quite different in magnitude. For example, when analyzing positive catch rates, estimates of intercept from GW-GLM doubled those from the SE model. Fortheringham et al. (1998) also revealed that GWR produced different parameter estimates in terms of magnitude from the expansion method, and even with the expansion method, the parameter estimates were different between linear expansion and quadratic expansion. Presumably, these differences in parameter estimates between candidate models is caused by the different ways of modeling spatial nonstationarity with different complexities. This finding highlights the importance of determining appropriate models to explore spatial nonstationarity, and the determination of such

models may depend on dataset, selection criterion, and the number of candidate models for comparison.

The parameter *h* has a great impact on model outcomes. Smaller values of *h* result in highly localized parameter estimates and larger values of *h* drives model towards the global model (Brunsdon et al., 1996; Fotheringham et al., 1998). However, I found that determination of *h* is highly sensitive to the selection criterion. For example, when analyzing the positive catch rates, I determined $h = 7.24$ according to the AIC value in this study while a value of 5.17 would be determined if the cross-validation sum of squares (Brunsdon et al., 1996) is used as the selection criterion. In this study, I determined $h = 1.01$ for the probability model based on the leave-one-out AUC value while the cross-validation sum of squares will give a value of 7.06. Thus, determination of the parameter *h* needs to be done with caution and the ecology of the studied species would be very helpful in this case.

In the positive catch rate model, only one parameter (i.e., target catch rate) showed significant spatial variation (*p*-value = 0.01, Table 1). The few positive observations in the seabird data may have limited the detection of spatial nonstationarity in positive catch rates.

Spatial patterns in the biological-environmental relationships shown in Figs. 2-5 appear to characterize fishing zones defined by the U.S. Atlantic pelagic longline fishery (Fig. 1), i.e., the three fishing zones in our study region (northeast coast, middle Atlantic bight and south Atlantic bight). The distinct oceanographical features in these three regions may have shaped the spatial patterns in biological-environmental relationships. For example, the south Atlantic bight is a wide and shallow shelf geomorphologically isolated from Middle Atlantic bight by Cape Hatteras and the Gulf Stream (Stefansson et al., 1971; Atkinson and Menzel, 1985). It has a water depth of 50m at shelf break, which is one third of the water depth at the middle Atlantic

Bight (Emery and Uchupi, 1972). It covers large latitudinal range from subtropical to temperate, which results in great latitudinal variation in climate and weather-related physical processes. Its physics, chemistry and biology are greatly influenced by the Gulf Stream (Atkinson and Menzel, 1985). Stream flow per length of coastline in middle Atlantic bight was estimated to be two times that in south Atlantic bight and northeast coast (Bumpus, 1973). Flow on continental shelf at northeast coast is tide-dominated year round whereas wind stress or buoyant outflow, depending on time of year, is the major force driving continental shelf flow at middle and south Atlantic bights (Epifanio and Garvine, 2001). This result suggests that longline fisheries management has a distinct biological-environmental relationship within each fishing zone, and these spatially varying relationships need to be taken into account when developing seabird bycatch reduction strategies. It seems that the current fishing zones defined for the U.S Atlantic longline fishery may also appropriately serve as management zones for seabird bycatch reduction because they match the zones identifying distinct relationships between seabird-bycatch and environmental factors.

In this study, I analyze the positive catch rates and the probability of catching a seabird in two separate models due to the high percentage of zero observations. This approach is the same as the delta model/hurdle model except that in the delta model, final estimates of the response variable are the product of the estimates from these two sub-models (Aitchison, 1955; Lo et al., 1992; Fletcher et al., 2005). To extend the application of GW-GLM to the two-step delta model, at least two questions need to be addressed. The first question is how to extrapolate the localized parameter estimates derived from the positive data to the entire study region. This extrapolation can be easily achieved by the SE model because each parameter is a function of locations. The second question is how to determine the parameter h for the entire delta model. In this study, I

determined values of *h* for the positive catch rate model and the probability model separately. However, in the delta model, a value of *h* to balance the selection criteria in both sub-models simultaneously is needed. One possible way may be to maximize the sum of log-likelihoods of both sub-models.

The current study can be extended in several ways. First, I did not test the sensitivity of parameter estimation to the choice of weighting matrix, which is necessary to fully understand the geographically weighted approach. Second, the weighting function can also vary over space (Brunsdon et al., 1996), which suggests a direction for future exploration of the methodology. Third, for a given weighting function, e.g., the Gaussian weighting function that I used in this study, I applied a global estimate of the parameter *h*, which assumed the values of *h* is constant over space. However, it is very likely that *h* also shows spatial variation. Thus, localized *h* estimates is worthwhile to explore (Brunsdon et al., 1996; Fotheringham et al., 1998). Forth, in this study, I fixed the set of environmental variables over space, but the set may also vary over space, i.e., one environmental variable may be important at one location but not at another location (Brunsdon et al., 1996). Localized sets of environmental variables may provide another direction for future exploration of the geographically weighted approach.

In conclusion, the geographically weighted approach is a flexible tool to explore spatial nonstationarity in ecological data. As an extension of the traditional generalized linear model (GLM), the geographically weighted generalized linear model (GW-GLM) allows parameter estimates to vary over space by assigning prior weights to observations during model fitting procedure. In the meanwhile, as an extension of the geographically weighted regression (GWR), GW-GLM models data with other statistical distributions in the exponential family beside normal distribution such as the binomial distribution in our seabird bycach example. With the

empirical example of longline seabird bycatch, I demonstrated the application of GW-GLM, and revealed great spatial variations in the relationships between seabird bycatch process and environmental factors. Compared with the global GLM and the spatial expansion (SE) method, GW-GLM performed the best in terms of goodness-of-fit and prediction.

3.6 Acknowledgements

This research was supported by the USDA Cooperative State Research, Education and Extension Service, Hatch project #0210510 to Dr. Y. Jiao, and a grant for spatial and temporal analysis and prediction of seabird bycatch of US Atlantic pelagic longline fleet awarded by the NOAA Southeast Fisheries Science Center. I also would like to thank Dr. Joan Browder from NOAA Southeast Fisheries Science Center for providing data-related information and valuable comments.

3.7 References

Aitchison, J., 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. Journal of the American Statistical Association 50, 901-908.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-723.

Atkinson, L.P., Menzel, D.W., 1985. Introduction: Oceanography of the southeast United States continental shelf. In: Atkinson, L.P., Menzel, D.W., Bush, K.A. (Eds), Oceanography of the southeastern US continental shelf. American Geophysical Union, Washington, D.C.

Belda, E., Sanchez, A., 2001. Seabird mortality on longline fisheries in the western Mediterranean: factors affecting bycatch and proposed mitigating measures. Biol Conserv 98, 357-363.

Bradley, A., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition 30, 1145-1159.

Brothers, N., Cooper, J., Løkkeborg, S., 1999a. The incidental catch of seabirds by longline fisheries: worldwide review and technical guidelines for mitigation. FAO Fisheries Circular No. 937, Rome.

Brothers, N., Gales, R., Reid, T., 1999b. The influence of environmental variables and mitigation measures on seabird catch rates in the Japanese tuna longline fishery within the Australian Fishing Zone, 1991-1995. Biol Conserv 88, 85-101.

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical analysis 28, 281-298.

Bumpus, D.F., 1973. A description of the circulation on the continental shelf of the east coast of the United States. Prog Oceanogr 6, 111-157.

Casetti, E., 1972. Generating Models by the Expansion Method: Applications to Geographical Research. Geographical analysis 4, 81-91.

Casetti, E., Jones III, J.P., 1992. Applications of the expansion method. Routledge, London, UK.

Charlton, M., Fotheringham, S., Brunsdon, C., 2009. Geographically weighted regression. White paper National Centre for Geocomputation National University of Ireland Maynooth.

Cressie, N., 1993. Statistics for spatial data. Wiley, New York.

Damalas, D., Megalofonou, P., Apostolopoulou, M., 2007. Environmental, spatial, temporal and operational effects on swordfish (*Xiphias gladius*) catch rates of eastern Mediterranean Sea longline fisheries. Fish Res 84, 233-246.

Emery, K.O., Uchupi, E., 1972. Western North Atlantic Ocean: topography, rocks, structure, water, life, and sediments. American Association of Petroleum Geologists, Tulsa, Oklahoma.

Epifanio, C., Garvine, R., 2001. Larval transport on the Atlantic continental shelf of North America: a review. Estuar Coast Shelf Sci 52, 51-77.

Fletcher, D., MacKenzie, D., Villouta, E., 2005. Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. Environ Ecol Stat 12, 45-54.

Foster, S.A., Gorr, W.L., 1986. An adaptive filter for estimating spatially-varying parameters: Application to modeling police hours spent in response to calls for service. Management Science 32, 878-889.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons, Chichester, England.

Fotheringham, A.S., Charlton, M., Brunsdon, C., 1997. Two techniques for exploring non-stationarity in geographical data. Geographical Systems 4, 59-82.

Fotheringham, A.S., Charlton, M.E., Brunsdon, C., 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment and Planning A 30, 1905-1927.

Furness, R., 2003. Impacts of fisheries on seabird communities. Scientia Marina 67, 33-45.

Goldstein, H., 1987. Multilevel models in education and social research. Oxford University Press, London, UK.

Klaer, N., Polacheck, T., 1998. The influence of environmental factors and mitigation measures on by-catch rates of seabirds by Japanese longline fishing vessels in the Australian region. Emu 98, 305-316.

Lee, D.W., Brown, C.J., 1998. Southeast Fisheries Science Center pelagic observer program data summary for 1992-1996. NOAA Technical Memorandum NMFS-SEFSC-408. National Marine Fisheries Service Southeast Fisheries Science Center, Miami, FL.

Li, Y., Jiao, Y., Reid, K., 2011. Assessment of landed and non-landed by-catch of walleye, yellow perch and white perch from the commercial gillnet fisheries of Lake Erie, 1994–2007. J Great Lakes Res 37, 325-334.

Lo, N., Jacobson, L., Squire, J., 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can J Fish Aquat Sci 49, 2515-2526.

Maunder, M., Punt, A., 2004. Standardizing catch and effort data: a review of recent approaches. Fish Res 70, 141-159.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman and Hall, London, UK.

Murray, K., 2004. Magnitude and distribution of sea turtle bycatch in the sea scallop (*Placopecten magellanicus*) dredge fishery in two areas of the Northwestern Atlantic Ocean, 2001-2002. Fish Bull 102, 671-681.

Rao, C.R., 1965. The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika 52, 447-458.

Stefansson, G., 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES J Mar Sci 53, 577-581.

Stefansson, U., Atkinson, L.P., Bumpus, D.F., 1971. Hydrographic properties and circulation of the North Carolina shelf and slope waters. Deep Sea Research 18, 383-420.

Swamy, P., 1971. Statistical inference in random coefficient regression models. Springer-Verlag, New York.

Tasker, M., Camphuysen, C., Cooper, J., Garthe, S., Montevecchi, W., Blaber, S., 2000. The impacts of fishing on marine birds. ICES Journal of Marine Science: Journal du Conseil 57, 531.

Winter, A., Jiao, Y., Browder, J.A., 2011. Modeling low rates of seabird bycatch in the U.S. longline fishery. Waterbirds 34, 289-303.

Zhang, C., Tang, Y., Xu, X., Kiely, G., 2011. Towards spatial geochemical modelling: Use of geographically weighted regression for mapping soil organic carbon contents in Ireland. Appl Geochem 26, 1239-1248.

Table 1

Results of the models for analyzing positive catch rate data. Three candidate models include the global generalized linear model (GLM), the spatial expansion model (SE) and the geographically weighted generalized linear model (GW-GLM). *u* and *v* index the parameters expanding along longitude and latitude in the SE model respectively.

Parameter	Global GLM	SE	GW-GLM	*p*-value for nonstationarity test
Intercept	1.9624	31.55	1.938 (1.7179, 2.4962)	0.34
Intercept (*u*)		0.8099		
Intercept (*v*)		-0.9169		
Water depth	-2.10×10^{-4}	-0.0038	-2.17×10^{-4} (-2.23×10^{-4}, 2.16×10^{-4})	0.99
Water depth (*u*)		-0.0001		
Water depth (*v*)		0.0001		
Water temperature	-0.0102	-0.4245	-0.0092 (-0.0162, -0.0061)	0.32
Water temperature (*u*)		-0.0101		
Water temperature (*v*)		0.0126		
Target catch rate	-0.0029	0.0963	-0.0042 (-0.0075, -0.0019)	0.01

Target catch rate (u)		3.67×10^{-4}		
Target catch rate (v)		-0.0028		
Wind speed	-0.0055	0.1143	-0.0081 (-0.0142, -0.0013)	0.52
Wind speed (u)		0.0095		
Wind speed (v)		-0.0057		
Wind direction	-0.0020	-0.0298	-0.0019 (-0.0027, -0.0015)	0.12
Wind direction (u)		-4.36×10^{-4}		
Wind direction (v)		7.88×10^{-4}		
Wave height	0.0617	1.6720	0.0689 (0.0454, 0.0873)	0.38
Wave height (u)		0.0278		
Wave height (v)		-0.0470		
AIC	133.2	141	120.71	

Table 2

Results of the models for estimating the probability of catching a seabird. See the caption of Table 1 for explanation of candidate models.

Parameter	Global GLM	SE	GW-GLM	*p*-value for nonstationarity test
Intercept	0.8305	-24.89	0.4946 (-15.5767, 3.5406)	0.01
Intercept (u)		-2.1		
Intercept (v)		1.2030		
Water depth	-6.16×10^{-5}	9.23×10^{-4}	-6.78×10^{-4} (-0.0107, 8.07×10^{-4})	<< 0.0001
Water depth (u)		-1.90×10^{-5}		
Water depth (v)		-2.16×10^{-5}		
Water temperature	-0.0771	0.2765	-0.0523 (-0.1155, 0.1452)	0.01
Water temperature (u)		0.0310		
Water temperature (v)		-0.0172		
Target catch rate	0.0061	-0.0923	0.0126 (-0.0036, 0.0247)	0.49
Target catch rate (u)		-0.0036		

Target catch rate (v)		0.0035		
Wind speed	0.0004	0.1270	-0.0083 (-0.2535, 0.0429)	0.23
Wind speed (u)		-0.0053		
Wind speed (v)		-0.0014		
Wind direction	0.0020	-0.0315	0.0012 (-0.0187, 0.0078)	0.11
Wind direction (u)		-2.79×10^{-4}		
Wind direction (v)		8.72×10^{-4}		
Wave height	0.0158	0.7646	0.0161 (-0.1990, 0.3745)	0.5
Wave height (u)		0.0278		
Wave height (v)		-0.0264		
AUC	0.69	0.73	0.84	

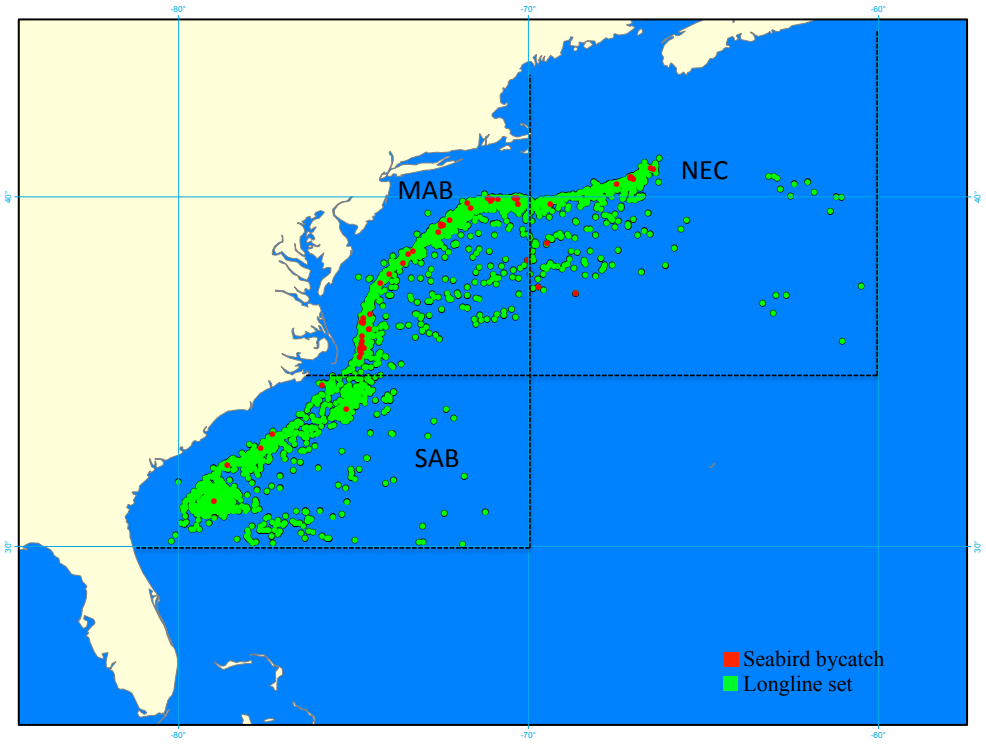

Figure 1. Spatial distribution of longline set and seabird bycatch in the study area. NEC-northeast coast, MAB-middle Atlantic bight, SAB-south Atlantic bight.

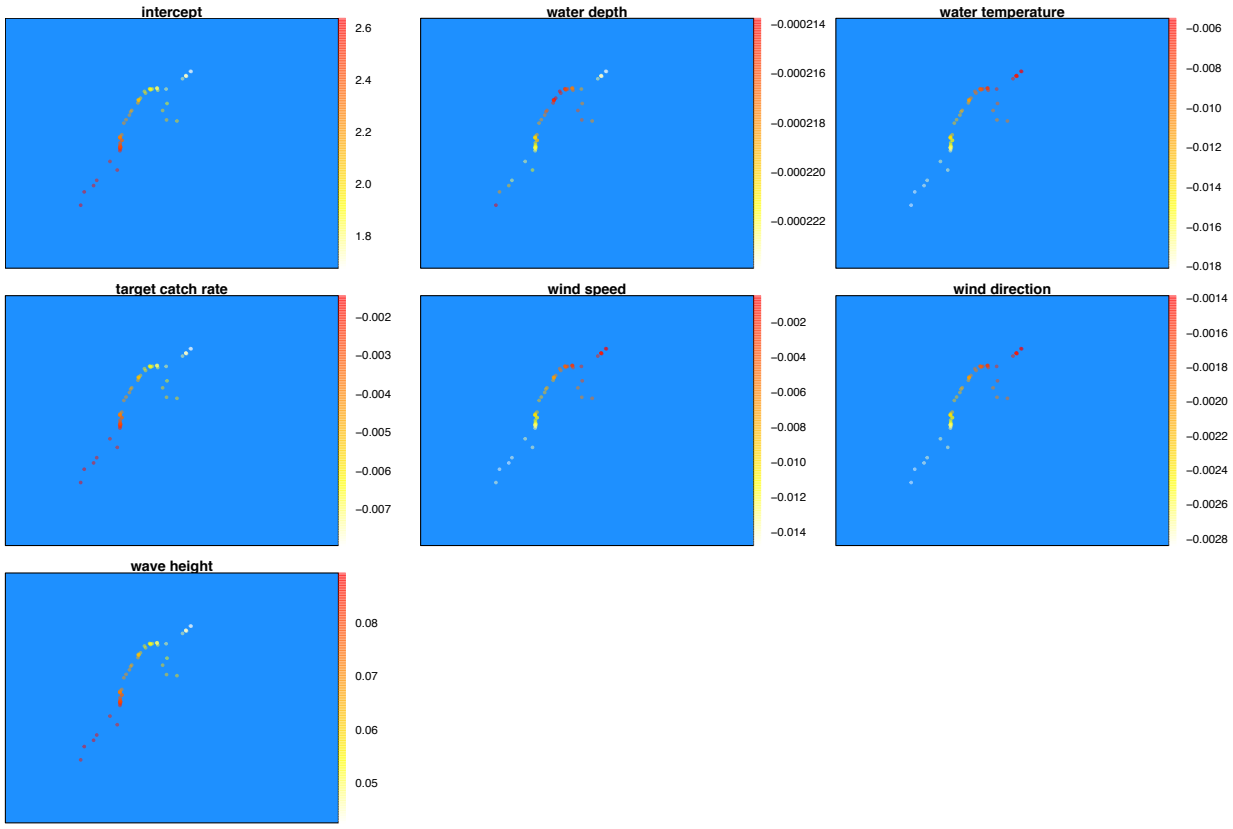

Figure 2. Spatial distribution of the parameters from the geographically weighted generalized linear model (GW-GLM) for analyzing positive catch rate data.

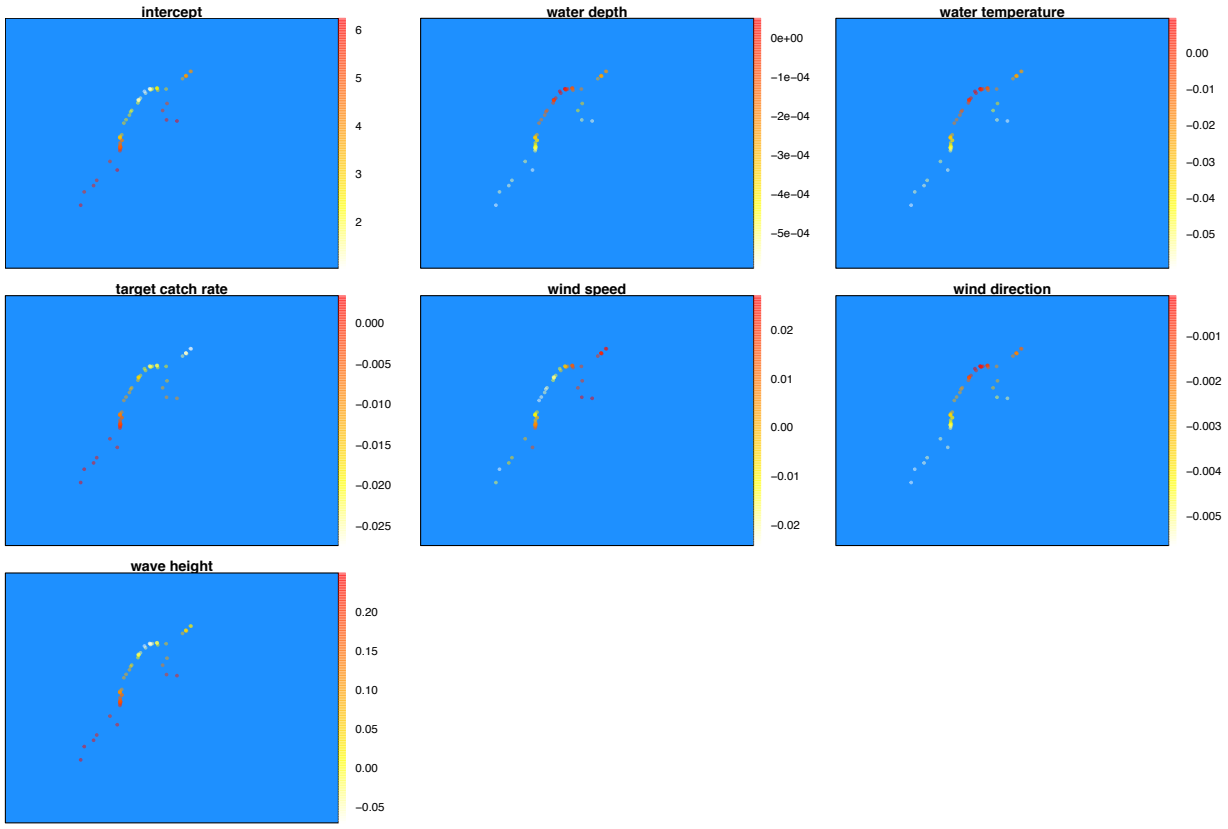

Figure 3. Spatial distribution of the parameters from the spatial expansion model (SE) for analyzing positive catch rate data.

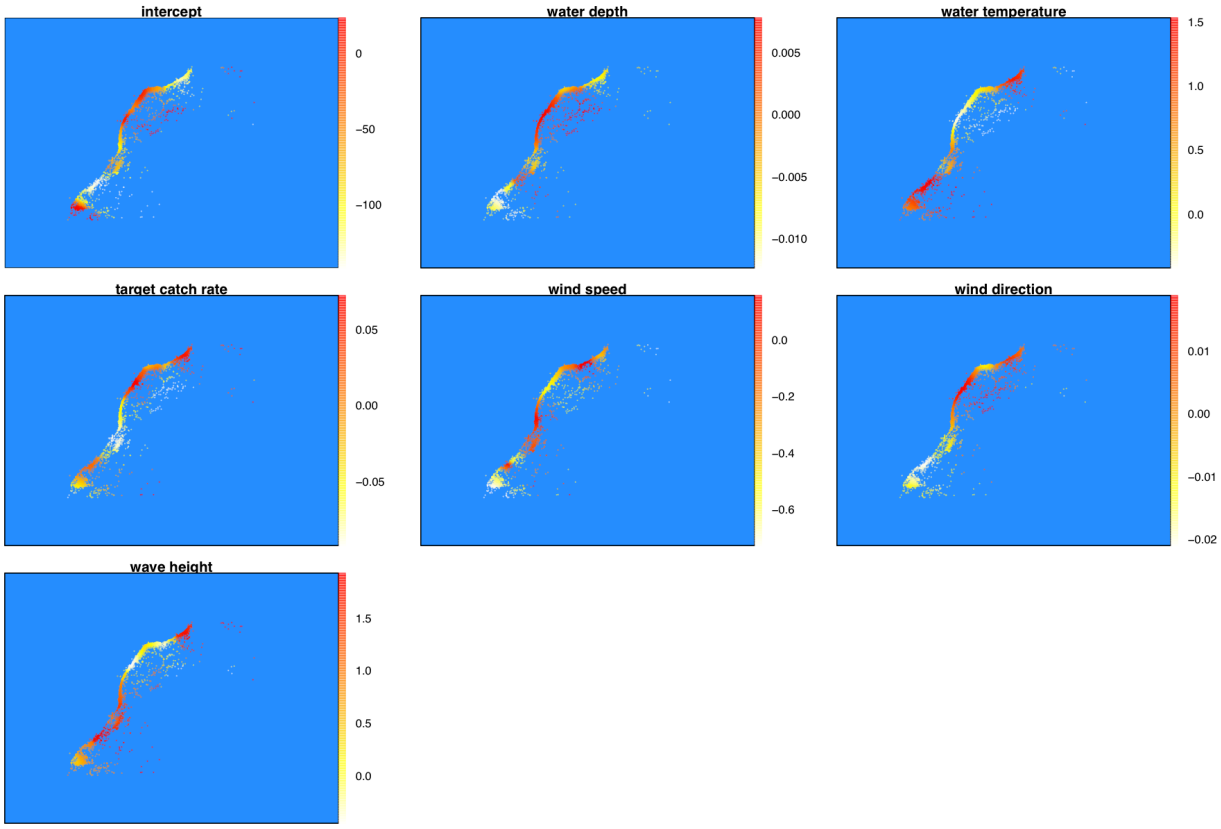

Figure 4. Spatial distribution of the parameters from the geographically weighted generalized linear model (GW-GLM) for estimating the probability of catching a seabird.

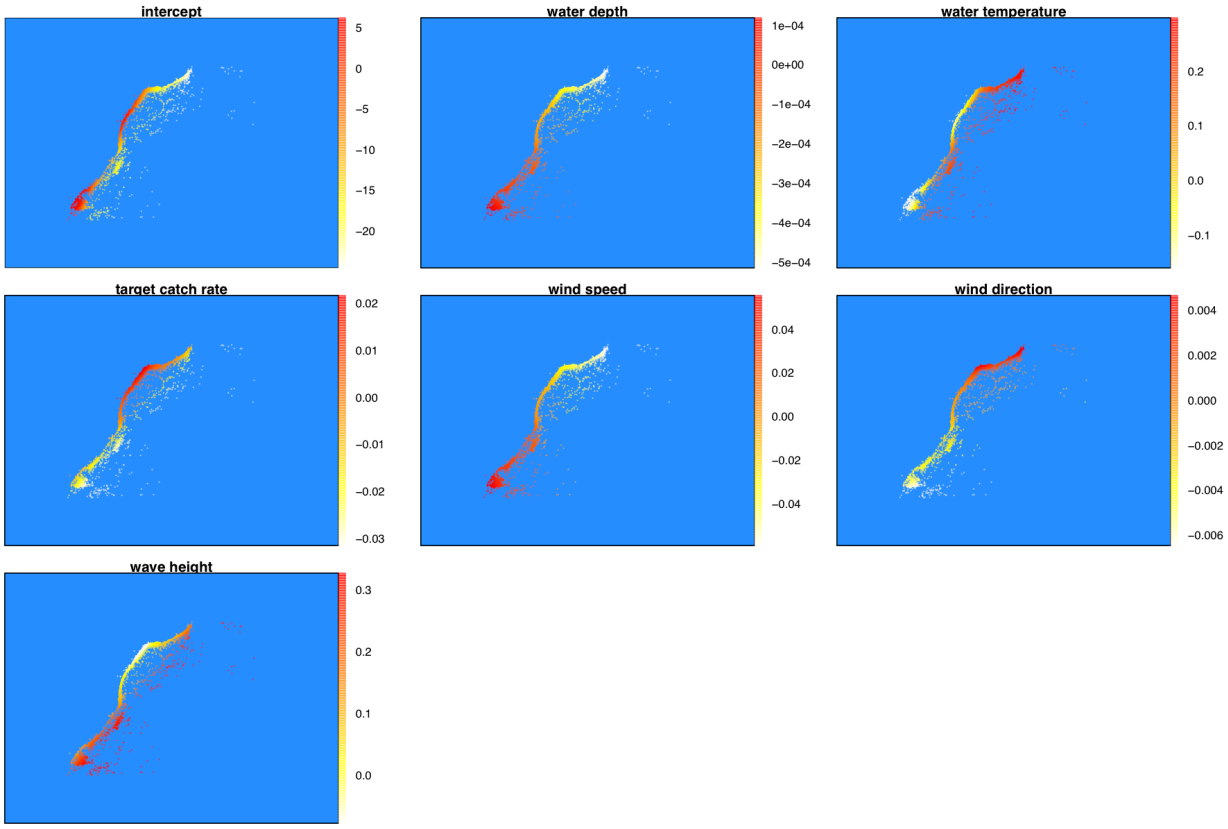

Figure 5. Spatial distribution of the parameters from the spatial expansion model (SE) for estimating the probability of catching a seabird

.

Chapter 4

Risk Assessment for Endangered White Abalone Using A Spatially Explicit Hierarchical Demographic Model

4.1 Abstract

Risk assessment of endangered white abalone (*Haliotis sorenseni*) has been challenging because of limited data. The present study demonstrated an application of the spatially explicit hierarchical demographic model for investigating population dynamics of white abalone and evaluating efficacy of hypothetical restoration strategies. The model represented demographic parameters as multi-level hierarchies, which accounted for uncertainty in parameter estimation due to limited data, and accounted for individual/sub-population variability in demographic traits in response to spatiotemporal heterogeneity. The model also accounted for the Allee effect (i.e., density-dependent fertilization success) observed in white abalone populations. In the calculation of fertilization success, I applied spatially explicit density estimates from three regions (Tanner Bank, Cortes Bank and San Clemente Island) to demonstrate an application of the spatially explicit approach in demographic models. Elasticity analyses showed that survival affected population growth rate more than fertility, and fertility and survival of large-size individuals (> 130 mm) had the most influence on growth rate; the mean elasticity to fertility decreased while the elasticity to survival increased as the population became less abundant. Sensitivity analyses showed that growth rate was sensitive to most parameters but to different degrees, and growth rate in the less abundant region (e.g., San Clemente Island) showed lower sensitivity than the growth rate in the more abundant region (e.g., Tanner Bank); growth rate was also sensitive to

determination of size categories; hierarchical models yielded similar estimates as non-hierarchical models, but had larger variance. Risk assessment suggested that restoration efforts directed at larger-size individuals may be more effective in increasing population density than efforts focusing on juveniles. The degree of increase depended on both initial population density and initial size structure; the increase became more prominent in regions with lower initial density than those with higher initial density when initialized with stable size structure. However, when initialized with surveyed size structure at Tanner Bank, the pattern reversed. However, stocking large-size individuals led to a decrease in the intermediate-size class (90-130 mm) proportion in most cases; efficacy of increasing intermediate-size class proportion heavily depended on initial population density. I suggest that restoration efforts may need to consider the spatially explicit density-dependence effect, and to balance stocking efficacy and laboratory expenses/time expenditure.

Keywords: white abalone; spatially explicit hierarchical demographic model; elasticity; sensitivity; risk assessment

4.2 Introduction

White abalone (*Haliotis sorenseni*) is the first marine invertebrate protected under the Endangered Species Act (Haaker, 1998; Hobday and Tegner, 2000). It is native to sand-surrounding rock or boulder habitat at depths of 20-60 m along the west coast of California and the northwest coast of Mexico (Shepherd, 1973; Tutschulte, 1976; Davis et al., 1996; Hobday and Tegner, 2000). It is a species with long life span (40 years, Cox, 1960), slow movement, external fertilization, and highly variable recruitment rates (Bartsch, 1940; Cox, 1960; Leighton, 1972; Hobday and Tegner, 2000). The white abalone fishery in California began in 1968, peaked

in 1972 and collapsed by 1978, driven by intensive exploitation and other factors such as environmental changes, diseases, hybridization, accidental mortality, predation and competition (Lundy, 1997; Hobday and Tegner, 2000). The abundance of white abalone has declined by greater than 99%, compared to the estimated pre-exploitation abundance (Hobday and Tegner, 2000). The most recent survey between 2002 and 2010 at Tanner Bank revealed dramatic and continued decline in both abundance and density, along with an increased proportion of large individuals and the vast majority being singletons (Stierhoff et al., 2012). At low density, white abalone populations may suffer from the Allee effect (Allee, 1931; Stoner and Ray-Culp, 2000; Gascoigne and Lipcius, 2004) because external fertilization success requires high densities of sperm and eggs provided by aggregation of remnant individuals (Cox, 1960; Hobday and Tegner, 2000; Stierhoff et al., 2012).

As the deepest-living and the latest harvested species among all abalone species along the west coast, the white abalone has been poorly studied (Hobday and Tegner, 2000). As concerns about its extinction arise, researchers and fishery managers have been taking actions to collect more information and to develop recovery plans for restoring the white abalone population. However, limited biological and ecological data complicate the assessment of its current status and extinction risk. Natural recovery without intervention seems unlikely to occur for white abalone (Stierhoff et al., 2012), and thus risk assessment under different restoration strategies, particularly under different stocking protocols will be essential. However, such risk assessment has been deficient relative to work on other endangered species. Rogers-Bennett and Leaf (2006) developed a deterministic demographic matrix model for white abalone and evaluated the impacts of recovery actions and potential threats on population growth, but did not account for uncertainty due to limited data, Allee effect or spatial heterogeneity. Thus, a risk assessment that

fully utilizes available information and accounts for uncertainty in parameter estimation due to limited data would better inform conservation of white abalone.

Demographic models are commonly used to understand population dynamics and to evaluate management tactics for threatened and endangered species (Beissinger and Westphal, 1998; Quinn and Deriso, 1999; Caswell, 2001). This type of model requires demographic information about the species. By representing demographic parameters as multi-level hierarchies, the hierarchical approach can incorporate demographic and environmental stochasticity by adding greater uncertainty to the model (Clark, 2003; Gelman et al., 2004; Jiao et al., 2009a; Jiao et al., 2009b). In the hierarchical model, each of the parameters that govern the probability distribution in the lower-level hierarchy is further determined by a probability distribution governed by parameters in the higher-level hierarchy (hereafter, termed "hyper parameters") (Clark, 2003; Gelman et al., 2004). Hierarchical models are preferred in cases where we have limited data or knowledge about the species, or we believe that the species has strong variability in population dynamics across individuals, sub-populations, regions and years (Jiao et al., 2009b), which is likely the case for white abalone. Ignorance of this variability in parameter estimation might lead to incomplete understanding of the population dynamics of the species and misleading management recommendations (Jiao et al., 2009b).

The spatially explicit approach is particularly appropriate for white abalone because its fertilization success is seriously hindered by low density due to the Allee effect (Clavier, 1992; Babcock and Keesing, 1999; Hobday and Tegner, 2000; Stierhoff et al., 2012), and its density is very likely to show spatial heterogeneity due to its limited movement (Leighton, 1972; Davis et al., 1998; Hobday and Tegner, 2000). After fertilization, the planktonic larvae of abalone disperse to suitable habitat for settlement. Larval dispersal has been shown to be very limited for

several Australian abalone species (Prince et al., 1987; McShane et al., 1988) and green abalone in California (Tegner and Butler, 1985a), but no information about larval dispersal distance exists for white abalone (Hobday and Tegner, 2000). The most recent survey using multibeam sonar mapping techniques revealed a significant difference in densities of white abalone among two offshore banks (Tanner Bank and Cortes Bank) and one island (San Clemente Island) off California coast (Butler et al., 2006). Additionally, its historic commercial catch data have shown spatial heterogeneity (Karpov et al., 2000).

The purpose of this study was to: (1) develop a spatially explicit hierarchical demographic model for white abalone; (2) conduct sensitivity and elasticity analyses to identify critical size classes and demographic parameters for population growth; (3) evaluate efficacy of hypothetical restoration strategies to aid in management of white abalone and to prioritize recovery efforts.

4.3 Methods

Hierarchical size-structured demographic model

I assumed a pre-breeding population for white abalone and applied the matrix model (Leslie, 1945; Caswell, 2001) because it has long been used to forward project population dynamics, to estimate the population growth rate and to determine the stable age or size structure and to conduct population viability analysis (Quinn and Deriso, 1999; Caswell, 2001). I constructed the matrix model based on size structure instead of age structure due to the difficulty and uncertainty in ageing white abalone and the distinct demographic characteristics across size classes. The size-structured matrix model comprises two components, the projection matrix and the initial population structure:

$$N_{y+1} = AN_y,$$

where $N_y = \left\{N_{y,1}, N_{y,2},..., N_{y,v}\right\}$ is a vector of size-specific population abundance at size class one to size class v in year y, in which v indexes the largest size class that I predetermined. The matrix A represents the projection matrix with a size of v x v. The projection matrix A takes the form:

$$A = \begin{bmatrix} s_1 p_{1\to 1} + e_1 & e_2 & \cdots & e_{v-1} & e_v \\ s_1 p_{1\to 2} & s_2 p_{2\to 2} & \cdots & 0 & 0 \\ s_1 p_{1\to 3} & s_2 p_{2\to 3} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_1 p_{1\to v} & s_2 p_{2\to v} & \cdots & s_{v-1} p_{(v-1)\to v} & s_v p_{v\to v} \end{bmatrix},$$

where e_i, $i = 1, 2, \ldots, v$, is the size-specific fertility every year, i indexes the ith size class, s_i is the probability of individuals surviving from size class i to size class (i+1) in one year, $p_{i\to i}$ is the probability of individuals staying within size class i and $p_{i\to j}$ is the probability of individuals growing from size class i to another size class j after one year. Values for s and p range from 0 to 1 and $\sum_{j=i}^{v} p_{i\to j} = 1$. Negative growth was not permitted in this study, and thus I set $j \geq i$ and the growth transition probabilities were standardized by dividing the probability of growing from a given size class to each size class by the summation of these probabilities of growing from a given class to all the size classes. The last size class was a plus group with the probability of staying in the same size being one (i.e., $p_{v\text{-}v} = 1$). Values for e are positive by definition.

The size-specific population abundance can be forward projected by using the projection matrix, and will eventually become stationary. This forward projection yields the growth rate λ, the stable size structure and the reproductive values, which are the dominant eigenvalues, the corresponding right and left eigenvectors of the projection matrix, respectively (Hilborn and

Walters, 1992; Quinn and Deriso, 1999; Caswell, 2001). In a non-hierarchical model, the projection matrix is further governed by parameters while in a hierarchical model, both parameters and hyperparameters determine the projection matrix (Caswell, 2001; Jiao et al., 2009b).

Size categories were constructed using biologically meaningful sizes. I constructed four size classes with unequal interval width. The first size class encompassed growth from the size at one year (25 mm) to the size before becoming incryptic (75 mm, Cox, 1960), and thus abalones within this size class stay cryptic. The first size class did not start with the size at settlement (0.1 mm) because of the assumption of prebreeding populations. Tutschulte (1976) estimated growth from a sample of 20 white abalones collected during search dives and a sample of 21 white abalones grown in the laboratory. The size at one year (25 mm) was determined from these analyses. The second size class ranged from 75.1 mm to 90 mm where white abalones stay incryptic before reaching maturity (Tutschulte, 1976). The third size class ranged from 90.1 mm to 130 mm, representing individuals reaching maturity and intermediate size (Davis et al., 1996; Hobday and Tegner, 2000). The Final White Abalone Recovery Plan (The Recovery Plan, NMFS, 2008) defines the intermediate-size individuals to be those within size class 90-130 mm, and sets conservation goals targeting this size class. Thus, I constructed this size class for evaluating hypothetical restoration strategies in the risk assessment. The fourth size class encompassed growth from 130.1 mm to 250 mm, the estimated maximum size (Cox, 1960).

Growth transitions were determined from a linear relationship between shell length increment and initial length (Tutschulte, 1976), and a density probability distribution of length increment (Quinn and Deriso, 1999):

$$\overline{\Delta L_i} = -0.0986 L_i + 24.841,$$

$$\Delta L_i \sim N\left(\overline{\Delta L_i}, \sigma^2_{\Delta L}\right),$$

where $\overline{\Delta L_i}$ is the mean length increment for individuals in the *i*th size class, L_i is the midpoint of the *i*th size class, and $N\left(\overline{\Delta L_i}, \sigma^2_{\Delta L}\right)$ is the normal distribution with a mean of $\overline{\Delta L_i}$ and a variance of $\sigma^2_{\Delta L}$ where $\sigma_{\Delta L} = 15.0$ for individuals smaller than 100 mm (i.e., $L_i \leq 100$ mm) and $\sigma_{\Delta L} = 7.8$ for those larger than 100 mm (i.e., $L_i > 100$ mm), which were estimated from a sample of 21 white abalones grown in the laboratory for one year (Tutschulte, 1976).

There are no survival estimates for juvenile and adult white abalone from the literature (Hobday and Tegner, 2000). Studies have shown that natural mortality generally declines with age for abalone (Tegner and Butler, 1985b; Prince et al., 1988; Shepherd and Daume, 1996). Thus I assumed the natural mortality of white abalone to be inversely proportional to length (Lorenzen, 1996; Lorenzen, 2000; Lorenzen, 2006):

$$M_i = cL_i^d,$$

where M_i is the natural mortality of individuals within size class *i*, $c > 0$ and $d < 0$ are constants and $d = -1$. The constant *c* can be determined by setting the area under the mortality curve equivalent to the area under the horizontal line of the predetermined constant natural mortality *M*:

$$\int_{L_{\min}}^{L_{\max}} cL^d \, dL = M\left(L_{\max} - L_{\min}\right),$$

where L_{max} and L_{min} are the maximum and minimum sizes that I used for constructing size classes, i.e., $L_{max} = 250$ mm and $L_{min} = 25$ mm. In the case of $d = -1$, by solving the above equation, I obtained:

$$c = \frac{L_{\max} - L_{\min}}{\ln L_{\max} - \ln L_{\min}} M \ .$$

The size-specific annual survival rates can be calculated as:

$$s_i = \exp(-M_i) \ .$$

In the hierarchical model, the M was further represented by hyperparameters m_1 and m_2:

$$\begin{aligned} M &\sim N\left(\bar{M}, \sigma_M^2\right), \\ \bar{M} &\sim U\left(m_1, m_2\right), \\ \sigma_M &= CV \times \bar{M}, \end{aligned}$$

where $U(m_1, m_2)$ represents a uniform distribution bounded by m_1 and m_2, and CV stands for the coefficient of variation. I set CV = 20 - 40% which has been used as a reasonable uncertainty level in fisheries data analyses (Walters, 1998; Jiao et al., 2009b). Due to lack of information on natural mortality of white abalone, I determined m_1 = 0.075 and m_2 = 0.413 based on estimates from empirical approaches (Table 1), including estimates from longevity (Hoenig, 1983; Hewitt and Hoenig, 2005), age of maturity (Jensen, 1996) and intrinsic growth rate in the von Bertalanffy growth model (Jensen, 1996).

In the development of size-specific fertility e_i, I considered size-specific fecundity (F_i, i.e., number of eggs produced by all individuals within size class i each year), spatially explicit density-dependent fertilization success (x_k, i.e., fertilization success given white abalone density in region k), larval survival rate before settlement (q), settlement success (r), and I also included the survival rate after settlement until reaching first year (s_0) because of the prebreeding assumption:

$$e_i = F_i \times x_k \times q \times r \times s_0 \ ,$$

where I obtained $q = 0.05$ and $r = 0.68$ (average over temperature treatments) from a laboratory study (Leighton, 1972). I set $s_0 = 0.00071$ to have the population growth rate approximately equal to one due to lack of information on larval survival for white abalone. The size-specific fecundity (F_i) was derived from a study on fecundity of wild female white abalones and a study on relationship between shell length and body weight (Tutschulte, 1976):

$$F_i = F \times W_i \times 0.5,$$

$$\ln(W_i) = 4.0243 \ln(L_i) - 14.5069,$$

where W_i is the body weight (g) of individuals within size class i, F is mean number of eggs per gram of female body weight each year, and the value 0.5 is used because I assumed a sex ratio of 1:1. In the hierarchical model, the F was further described by hyperparameters f_1 and f_2:

$$F \sim N\left(\bar{F}, \sigma_F^2\right),$$
$$\bar{F} \sim U\left(f_1, f_2\right),$$
$$\sigma_F = CV \times \bar{F},$$

where I set CV = 20 - 40%, and $f_1 = 7170$ and $f_2 = 10620$ were determined from a four-year fecundity study with 197 wild female white abalones (Tutschulte, 1976).

To incorporate the Allee effect, I borrowed a model of fertilization success obtained for the sea urchin *Diadema antillarum* Philippi (Levitan, 1991) :

$$\ln(x_k) = 0.72 \ln(h_k) + 0.49,$$

where h_k is the population density (number of individuals per square meter) in region k. I used this model for white abalone for three major reasons. First, there is no direct measure on the relationship between fertilization success and density available for abalone. Second, abalone and sea urchins live in similar habitats and aggregate at a similar spatial scale for spawning (Shepherd and Partington, 1995). For example, optimal sperm concentrations for a fertilization

success greater than 90% are around 10^6 to 10^9 per liter for both sea urchins (Pennington, 1985; Young et al., 1992; Levitan, 2002) and abalone (Clavier, 1992; Encena et al., 1998; Babcock and Keesing, 1999). Third, Shepherd and Partington (1995) successfully applied the same model for the development of a spawner-recruitment model for the greenlip abalone *Haliotis laevigata* in South Australia.

Spatially explicit approach

To demonstrate the spatially explicit approach, I derived fertilization success, and thus developed models and conducted the following analyses for three regions off the California coast based on the most recent surveys on white abalone, i.e., the 2002 survey for Tanner Bank, the 2003 survey for Cortes Bank and the 2004 survey for San Clemente Island (Butler et al., 2006). The mean densities for these three regions were estimated to be 13, 8 and 1.5 abalones per hectare respectively. I decided to use density estimates from these surveys for two major reasons. First, these surveys were conducted most recently using the multibeam sonar mapping techniques and therefore provided more accurate density estimates. Second, these surveys revealed a significant difference in densities among these three regions, which makes a good illustration of the spatially explicit approach.

Elasticity analysis and sensitivity analysis

Elasticity is defined as the proportional change in model results (i.e., the growth rate λ in our case) in response to proportional change in model parameters θ (Caswell, 2001). I used a simulation-approximation approach to calculate elasticity (Jiao et al., 2009b):

$$\frac{\theta}{\lambda}\frac{\partial\lambda}{\partial\theta} = \frac{\theta}{\lambda}\frac{\Delta\lambda}{\Delta\theta},$$

where θ can be survival and fertility of each size class. Therefore, by simulating the same propotional change in parameters for each size class in the elasticity analysis, I can compare the relative importance of different size classes to the population growth rate for each of the three regions (Caswell, 2001).

In this study, I derived growth transition, survival and fertility based on other parameters, including the shape parameter d, the m_1 and m_2 that determined the constant natural mortality M, the f_1 and f_2 that determined the constant fecundity F, larval survival rate before settlement q, settlement success r, the survival rate after settlement until reaching first year s_0, the intercept and slope parameters that determined the relationships of $\overline{\Delta L}$, $ln(w)$ and $ln(x)$, and the CV. I also calculated the sensitivity of λ with respect to these parameters for each of the three regions, where sensitivity is defined as the absolute change in model results in response to infinitesimal change in model parameters (Caswell, 2001). Additionally, I examined the sentitivity of λ to population density h, where I set the upper bound of h to be 2000 individuals per hectare according to the Recovery Plan.

To address the uncertainty introduced by determining size categories, I additionally constructed the size classes using two other approaches and compared the population growth rates from these two approaches with the one that I employed (I) for each of the three regions. The size classes that I used (I) covered a large range of sizes because of limited data, but may ignore the individual variations within a class. Thus, in method II, I constructed seven size classes with unequal interval width by splitting each of the size classes (except the second one) in method I into two finer size classes from the middle point. In method III, I constructed six size

classes with equal interval width of 35 mm as applied in Rogers-Bennett and Leaf (2006). The first size class started with 35 mm and the last size class stretched to 245 mm to obtain an equal interval width.

To examine the difference between hierarchical and non-hierarchical models, I compared the population growth rate estimates, percentage of reproductive contribution, and elasticities to fertility and survival from these two types of models for each of the three regions. In the non-hierarchical model, the hierarchical structures in M and F were ignored:

$$M \sim U(m_1, m_2),$$
$$F \sim U(f_1, f_2).$$

Risk assessment

The Recovery Plan (NMFS, 2008) suggests a minimum density of 2000 individuals per hectare and a minimum proportion of 85% intermediate-size individuals (90-130 mm) for white abalone to be downlisted. According to the Recovery Plan, I developed a series of hypothetical stocking strategies and assessed their efficacy: for each of the three regions, stocking up to 1500 individuals within a size class each year for 10 years, and implementing this strategy for each size class. The increase in population density after 10 years relative to initial density (h_{10} / h_0), and the increase in proportion of intermediate-size individuals (90-130 mm) after 10 years relative to initial proportion (%intermediate-size$_{10}$ / %intermediate-size$_0$) were evaluated. I focused on stocking strategies in this study because the strong degree of spawning synchrony and high fecundity of white abalone make captive breeding and outplanting feasible (Stierhoff et al., 2012). I used a 10-year time period for two reasons. First, it is moderate relative to the longevity of white abalone (35-40 years, Cox, 1960). Second, Morris and Doak (2004) has suggested a

minimum number of 10 years to examine the population trend in a population viability analysis. I used the stable size structure from the matrix model as the initial size structure (Table 2). I assumed the total habitat area for Tanner Bank, Cortes Bank, and San Clemente Island to be 1619, 1138, and 889 hectare, respectively (Butler et al., 2006). Results were obtained from 10000 Monte Carlo simulation runs.

The most recent survey at Tanner Bank (Stierhoff et al., 2012) updated the density estimates and size distributions for 2002 and 2004, and additionally provided this information for 2008 and 2010. With the updated information, I repeated the risk assessment for Tanner Bank using the actual surveyed size structure as the initial structure for 2002, 2004, 2008 and 2010. I determined the densities for each of the four years by dividing the total number of sightings each year across all depths by the total areas surveyed, i.e., 21 abalones per hectare for 2002, 5 abalones per hectare for 2004 and 2010, and 10 abalones per hectare for 2008. I roughly estimated the size structure (Table 2) for each of the four years by digitizing the size distribution curves in Stierhoff et al. (2012). Our purpose was to explore which stocking strategy is the most effective under the current population structure instead of the hypothetical stable structure.

4.4 Results

The estimated population growth rate decreased as white abalone population became less abundant (Table 3), with the mean growth rates in Tanner Bank and Cortes Bank being 10% and 7% higher than the one in San Clemente Island, respectively. For all three regions (Fig. 1), large-size individuals (130-250 mm) contributed to reproduction the most (mean 44-51%), followed by intermediate-size individuals (90-130 mm, mean 25-26%), where individuals in higher density region (e.g, Tanner Bank) contributed greater than that in lower density region (e.g., San

Clemente Island). Population growth rate was most sensitive to fertility and survival of large-size individuals, with the mean elasticity to fertility decreasing while the mean elasticity to survival increasing as the population became less abundant (fertility: Tanner Bank 0.0782 > Cortes Bank 0.0692 > San Clemente Island 0.0397; survival: Tanner Bank 0.5003 < Cortes Bank 0.5504 < San Clemente Island 0.7274). For all size classes in the three regions, survival affected growth rate more than fertility, represented by larger elasticity values to survival than those to fertility. For the stabilized size structure (Table 2), the proportion of Size 1 individuals on San Clemente Island is only half of the proportions on the other two offshore banks. The low density on San Clemente Island may have led to low recruitment through the Allee effect.

I conducted sensitivity analysis for all the parameters involved in constructing the projection matrix. Figs. 2 and 3 only present results for those parameters to which population growth rate was shown to be highly sensitive, including two growth-related parameters (the intercept and slope of $\overline{\Delta L}$ relationship), three survival-related parameters (natural mortality shape parameter d, m_1 and m_2 that determined the constant natural mortality M) and six fertility-related parameters (the intercept and slope of $ln(w)$ relationship, and the slope of $ln(x)$ relationship, larval survival rate before settlement q, settlement success r, and the survival rate after settlement until reaching first year s_0). Inflating the slope and intercept of $\overline{\Delta L}$ relationship increased the mean of growth rate (Fig. 2). As the size-dependent natural mortality curve became less concaved down (i.e., d increased) until it leveled up ($d = 0$), the mean of growth rate decreased while its variance increased. Inflating the hyperparameters m_1 and m_2 that determined the constant natural mortality M also decreased the mean of growth rate. Growth rate was most sensitive to the intercept and slope of $ln(w)$ relationship among all the parameters in the sensitivity analysis (Fig. 3). Inflating the intercept and slope of $ln(w)$ relationship increased the

mean and variance of growth rate, and this increase became more dramatic after the intercept reached -14 or the slope reached 4. Inflating the slope of $ln(x)$ relationship decreased the mean and variance of growth rate. Inflating the larval survival rate before settlement q, settlement success r, and the survival rate after settlement until reaching first year s_0 increased the mean of growth rate. In the sensitivity analysis, growth rate in less abundant region (e.g., San Clemente Island) showed lower sensitivity to all parameters than the one in the more abundant region (e.g., Tanner Bank), represented by smaller changes in growth rate in San Clemente Island as each parameter increased (Figs. 2 and 3). Both mean and variance estimates of population growth rate increased as the population became more abundant, but this increase became slower after population density reached approximately 250 individuals per hectare (Fig. 4).

Growth rate was sensitive to the way of determining size categories (Table 3). I obtained a smaller estimate of mean growth rate when applying method II to determine size categories compared to method I, suggesting that growth rate decreased with smaller size intervals and this decrease was more prominent in abundant regions, i.e., growth rate in Tanner Bank decreased the most (5%), followed Cortes Bank (4%) and San Clemente Island (2%). Compared to method II, using method III yielded larger growth rate, indicating that growth rate may increase if using equal size class instead of an unequal one; this conclusion needs to be further justified because methods II and III were not restrictively comparable in this study. Given a certain way of determining size categories, hierarchical and non-hierarchical models yielded similar estimates for the mean growth rate (Table 3), percentage of reproductive contribution, and elasticities to fertility and survival (results not shown), but different estimates for variance, with the 95% confidence interval from the hierarchical model being wider.

Regardless of initial population density and initial size structure, stocking larger-size individuals, and for a given size class, stocking more individuals every year, were more effective in increasing population density after ten years (Figs. 5 and 6). However, the degree of increase depended on both initial population density and initial size structure. When the population was initially dominated by young individuals (i.e., initialized with stable size structure, Fig. 5), the increase in density became more prominent in regions with lower initial density (e.g., San Clemente Island) than those with higher initial density (e.g., Tanner Bank and Cortes Bank). By contrast, when population was initially dominated by old individuals (i.e., initialized with surveyed size structure at Tanner Bank, Fig. 6), I observed more obvious increase in years with higher initial density (e.g., year 2002) than those with lower initial density (e.g, years 2004 and 2010).

Regardless of initial population density and initial size structure, stocking large-size individuals led to a decrease in intermediate-size class proportion (Figs. 5 and 6). Stocking other three size classes raised the proportion and stocking emerging-size individuals (75-90 mm) were more effective in all cases except the one initialized with surveyed size structure at Tanner Bank in 2002. The increase in intermediate-size proportion became more obvious as initial density became lower, and tended to reduce or remain unchanged as more individuals were stocked.

4.5 Discussion

The Recovery Plan recommends models for white abalone to incorporate demographic and environmental stochasticity. Hierarchical demographic models provide an alternative to account for not only demographic and environmental stochasticity, but also great uncertainty due to limited data. For endangered species such as white abalone, an extremely low number of

extant individuals dominated by larger-size class makes it difficult to collect and update demographic information for a wild population, especially the survival estimates. The growth and reproduction estimates that I used in this study are 37 years old, and white abalone population dynamics may have dramatically changed over such a long period because density estimates declined from 2300 abalones per hectare in 1969-1972 surveys (Tutschulte, 1976) to 1.5-13 abalones per hectare in 2002/2004 surveys (Butler et al., 2006), and its demographic traits (i.e., fertilization success) may greatly depend on density (Clavier, 1992; Babcock and Keesing, 1999; Hobday and Tegner, 2000). Other estimates that I used, such as for settlement success, are either old laboratory data or rely on captive populations. With urgent needs for assessing the current status of white abalone and the risk of applying possible restoration strategies, but with great difficulties for acquiring more data, conservation efforts can take advantage of the modeling approach, such as the hierarchical demographic model, that can not only fully utilize limited data for this species but also properly incorporate great uncertainty due to such data limitation.

It is always debatable how much or even whether uncertainty should be incorporated in models for species with limited data (Beissinger and Westphal, 1998). On one hand, without sufficient information, it is difficult to obtain variance estimates other than mean estimates for key demographic parameters. On the other hand, little information on these parameters introduces great uncertainty in modeling population dynamics of this species. In this study, I preferred the hierarchical demographic model for white abalone rather than non-hierarchical or deterministic models for three reasons. First, other than data limitation, individual/sub-population variability in demographic traits due to genetic diversity and environmental stochasticity also encourage the modeling approaches that can account for this variability, such

as the hierarchical demographic model. For example, the fecundity of female white abalone can range from 7170 to 10620 eggs per gram female body weight per year (Tutschulte, 1976); its larvae development, growth and settlement success are sensitive to water temperature, with settlement success ranging from 0% to 100% under different temperatures from10 ºC to 20 ºC, and water temperature is likely to show spatiotemporal heterogeneity. Second, to address the concerns associated with limited data, I borrowed information from white abalone's congeners and closely-related species for some parameters; as suggested by Beissinger and Westphal (1998), I used a relative instead of an absolute approach to interpret results and draw conclusions, i.e., I focused on the relative changes in growth rate rather than its absolute estimates. Third, compared with the non-hierarchical model, the hierarchical model incorporated greater uncertainty and variability but yielded similar mean estimates of growth rate, reproduction contribution, and elasticities to fertility and survival. Morris (1999) suggested that even with limited data, a more complicated model can still be informative if it confirms the conclusions of a simpler model, and I can place greater confidence in the results and their implications for conservation planning.

This study confirmed that population dynamics (i.e., the response of population growth rate to demographic parameters and restoration strategies) of white abalone is sensitive to its density, and thus restoration effort focusing on increasing its density could be a priority. For example (Figs. 2 and 3), growth rate in a region where it is less abundant (e.g., San Clemente Island) showed lower sensitivity than in a region where it is more abundant (e.g., Tanner Bank). However, I need to clarify that density does not necessarily represent separation distance between individuals, and density along with spatial distribution together can better characterize this Allee effect in white abalone populations. For example, given the same density, a spatially

evenly-distributed population may show totally different population dynamics (e.g., fertilization success) from a spatially patchily-distributed one. Thus, it is also important to consider spatial distribution while setting restoration goals for increasing density. Butler et al. (2006) and Stierhoff et al. (2012) provided important information on aggregations of current white abalone populations, and the majority of individuals were observed alone. With such field data, a simulation study that can incorporate not only density-dependence but also spatial-distribution would provide essential information for setting restoration goals to increase both density and population growth.

This study showed that different methods of assigning size classes may affect outcomes of a demographic model (Table 3), which has rarely been tested in previous studies related to population viability analysis. With too wide size intervals, I may increase the degree of ignoring variability within each size class and even some critical features of a certain group of individuals in a size class (called the distribution error in Vandermeer (1978)). For example, results showed that the largest size class (130-250 mm) contributed the most to reproduction under method I (Fig. 1). In reality, not all individuals in this size class would contribute equally to reproduction; instead, the smaller individuals in this size class may contribute much more than larger ones. In this case, method II that further split each size class into two finer intervals may be more appropriate. With too narrow size intervals, data may not be sufficient enough to provide information for each size class (called the sampling error in Vandermeer (1978)). Rogers-Bennett and Leaf (2006) applied the Maloney-Vandermeer algorithm (Caswell, 2001) to minimize both the problem with too few individuals sampled within a size class and the problem with treating all individuals within a size class equally, although this procedure contributed little to their final decision on size class because they used very large size intervals. Vandermeer

(1978) also indicated that such a procedure is not meant to apply to those situations where size categories are defined biologically. Thus, I did not apply such algorithms to aid in our decision on size class. Instead, I conducted a sensitivity analysis to examine the response of growth rate to different classification methods, and I would recommend such a sensitivity analysis for size-based demographic models to obtain less biased information.

Extensive studies have been conducted on the survival of outplanted individuals of its congeners across a wide range of size classes from larvae to large adults (150 mm), and these information are summarized in McCormick (1994) and the Recovery Plan (NMFS, 2008). In this study, I determined the lower (m_1 = 0.075) and upper (m_2 = 0.413) bounds of mean natural mortality for white abalone based on its demographic information through empirical approaches. I validated this upper and lower bounds using the field data with outplanted abalones (Fig. 7). I estimated the 95% confidence intervals of size specific natural mortality by taking the lower and upper bounds of mean natural mortality, respectively. I used the *CV*=40% to encompass the greatest uncertainty. Results confirmed that the majority of the field observations fell within the range that I obtained based on this upper and lower bounds of the mean natural mortality. Thus, although great uncertainty is embedded in natural mortality estimation for white abalone due to lack of information on wild populations, the range of mean natural mortality that I applied in this study was fairly appropriate.

The sensitivity analysis on growth rate relative to population density showed a faster increase in growth rate before population density reached approximately 250 individuals per hectare and the increase became slower afterwards (Fig. 4). This result confirmed that the population growth rate may suffer from the Allee effect at low density, and also may start to be limited by available resources when density reaches a certain level. Thus, it is important to

consider both the Allee effect and the effect of intraspecific competition when setting conservation goals based on density.

In this study, I assumed a 1:1 sex ratio, and did not account for the possibility of hermaphroditism in white abalone. Ghiselin (1969) reviewed three conditions under which hermaphroditism is likely to evolve in animals: (1) where it is difficult to find a mate due to low density; (2) where one sex benefits from being larger or smaller than the other; or (3) gene dispersal is restricted due to small and genetically isolated populations. Although hermaphroditism (simultaneous hermaphroditism in most cases) has been found among the Gastropoda (Ghiselin, 1966; Ghiselin, 1969), no solid evidence has been revealed among abalones. Giorgi and DeMartini (1977) observed no historical evidence of hermaphroditism in red abalone *Haliotis rufescens* Swainson in California. Girard (1972) noted a hermaphroditic specimen of the ormer (*Haliotis tuberculata*), and Murayama (1935) observed one of the Japanese species *Haliotis gigantean*. Hayashi (1980) studied the reproduction of the ormer in Guernsey but did not find any trace of hermaphroditism. Without further investigation, these instances of hermaphroditism among abalone are unexplained and may be considered as an uncommon phenomenon. Nevertheless, according to the three conditions above, the possibility exists for white abalone to evolve from the gonochoric to the hermaphroditic state given its current low density and spatially heterogeneous distribution. Further investigation on the physiology and reproductive biology of white abalone is needed to explore this hypothesis.

Focusing on the absolute values of growth rate for different population densities (Fig. 4), it seems that with a density of as low as 50 individuals per hectare, I will be able to achieve the growth rate of one, i.e., preventing the population from continued decline. Also, looking at the absolute values of density ratio corresponding to different stocking strategies (Fig. 5), it appears

to suggest that even without any stocking, the density of white abalone will be increased and even doubled (e.g., Tanner Bank) in 10 years. These implications seem unrealistic for white abalone given its current population status. To properly understand and interpret these results, recall that when I constructed the projection matrix, I adjusted the survival rate before reaching the first year (s_0) to make the population growth rate approximately equal to one; I borrowed information from sea urchins for the density-dependent fertilization success; and I roughly estimated the mean natural mortality through empirical methods due to lack of data on white abalone. Thus, as discussed above, given the limited information on white abalone, I based our conclusions on the relative changes in the quantities of interest instead of their absolute values.

The Recovery Plan has set up goals based on both density (e.g, density > 2000 individuals per hectare for downlisting) and size structure (e.g., intermediate-size (90-130 mm) proportion ≥ 85%; large-size (>130 mm) proportion ≤ 15% for downlisting). I have three concerns associated with these conservation goals. First, based on this study, stocking large-size (>130 mm) individuals may be most effective in increasing density but may lead to decrease of intermediate-size proportion. Thus, restoration efforts focusing only on increasing density or population size may not be sufficient to meet both goals. Second, with effective stocking strategies, the degree of increase in density depended on both initial population density and initial size structure whereas the degree of increase in intermediate-size proportion only depended on initial density (Figs. 5 and 6). With the observed spatial heterogeneity in white abalone populations, it would be more appropriate to have spatially explicit restoration plans and goals based on current initial population status. Third, results suggested that stocking large-size individuals was most effective in increasing white abalone density, which agreed with the findings (140-175 mm) from Rogers-Bennett and Leaf (2006). However, given the laboratory

expenses and time consumption raising abalones in laboratory until adults, concerns may rise when it comes to practice. For example, with a growth rate estimate of 10 mm per year for adults (Tutschulte, 1976; Tutschulte and Connell, 1981) and a stocking size of 130 mm, the white abalones available for stocking will be around 9-13 years old. Based on a review provided by McCormick et al. (1994), outplanting abalones at sizes greater than 40 mm seemed to have the highest survival rates, and it is believed that outplanting younger animals may have higher reproductive potential than older ones (NMFS, 2008). However, these studies were based on the assumption that captive abalones introduced into the wild have the same reproductive fitness and survival as wild white abalone; this assumption still needs to be justified. Thus, more laboratory and simulation studies are needed to aid in the design of stocking programs for white abalone that can balance stocking efficacy and laboratory expenses/time consumption. For example, laboratory measurement on size-specific survival, first year survival, and relationship between fertilization success and density/separation distance, or lab/field data on its suitable habitat will be very helpful. An individual-based simulation model that can track individual performance given a certain level and a certain pattern of aggregation, or a simulation that links ecosystem (e.g., ocean current) to its population dynamics (e.g., larval dispersal, settlement, or growth) would be very helpful to guide stocking of this species.

In summary, hierarchical demographic models provide a powerful tool for incorporation of uncertainty due to data limitation and individual/sub-population variability due to spatiotemporal heterogeneity, which is particularly useful for endangered species such as white abalone. The spatially explicit approach is appropriate for white abalone due to its density-dependent fertilization success and its current heterogeneous spatial distribution. Density has a great influence on degree and/or direction of the response of population growth rate to

demographic parameters and restoration strategies. Stocking large-size (>130 mm) abalones may be most effective in increasing density but may lead to decrease of intermediate-size proportion. Thus, to meet the conservation goals in the Recovery Plan, restoration efforts may need to consider the spatially explicit density-dependence effect and to balance stocking efficacy and laboratory expenses/time consumption.

4.6 Acknowledgements

This research was supported by the USDA Cooperative State Research, Education and Extension Service, Hatch project #0210510 to Dr. Y. Jiao, and a grant for teaching assistantship from Department of Fish and Wildlife Conservation, Virginia Tech.

4.7 References

Allee, W.C., 1931. Animal aggregations: a study in general sociology. University of Chicago Press, Chicago.

Babcock, R., Keesing, J., 1999. Fertilization biology of the abalone *Haliotis laevigata*: laboratory and field studies. Can J Fish Aquat Sci 56, 1668-1678.

Bartsch, P., 1940. The West American *Haliotis*. Proceedings of the United States National Museum 89, 49-58.

Beissinger, S.R., Westphal, M.I., 1998. On the use of demographic models of population viability in endangered species management. The Journal of Wildlife Management 62, 821-841.

Butler, J., Neuman, M., Pinkard, D., Kvitek, R., Cochrane, G., 2006. The use of multibeam sonar mapping techniques to refine population estimates of the endangered white abalone (*Haliotis sorenseni*). Fish Bull 104, 521-532.

Caswell, H., 2001. Matrix population models. Sinauer Associates, Sunderland.

Clark, J.S., 2003. Uncertainty and variability in demography and population growth: a hierarchical approach. Ecology 84, 1370-1381.

Clavier, J., 1992. Fecundity and optimal sperm density for fertilization in the ormer (*Haliotis tuberculata L.*). In: Shepherd, S.A., Tegner, M.J., Proo, S.A.G.d. (Eds), Abalone of the world; Biology, fisheries and culture. Fishing News Books (Division of Blackwell Scientific Publicaitons Ltd.), Oxford, UK.

Cox, K.W., 1960. Review of the abalone of California. California Fish and Game Bulletin 46, 381-406.

Davis, G.E., Haaker, P.L., Richards, D.V., 1996. Status and trends of white abalone at the California Channel Islands. Trans Am Fish Soc 125, 42-48.

Davis, G.E., Haaker, P.L., Richards, D.V., 1998. The perilous condition of white abalone *Haliotis sorenseni*, Bartsch, 1940. J Shellfish Res 17, 871-876.

Encena, V., Capinpin, E., Bayona, N., 1998. Optimal sperm concentration and time for fertilization of the tropical abalone, *Haliotis asinina* Linne 1758. Aquaculture 165, 347-352.

Gascoigne, J., Lipcius, R.N., 2004. Allee effects in marine systems. Mar Ecol Prog Ser 269, 49-59.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian data analysis. Chapman and Hall, Boca Raton, Florida.

Ghiselin, M.T., 1966. Reproductive function and the phylogeny of opisthobranch gastropods. Malacologia 3, 327-378.

Ghiselin, M.T., 1969. The evolution of hermaphroditism among animals. Q Rev Biol 44, 189-208.

Giorgi, A.E., DeMartini, J.D., 1977. A study of the reproductive biology of the red abalone, *Haliotis rufescens* Swainson, near Mendocino, California. Calif Fish Game 63, 80-94.

Girard, A., 1972. La reproduction de l'Ormeau *Haliotis tuberculata* L. Revue des Travaux de l'Institut des Peches Maritimes 36, 163-184.

Haaker, P.L., 1998. White abalone-off the deep end forever? Outdoor California January-February, 17-20.

Hayashi, I., 1980. The reproductive biology of the ormer, *Haliotis tuberculata*. J Mar Biol Assoc UK 60, 415-430.

Hewitt, D.A., Hoenig, J.M., 2005. Comparison of two approaches for estimating natural mortality based on longevity. Fish Bull 103, 433-437.

Hilborn, R., Walters, C.J., 1992. Quantitative fisheries stock assessment. Choice, dynamics and uncertainty. Chapman & Hall, London, UK.

Hobday, A.J., Tegner, M.J., 2000. Status review of white abalone (*Haliotis sorenseni*) throughout its range in California and Mexico. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southwest Region Office, Long Beach, California.

Hoenig, J.M., 1983. Empirical use of longevity data to estimate mortality rates. Fish Bull 82, 898-903.

Jensen, A., 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. Can J Fish Aquat Sci 53, 820-822.

Jiao, Y., Hayes, C., Cortes, E., 2009a. Hierarchical Bayesian approach for population dynamics modelling of fish complexes without species-specific data. ICES J Mar Sci 66.

Jiao, Y., Lapointe, N.W.R., Angermeier, P.L., Murphy, B.R., 2009b. Hierarchical demographic approaches for assessing invasion dynamics of non-indigenous species: An example using northern snakehead (*Channa argus*). Ecol Model 220, 1681-1689.

Karpov, K.A., Haaker, P.L., Taniguchi, I.K., Rogers-Bennett, L., 2000. Serial depletion and the collapse of the California abalone (*Haliotis spp.*) fishery. In: Campbell, A. (Ed), Workshop on Rebuilding Abalone Stocks in British Columbia. NRC Research Press, Ottawa, Ontario, Canada.

Leighton, D.L., 1972. Laboratory observations on the early growth of the abalone, *Haliotis sorenseni*, and the effect of temperature on larval development and settling success. Fish Bull 70, 373-380.

Leslie, P.H., 1945. On the use of matrices in certain population mathematics. Biometrika 33, 183-212.

Levitan, D.R., 1991. Influence of body size and population density on fertilization success and reproductive output in a free-spawning invertebrate. The Biological Bulletin 181, 261-268.

Levitan, D.R., 2002. The relationship between conspecific fertilization success and reproductive isolation among three congeneric sea urchins. Evolution 56, 1599-1689.

Lorenzen, K., 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. J Fish Biol 49, 627-642.

Lorenzen, K., 2000. Allometry of natural mortality as a basis for assessing optimal release size in fish-stocking programmes. Can J Fish Aquat Sci 57, 2374-2381.

Lorenzen, K., 2006. Population management in fisheries enhancement: gaining key information from release experiments through use of a size-dependent mortality model. Fish Res 80, 19-27.

Lundy, A.L., 1997. The California abalone industry, a pictorial history. Best Publishing Company, Flagstaff, Arizona.

McCormick, T., Herbinson, K., Mill, T., Altick, J., 1994. A review of abalone seeding, possible significance and a new seeding device. Bull Mar Sci 55, 2-3.

McShane, P.E., Black, K.P., Smith, M., 1988. Recruitment processes in *Haliotis rubra* (Mollusca: Gastropoda) and regional hydrodynamics in southeastern Australia imply localized dispersal of larvae. J Exp Mar Biol Ecol 124, 175-203.

Morris, W., Doak, D.F., Groom, M., Kareiva, P., Firberg, J., Gerber, L., Murphy, P., Thomson, D., 1999. A practical handbook for population viability analysis. Nature Conservancy.

Morris, W.F., Doak, D.F., 2004. Buffering of life histories against environmental stochasticity: accounting for a spurious correlation between the variabilities of vital rates and their contributions to fitness. The American Naturalist 163, 579-590.

Murayama, S., 1935. On the development of the Japanese abalone, *Haliotis gigantea*. Journal of the College of Agriculture Tokyo Imperial University 23, 227-233.

National Marine Fisheries Service (NMFS), 2008. Final white abalone recovery plan (*Haliotis sorenseni*). National Marine Fisheries Service, Long Beach, CA.

Pennington, J.T., 1985. The ecology of fertilization of echinoid eggs: the consequences of sperm dilution, adult aggregation, and synchronous spawning. The Biological Bulletin 169, 417-430.

Prince, J.D., Sellers, T.L., Ford, W.B., Talbot, S.R., 1987. Experimental evidence for limited dispersal of haliotid larvae (genus *Haliotis*: Mollusca: Gastropoda). J Exp Mar Biol Ecol 106, 243-264.

Prince, J.D., Sellers, T.L., Ford, W.B., Talbot, S.R., 1988. Recruitment, growth, mortality and population structure in a southern Australian popuation of *Haliotis rubra* (Mollusca: Gastropoda). Mar Biol 100, 75-82.

Quinn, T.J., Deriso, R.B., 1999. Quantitative fish dynamics. Oxford University Press, New York.

Rogers-Bennett, L., Leaf, R.T., 2006. Elasticity analyses of size-based red and white abalone matrix models: management and conservation. Ecol Appl 16, 213-224.

Shepherd, S., Partington, D., 1995. Studies on southern Australian abalone (Genus *Haliotis*). XVI. Recruitment, habitat and stock relations. Mar Freshwat Res 46, 669-680.

Shepherd, S.A., 1973. Studies on southern Australian abalone (Genus *Haliotis*) I: ecology of five sympatric species. Australian Journal of Marine and Freshwater Research 24, 217-257.

Shepherd, S.A., Daume, S., 1996. Ecology and survival of juvenile abalone in a crustose coralline habitat in South Australia. In: Watanabe, Y., Yamashita, Y., Oozeki, Y. (Eds), Survival strategies in early life stages of marine resources; International Workshop, Yokohama, Japan. A. A. Balkema, Rotterdam, Netherlands.

Stierhoff, K.L., Neuman, M., Butler, J.L., 2012. On the road to extinction? Population declines of the endangered white abalone, *Haliotis sorenseni*. Biol Conserv 152, 46-52.

Stoner, A.W., Ray-Culp, M., 2000. Evidence for Allee effects in an over-harvested marine gastropod: density-dependent mating and egg production. Mar Ecol Prog Ser 202, 297-302.

Tegner, M.J., Butler, R.A., 1985a. Drift-tube study of the dispersal potential of green abalone (*Haliotis fulgens*) larvae in the southern California Bight: implications for recovery of depleted populations. Mar Ecol Prog Ser 26, 73-84.

Tegner, M.J., Butler, R.A., 1985b. The survival and mortality of seeded and native red abalones, *Haliotis rufescens*, on the Palos Verdes Peninsula [California, USA]. California Fish and Game Bulletin 71, 160-163.

Tutschulte, T.C., 1976. The comparative ecology of three sympatric abalones. University of California at San Diego, San Diego, California.

Tutschulte, T.C., Connell, J.H., 1981. Reproductive biology of three species of abalones (*Haliotis*) in southern California [USA]. Veliger 23, 195-206.

Vandermeer, J., 1978. Choosing category size in a stage projection matrix. Oecologia 32, 79-84.

Walters, C., 1998. Evaluation of quota management policies for developing fisheries. Can J Fish Aquat Sci 55, 2691-2705.

Young, C., Tyler, P., Cameron, J., Rumrill, S., 1992. Seasonal breeding aggregations in low-density populations of the bathyal echinoid *Stylocidaris lineata*. Mar Biol 113, 603-612.

Table 1

A summary of estimates for natural mortality using empirical approaches.

Parameter	Empirical approach	Estimate for natural mortality	Reference for parameter	Reference for empirical approach
Longevity t_{max}=40 years	$3/t_{max}$	0.075	Cox, 1960; Leighton, 1972	Hoenig, 1983
Longevity t_{max}=40 years	$4.22/t_{max}$	0.1055	Cox, 1960; Leighton, 1972	Hewitt and Hoenig, 2005
Age at maturity t_m=4-6 years	$1.65/t_m$	0.275-0.4125	Tutschulte, 1976	Jensen, 1996
Growth rate coefficient k=0.111	$1.5k$	0.1665	Estimated by fitting the von Bertalanffy growth model using data in Tutschulte (1976)	Jensen, 1996

Table 2

Stable size structure (%) from the hierarchical demographic models for Tanner Bank, Cortes Bank and San Clemente Island, and surveyed size structure (%) at Tanner Bank in 2002, 2004, 2008 and 2010. The surveyed size structure is digitized from Stierhoff et al. (2012).

	Size 1 25-75mm	Size 2 < 90mm	Size 3 < 130mm	Size 4 < 250mm	Density (individual/ha)
Stable size structure (%)					
Tanner Bank	45.37	9.14	19.31	26.18	13
Cortes Bank	40.54	8.42	18.88	32.16	8
San Clemente Island	23.79	5.42	14.61	56.18	1.5
Surveyed size structure (%) at Tanner Bank					
2002	0	0	49.31	50.69	21
2004	0	0	37.98	62.02	5
2008	0	0	18.43	81.57	10
2010	0	0	26.20	73.80	5

Table 3

Population growth rate from hierarchical and non-hierarchical models. The 95% confidence interval is indicated in parenthesis.

Size[a]	Tanner Bank		Cortes Bank		San Clemente Island	
	Hierarchical	Non-hierarchical	Hierarchical	Non-hierarchical	Hierarchical	Non-hierarchical
I	1.0329	1.0321	1.0036	1.0033	0.9361	0.9334
	(0.8233, 1.2187)	(0.9067, 1.1800)	(0.8125, 1.1740)	(0.8854, 1.1397)	(0.7843, 1.0661)	(0.8394, 1.0419)
II	0.9812	0.9788	0.9618	0.9587	0.9196	0.9166
	(0.7986, 1.1610)	(0.8589, 1.1224)	(0.7988, 1.1234)	(0.8498, 1.0922)	(0.7940, 1.0376)	(0.8386, 1.0139)
III	1.0171	1.0169	0.9956	0.9941	0.9409	0.9399
	(0.8221, 1.1929)	(0.8958, 1.1593)	(0.8087, 1.1598)	(0.8800, 1.1266)	(0.7912, 1.0693)	(0.8465, 1.0471)

[a]: Different methods to determine the size categories, i.e., I-four uneven size classes, II-seven uneven size classes, III-six even size classes.

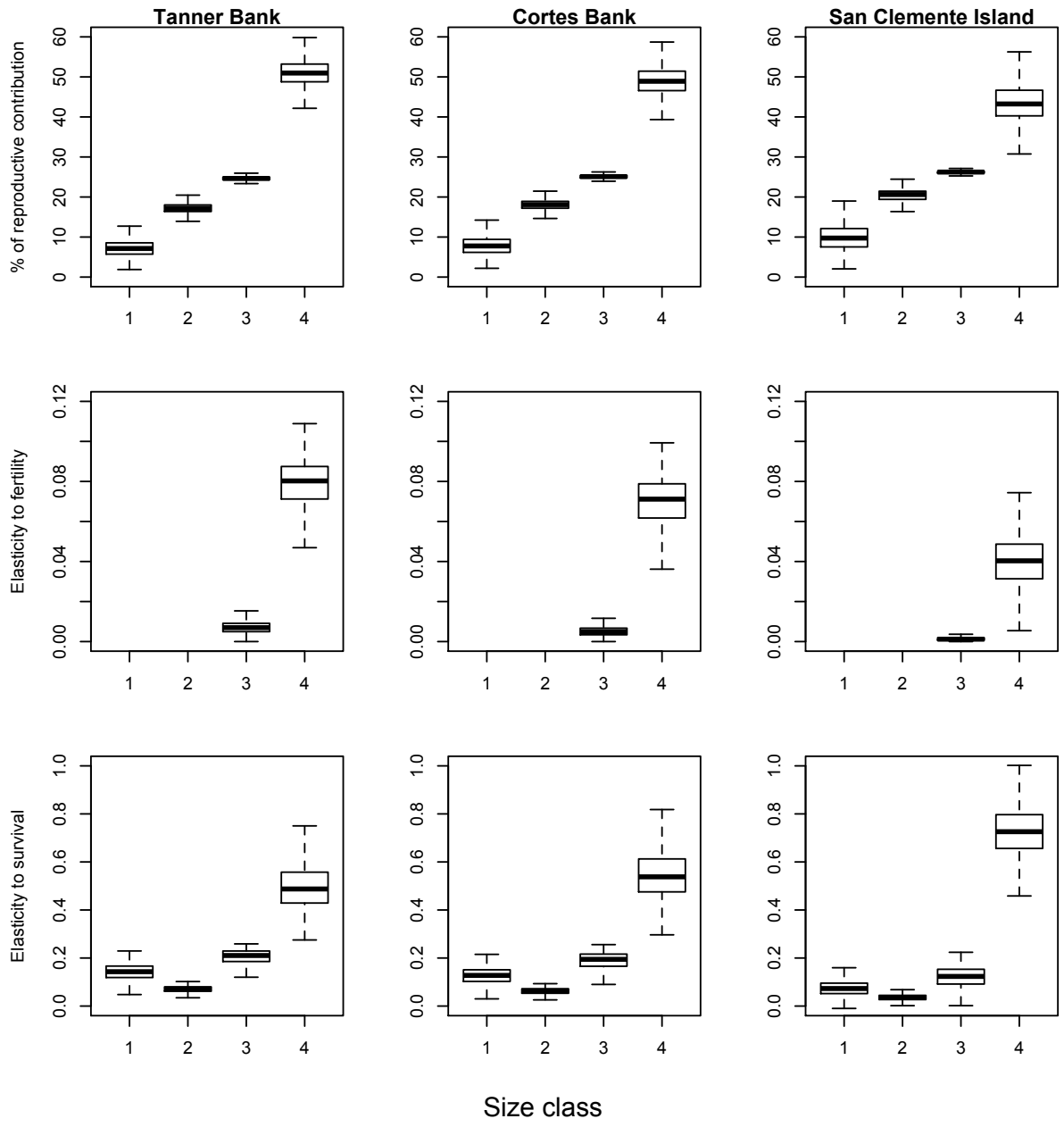

Figure 1. Percentage (%) of reproductive contribution, elasticities of population growth rate to fertility and survival. Upper and lower bars represent data range; upper and lower boundaries of the box represent the third and the first quartiles of the data; line in the middle of the box represents median of the data.

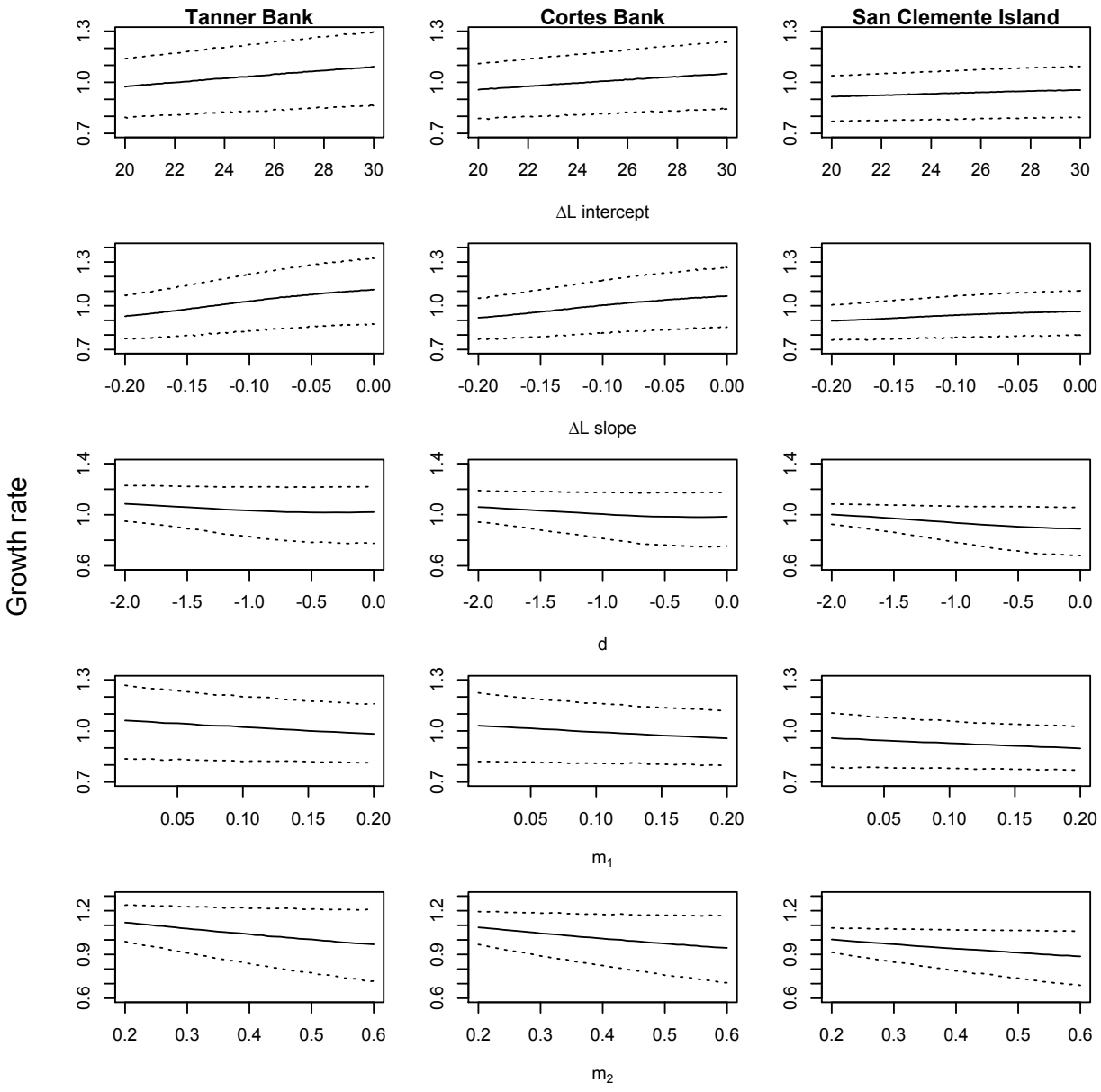

Figure 2. Sensitivity of population growth rate to the growth- and survival-related parameters in projection matrix. Solid lines represent mean estimates of growth rate under different parameter values, and dotted lines represent 95% confidence intervals.

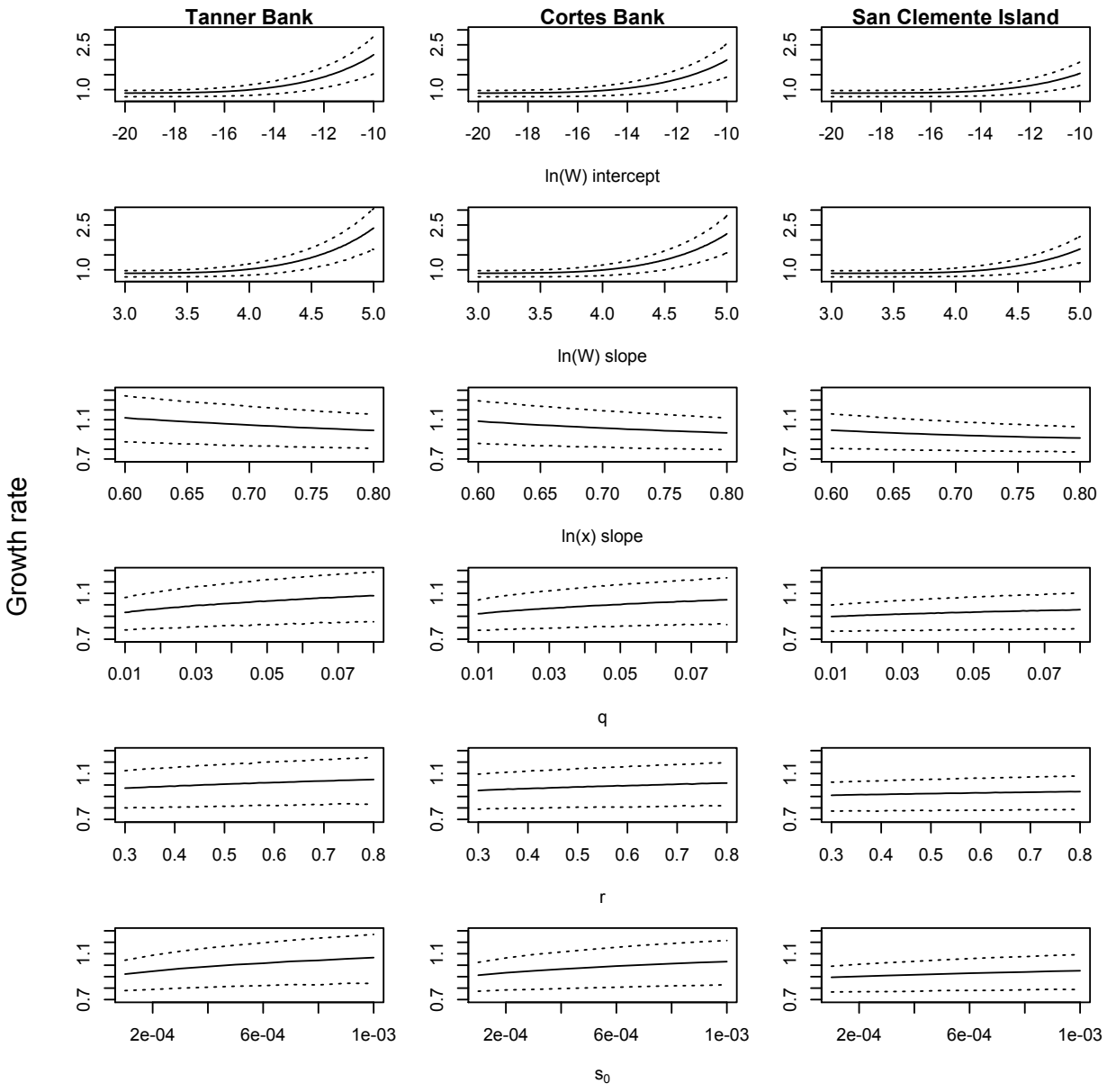

Figure 3. Sensitivity of population growth rate to the fertility-related parameters in projection matrix. See Fig. 2 for the explanation of plots.

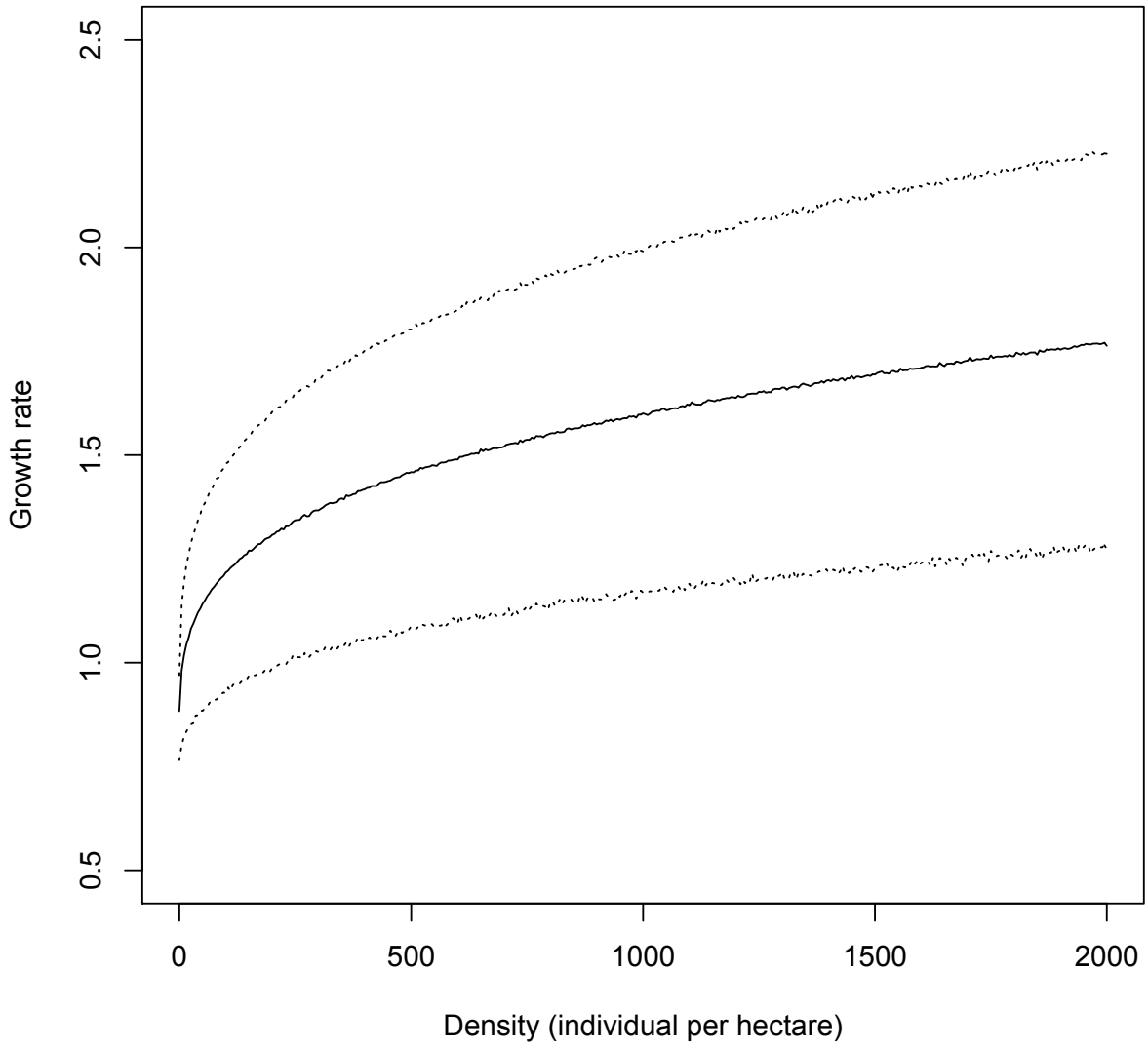

Figure 4. Sensitivity of population growth rate to population density. See Fig. 2 for the explanation of plots.

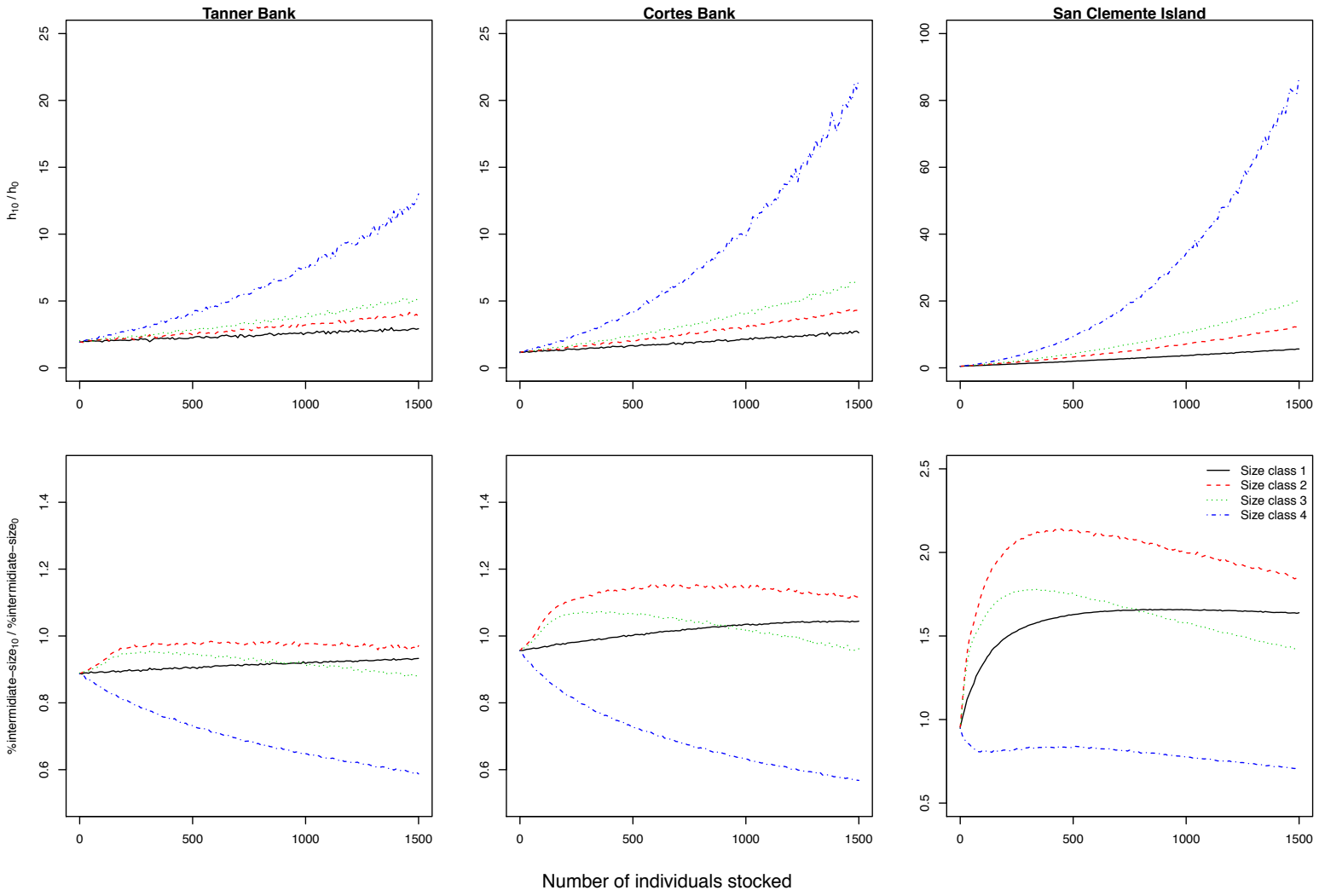
Tanner Bank
Cortes Bank
San Clemente Island
h_{10}/h_0
%intermidiate−size$_{10}$ / %intermidiate−size$_0$
Size class 1
Size class 2
Size class 3
Size class 4
Number of individuals stocked

Figure 5. Hypothetical restoration strategies and their efficacy, shown as the increase in population density after 10 years relative to initial density (h_{10} / h_0), and the increase in proportion of intermediate-size individuals (90-130 mm) after 10 years relative to initial proportion (%intermediate-size$_{10}$ / %intermediate-size$_0$). Initial population size structure is assumed to be the stable size structure. Lines represent mean estimates from 10000 iterations.

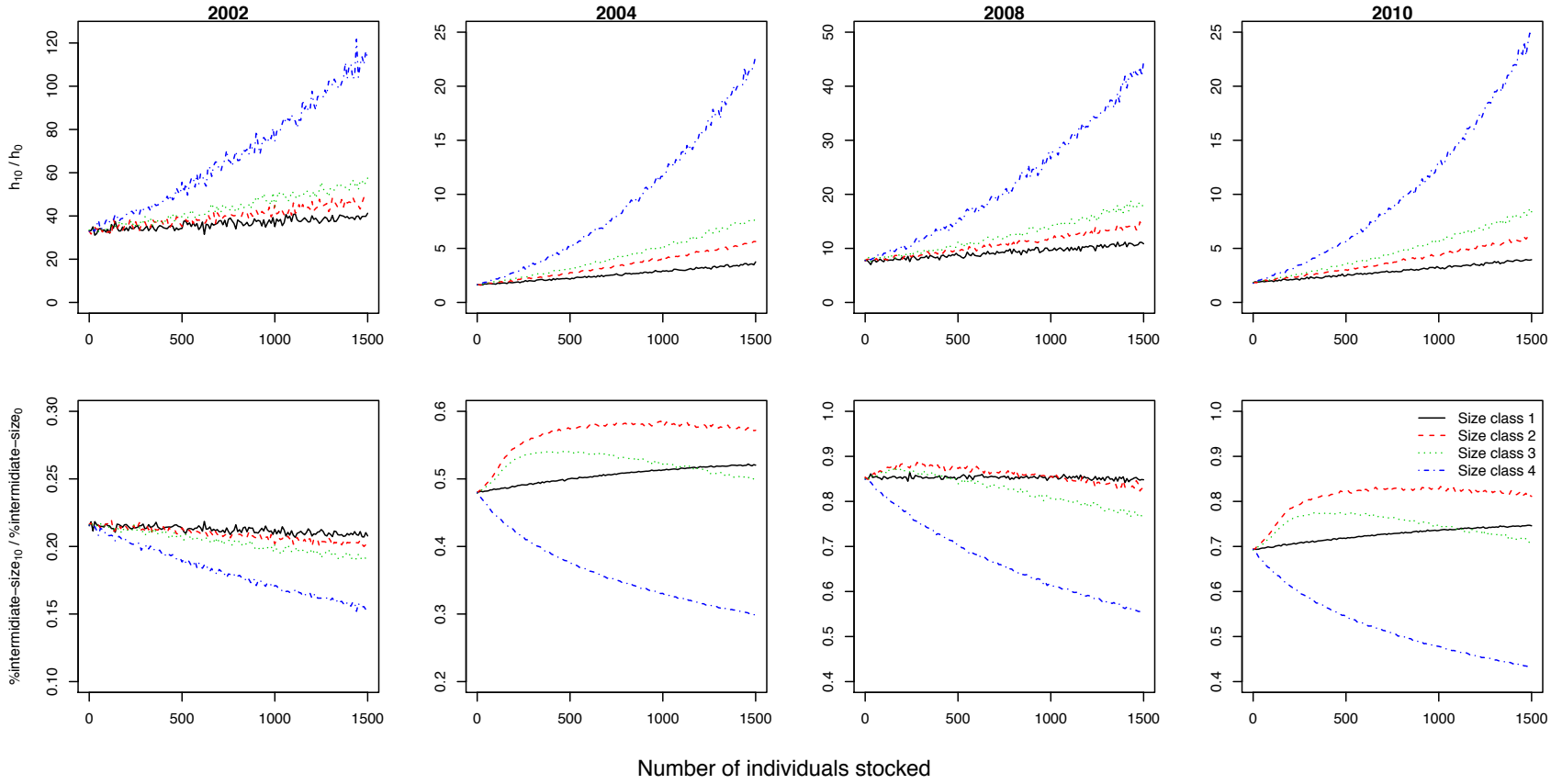

Figure 6. Hypothetical restoration strategies and their efficacy for Tanner Bank when its initial population size structure is the actual surveyed one in 2002, 2004, 2008 and 2010. See Fig. 5 for explanations of y-axis labels.

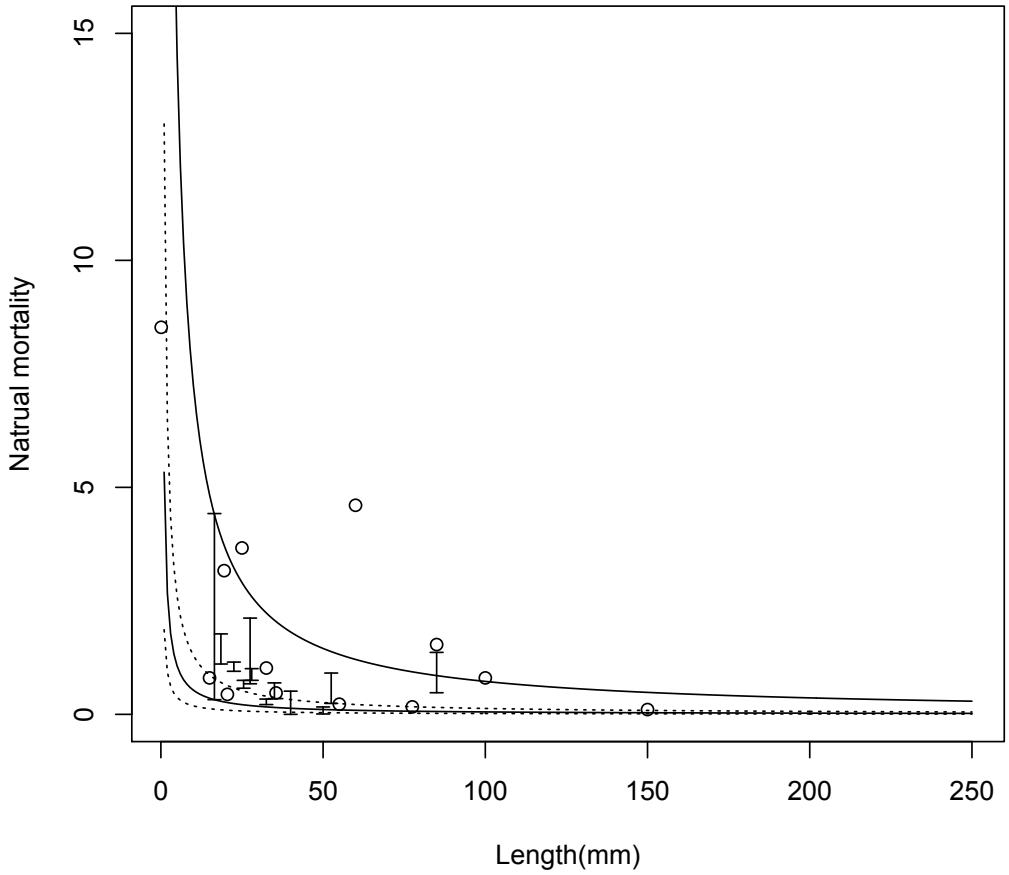

Figure 7. Validation of the lower and upper bounds of mean natural mortality with field outplanting data from literature. Dotted lines and solid lines outline the 95% confidence intervals of size specific natural mortality estimated by taking the lower (m_1 = 0.075) and upper (m_2 = 0.413) bounds of mean natural mortality. I used *CV*=40% for estimation. Circles represent observed mean natural mortality and bars represent observed range of natural mortality for abalones of a certain size in literature.

Chapter 5

Spatial Nonstationarity And Decline of Endangered White Abalone

5.1 Abstract

White abalone (*Haliotis sorenseni*) has been a great concern of fisheries management and marine conservation because its fishery collapsed after only one-decade of exploitation and its population has declined by greater than 99%. With the commercial catch data from 1969 to 1995, I estimated the spatially explicit decline rate for white abalone catch by assuming an exponential decline. The catch was estimated to decline at a rate of -0.2727 in median with a 95% confidence interval between -0.5790 and 0.3788. The decline rate showed a complicated mosaic spatial pattern where more severe decline occurred when moving away from coast in the south of the study region, and less severe decline was found along coast between Point Conception and Santa Barbara. I linked the decline rate to environmental factors including water depth, distance to California coast, distance to land, sea surface temperature and chlorophyll concentration, and I did not detect any significant associations between decline rate and these five environmental factors. This study is limited by the few fishing sites with available data and the limited availability of information on environmental factors. Although I did not find significant associations of catch decline with these five environmental factors and I only analyzed the commercial catch data which generally cannot be used to infer population status without fishing effort information. Thus, through this study I am hoping to emphasize that spatial nonstationarity is likely to exist for white abalone and should be taken into account when developing restoration strategies.

Keywords: Spatial nonstationarity; white abalone, decline, commercial catch, environmental factor

5.2 Introduction

White abalone (*Haliotis sorenseni*), along with four other abalone species once supported a valuable commercial and recreational abalone fishery in California but the whole fishery collapsed by 1997 following a serial depletion (Haaker, 1994; Hobday and Tegner, 2000). The white abalone fishery began in 1968, peaked in 1972, collapsed by 1978 and became the first marine invertebrate protected under the Endangered Species Act in 2001. With respect to the estimated pre-exploitation abundance, abundance of white abalone has declined by greater than 99%, and the decline continues (Hobday and Tegner, 2000; Butler et al., 2006; Stierhoff et al., 2012). White abalone has external fertilization, which requires aggregation of individuals to provide a sufficient concentration of sperm and eggs (Cox, 1960; Hobday and Tegner, 2000; Stierhoff et al., 2012). At such an extremely low density, the remnant abalones may not be able to reproduce successfully, which is known as the Allee effect (Allee, 1931; Stoner and Ray-Culp, 2000; Gascoigne and Lipcius, 2004).

Population dynamics of white abalones are likely to show spatial nonstationarity, i.e., variation in population dynamics over space. First, abalones are known to have limited movement after larvae disperse and settle to a suitable habitat, and their suitable habitat extends along both coastline and depth (Shepherd, 1973; Tutschulte, 1976; Hobday and Tegner, 2000). Second, the most recent survey using multibeam sonar mapping techniques showed that densities among two offshore banks (Tanner Bank and Cortes Bank) and one island (San Clemente Island) off the California coast differed greatly (Butler et al., 2006). Third, its historic commercial catch

data have shown spatial heterogeneity (Karpov et al., 2000). Thus, conservation and management of white abalone will benefit from a more complete understanding of the spatial variation in population dynamics.

Decline of the white abalone population is believed to be driven by intensive exploitation, along with environmental changes, diseases, hybridization, accidental mortality, predation and competition (Lundy, 1997; Hobday and Tegner, 2000). Impacts of environmental factors such as water temperature on abalone population dynamics (e.g., growth and survival) have been documented (Haaker et al., 1998; Shepherd et al., 1998; Hobday and Tegner, 2002; Vilchis et al., 2005). However, how the environmental conditions are related to population decline of white abalone is still understudied, and these environmental conditions may also vary over space.

The purposes of this study were to: (1) explore spatial nonstationarity in the decline of white abalone commercial catch; (2) relate the decline of catch to environmental factors, including water depth, distance to California coast, distance to land, sea surface temperature and chlorophyll concentration; (3) provide recommendations for white abalone restoration.

5.3 Methods

Data

Commercial catch data of white abalone from 1969-1995 are provided by California Department of Fish and Game (CDFG). The entire fishing area was divided into sites with a size of 10 minute by 10 minute and catch data were collected from each site over years. Data for 52 sites were available, but I only used data for 27 sites to analyze the decline of commercial catch. I excluded those sites that contained too few catch data (i.e., catch for only one or two years) to

detect decline trend of catch, and those sites with missing information on location. Among the 27 sites, there was one site (site 897 near Cortes Bank) with a size of 10'(latitude)× 30'(longitude).

Spatial variation in the decline of commercial catch

I assumed an exponential decline for white abalone commercial catch:

$$C_t = C_0 e^{kt} \quad (1)$$

where C_t is the catch (lb) at year t, C_0 is the initial catch, and k is the decline rate. A k value of zero indicates constant catch over years, a k value greater than zero indicates increasing catch and a k value less than zero indicates declining catch. I calculated the decline rate (k) for each of the 27 sites using the annual catch data from the corresponding site.

Correlation between the decline of catch and environmental factors

In this study, I compared the spatially explicit decline rate of white abalone catch with water depth, distance to California coast, distance to land, sea surface temperature and chlorophyll concentration. Water depth at each site referred to the mean depth (m) within the site and was estimated by the U.S. Coastal Relief Model (2003) available at National Oceanic and Atmospheric Administration (NOAA) National Geophysical Data Center. Distance to California coast (km) was defined as the shortest distance from center of each site to California coastline, which may provide insights on the impact of human disturbance on white abalone population. Distance to land (km) was defined as the shortest distance from each site to its nearest land (either the California coastline or an island), in which information about habitat type, ocean current and water depth may be imbedded. These two distances were calculated in ArcGIS (Version 10.1). Studies have shown that abalone growth and survival are sensitive to water

temperature (Haaker et al., 1998; Shepherd et al., 1998; Hobday and Tegner, 2002; Vilchis et al., 2005). Butler et al. (2006) surveyed Tanner Bank and found presence of white abalone associated with algal coverage. Thus, I hypothesized that decline of catch may also be influenced by water temperature and algal coverage. Data of sea surface temperature (ºC) and chlorophyll concentration (mg/m^3) were available from National Aeronautics and Space Administration (NASA) Earth Observatory Program, and I used one-year (2013) average for these two factors.

5.4 Results

The decline rate and trend of white abalone commercial catch were calculated for each site (Fig. 1). Catch at most sites (e.g., sites 689, 684, 829 and 861) exhibited an exponential decline, catch at three sites (i.e., sites 655, 842 and 849) showed exponential increase, and catch at two other sites (i.e., sites 654 and 700) fluctuated dramatically over time. The median decline rate was estimated to be -0.2727 with a 95% confidence interval (CI) between -0.5790 and 0.3788. The negative median value of this decline rate suggested that a severe decline in commercial catch of white abalone has occurred in most sites, and the wide 95% CI implied that the decline of catch varied greatly over space. The site off shore of San Diego (site 861, Fig. 2) suffered the most severe decline in commercial catch of white abalone since 1969 with a decline rate of k = -0.8714. By contrast, the site near central San Clemente Island (site 849, Fig. 2) experienced the greatest increase in commercial catch with a rate of k = 0.4911.

The decline rate (k) showed a complicated mosaic spatial pattern (Fig. 2). In the north of the study region, more positive rates were found along the coast between Point Conception and Santa Barbara. In the water around Channel Islands, more severe declines occurred when moving towards the east and west sides of the islands. In the south of the study region (i.e., south to Point

Dume), the decline rate appeared to be more negative towards open sea, suggesting more severe decline in commercial catch when moving away from coast. Santa Catalina Island experienced less severe decline in catch than other offshore islands. I did not detect any significant correlation between decline rate and the five environmental factors that I examined, i.e., water depth, distance to California coast, distance to land, sea surface temperature and chlorophyll concentration (Fig. 3).

5.5 Discussion

Decline of white abalone catch did show great spatial variation, which confirmed the existence of spatial nonstationarity in white abalone catch data. Theoretically catch data without fishing effort information cannot be used to infer population size or density, and commercial catch is usually subject to fisheries management. I only analyzed the commercial catch data because spatially explicit information on fishing effort was not available. Through such analysis, I am hoping to emphasize that spatial nonstationarity is likely to exist for white abalone and should be taken into account when developing restoration strategies.

Catch at three sites (e.g., sites 655, 842 and 849 in Fig. 1) increased exponentially, which is counterintuitive because the overall catch of white abalone declined dramatically after 1978 due to depletion of this species caused by intensive exploitation. There are two possible reasons for observing such increasing trend. First, only three or four years' catch data were available for these three sites, making it difficult to capture the long-term trend. Second, the available data for these three sites were for a short period before 1980 when the overall catch of white abalone stayed relatively high. I also observed great fluctuations at two other sites (i.e., sites 654 and 700). Recall that catch data cannot infer species abundance without information on fishing effort,

and are subject to changes in fisheries management. The white abalone commercial fishery in California was historically managed based on minimum size limits, a closed season during spawning, and license fees (Davis et al., 1998; Hobday and Tegner, 2000). A minimum size limit of 120 mm (in diameter) was established for all abalone species in California in 1917, and the size limit increased to 159 mm in 1970. This increase in size limit led to a drop in harvest effort on more popular species such as pink abalone but an increase in harvest effort on less exploited species such as white abalone (Hobday and Tegner, 2000), which may partially contribute to the increasing trend of white abalone catch after 1970 that I observed at all sites with positive decay rates including these two. A commercial license for divers cost $100 in 1970, increased to $200 in 1975 and again to $250 in 1986, and then went to $330 in 1991 (Burge et al., 1975; Lundy, 1997; Hobday and Tegner, 2000). Compare to the license fee in 1970, the higher license fee in 1975 may discourage divers harvesting white abalone and contribute to the lower catch in 1975 at some sites (e.g., sites 654, 690, 700 and 709).

Studies have shown that abalone population dynamics (e.g., growth, survival and density) were closely related to environmental conditions such as water temperature (Haaker et al., 1998; Shepherd et al., 1998; Hobday and Tegner, 2002; Vilchis et al., 2005) and depth (Butler et al., 2006). However, contrary to our expectation, decline rate of white abalone catch was not significantly associated with any of the five environmental factors that I examined in this study (Fig. 3). One possibility is that I investigated too few sites to be able to detect their correlations. The other possibility is that in the years earlier than 1980 when white abalones were under intensive exploitation, impacts of environmental factors were overridden by intensive exploitation. When pooling data from all years together, it is difficult to detect the significant association with environmental factors. The third possible reason is that the significant

correlation between decline rate and environmental factors may exist at a local scale but is negligible at a global scale. For example, just looking at the southeast of the study region, catch decline appeared to be associated with distance to California coast (Fig. 2). Particularly for sea surface temperature and chlorophyll concentration, there are two more possible reasons for not showing any significant correlation with catch decline. First, these two factors may vary dramatically from year to year. Thus, one-year average seems not sufficient to represent the characteristics of these two factors and long-term data will be needed to fully explore their influences on catch decline. Second, white abalone is known as the deepest living one among the five abalone species in California, and was commonly found on sand-surrounding rock or boulder at depths of 20-60 m (Shepherd, 1973; Tutschulte, 1976; Davis et al., 1996; Hobday and Tegner, 2000). Thus, factors measured near water surface such as sea surface temperature and chlorophyll concentration may not be good indicators for population dynamics of white abalone. Additionally, chlorophyll concentration is primarily used to indicate phytoplankton abundance, which is not the dominant food source for large adult abalones. Large white abalones were found to feed on attached or drifting brown algae (Tutschulte, 1976; Hobday and Tegner, 2000).

Ocean current and habitat may also be important factors and worth investigating in future studies. Ocean current may influence the efficiency of feeding on drifting algae for white abalone, and the California coast is heavily influenced by north Pacific currents that bring waters towards shore. White abalone appears to be a habitat specialist which prefers sand-surrounding rock or boulder covered with algae (Shepherd, 1973; Tutschulte, 1976; Davis et al., 1996; Hobday and Tegner, 2000Butler et al., 2006). Sand is believed to be important in forming channels for movement and aggregation of drifting algae (Shepherd, 1973; Hobday and Tegner, 2000). Butler et al. (2006) also found the presence of white abalone at Tanner Bank to be

positively associated with flat sea floor, brown algae abundance and habitat deformed with sand. Thus, I hypothesize that habitat may be an important factor associated with catch decline. More detailed information on habitat (e.g., rock coverage, sand coverage, algae abundance especially brown algae abundance, and habitat stability) will help better our understanding of white abalone decline.

In conclusion, decline of white abalone catch showed great spatial variation, but I did not find significant evidence for the decline rate to be associated with water depth, distance to California coast, distance to land, sea surface temperature and chlorophyll concentration. This study is limited by the few fishing sites with available data and availability of information on environmental factors. Although I did not find significant associations of catch decline with these five environmental factors and I only analyzed the commercial catch data that generally cannot be used to infer population status without fishing effort information. Through this study I am hoping to emphasize that spatial nonstationarity is likely to exist for white abalone and should be taken into account when developing restoration strategies. Collection of spatially explicit information on other environmental factors such as ocean current and habitat is warranted to further explore the possible driving factors for catch decline and population dynamics of white abalone.

5.6 Acknowledgements

This research was supported by the USDA Cooperative State Research, Education and Extension Service, Hatch project #0210510 to Dr. Y. Jiao, and a grant for teaching assistantship from Department of Fish and Wildlife Conservation, Virginia Tech.

5.7 References

Allee, W.C., 1931. Animal aggregations: a study in general sociology. University of Chicago Press, Chicago.

Burge, R., Schultz, S., Odemar, M., 1975. Draft report on recent abalone research in California with recommendations for management. California Department of Fish and Game, Bodega, CA.

Butler, J., Neuman, M., Pinkard, D., Kvitek, R., Cochrane, G., 2006. The use of multibeam sonar mapping techniques to refine population estimates of the endangered white abalone (*Haliotis sorenseni*). Fish Bull 104, 521-532.

Cox, K.W., 1960. Review of the abalone of California. California Fish and Game Bulletin 46, 381-406.

Davis, G.E., Haaker, P.L., Richards, D.V., 1996. Status and trends of white abalone at the California Channel Islands. Trans Am Fish Soc 125, 42-48.

Davis, G.E., Haaker, P.L., Richards, D.V., 1998. The perilous condition of white abalone *Haliotis sorenseni*, Bartsch, 1940. J Shellfish Res 17, 871-876.

Gascoigne, J., Lipcius, R.N., 2004. Allee effects in marine systems. Mar Ecol Prog Ser 269, 49-59.

Haaker, P.L., 1994. Assessment of abalone resources at the Channel Islands. in: Halvorson, W.L., Maender, G.J., eds. The Fourth California Islands Symposium: Update on the Status of Resources. Santa Barbara, CA: Santa Barbara Museum of Natural History;

Haaker, P.L., Parker, D.O., Barsky, K.C., Chun, C.S., 1998. Growth of red abalone, *Haliotis rufescens* (Swainson), at Johnsons Lee, Santa Rosa Island, California. J Shellfish Res 17, 747-754.

Hobday, A.J., Tegner, M., 2002. The warm and the cold: influence of temperature and fishing on local population dynamics of red abalone. Rep CA Coop Ocean Fish Invest 43, 74-96.

Hobday, A.J., Tegner, M.J., 2000. Status review of white abalone (*Haliotis sorenseni*) throughout its range in California and Mexico. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southwest Region Office, Long Beach, California.

Karpov, K.A., Haaker, P.L., Taniguchi, I.K., Rogers-Bennett, L., 2000. Serial depletion and the collapse of the California abalone (*Haliotis spp.*) fishery. In: Campbell, A. (Ed), Workshop on Rebuilding Abalone Stocks in British Columbia. NRC Research Press, Ottawa, Ontario, Canada.

Lundy, A.L., 1997. The California abalone industry, a pictorial history. Best Publishing Company, Flagstaff, Arizona.

Shepherd, S., Turrubiates-Morales, J., Hall, K., 1998. Decline of the abalone fishery at La Natividad, Mexico: Overfishing or climate change? J Shellfish Res 17, 839-846.

Shepherd, S.A., 1973. Studies on southern Australian abalone (Genus *Haliotis*) I: ecology of five sympatric species. Australian Journal of Marine and Freshwater Research 24, 217-257.

Stierhoff, K.L., Neuman, M., Butler, J.L., 2012. On the road to extinction? Population declines of the endangered white abalone, *Haliotis sorenseni*. Biol Conserv 152, 46-52.

Stoner, A.W., Ray-Culp, M., 2000. Evidence for Allee effects in an over-harvested marine gastropod: density-dependent mating and egg production. Mar Ecol Prog Ser 202, 297-302.

Tutschulte, T.C., 1976. The comparative ecology of three sympatric abalones. University of California at San Diego, San Diego, California.

Vilchis, L.I., Tegner, M.J., Moore, J.D., Friedman, C.S., Riser, K.L., Robbins, T.T., Dayton, P.K., 2005. Ocean warming effects on growth, reproduction, and survivorship of southern California abalone. Ecol Appl 15, 469-480.

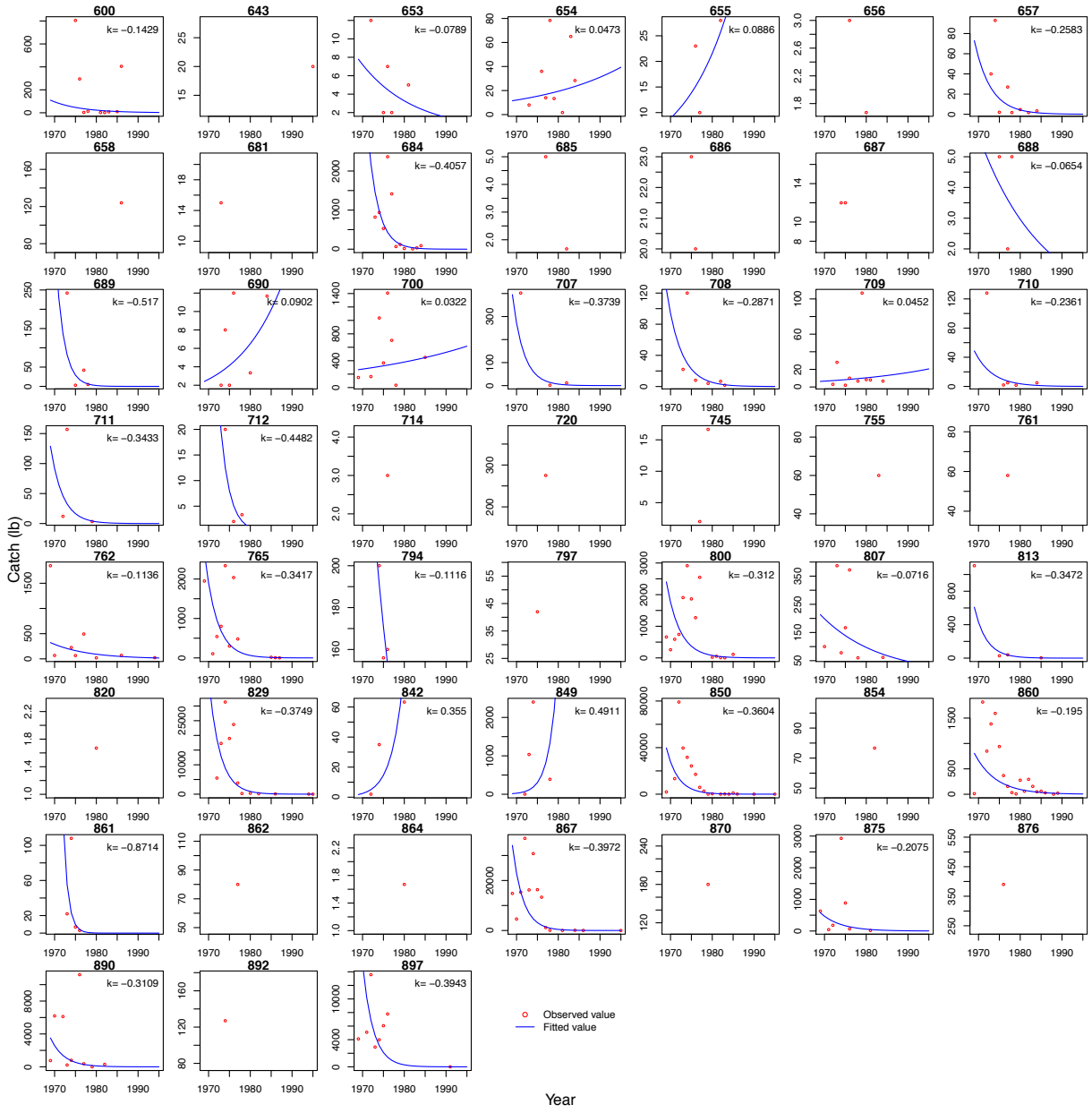

Figure 1. Spatially explicit decline rate (*k*) and decline trend of white abalone commercial catch. A total of 52 sites are plotted including those sites without decline rate estimated. The number over each plot indicates the ID number of each site from California Department of Fish and Game. Refer to Fig. 2 for the location of each site.

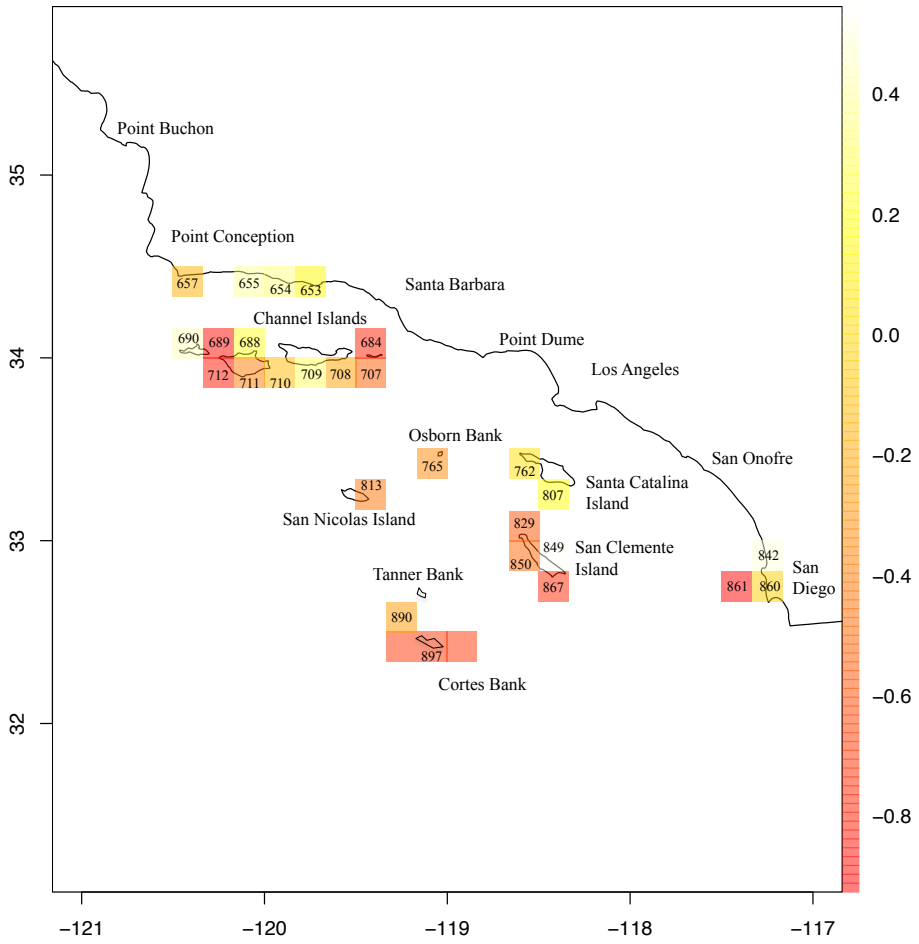

Figure 2. Spatial distribution of decline rate (k) of white abalone commercial catch. Results of the 27 sites for which decline rate were estimated are plotted. The number on each site denotes the ID number from California Department of Fish and Game.

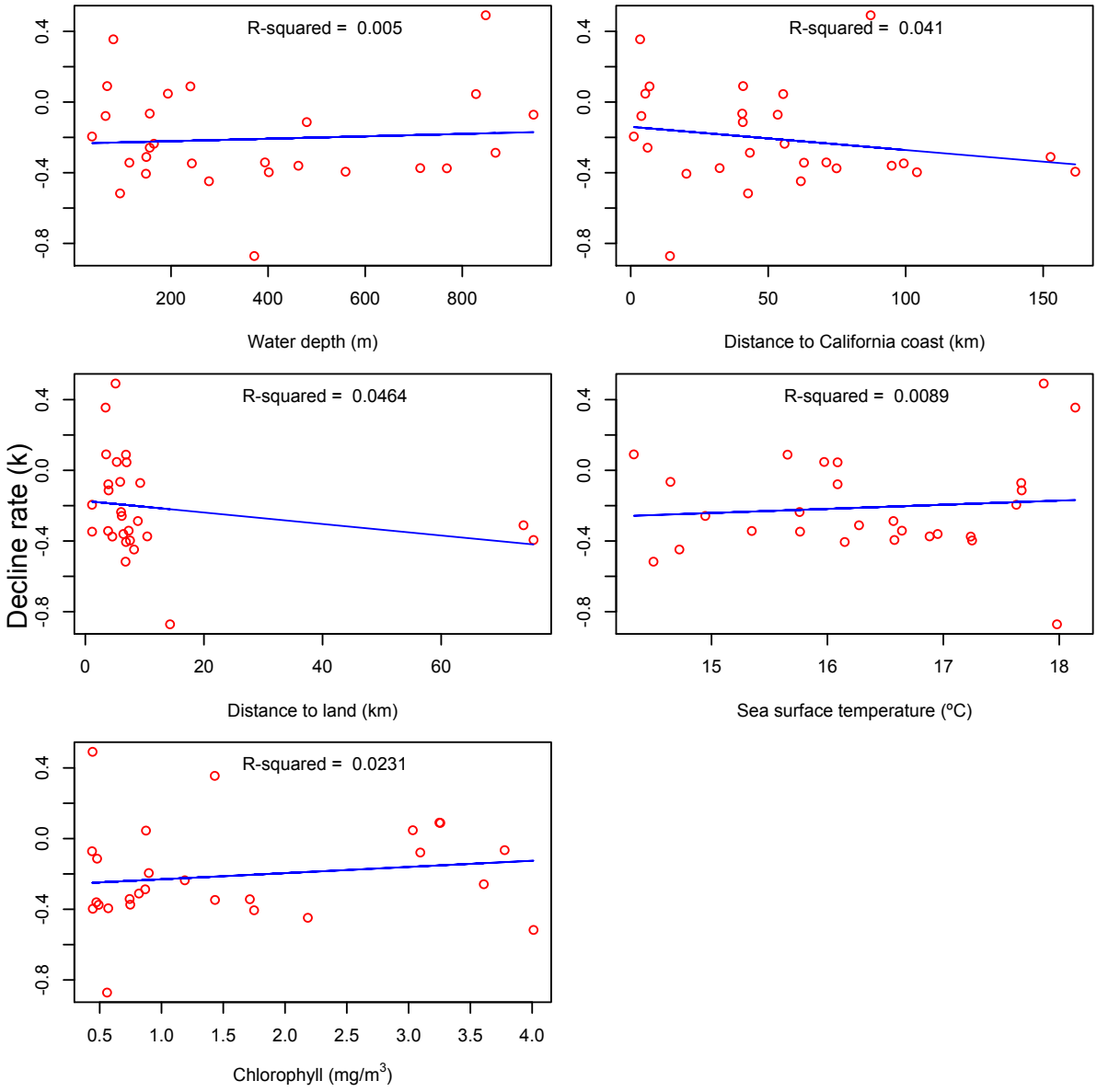

Figure 3. Correlation between decline rate (k) and environmental factors.

Chapter 6

Conclusions

With the two empirical examples of longline seabird bycatch and endangered white abalone, I demonstrated four methods to explore spatial dynamics of data-poor species, and highlighted the implications of such spatial analysis in conservation and management (Table 1). The four methods included the eigenfunction-based spatial filtering, the geographically weighted generalized linear model, the spatially explicit demographic model and the nonstationarity analysis of white abalone catch data. With the example of longline seabird bycatch, the eigenfuncion-based spatial filtering and the geographically weighted approach provided flexible tools to explore spatial dynamics in ecological data by integrating new techniques with those existing modeling approaches that deal with zero-inflated data. Both methods showed superior performance over the candidate models that I compared and revealed great spatial variation in estimated seabird bycatch or in the relationships between seabird bycatch and environmental factors. With the example of endangered white abalone, the spatially explicit approach combined with hierarchical demographic models was shown to be a competitive alternative to incorporate spatial patterns and high uncertainty due to data limitation in population dynamics modeling. The spatial nonstationarity analysis of white abalone catch data provided a valuable tool to explore spatial variation in the decline of white abalone, and showed that such a variation was very likely to exist and should be taken into account when developing restoration strategies for white abalone.

In this study, these four methodologies were modified and tailored for data-poor species. As relatively new approaches applied in ecological studies, they showed advantages that they are

flexible in customizing spatial dependence based on data and research interest, easy to modify and integrate with existing modeling approaches for data-poor species, helpful in exploration of spatial patterns in the data. Results confirmed that these four methods outperformed the traditional models and helped capture the spatial patterns. Given the high concerns on data-poor species and challenges with modeling their spatial patterns, this study provided not only implications for their conservation and management, but also alternative tools that can benefit the development of spatial models for these data-poor species.

Conclusions from this study may be limited by selection criterion, candidate models compared and data structure (e.g., the percentage of zero observations in the data or the number of observations analyzed). Although there are many other methodologies available to model spatial patterns, to deal with zero-inflated data, or to analyze data with limited information and I only explored four of them, I hope that these results will provide insights on applying or adapting existing methods for modeling spatial dynamics of data-poor species, and on utilizing information from such analyses to aid in conservation and management of these species.

Table 1

A summary of the methodologies, major conclusions and implications from this study.

Data-poor species	Data type	Example	Method	Advantage	Major conclusion and implication
Species in rare events	Zero-inflated; large dataset	Longline seabird bycatch	Spatial filtering	Flexible tool; easily modified for existing models; visualize spatial patterns; reduce computation cost	Improved model performance; provided more detailed spatial patterns
Species in rare events	Zero-inflated	Longline seabird bycatch	Geographically weighted approach	Flexible tool; easily modified for existing models; visualize spatial patterns; allow parameter to vary	Improved model performance; relationships between biological processes and environmental factors varied over space

				over space	
Species with low density	Limited information; spatially varied population dynamics	Endangered white abalone	Spatially explicit hierarchical demographic model	Incorporate high uncertainty; account for Allee effect	Density greatly affected population growth; stocking larger abalones may increase density but decrease intermediate-size proportion; restoration needs to consider Allee effect at low density and carrying capacity at high density; needs to balance the increase in density and the increase in intermediate-size proportion; needs to balance lab cost and efficacy of outplanting
Species with low density	Limited information; spatially varied population	Endangered white abalone	Nonstationary analysis	Allow parameter to vary over space; easily applied	Decline of catch varied over space; no significant association between decline rate and environmental factors; restoration needs to consider spatial nonstationarity; further exploration of

dynamics	the association between decline rate and environmental factors at both local and global scale is warranted

Appendix A

Table 1

Performace of the 108 spatial weighting matrices and the PCNM method for the POP data. The best spatial weighting matrics selected are shown in bold.

				Positive sub-model					Probability sub-model					
B[a]	A[b]	m	# pos[c]	# sel[d]	AIC	Obs MC[e]	BL MC[f]	SF MC[g]	# sel	AIC	Obs MC	BL MC	SF MC	AUC
MST	F1	-	4422	0	142.342	0.008	0.009	0.009	3	748.275	0.010	0.003	≈ 0	0.894
	F2	0.5	9010	0	142.342	0.011	0.013	0.013	3	748.741	0.010	0.003	≈ 0	0.894
		1	5862	0	142.342	0.008	0.010	0.010	3	748.476	0.010	0.003	≈ 0	0.894
		2	4438	0	142.342	0.008	0.009	0.009	3	748.294	0.010	0.003	≈ 0	0.894
		3	4422	0	142.342	0.008	0.009	0.009	3	748.276	0.010	0.003	≈ 0	0.894
	F3	0.5	9699	0	142.342	0.034	0.037	0.037	3	749.593	0.011	0.003	0.001	0.894

	1	9536	0	142.342	0.040	0.045	0.045	2	748.441	0.011	0.004	0.001	0.895
	2	5284	1	139.844	0.048	0.056	0.038	4	746.894	0.013	0.004	0.001	0.898
	3	3919	3	140.483	0.049	0.059	0.029	9	746.255	0.014	0.005	0.001	0.905
F4	0.5	1938	2	141.624	0.035	0.040	0.020	5	749.652	0.011	0.004	≈ 0	0.897
	1	2563	1	140.910	0.066	0.080	0.064	8	748.842	0.016	0.006	0.001	0.901
	2	3812	2	139.849	0.150	0.172	0.103	8	755.415	0.025	0.013	0.004	0.893
	3	4867	2	140.811	0.213	0.237	0.190	10	748.114	0.036	0.023	0.008	0.899
F5	0.5	9781	0	142.342	0.038	0.044	0.044	5	744.424	0.011	0.004	≈ 0	0.901
	1	11906	1	140.305	0.074	0.084	0.067	7	745.054	0.014	0.005	0.001	0.903
	2	**11909**	**1**	**139.478**	**0.141**	**0.158**	**0.118**	**8**	**752.745**	**0.020**	**0.010**	**0.004**	**0.898**
	3	11914	1	140.767	0.199	0.219	0.199	8	753.287	0.027	0.016	0.008	0.896
F6	0.5	8970	2	141.099	-0.004	≈ 0	-0.011	2	738.064	0.013	0.004	-0.001	0.899

		1	8948	1	141.650	-0.009	-0.006	-0.014	2	737.139	0.012	0.004	-0.001	0.901
		2	8935	1	141.673	-0.009	-0.007	-0.014	2	737.326	0.012	0.003	-0.001	0.901
		3	8961	1	141.678	-0.009	-0.007	-0.014	2	737.336	0.012	0.003	-0.001	0.899
SVG	F1	-	4034	0	142.342	0.003	0.002	0.002	1	760.153	0.002	0.001	≈ 0	0.882
	F2	0.5	9408	0	142.342	0.007	0.007	0.007	2	757.564	0.004	0.001	≈ 0	0.885
		1	5986	0	142.342	0.004	0.003	0.003	2	758.823	0.003	0.001	≈ 0	0.884
		2	4043	0	142.342	0.003	0.002	0.002	1	759.561	0.003	0.001	≈ 0	0.882
		3	4039	0	142.342	0.003	0.002	0.002	1	759.995	0.002	0.001	≈ 0	0.882
	F3	0.5	10284	1	142.217	0.026	0.028	0.025	2	756.273	0.005	0.002	≈ 0	0.887
		1	10390	1	142.309	0.035	0.039	0.037	2	741.521	0.008	0.003	≈ 0	0.899
		2	4890	1	139.978	0.047	0.055	0.038	4	746.869	0.012	0.004	0.001	0.898
		3	1934	3	140.444	0.048	0.059	0.029	9	746.146	0.013	0.005	0.001	0.905

F4	1/20	4065	1	142.037	0.006	0.005	0.001	2	754.536	0.004	0.001	≈ 0	0.887
	0.5	1869	2	141.624	0.035	0.040	0.020	5	749.653	0.011	0.004	≈ 0	0.897
	1	2563	1	140.910	0.066	0.080	0.064	8	748.842	0.016	0.006	0.001	0.901
	2	3812	2	139.849	0.150	0.172	0.103	8	755.415	0.025	0.013	0.004	0.893
	3	4867	2	140.811	0.213	0.237	0.190	10	748.114	0.036	0.023	0.008	0.899
F5	1/20	8554	1	142.112	0.007	0.007	0.003	2	755.303	0.005	0.001	≈ 0	0.886
	0.5	11905	0	142.342	0.038	0.043	0.043	5	745.237	0.011	0.004	≈ 0	0.900
	1	11906	1	140.326	0.074	0.084	0.067	7	744.983	0.014	0.005	0.001	0.903
	2	**11909**	**1**	**139.478**	**0.141**	**0.158**	**0.118**	**8**	**752.745**	**0.020**	**0.010**	**0.004**	**0.898**
	3	11914	1	140.767	0.199	0.219	0.199	8	753.287	0.027	0.016	0.008	0.896
F6	1/20	11015	1	142.331	0.011	0.011	0.007	1	746.502	0.008	0.002	≈ 0	0.892
	0.5	9299	0	142.342	-0.065	-0.072	-0.072	2	786.973	-0.007	-0.002	-0.001	0.864

		1	9284	0	142.342	-0.063	-0.069	-0.069	2	786.998	-0.007	-0.002	-0.001	0.864
		2	9303	0	142.342	-0.062	-0.068	-0.068	2	787.009	-0.007	-0.002	-0.001	0.864
		3	9341	0	142.342	-0.062	-0.068	-0.068	2	787.011	-0.007	-0.002	-0.001	0.864
DEL	F1	-	7052	2	140.965	0.817	0.802	0.823	32	661.131	0.170	0.165	0.109	0.959
	F2	0.5	7119	1	140.310	0.819	0.804	0.818	38	659.871	0.175	0.170	0.109	0.963
		1	7054	3	140.864	0.817	0.802	0.801	40	665.163	0.170	0.165	0.097	0.959
		2	7052	2	141.008	0.817	0.802	0.819	41	663.974	0.170	0.165	0.098	0.963
		3	7052	2	141.210	0.817	0.802	0.821	36	659.887	0.170	0.165	0.106	0.962
	F3	0.5	7671	4	141.806	0.834	0.820	0.853	32	683.946	0.205	0.199	0.137	0.942
		1	7305	1	142.331	0.821	0.806	0.802	33	670.105	0.186	0.180	0.113	0.948
		2	7158	1	142.201	0.817	0.802	0.823	32	655.687	0.177	0.171	0.101	0.954
		3	7130	1	142.219	0.817	0.802	0.819	34	657.338	0.175	0.170	0.101	0.957

F4	0.5	7098	1	141.290	0.817	0.802	0.821	34	661.477	0.174	0.168	0.108	0.955
	1	7181	1	142.226	0.817	0.802	0.824	38	664.522	0.178	0.172	0.105	0.956
	2	7400	0	142.342	0.818	0.803	0.803	37	680.159	0.187	0.181	0.106	0.943
	3	7602	2	142.076	0.819	0.804	0.854	39	688.248	0.197	0.190	0.116	0.945
F5	0.5	7203	0	142.342	0.819	0.804	0.804	40	655.171	0.180	0.175	0.104	0.962
	1	7366	1	142.086	0.821	0.806	0.803	39	682.412	0.188	0.182	0.110	0.943
	2	7681	0	142.342	0.824	0.810	0.810	33	684.274	0.202	0.195	0.127	0.939
	3	7982	0	142.342	0.828	0.814	0.814	39	672.437	0.214	0.207	0.128	0.944
F6	0.5	7277	0	142.342	0.820	0.805	0.805	0	793.550	0.044	0.041	0.041	0.850
	1	7524	0	142.342	0.823	0.808	0.808	0	793.550	0.012	0.010	0.010	0.850
	2	7917	0	142.342	0.829	0.815	0.815	0	793.550	0.006	0.005	0.005	0.850
	3	8182	0	142.342	0.835	0.821	0.821	0	793.550	0.006	0.004	0.004	0.850

GAB	F1	-	7562	2	141.548	0.791	0.778	0.759	31	655.610	0.236	0.229	0.130	0.942
	F2	0.5	7643	2	141.590	0.793	0.781	0.761	28	655.978	0.240	0.234	0.135	0.937
		1	7571	2	141.549	0.791	0.778	0.759	30	656.530	0.236	0.229	0.134	0.944
		2	7562	2	141.548	0.791	0.778	0.759	32	655.405	0.236	0.229	0.129	0.940
		3	7562	2	141.548	0.791	0.778	0.759	31	655.610	0.236	0.229	0.130	0.942
	F3	0.5	8150	2	141.666	0.808	0.797	0.780	32	650.798	0.272	0.265	0.150	0.937
		1	7791	2	141.609	0.794	0.782	0.763	27	652.249	0.250	0.243	0.143	0.936
		2	7649	2	141.548	0.791	0.778	0.759	29	661.572	0.240	0.234	0.136	0.936
		3	7618	2	141.558	0.791	0.778	0.759	30	656.746	0.239	0.232	0.134	0.940
	F4	0.5	7604	2	141.556	0.791	0.778	0.759	32	656.092	0.238	0.231	0.134	0.945
		1	7663	2	141.552	0.791	0.778	0.759	30	661.928	0.241	0.234	0.139	0.939
		2	7822	2	141.590	0.791	0.779	0.759	27	649.998	0.249	0.242	0.136	0.935

		3	7982	2	141.460	0.792	0.780	0.761	23	652.787	0.257	0.250	0.144	0.932
	F5	**0.5**	**7702**	**2**	**141.392**	**0.793**	**0.780**	**0.767**	**31**	**646.898**	**0.244**	**0.237**	**0.132**	**0.940**
		1	7832	2	141.615	0.794	0.782	0.763	27	651.404	0.251	0.244	0.138	0.936
		2	8076	1	141.731	0.798	0.786	0.780	30	664.559	0.264	0.257	0.153	0.932
		3	8308	2	141.583	0.801	0.790	0.772	36	656.903	0.276	0.268	0.154	0.940
	F6	0.5	7779	0	142.342	0.794	0.781	0.781	1	786.021	0.245	0.238	0.236	0.859
		1	7977	0	142.342	0.796	0.784	0.784	0	793.550	0.218	0.212	0.212	0.850
		2	8294	0	142.342	0.801	0.790	0.790	0	793.550	0.100	0.095	0.095	0.850
		3	8523	0	142.342	0.807	0.796	0.796	0	793.550	0.040	0.037	0.037	0.850
REL	F1	-	7936	1	141.490	0.773	0.768	0.763	13	765.637	0.276	0.268	0.244	0.883
	F2	0.5	7997	1	141.491	0.775	0.770	0.766	10	765.188	0.280	0.272	0.248	0.879
		1	7940	1	141.490	0.773	0.768	0.763	13	765.723	0.276	0.268	0.244	0.883

	2	7936	1	141.490	0.773	0.768	0.763	13	765.637	0.276	0.268	0.244	0.883
	3	7936	1	141.490	0.773	0.768	0.763	13	765.637	0.276	0.268	0.244	0.883
F3	0.5	8466	2	140.020	0.792	0.787	0.781	10	768.321	0.312	0.304	0.279	0.880
	1	8119	1	141.489	0.776	0.771	0.767	8	766.813	0.288	0.280	0.256	0.877
	2	7995	1	141.489	0.773	0.768	0.763	12	765.244	0.279	0.271	0.247	0.882
	3	7976	1	141.489	0.773	0.768	0.763	13	765.918	0.278	0.270	0.246	0.883
F4	0.5	7959	1	141.489	0.773	0.768	0.763	12	766.076	0.277	0.269	0.245	0.881
	1	8004	1	141.489	0.773	0.768	0.763	12	765.239	0.279	0.271	0.247	0.882
	2	8114	1	141.489	0.773	0.768	0.764	12	765.025	0.285	0.277	0.250	0.883
	3	8252	1	141.488	0.774	0.768	0.764	11	763.585	0.292	0.283	0.258	0.883
F5	0.5	8039	1	141.489	0.775	0.769	0.765	11	764.429	0.283	0.275	0.251	0.882
	1	8147	1	141.489	0.776	0.771	0.767	9	766.577	0.289	0.281	0.256	0.878

		2	8361	1	141.491	0.779	0.774	0.770	12	766.913	0.300	0.292	0.268	0.882
		3	8545	2	141.166	0.783	0.777	0.773	10	766.719	0.311	0.302	0.276	0.879
	F6	0.5	8108	1	141.489	0.775	0.770	0.766	5	786.117	0.285	0.277	0.268	0.866
		1	8277	0	142.342	0.778	0.773	0.773	0	793.550	0.275	0.267	0.267	0.850
		2	8562	0	142.342	0.783	0.778	0.778	0	793.550	0.180	0.175	0.175	0.850
		3	8761	0	142.342	0.788	0.783	0.783	0	793.550	0.088	0.085	0.085	0.850
PCNM	-	-	2271	0	142.342	-0.015	-0.014	-0.014	3	748.497	0.010	0.002	≈ 0	0.894

[a]: The *B* refers to the connectivity matrix.

[b]: The *A* refers to the spatial weighting function matrix.

[c]: the number of positive spatial filters generated from a spatial weighting matrix.

[d]: the number of spatial filters selected into the sub-model.

[e]: the Moran's I coefficient of the observations.

[f]: the Moran's I coefficient of the response residuals for the baseline model without spatial dependence considered.

[g]: the Moran's I coefficient of the response residuals for the model modified with spatial filters.

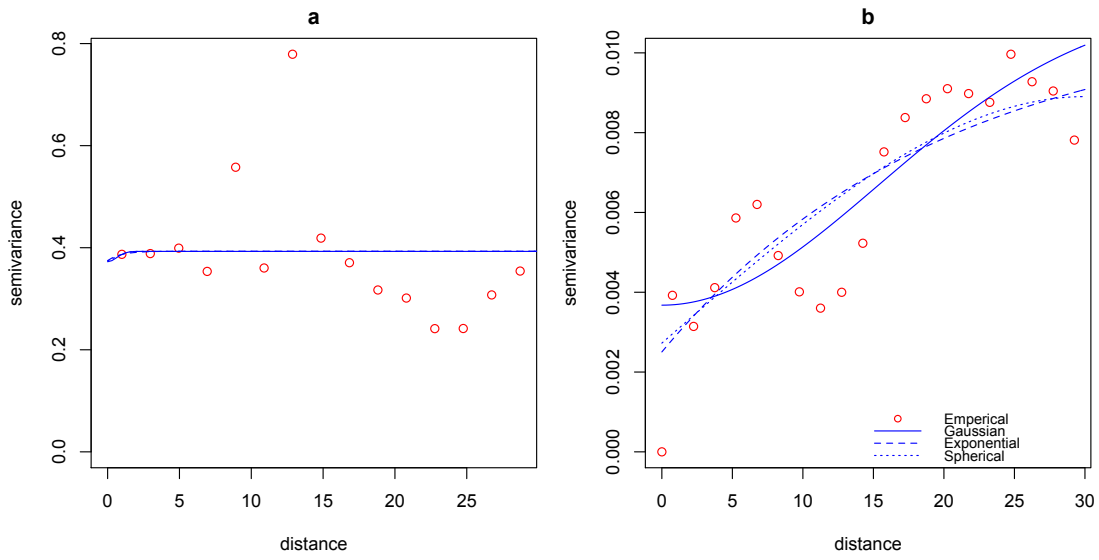

Figure 1. Empirical and fitted variograms for logrithm of the positive catch data (a), and the presence/absence data (b) of seabird bycatch. Distance (10^2 km) is the Euclidiean distance between two locations.