5-2010

Identifying Critical Fish Habitat and Long-term Trends in Fish Abundances in the Hudson River Estuary

Megan Patricia O'Connor

University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/open_access_dissertations

Part of the Aquaculture and Fisheries Commons

Recommended Citation

https://scholarworks.umass.edu/open_access_dissertations/249

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
IDENTIFYING CRITICAL FISH HABITAT AND LONG-TERM TRENDS IN FISH ABUNDANCES IN THE HUDSON RIVER ESTUARY

A Dissertation Presented

by

MEGAN P. O’CONNOR

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2010

Wildlife and Fisheries Conservation
IDENTIFYING CRITICAL FISH HABITAT AND LONG-TERM TRENDS IN FISH ABUNDANCES IN THE HUDSON RIVER ESTUARY

A Dissertation Presented

by

MEGAN P. O’CONNOR

Approved as to style and content by:

Francis Juanes, Co-Chair

Charles M. Schweik, Co-Chair

Kevin McGarigal, Member

Robert Newton, Member

Paul Fisette, Department Head
Natural Resources Conservation
DEDICATION

For my great aunt and guide, Sister Mary Crescentia O’Connor
ACKNOWLEDGMENTS

Many people contributed to this research, and I wish to extend my heartfelt thanks to all of them. The presented work would not have been possible without the support of my co-advisors Francis Juanes and Charlie Schweik. Specifically, Dr. Juanes helped me procure individual funding and encouraged me to write proposals for financial support throughout this research endeavor. Dr. Schweik opened the world of data management and Geographical Information Systems (GIS) and my research took on new dimension. I would also like to thank my committee members Kevin McGarigal and Robert Newton for their invaluable contributions. Dr. McGarigal was paramount in teaching me and guiding me through the ups and downs of applying and understanding multivariate statistics. And Dr. Newton helped me understand hydrology and also provided a quiet place for me to work at Smith College. Jon Caris, the GIS expert from Smith College, saved me from the black hole of ESRI software and was my geostatistical and cartographic critic (and he always found me space on a server!). Lastly, the graduate students in the Department of Natural Resources Conservation were also extremely supportive and shared their statistical knowledge with me- Sandra Haire, Jenn Seavey, Ethan Plunkette, Bill DeLuca, and Richard Chandler.

Many people and agencies contributed necessary data to this study and I am truly grateful because the presented research relied solely on secondary data. The Poughkeepsie water Treatment facility provided daily water temperatures from 1974 to 2005. And John Young from ASA Inc. provided me with the bulk of the necessary data that included the utilities’ year class survey from 1974 to 2005. John also answered all of my questions regarding data issues and concerns associated with the survey.
I am also very grateful for the funding provided through NOAA’s NERR Graduate fellowship program and Hudson River Foundation Graduate Research Fellowship program.

Finally, I thank my husband and fellow graduate student, Steve Gaurin, for his support and I especially thank my son, Nicholas, for helping me keep it all in perspective. I love you both more than words can say.
The Hudson River estuary (HRE) is a well monitored aquatic resource and much secondary data exist for this system. We developed two objectives based on accessible HRE aquatic data. The first objective was to determine if changes in HRE fish community over the time period (1974 to 2005) years are correlated to local and regional climate. We addressed this objective by employing a multivariate statistical approach. We confirmed that the HRE fish community structure has changed over the time period (1974 to 2005). These changes are correlated with local hydrology (freshwater flow and water temperature) and regional climate (Atlantic Multidecadal Oscillation or AMO and North Atlantic Oscillation or NAO). We found that abundances of striped bass larval stages are positively correlated with high freshwater flows and juvenile shad abundances are negatively correlated with the AMO or warmer sea surface temperatures (SST). This finding suggests that climate-related variability affects HRE juvenile shad abundances and current management strategies for this declining species should include the implications of climate change.
The second objective was to examine whether factors such as sediment type, water characteristics and distance to nearest submerged aquatic vegetation (SAV) beds affect the occurrence or presence/absence of juvenile American shad (*Alosa sapidissima*) and juvenile striped bass (*Morone saxatilis*) in the HRE during the fall. We addressed this objective by applying geostatistics and general linear mixed effects models. We found the probability of presence for both species were commonly driven by spatial dependence or river mile, Julian day and salinity. Our results include maps depicting probability of occurrence (or presence) for both species throughout the HRE. We found the highest predicted probabilities of juvenile American shad presence are found in the Upper HRE. Conversely, highest predicted probabilities of juvenile striped bass presence are found in the Lower HRE. Habitat partitioning between these two species is present during the fall in this system but the mechanism is unclear. Future studies could address a possible predator-prey or competitive relationship between juvenile American shad and juvenile striped bass.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xii</td>
</tr>
</tbody>
</table>

## CHAPTER

1. INTRODUCTION ........................................................................................................ 1

2. ADVANTAGES, CHALLENGES AND LIMITATIONS TO USING SECONDARY DATA IN AQUATIC RESEARCH ........................................ 7  
   2.1 Abstract ........................................................................................................ 7  
   2.2 Introduction ................................................................................................... 7  
   2.3 A Theoretical Framework for Data Access Models ...................................... 11  
   2.4. Overview of utilizing secondary data in hypothesis-driven research ........ 15  
      2.4.1 Background ...................................................................................... 15  
      2.4.2. Obtaining Available Secondary Data ............................................ 15  
      2.4.3. Understanding Available Secondary Data ...................................... 16  
      2.4.4. Utilizing Available Secondary Data ................................................ 17  
   2.5. Case Study: Hudson River Estuary Aquatic Data .......................................... 18  
      2.5.1. Background ..................................................................................... 18  
      2.5.2. Obtaining Available Secondary Data on the HRE .................................. 19  
      2.5.3. Understanding Available Secondary Data on the HRE .......................... 22  
      2.5.4 Utilizing Secondary Data on the HRE and Challenges Encountered .......... 23  
   2.6 Closing Remarks ............................................................................................. 25  

3. LONGTERM EFFECTS OF LOCAL HYDROLOGY AND REGIONAL CLIMATE ON THE HUDSON RIVER ESTUARINE FISH ASSEMBLAGE .. 32  
   3.1 Abstract ........................................................................................................ 32  
   3.2 Introduction ................................................................................................... 33  
   3.3 Methods ........................................................................................................... 35
3.3.1 Study Area ................................................................. 35
3.3.2 Fish Community Data ............................................... 36
3.3.3 Climate indices .......................................................... 37
3.3.4 Hydrological Variables ........................................... 39
3.3.5 Data Analysis .......................................................... 40

3.4 Results ........................................................................... 41
3.5 Discussion ...................................................................... 43

4. DISTRIBUTION AND HABITAT UTILIZATION OF JUVENILE AMERICAN SHAD AND STRIPED BASS IN THE HUDSON RIVER ESTUARY ............. 55

4.1 Abstract ........................................................................... 55
4.2 Introduction ..................................................................... 56
4.3 Methods .......................................................... 58
   4.3.1 Study Area ........................................................... 58
   4.3.2 Fish Data .......................................................... 58
   4.3.3 Environmental Data .............................................. 59
   4.3.4 Analytical Approach ........................................... 61

4.4 Results ........................................................................... 62
4.5 Discussion ...................................................................... 65

5. A CRITICAL EXAMINATION OF FINDINGS, AND PROPOSED FUTURE DIRECTIONS ..................................................... 82

5.1 Climate and Hudson River American shad ..................................................... 82
5.2 Relationship between Juvenile American shad and striped bass in the Hudson River Estuary ............................................. 84

LITERATURE CITED ......................................................................... 90
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Two by two matrix denoting the theory of goods, based on Ostrom, Rules, Games, and Common Pool Resources, University of Michigan Press, 1994</td>
</tr>
<tr>
<td>2.2. Secondary data used to address the research question, ‘How does the HRE fish assemblage respond to climate and hydrological variables over time (1974-2005) (Chapter 3’s research question)?’</td>
</tr>
<tr>
<td>2.3. Secondary data used to address the research question, ‘Do abiotic factors such as sediment type, distance to nearest submerged aquatic vegetation (SAV) bed, salinity, water temperature and dissolved oxygen influence the distribution of juvenile American shad and juvenile striped bass in the Hudson River estuary (Chapter 4’s research question)?’</td>
</tr>
<tr>
<td>3.1. Fish Abundance Indices from 1974 to 2005 used in the multivariate analyses.</td>
</tr>
<tr>
<td>4.1. Pearson correlation coefficients between all pairs of environmental variables.</td>
</tr>
<tr>
<td>4.2. Parameter estimates from the global generalized mixed models (GLMM).</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. Map of the study area, Hudson River estuary (New York, USA). Study area, Hudson River estuary from the Battery to Albany.</td>
<td>6</td>
</tr>
<tr>
<td>2.1. An example of data layers used in data processing and data analysis for addressing the second research question (Chapter 4).</td>
<td>31</td>
</tr>
<tr>
<td>3.1. Map of the study area, Hudson River estuary (New York, USA). Study area, Hudson River estuary from the Battery to Albany.</td>
<td>50</td>
</tr>
<tr>
<td>3.2. Mantel correlogram showing the correlation in the fish community structure in relation to the time (years) between observations. Each point shows the correlation between the fish community at different time intervals (represented by the distance class); solid circles represent significant correlations. The monotonically decreasing correlation over time indicates a long-term trend; i.e., fish community similarity decreases steadily over time.</td>
<td>51</td>
</tr>
<tr>
<td>3.3. Constrained ordination triplot based on redundancy analysis depicting the first and second constrained axes (p&lt;0.01). The length and direction of the vectors reveals the strength and direction, respectively, of maximum rate of change in the corresponding variable. Environmental variables are depicted with dashed vectors and species-stage variables are depicted with solid vectors. Species codes are listed in Table 3.1. AMOextr and NAO extr denote the respective extreme yearly climate index. And FlowMax and TempMax denote maximum yearly values of freshwater flow and water temperature in the HRE.</td>
<td>52</td>
</tr>
<tr>
<td>3.4. Enhanced ordination plot depicting the relative abundance of juvenile shad each year from 1974-2005. The solid circles represent years; the size of the circle is proportional to the relative abundance of juvenile shad. Superimposed on the ordination plot is a fitted surface of the Atlantic Multidecadal Oscillation (AMO) based on a generalized additive model. The regular and straight contours indicate that a linear model was optimal. The contour labels represent the predicted values of AMO.</td>
<td>53</td>
</tr>
<tr>
<td>3.5. Venn diagram showing the variance partitioning of the constrained ordination. The non-overlapping portion of each circle represents the percentage of “variance” in the fish community explained independently by the corresponding constraining variables (e.g.,</td>
<td></td>
</tr>
</tbody>
</table>
hydrology) after accounting the other constraining variables. The overlapping portion of the circles represents the confounded portion of the explained variance; i.e., the portion that cannot be distinguished between the two sets of constraining variables. The “residual” is the unconstrained inertia or the total “variance” in the species data set left unexplained or unaccounted for by the environmental constraints; it is equal to the sum of the unconstrained eigenvalues.


4.2. Plot of the probability of juvenile American shad and juvenile striped bass presence versus Julian day in the HRE based on the global GLMMs.................................................................74

4.3. Plot of the probability of juvenile American shad and juvenile striped bass presence versus river mile in the HRE based on the global GLMMs.................................................................75

4.4. Plot of the probability of juvenile American shad and juvenile striped bass presence versus salinity in the HRE based on the global GLMMs........76

4.5. Plot of the probability of juvenile striped bass presence versus water temperature in the HRE based on the global GLMMs..........................77

4.6. Calibration plot for juvenile American shad global GLMM....................78

4.7. Calibration plot for juvenile striped bass global GLMM...........................79

4.8. Map of the probability of juvenile American shad presence in the HRE during the fall based on the global GLMM (2000-2005). Higher probabilities are depicted in red and lower probabilities are depicted in blue. ......................................................................................80

4.9. Map of the probability of juvenile striped bass presence in the HRE during the fall based on the global GLMM (2000-2005). Higher probabilities are depicted in red and lower probabilities are depicted in blue. ......................................................................................81

5.1. Cross-correlation function (CCF) plot for HRE juvenile American shad abundance indices and the AMO over the time period, 1974 to 2005.................................................................89
5.2. Conceptual model of population-level variation in reproductive characteristics for adult American shad under the influence of increased water temperatures. ................................................................. 89

5.3. Map of the probability of juvenile American shad presence in the HRE during the summer based on the global GLMM (2000-2005). Higher probabilities are depicted in red and lower probabilities are depicted in blue. ...................................................................................... 89

5.4. Map of the probability of juvenile striped bass presence in the HRE during the summer based on the global GLMM (2000-2005). Higher probabilities are depicted in red and lower probabilities are depicted in blue. ...................................................................................... 89
CHAPTER 1

INTRODUCTION

The Hudson River extends 507 km from its source, Lake Tear of the Clouds in the Adirondack Mountains, to New York Harbor. It is tidally influenced up to a dam located near Troy, NY (250 km) and thus defines Hudson River estuary (HRE) (Figure 1.1). The New York State Department of State has designated large portions of the HRE as a Significant Coastal Fish and Wildlife Habitat (USFWS 1997). Four coastal wetlands within the HRE are designated by the National Oceanic and Atmospheric Agency (NOAA) as National Estuarine Research Reserve System (NERRS) sites. The Hudson River is one of the most studied aquatic systems in North America and much research has been conducted to identify population trends in its fish species (Pace et al. 1993, Daniels 1995, Strayer et al. 2004, Hurst et al. 2004). The estuarine portion of the Hudson River is composed of a diverse assemblage of marine, freshwater and estuarine fishes. Beebe and Savidge (1988) report that 140 species have been collected in the HRE and approximately one dozen of these species are diadromous (Waldman 2006).

There is much evidence supporting the observations that conventional fish abundance and distributions are changing or shifting in the HRE but the mechanisms are unclear (Pace et al. 1993). Daniels (1995) examined the changes in abundance and distribution of nearshore fish assemblages from fish collected 1936 to 1990 throughout the HRE. He concluded that during the past 60 years the component species have changed, and species diversity is decreasing. There is strong evidence that exotic species, mainly the zebra mussel (Dreissena polymorpha), have altered the fish communities in the Upper HRE (USFWS 1997, HNERR, Schmidt 2001-2004, and Strayer et al. 2004).
Strayer et al. (2004), while studying the effects of zebra mussels in the HRE, found that populations of open-water species (e.g., \textit{Alosa} spp.) had shifted downriver away from the zebra mussel populations and decreased in abundance. Conversely, populations of littoral fish species (e.g., Centrarchidae) had shifted upriver and increased in abundance. Hurst et al. (2004) examined nearshore fish community structure in the mesohaline region (Tappan Zee to Haverstraw Bay or rkm 36 to 64) of the HRE. They identified a long term decline in fish community diversity and stability after analyzing 21 years of fish data.

The Hudson River Estuary Management Action Plan (HREMP) was enacted in May 1996, and focused on an ecosystem approach to managing the estuary. A specific priority of the HREMP is to manage fisheries for sustainable use and to identify habitat needs to preserve biodiversity (HREMP 1996). The first HREMP was updated in December 2005 and includes the Hudson River Action Agenda for 2005 to 2009. Its goals are “to restore the signature fisheries to their full potential” and to conserve and protect critical river habitat to assure that these characteristic species are supported throughout their life stages. The HREMP states that currently declining species including American shad (\textit{Alosa sapidissima}), Atlantic sturgeon (\textit{Acipenser oxyrhynchus}), river herring (\textit{Alosa pseudoharengus} and \textit{Alosa aestivalis}), American eel (\textit{Anguilla rostrata}) and largemouth bass (\textit{Micropterus salmoides}) and the recovering striped bass fishery (\textit{Morone saxatilis}) need to be managed through an ecosystem approach. Evaluating fish habitat suitability is goal of the HREMP, however no procedures were described on how this goal was to be achieved (NYSDEC 2005).
The primary goals of this dissertation were to elucidate possible abundance trends in HRE fish community over the time period 1974 to 2005 and to identify important fish habitat for juvenile American shad and juvenile striped bass in the HRE. Two approaches were used to address the primary goals or objectives of this dissertation research. The first approach is strictly temporal and non-spatial and includes elucidating trends in HRE fish abundances due to climate change over the time period 1974 to 2005. The second approach is spatial and focuses on identifying important fish habitat in the HRE using Geographical Information Systems (GIS) and associated spatial analyses. But first we examine the feasibility of using secondary data to address these goals because of the large spatio-temporal scale necessary to deal with this approach.

The dissertation is organized into five chapters. The second Chapter is a reflection on the advantages, challenges and limitations to using secondary data in aquatic research. We used only secondary data in the research presented in Chapters 3 and 4. Acquiring and utilizing secondary data were challenging due to data access constraints and also due to varying and often disparate spatio-temporal data scales. Much secondary data exist for the HRE and they include many abiotic and biotic measurements from various surveys. Specifically, there has been an annual intensive multispecies fishery-independent survey since 1974 when year class reports were first prepared for the electric utility generating stations. The goal of these reports is to determine the environmental impact that these utility companies have on the HRE fish populations. The study area begins at the Battery (km 0) and ends in Albany (km 240) (ASA 2004). Over thirty years of spatio-temporal fish abundance data exist due to this survey. The utilities’ survey consists of the Long River Survey (LRS), designed to assess egg and larval densities; the Fall Shoals Survey
to assess juvenile densities offshore; and the Beach Seine Survey (BSS), designed
to assess nearshore fish communities and abundance (Limburg et al. 2006). Many studies
have utilized these fishery-independent survey data including Pace et al. (1993), Daniels

The third Chapter is an investigation of the effects of local hydrology and regional
climate on the HRE fish community over the time period 1974 to 2005. The effects of
climate change are a central contemporary issue that is at the forefront of many
ecological research endeavors (Walther et al. 2002). The HRE has been affected by
changes in climate and sea level but there have been no studies attempting to associate
these phenomena with observable variations within the fish community (Waldman et al.
2006). Addressing these effects will change how we manage our resources presently and
in the future. Quantifying the effects of fluctuation in these patterns on fish population
dynamics could help Hudson River fisheries managers better manage the resource.

The fourth Chapter describes habitat suitability for juvenile American shad and
striped bass during the fall season in the HRE. My approach includes performing
geostatisical analyses and applying generalized mixed models (GLMMs) on available
HRE data existing as GIS, including sediment type and submerged aquatic vegetation
characteristics (CUGIR and NYSDEC), water quality parameters (ASA 2004) and fish
distribution and creating quantitative maps to evaluate fish habitat suitability in the HRE.
These habitat suitability maps can be directly applied to identify critical fish habitat, a
goal of the HREMP (NYSDEC 2005). In order to conserve habitat and produce
sustainable fish resources critical fish habitat needs to be identified throughout the
estuary.
The fifth and final Chapter includes a brief summary of the previous three chapters that include a critical examination of findings. Overall, we found that the HRE fishes are influenced by regional climate and local hydrology over the time period 1974 to 2005. Specifically, we confirmed that larval striped bass have a strong positive relationship with freshwater flow and juvenile striped bass have a strong negative relationship with the AMO. These two species also utilize the estuary in different ways during the juvenile stage, American shad primarily occur north of Poughkeepsie, New York and striped bass primarily occur south of Poughkeepsie during the fall season. This chapter also proposes future research approaches and includes specific details regarding fulfilling the proposed studies.
Figure 1.1: Map of the study area, Hudson River estuary (New York, USA). Study area, Hudson River estuary from the Battery to Albany.
CHAPTER 2

ADVANTAGES, CHALLENGES AND LIMITATIONS TO USING SECONDARY DATA IN AQUATIC RESEARCH

2.1 Abstract

Secondary analyses are useful for aquatic research that encompasses large time scales and spatial scales. We used this approach while researching Hudson River estuary (HRE) fish ecology concerns with time series and spatial statistical methods. In this paper, we report on data access and data utilization issues in a general and specific framework so that others may gain insight to this approach.

2.2 Introduction

Much data exist due to extensive monitoring of aquatic resources. For this reason, conducting comprehensive aquatic ecological analyses could be executed by acquiring or referencing available databases in preference to conducting extensive field or lab studies. Available databases or data that have already been collected are sometimes referred to as “secondary data” and subsequent analyses are termed “secondary analyses” (Kiecolt and Nathan 1985). The use of secondary data is customary in many fields of research including the social sciences and economics\(^1\) but the practice is just becoming more popular in the aquatic sciences.

The increase in long-term data gathering projects in the United States concerning aquatic ecosystems was driven, at least in part, as a response to the enactment of the

\(^1\) See, for example, the Inter-University Consortium for Political and Social Research at the University of Michigan that acts as a curator for many social science datasets, in part, to encourage reuse (http://www.icpsr.umich.edu/icpsrweb/ICPSR/).
National Environmental Policy Act of 1969 
(http://www.nepa.gov/nepa/regs/nepa/nepaeqia.htm) which requires Environmental Impact Statements and Assessments to be developed for projects that receive federal funding or are performed by federal agencies. Because monitoring studies or surveys are usually conducted at regular intervals (e.g. annually) over many years under a standard sampling design and can encompass an entire ecological system, they can add to understanding the environment in ways never intended through subsequent hypothesis-driven research, a type of secondary analysis. Hypothesis-driven research allows the investigator to focus on specific questions and some granting agencies require this type of approach (O’Malley et al. 2009). This type of research differs from exploratory or data-driven research, another useful type of secondary analysis, which is typically used to analyze secondary data. Therefore, available monitoring data or secondary data are valuable because they provide limitless opportunities for the replication, re-analysis and re-interpretation of existing research (Smith 2008).

Increasingly, we are moving toward a situation where available data exist online that others are free to use despite the ongoing issues and pressures around “enclosure” of information (e.g., Boyle 2003) that the “open-access” movement is pushing to stop. There are two examples of advances being made, one technical, and the other more information policy-related. First, from a technological perspective, there is the growing use of XML and technologies using “web services” where data are provided online and are accessible in real time over the Internet (Dangermond, 2009). As continued developments are made in environmental monitoring devices with data loggers that are connected to the Internet there will be even more opportunities for the use of secondary
data collected by others. Of course, this kind of situation is already here. For example, oceanographers and meteorologists have access to data in real time through the National Oceanic and Atmospheric Administration (NOAA) nowCoast program (http://nowcoast.noaa.gov/).

Second, from an information policy perspective, currently there is a serious push, at least at the U.S. Federal government level, toward data sharing and data transparency in the public sector. “Data produced by government agencies are often hard to find or are published in proprietary formats of limited utility. As a result, a wealth of information remains untapped by the ingenuity and creativity of the American people” (http://www.whitehouse.gov/open/about/). Therefore, the Obama Administration has recently launched the Open Government Initiative (http://www.whitehouse.gov/Open/) and, associated with that, the data.gov website which provides access to non-sensitive federal government data sets in hopes of stimulating economic, scientific and educational innovation. Government funding agencies such as the National Science Foundation (NSF) are mandating that data are made available after some reasonable time period for publishing of results (NSF 2009).

Back in the mid-1980s, Kiecolt and Nathan (1985) noted that new data collection by individual researchers had become difficult due to the current economic climate and declining resources for research in the social sciences. Therefore they emphasized that social scientists should rely on secondary data. This same sentiment can also be applied to other fields of research especially during our current depressed national and global economic environment, the late 2000’s recession due to the national housing market and global banking collapses. The social sciences have made great strides to using secondary
data over the past 20 plus years (Smith 2008) and it is thought that other fields such as ecology could see similar gains if data are made available for their particular area of study (Parr and Cummings 2005).

Some in the field of ecology are working to address these needs through the Knowledge Network for Biocomplexity (KNB) (http://knb.ecoinformatics.org/index.jsp). The KNB is a national network that facilitates the discovery, access, interpretation, integration, and analysis of complex ecological data collected from a highly distributed set of field stations, laboratories, research sites, and individual researchers. Their products include an XML-based Ecological Modeling Language that is used to better articulate ecological data in part to assist in the implementation of web data services and, more recently, could assist in the enhancing the creation of ecological data “mashups.” This relatively new concept describes situations where an interested party combines data from multiple sources (often two different sets of government-produced data) to create new information using the use of so-called Web 2.0 web-database technologies. Most current mashups are community based either in politics or demographics (e.g., Google Maps mashups and the Sunlight Foundation), but, like web services, they point to where much of this data-sharing is headed. Our research approach described later in the paper could be considered one form of a mashup.

As more organizations and individuals make their data available over the Internet, the question will turn to the issues and challenges of pulling these data together to support sound and robust analysis. In this paper, we report on some of the advantages, challenges and limitations to using secondary aquatic data in hypothesis-driven research. This discussion is based on our experiences encountered while researching Hudson River
estuary (HRE) fish ecology concerns. Much data exist concerning the HRE, including both biotic and abiotic data sets that encompass over 30 years and also encompass the entire estuary. These data exist because the electric utility plants operating on the HRE are mandated by law to perform these yearly surveys. Because of these extensive secondary data there was an opportunity to develop research questions concerning the effects of local and regional climate on the fish assemblage over the time period 1974 to 2005, as well as separate research questions focusing on factors that influence suitable habitat for juvenile striped bass (*Morone saxatilis*) and juvenile American shad (*Alosa sapidissma*) (Chapter 3 and Chapter 4). Driven by this experience, and given the increasing opportunities emerging for the use of secondary data in the areas of fish ecology (as well as in other areas of science) it is useful to reflect on this experience.

This paper is organized into three parts. First, we describe issues concerning access to secondary data. Second, we discuss some advantages as well as challenges and limitations of using secondary data in research and use our HRE fish ecology research as an example. Third, we conclude the paper with insights from this effort that may be of interest to others pursuing this approach.

### 2.3 A Theoretical Framework for Data Access Models

The arguments associated with data sharing or data access are based on the differences between two data types, open-access and closed-access data. However, data access is not binary but functions on a gradient or spectrum. Because data access issues are very complex, it may be helpful if we first look at the “theory of goods” concept (Samuelson, 1954; Ostrom and Ostrom, 1977; Ostrom, Gardner and Walker, 1994) and how it relates to data access. The theory of goods is represented a 2-by-2 matrix (Table
2.1) that is organized by two parameters: (1) Excludability and (2) Subtractability. Excludability refers to the question of whether it is difficult to prohibit people from data access (yes, no); Subtractability is the question of whether the data are diminishable (yes, no) (Table 2.1).

Depending on the situation, digital data could be treated as a private good (closed-access) or a public good (open-access). But an organization could also limit how the data are shared (a nonsubtractable good) and exclude people from using them or only allow some organizations in the “club” access to them (a club good). Therefore, digital information can fall under three of these categories (private, public, and club, shown in Table 2.1) depending on the policies and procedures of the organization. The fourth category (common-pool) does not usually apply to digital data because they are not really a subtractable resource (unless they are only provided on, say for example, a CD where they can’t be copied – then it moves more to a private good).

The data access spectrum, defined by the “theory of goods”, is demonstrated by Creative Commons (http://creativecommons.org), a nonprofit organization founded in 2001 that provides a tool for people to grant copyright permission for their work. Creative Commons makes the argument that “content” created by people and their rights to it do not fall in the binary choice of full copyright/all rights reserved/closed-access versus public domain/no rights reserved/open-access, but rather a spectrum of rights/restrictions or “some rights reserved”. For example, an individual data developer might make data open-access but put restrictions on what can be done with them.

In the work reported here, we realized the complexities associated with data access and were fortunate to find ourselves at opposite ends of the data access spectrum
by procuring only two types of data, open-access and closed-access. We did not use open-access data that had requirements or use restrictions attached to them and were able to develop simple definitions for data access. We define open-access data as secondary data that are accessed without permission and without costs (Suber 2007). Using our definition, open-access is equivalent to the public-good cell in Table 2.1. We define closed-access data as secondary data that are accessed only with permission from the data owner. Permission could encompass a wide range of scenarios but we will define it as merely obtaining formal consent or authorization requiring some form of communication with the data owner and/or creator. These would fall under the private or club good categories in Table 2.1.

The differences between open-access and closed-access data can have direct impacts on conducting hypothesis-driven aquatic research. Many aquatic ecologists depend on open-access, secondary data to conduct or support their research. Our research endeavor, which we discuss more in following sections, is a case in point, where we depended on access to HRE fish survey data as well as other aquatic survey data collected throughout the HRE. We found that much of the open-access data are products of government funded research and surveys, whereas the closed-access data are indicative of private sector survey projects. Open-access data were viewed as an impetus while closed-access data were viewed as an impediment to meeting our research goals.

We also want to emphasize the importance of data access not only from the researchers’ view but also from the public’s view. First, we must mention that there are often policies in place for data access. For example, State Governments can exercise a copyright over their works, but the New York State Senate believes in the adoption of
open licenses (e.g., http://creativecommons.org/licenses/by-nc-nd/3.0/) for the public to use and share the content on their website provided it is not for political fund raising purposes (http://www.nysenate.gov/). Because there are restrictions associated with this Creative Commons license it can also serve as an example of digital information that exists on the data access spectrum.

Interestingly, there are many existing policy models for data access, most of which originate from the international world (e.g. CODATA, Committee on Data for Science and Technology and part of the International Council for Science). But it is important for the public to know that these are not compulsory and it is up to the data owner to enter into these practices (Uhlir and Schroder 2007). The need for sharing data in ecology is directly addressed in The National Academy of Science (NAS) report (2003), “Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences”. This document is meant to give support to those who are frustrated by a lack of adequate sharing or access. Community Standards or practices for sharing data, software and materials are laid out in the report and were built upon the idea of the uniform principle for sharing data in a timely manner. But these were merely guidelines and have not been translated to government policy.

Lastly, Arzberger et al. (2004b) state “Ensuring research data are easily accessible, so that they can be used as often and as widely as possible, is a matter of sound stewardship of public resources.” This opinion is repeated by other critics who further point out that publicly funded data should be openly accessible to other researchers to ensure that the public receive optimum returns on their investment in science (Maurer et al. 2000, Arzberger et al. 2004a, Parr and Cummings 2005, Uhlir and
Schroder 2007). Parr and Cummings (2005) feel that the field of Ecology has not experienced the same rapid advancement as fields such as Molecular Biology and Genomics because of the lack of data sharing or access. This opinion provides justification for exploring the advantages and limitations of using secondary aquatic data in hypothesis-driven research -- the primary goal of this paper.

2.4. Overview of utilizing secondary data in hypothesis-driven research

2.4.1 Background

The use of secondary data in hypothesis driven research can bypass many obstacles that can arise during primary data collection (e.g., sampling and measurement in the field) or experimental activities. One obstacle concerns funding or the lack of funding; this problem can impede research. The uses of secondary data are usually monetarily less expensive, especially in aquatic research and specifically in fish collection where use of a boat and associated technology is required. For example, a great deal of resources that require financial support is invested in the utilities’ HRE fish and water surveys; we discuss the relevance of this survey in the next section. But the result of this investment is an extremely useful annual fish survey encompassing the entire estuary that employs standard sampling techniques and design. This translates to consistency which is important when considering using secondary data collected over many years.

2.4.2. Obtaining Available Secondary Data

Data access seems to be the main challenge facing most researchers pursuing secondary analyses. There are several possible reasons for this obstacle. Parr (2003) says
the chief complaints associated with data sharing are, “getting scooped, left out of collaborations, or unrewarded by other researchers.” Providing incentives to sharing data (Schweik, Stepanov and Grove, 2005) seems like an obvious solution but proves to be difficult because there are many different types of incentives, ranging from monetary exchange to written recognition and acknowledgement in a final report or article. Collaborative efforts also appear to be a reasonable approach, but they can be challenging for many reasons such as incurring time and distance constraints. In our study (described in detail in section 2.5, below) we realized that there should be some incentive to sharing data but we were not able to offer more than a collaborative effort and written recognition and acknowledgment for all products produced from the source data, because funding for the research was minimal.

2.4.3. Understanding Available Secondary Data

Data management issues are a primary limitation to using secondary data. Metadata is an important component of data management (Gotelli and Ellison 2004, Ellison et al. 2006) but is commonly lacking. Associated documentation is necessary when sharing data, and Ellison et al. (2006) define two types of metadata, process metadata (analytical processing) and descriptive metadata (describing datasets), both of which should be produced with all research endeavors. Because data integration from multiple sources can be cumbersome, metadata has to be very thorough and complete.

A strong argument supporting open-access data policy concerns data management. Data can become lost when they are not managed properly, but when they are open accessed they have a lesser chance of being lost (Griffin 2005). Much data are stored in spreadsheets due to their convenience, but these are often mismanaged because
they usually lack associated metadata and are also usually stored on only one computer (Jones et al. 2006). Schweik et al. (2005) describes this type of data management as a “file cabinet problem” where individual office file cabinets house valuable research products and data but, because the filing system is personal and haphazard, become unusable or even lost when employees retire, die, or leave the organization. Aquatic secondary data have a better opportunity of being viable long-term if they are open-access, and this opportunity is available due to existing metadata-driven databases like the Knowledge Network for Biocomplexity (KNB) (http://knb.ecoinformatics.org/index.jsp) or CUGIR for New York state data (see section 2.5.3 below).

2.4.4. Utilizing Available Secondary Data

The value of aquatic secondary data is further increased when they are integrated and synthesized and used for purposes other than those for which they were originally intended, especially when they are involved in integrated management strategies of natural resources (Jones et al. 2006). Jones et al. (2006) also mention that data synthesis, i.e. using secondary data, allows a broader perspective over time and space, and across many disciplines, than is possible from one or a few studies. Researchers are able to address questions that were unknown or unapproachable at the time the data were collected. This view leads to another related advantage to using aquatic secondary data, that the researcher can devote more time to developing and testing hypotheses rather than being concerned with the cost and time involved with conducting experiments or other primary data collection.
Of course, there are some negative aspects to using secondary data. These include data quality or quality assurance issues, temporal or spatial sampling issues and mismatch of primary and secondary analysis research objectives (Kiecolt and Nathan 1985). In the case study below we did not have quality assurance issues because each data set was accompanied by metadata that addressed this concern. But we did encounter a mismatch of research objectives that were caused by unforeseen temporal and spatial sampling issues. This caused us to augment and change our analytical approach for a portion of our research.

2.5. Case Study: Hudson River Estuary Aquatic Data

2.5.1. Background

During our study on the HRE fish community and habitat (Chapter 3 and Chapter 4), we searched for secondary data that encompassed the entire estuary over the longest time period for many aquatic variables that would help us address our research questions. While we encountered many challenging issues including licensing/allowable use of the data and spatiotemporal scale differences, the advantages to this approach clearly outweigh the limitations to using secondary data in aquatic research. Using data we discovered, we were able to address two primary research questions:

1. How does HRE fish assemblage respond to climate and hydrological variables over time (1974-2005) (Chapter 3’s research question)?

2. Do abiotic factors such as sediment type, distance to nearest submerged aquatic vegetation (SAV) bed, salinity, water temperature and dissolved oxygen influence the distribution of juvenile American shad
and juvenile striped bass in the Hudson River estuary (Chapter 4’s research question)?

2.5.2. Obtaining Available Secondary Data on the HRE

Finding the data was not difficult because, after reading much HRE literature and working for some time in this area of fisheries research we simply knew from experience where the majority of the important data were. However, the most challenging situation we faced was obtaining access to closed-access secondary data. Tables 2.2 and 2.3 list all data used to address our research questions. The first research question uses an approach that is strictly temporal and includes elucidating abundance trends in HRE fish community due to climate change over the time period 1974 to 2005. The second approach is spatial and focuses on identifying important fish habitat in the HRE using a hierarchical modeling approach and associated spatial analyses. If the dataset was a time series from 1974 to 2005, then it could be used to address our first research question (Table 2.2). Only New York State Department of Environmental Conservation (NYSDEC) and utilities’ data sets were used to address our second research question (Table 2.3) because their surveys involved sampling the entire HRE.

Some of the data we needed were open-access via Cornell University Geospatial Information Repository (CUGIR) (http://cugir.mannlib.cornell.edu/index.jsp). CUGIR provides geospatial data and metadata for New York State, but differs from the New York State GIS Clearinghouse (http://www.nysgis.state.ny.us/) by placing a special emphasis on natural features relevant to agriculture, ecology, natural resources, and human-environment interactions. All data on CUGIR can also be accessed via the National Spatial Data Infrastructure (NSDI) network (http://www.fgdc.gov/) which promotes
the development and sharing of national digital geographic information resources. All datasets from CUGIR had the same rules for use and are open-access; they do not restrict access to any of the data in the repository and data requiring restrictions to access are not distributed via CUGIR. However, potential datasets are screened for security risks per the Rand Corporation’s report, “Mapping the risks: Assessing the Homeland Security Implications of Publicly Available Geospatial Information” (http://www.rand.org/pubs/monographs/MG142/index.html) as described in CUGIR's Security Assessment Procedure. CUGIR does not impose any access or use constraints on data except for security-related requests made by data providers regarding their datasets. They point out that data providers may stipulate use constraints in the metadata for their datasets, as long as they do not interfere with their open-access philosophy. Moreover, CUGIR and Cornell University assert no intellectual property rights on data distributed through CUGIR. They state that the data provider assures they have made the necessary agreements with the original owners of the data. Finally, CUGIR declares the following statement freeing them from liability

(http://cugir.mannlib.cornell.edu/CugirDataMgmtPolicy.20060828.pdf):

"Cornell University provides these geographic data ‘as is.’ Cornell makes no guarantee or warranty concerning the accuracy of information contained in the geographic data. Cornell further makes no warranty, either expressed or implied, regarding the condition of the product or its fitness for any particular purpose. The burden for determining fitness for use lies entirely with the user. Although these data have been processed successfully on computers at Cornell University, no warranty is made by Cornell regarding the use of these data on any other system, nor does the fact of distribution constitute or imply any such warranty."

Specifically, CUGIR provided us with two important data sets for our study. These included: (1) location and extent of submerged aquatic vegetation (SAV) beds
throughout the HRE (Nieder et al. 2004 and CUGIR 2009); and (2) sediment type based on grain size and river bottom characteristics for the entire estuary (Nitsche et al. 2004 and CUGIR 2009). The New York State Department of Environmental Conservation (NYSDEC) sponsored both data collection efforts.

Freshwater flow data for the HRE during 1974 to 2005, crucial for this study, was measured by the U.S. Geological Survey (USGS) at the federal dam gauge at Green Island, New York. This gauge is typically used to represent freshwater flow into the HRE (Abood 1992). We quantified freshwater flow as the maximum monthly mean net discharge per year. We did not use maximum annual peak flows because they can be unduly influenced by regulation or diversion. These data are also open-access (http://ny.water.usgs.gov/projects/dialer_plots/Hudson_R_at_Green_Island_Freshwater_Discharge.htm).

National Oceanic and Atmospheric Agency (NOAA) provided us with essential climate indices over the time period 1974 to 2005 (www.cdc.noaa.gov). We used two climate indices – the Atlantic Multidecadal Oscillation (AMO) and the North Atlantic Oscillation (NAO). The AMO is a relatively newly identified climate index based on a multi-decadal oscillation in North Atlantic sea surface temperature (SST), first recognized by Schlesinger and Ramankutty (1994). And the NAO is the predominant mode of climate variability in the northern hemisphere Atlantic Ocean region, on annual-to-decadal time scales (Hurrell et al. 2003).

The Poughkeepsie Water Treatment Facility provided us with year round water temperature measurements. We obtained daily recorded water temperatures and used the maximum annual temperature for each year (1974 to 2005) from the Poughkeepsie Water
Works located 122 km north of the river mouth. These data are closed-access and required permission for use.

In terms of necessary fish and water characteristic data needed for our proposed hypothesis-driven research, the HRE is unique in that several monitoring surveys exist and provide an ongoing time series of over 30 years of secondary data that include measurements of fish abundance, water temperature, dissolved oxygen and salinity throughout the estuary. These data are collected annually by the electric utility generating stations since 1974. The utilities’ survey is an example of privately funded research, driven by law to be collected, and with a management policy of closed-access requiring permission for use. This situation is commonly seen in other fields, which rely on privately funded research, such as genomics. We obtained the utilities’ data set (1974-2005) after some back-and-forth related to the appropriate use for the data and then were able to proceed with our proposed research.

2.5.3. Understanding Available Secondary Data on the HRE

After the data above were acquired to test our research questions several limitations were encountered that were inherent to the data or to the data collection methods/sampling design; these are described in section 2.5.4. Metadata, however, were not a limiting factor. All data we obtained were accompanied by metadata; some documentation initially seemed sparse or vague but the data owners were available and answered all questions fully. In comparison, the utilities’ data were accompanied by volumes of metadata including a data dictionary because their data were numerically coded. Also included with the utilities’ metadata were historical reports and standard operating procedures for sampling. This comprehensive metadata enabled us to avoid the
pitfalls commonly associated with understanding the data and made our approach possible.

2.5.4 Utilizing Secondary Data on the HRE and Challenges Encountered

Realizing the many significant advantages of utilizing secondary data, we adapted this approach to our HRE research despite the known challenges and unforeseen challenges. First, we were able to address our first research question concerning the effects of local and regional climate on the HRE fish community straightforwardly and take a time series approach because the utilities’ survey encompassed over 30 years (1974-2005) and were collected annually. The climate indices (AMO and NAO), freshwater flow and water temperature data used in the analysis also encompassed this time period. And by employing a multivariate approach, we confirmed that the HRE fish community has changed over this time period and these changes are correlated with local hydrology (freshwater flow and water temperature) and regional climate (AMO and NAO). We also found that juvenile shad abundances are negatively correlated with the AMO or warmer sea surface temperatures. This finding suggests that climate-related variability affects HRE juvenile shad abundances and current management strategies of this species should include the implications of climate change and regional climate on HRE juvenile shad abundances (Chapter 3).

Our second research question, ‘Do abiotic factors such as sediment type, distance to nearest submerged aquatic vegetation (SAV) bed, salinity, water temperature and dissolved oxygen influence the distribution of juvenile American shad and striped bass in the Hudson River estuary?’, undertaken in Chapter 4, could not be investigated without using secondary data either because of study’s large spatial extent, the entire HRE
Addressing this research question and creating habitat maps for juvenile American shad and striped bass required more data processing because we encountered a mismatch of primary and secondary analysis research objectives. In our case these constraints were primarily spatiotemporal scale differences inherent to the fish and water survey data because the samples were not co-located therefore they were not collected at the same time or in the same location (Table 2.3). First, we quickly noticed that the water samples were not georeferenced but did have accompanied spatial information. This spatial information included a river mile location and within river location (beach, shoal, channel). We created a “thalweg,” a line defining the deepest channel, for the HRE (per Merwade et al. 2005) to measure distance and to assign approximate position coordinates to the water sample locations (see Figure 2.2). Next, we understood that the goal of the utilities’ survey was to describe large-scale spatio-temporal/interannual changes in early life stages of selected fish species and water characteristics in the HRE with a focus on estimating the environmental impact of five Hudson River generating stations, although we were surprised when we realized that the fish samples in the utilities’ dataset were not co-located with the water samples in the utilities’ dataset. These data issues at first seemed to be a limitation but were overcome by taking a different and more creative analytical approach that allowed us to be less specific both spatially and temporally. Specifically, we took a seasonal point of view and were able to integrate both datasets by using universal kriging for each water characteristic or abiotic variable (dissolved oxygen, salinity, water temperature). The results were seasonal surfaces, or data layers, of each abiotic variable. And by overlaying the fish sample locations and extracted water measurements for each fish sample we were able to artificially co-locate these
samples/variables. Lastly, we processed the other variables needed for this analysis by extracting sediment type for each fish sample location and by calculating the Euclidean distance from each fish sample location to nearest SAV bed (Figure 2.2).

When secondary data are used in research they can take on a new dimension through additional analyses, and thus the dataset grows adding more intrinsic value. Through our analysis, we georeferenced and interpolated seasonal water characteristic (dissolved oxygen, water temperature, salinity) data at three depths seasonally for the time period 2000 to 2005. These generated data can be an added benefit to conducting additional hypothesis-driven secondary analyses because they add to the pool of available secondary data for the HRE. By using secondary data we have exacted an additive and beneficial influence on the original dataset and created a form of (non-web-based) data mashup. We hope to help other researchers avoid redundancy by publishing our resulting data in CUGIR. This would help further our understanding of the HRE aquatic processes by providing information for others to discern and hopefully lead to additional hypothesis creation.

2.6 Closing Remarks

Recognizing the large amount of data collected during routine aquatic or fish surveys and realizing the emphasis is often still on new data collection in fisheries research, we explored the possibility of using secondary data in hypothesis-driven research. We conclude the paper with insights from this effort that may be of interest to others pursuing this approach.

Overall we found that there is a benefit to using secondary data in aquatic research. The primary advantage we gained was the ability to develop and address two
relevant research questions concerning the HRE fish resources (Chapters 3 and 4). We could not have investigated the effects of climate on the HRE fish assemblage if secondary data were not available because a long-term dataset (eg. thirty plus years) was necessary.

Finding all data was easy but acquiring the data was time consuming in some cases, because of access issues. This only occurred with the utilities’ data set because the data are managed by a private consulting firm and permission must come from the utilities’ authorities. Other studies pertaining to the HRE fish community have used some of the same secondary data, specifically the utilities’ fish survey data (eg. Pace et al. 1993, Strayer et al. 2004). The utilities’ database is extensive and there are many more opportunities for future HRE hypothesis-driven research using this source.

While understanding the data was made easier because of the detailed associated metadata with each data set, utilizing the data proved to be more challenging because of the added data processing. The possibility of encountering these types of problems, such as spatiotemporal inconsistencies, should not deter the researcher because there are many analytical techniques that can be used to overcome these types of issues. In our case interpolation, specifically kriging, helped us to reach our research goals and provide useful information to HRE fisheries managers. This less accurate, estimating technique was “better than nothing” and less expensive and time consuming than collecting the data ourselves.

Finally, much financial support is necessary when the emphasis is on new data collection; in light of today’s current economic situation it is imperative to be frugal wherever possible. Therefore, there should be more emphasis on data sharing and
secondary analysis with a focus on contributing to existing open-access data repositories in the aquatic sciences (e.g., KNB and NSDI). The current enclosure movement, which seeks to limit secondary data access, impresses a subtractive view regarding data sharing instead of an additive one (Hess and Ostrom 2007). In this paper we demonstrate that data sharing can promote new data creation as shown by the seasonal dissolved oxygen, salinity and water temperature GIS layers we produced. We hope our experience encourages others to contribute and/or support the open-access to data movement.
Table 2.1: Two by two matrix denoting the theory of goods, based on Ostrom, Rules, Games, and Common Pool Resources, University of Michigan Press, 1994.

<table>
<thead>
<tr>
<th>Excludable</th>
<th>Yes</th>
<th>No</th>
<th>Common Pool Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtractable</td>
<td>No</td>
<td>Yes</td>
<td>Private good</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Club Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Public good</td>
</tr>
</tbody>
</table>
Table 2.2: Secondary data used to address the research question, ‘How does HRE fish assemblage respond to climate and hydrological variables over time (1974-2005) (Chapter 3’s research question)?’

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Data Content</th>
<th>Spatial Extent</th>
<th>Data Type</th>
<th>Years Sampled</th>
<th>Data Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>USGS</td>
<td>River flow</td>
<td>Green Island Dam</td>
<td>Vector/Point</td>
<td>1974-2005</td>
<td>Open</td>
</tr>
<tr>
<td>Poughkeepsie</td>
<td>Water Temperature</td>
<td>Poughkeepsie</td>
<td>Vector/Point</td>
<td>1974-2005</td>
<td>Closed</td>
</tr>
<tr>
<td>Water Treatment Facility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOAA</td>
<td>Climate Indices</td>
<td>Atlantic Ocean</td>
<td>Raster</td>
<td>1974-2005</td>
<td>Open</td>
</tr>
<tr>
<td>Utilities’</td>
<td>Fish abundance</td>
<td>The Battery to Albany</td>
<td>Vector/Point</td>
<td>1974-2005</td>
<td>Closed</td>
</tr>
</tbody>
</table>
Table 2.3: Secondary data used to address the second research question, ‘Do abiotic factors such as sediment type, distance to nearest submerged aquatic vegetation (SAV) bed, salinity, water temperature and dissolved oxygen influence the distribution of juvenile American shad and juvenile striped bass in the Hudson River estuary (Chapter 4’s research question)’

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Data Content</th>
<th>Spatial Extent</th>
<th>Data Type</th>
<th>Years Sampled*</th>
<th>Data Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilities’</td>
<td>Water Characteristics</td>
<td>The Battery to Albany</td>
<td>Vector/Point</td>
<td>2000-2005</td>
<td>Closed</td>
</tr>
<tr>
<td>NYSDEC/ CUGIR</td>
<td>SAV* Beds</td>
<td>The Battery to Albany</td>
<td>Vector/Polygon</td>
<td>1995-1997</td>
<td>Open</td>
</tr>
<tr>
<td>NYSDEC/ CUGIR</td>
<td>Sediment Type</td>
<td>The Battery to Albany</td>
<td>Vector/Polygon</td>
<td>1998-2003</td>
<td>Open</td>
</tr>
<tr>
<td>NYSDEC</td>
<td>Bathymetry</td>
<td>The Battery to Albany</td>
<td>Vector/Line</td>
<td>1998-2003</td>
<td>Open</td>
</tr>
<tr>
<td>Utilities’</td>
<td>Fish abundance</td>
<td>The Battery to Albany</td>
<td>Vector/Point</td>
<td>2000-2005</td>
<td>Closed</td>
</tr>
</tbody>
</table>

Notes: For the utilities’ data we have data for each year (2000 to 2006) for all sampling locations throughout the HRE. The NYSDEC data do not differ over the time (2000 to 2006) because these data were sampled only once through the HRE but took several years to complete. So these data are actually static variables for each sampling location. *SAV denotes submerged aquatic vegetation
Figure 2.1: An example of data layers used in data processing and data analysis for addressing the second research question (Chapter 4).
CHAPTER 3
LONGTERM EFFECTS OF LOCAL HYDROLOGY AND REGIONAL CLIMATE ON THE HUDSON RIVER ESTUARINE FISH ASSEMBLAGE

3.1 Abstract

It has been hypothesized that climate change is an underlying factor in determining fish abundances in the Hudson River estuary (HRE). To study the effects of hydrology and climate on the HRE fish assemblage, we applied multivariate statistical methods to 20 species-life history stages collected from 1974 to 2005. We confirm that the HRE fish assemblage has changed over this time period. These changes are correlated with local hydrology (freshwater flow and water temperature) and regional climate (Atlantic Multidecadal Oscillation or AMO and North Atlantic Oscillation or NAO). We found that abundances of striped bass larval stages are positively correlated with high freshwater flows and juvenile shad abundances are negatively correlated with the AMO or warmer sea surface temperatures (SST). Our findings suggest that climate-related variability affects HRE juvenile shad abundances and current management strategies for this declining species should include the implications of climate change.
3.2 Introduction

In the complex web of environmental variables to which fish respond, regional climate is perhaps the least well understood and probably the least well studied. There are documented correlations among fish abundances/distributions and climate signals, but most studies concern fish populations in open-ocean settings such as the North Sea, the Bering Sea, the north-east Atlantic and the north-west Pacific (e.g. Brander 2007; Brunel and Boucher 2007). Recently however, there has been a focus on examining the effects of climate change on coastal fish populations in the eastern United States. For example, Condron et al. (2005) found that North American Atlantic salmon (*Salmo salar*) abundances fluctuate with a sea surface temperature-based climate index known as the Atlantic Multidecadal Oscillation (AMO). They concluded that continued warming near the Grand Banks of Newfoundland will negatively impact the already depleted salmon stock regardless of the reduction in commercial fishing. Similarly, Hare and Able (2007) found that large-scale variability in Atlantic croaker (*Micropogonias undulatus*) abundances along the east coast of the United States are also driven by climatic processes, such that an increase in croaker abundance or larger year classes is correlated with warm winters, which are associated with the positive phase of the North Atlantic Oscillation (NAO) (Hurrell et al. 2003).

However, the effects of climate are not linear, and complex relationships between climate indices and abiotic factors such as discharge or streamflow have been identified in some aquatic systems in the eastern United States. Specifically, the AMO has been linked to rainfall and streamflow variability in Florida, where the inflow to Lake Okeechobee varies by 40% between AMO warm and cool phases (Enfield et al. 2001).

The Hudson River estuary (HRE) fish assemblage is an ideal model to test how local hydrology (e.g., water temperature and streamflow) and regional climate influence fish abundance patterns over time. It has a relatively long record of both fish abundance and abiotic data. A long-term decline in HRE fish community diversity and stability has been identified in nearshore fish assemblages by Daniels (1995) and Hurst et al. (2004). Also, there is strong evidence that an exotic species, the zebra mussel (*Dreissena polymorpha*), has altered the fish communities in the upper HRE since its invasion in 1991 (Strayer et al. 2004). Although historical fish abundance and distributions are changing or shifting in the HRE, the mechanisms responsible for the shift remain unclear (Pace et al. 1993a).

Regional climate change is broadly considered to be an underlying factor in determining fish abundance in the HRE, but a quantifiable relationship between climate phenomena and observable variations within the fish community has not yet been established (Waldman et al. 2006). For example, Daniels et al. (2005) state that the fish assemblages of the HRE are changing, note that causes for these changes are varied, and speculate that global warming may have caused the disappearance of rainbow smelt (*Osmerus mordax*) in the HRE.

The purpose of this study is to examine effects of local hydrology and regional
climate on the HRE fish assemblage. Identifying possible correlations among climate indices, freshwater flow, water temperature and fish abundances over time allows us to examine the effects of climate at both local and regional scales. Our first objective is to confirm that the HRE fish assemblage has changed significantly over the past 32 years (1974-2005). Our second objective is to examine how the HRE fish assemblage responds to climate and hydrological variables over this time period through constrained ordination analysis. Our third and final objective is to evaluate the relative importance of the environmental variables (regional climate and local hydrology) on HRE fish assemblage abundance patterns.

3.3 Methods

3.3.1 Study Area

The Hudson River extends 507 km from its source, Lake Tear of the Clouds in the Adirondack Mountains, to New York Harbor (Figure 3.1). It is tidally influenced up to a dam located near Troy, NY (250 river kilometer (rkm)). The New York Department of State has designated large portions of the Hudson River estuary as a Significant Coastal Fish and Wildlife Habitat (USFWS 1997). Four coastal wetlands within the HRE are designated by the National Oceanic and Atmospheric Agency (NOAA) as National Estuarine Research Reserve System (NERRS) sites. The estuarine portion of the Hudson River is composed of a diverse assemblage of marine, freshwater and estuarine fishes. Beebe and Savidge (1988) report that 140 species have been collected in the Hudson River estuary and approximately one dozen of these species are diadromous (Waldman 2006).
After examining studies concerning spatio-temporal fish abundance trends in nearshore areas (Daniels 1995, Hurst et al. 2004) and the entire HRE (Pace et al. 1993a, ASA 2007, Strayer et al. 2004), we decided to focus this study on the entire HRE defined as the Battery (rkm 0) north to Troy (rkm 248) (Figure 3.1).

### 3.3.2 Fish Community Data

In the HRE, there has been an annual intensive multi-species survey since 1974 when year class reports were first prepared for the electric utility generating stations. The goal of these reports is to determine the environmental impact that these utility companies have on the HRE fish populations. The utilities’ survey consists of three annual surveys, the Long River Survey (LRS), designed to assess egg and larval densities; the Fall Juvenile Survey (FJS), to assess juvenile densities offshore; and the Beach Seine Survey (BSS), to assess nearshore fish communities and abundance. The survey area begins at the Battery and ends in Albany (ASA 2007). Over thirty years of spatio-temporal fish abundance data exist due to this survey. The utilities’ fish survey data have been subjected to the rigors of complete QA/QC protocol and have been used in numerous analyses (e.g., Pace et al. 1993b, Strayer et al. 2004, Dunning et al. 2006).

Our analyses were conducted using the fish abundance indices calculated in the annual utilities’ survey report for the period 1974-2005 (Table 3.1). Full details of the abundance index calculation can be found in Chapter 2 and Appendix D of the 2005 Year Class Report (ASA 2007) and are only briefly summarized here. The numbers of fish collected in the survey each year are adjusted to an index representing the number caught under a standardized sampling effort. The indices are based on fish densities for the LRS and FJS, or by catch per unit effort (CPUE) for the BSS, and are calculated for the entire...
estuary. To allow comparisons of data across years, only data from samples collected from common weeks or consistently sampled over the time period 1974-2005 for each survey were used for index calculations.

A variety of resident, marine, freshwater and anadromous fish species were chosen to represent the HRE assemblage (Table 3.1). Fish species and life history stages were selected based on recommendations found in the 2005 Year Class Report. Some fish species are selectively sampled during these surveys based on gear design and sampling location. The utility companies selected these species based on the New York State Department of Environmental Conservation’s (NYSDEC) concern for the ecological effects of the utilities’ practices.

Our fish assemblage data set was a matrix containing 32 years (rows) and 20 species-life history stages (columns: Table 3.1). In many ecological data sets, especially community data sets involving species abundances, it is often quite useful to standardize (or relativize) the data before conducting subsequent analyses (Legendre and Gallagher 2001). Based on the recommendations of Legendre and Gallagher (2001), we used a row total standardization to adjust for differences in absolute fish abundances among years (rows) and place the emphasis on relative abundance profiles within years. We evaluated other row standardizations as recommended by Legendre and Gallagher (2001) and the results were consistent.

3.3.3 Climate indices

We used two climate indices in this study – the AMO and the NAO. The AMO is a relatively recently identified climate index based on a multi-decadal oscillation in North Atlantic sea surface temperature (SST), first recognized by Schlesinger and Ramankutty
It is thought to be closely tied to ocean thermohaline circulation (Kerr 2000, Delworth and Mann 2000, Sutton and Hodson 2005). Over the last ~150 years, observational annual mean basin-averaged North Atlantic SST data appear to have undergone a statistically significant warm-cold oscillation, with overall warm periods covering the late 19th century, 1931-1960, and 1991-present; and with overall cold periods covering 1905-1925 and 1965-1990 (Sutton and Hodson 2005, Enfield et al. 2001). The AMO climate index used in this report is an unsmoothed time series of monthly SST anomaly data for the North Atlantic region (Enfield et al. 2001) taken from the Kaplan SST version 2 global dataset (www.cdc.noaa.gov). These data cover the period 1856-present and consist mainly of ship observations and buoy data, coupled with remote sensing data from 1981 on. The dataset is a 5°x5° set of SST anomalies from 1856-present with a base period of 1951-1980. We specify the AMO in this paper as the unsmoothed, detrended time series of area-weighted SST anomaly data in the North Atlantic basin from the equator to 70°N, covering the period from 1974 to 2006.

The NAO is the predominant mode of climate variability in the northern hemisphere Atlantic Ocean region on annual-to-decadal time scales (Hurrell et al. 2003). The NAO is a quasi-periodic oscillation in the winter season atmospheric pressure difference in the North Atlantic region, between the subtropical high (the “Bermuda” or “Azores” high) and the subpolar low (the “Icelandic” low). In the positive NAO phase, this pressure difference is great, mid-latitude winds are stronger than normal, the northwest Atlantic regions receive cold Canadian winds, and trans-Atlantic storms tend toward northern Europe as they make landfall; in the negative NAO phase, the pressure difference is less, and reverse conditions occur (Hurrell et al. 2003, Visbeck et al. 2003,
Visbeck 2002). The NAO climate index we used here is the time series corresponding to one of ten leading atmospheric “teleconnection patterns,” or recurring large-sale patterns of low-frequency (long time-scale) variability in atmospheric pressure and circulation fields for the northern hemisphere, identified by researchers at NOAA’s Climate Prediction Center (CPC).

For both the AMO and NAO in this analysis, we used the monthly value from each year with the highest magnitude, or the “extreme” value. In other words, the “AMO extreme” and “NAO extreme” for each year were chosen as the single monthly data point with the greatest absolute value. We assumed that the highest negative or positive monthly value from each year would have the greatest chance of influencing fish behavior and distribution.

3.3.4 Hydrological Variables

Yearly maxima of water temperature and freshwater flow in the HRE were assembled from various sources. Freshwater flow during 1974 to 2005 was measured by the U.S. Geological Survey (USGS) at the federal dam gauge at Green Island, New York. This gauge is typically used to represent freshwater flow into the HRE (Abood et al. 1992). We quantified freshwater flow as the maximum monthly mean net discharge per year. We did not use maximum annual peak flows because they can be unduly influenced by regulation or diversion. We obtained daily recorded water temperatures and used the maximum annual temperature for each year (1974 to 2005) from the Poughkeepsie Water Works (122 rkm).
3.3.5 Data Analysis

We used multivariate statistical methods to assess changes in fish assemblage abundance profiles over the period 1974 to 2005 in relation to hydrology and climate change. First, we constructed a Mantel correlogram to examine changes in fish assemblage profiles over time. In this context, the Mantel correlogram measures the how the strength of correlation between fish assemblage profiles in different years varies as the number of years between profiles increases. The Mantel test operates on distance matrices. We transformed the fish assemblage data set into a distance matrix using the euclidean measure of dissimilarity, where each entry represented the dissimilarity in fish assemblage profiles between two different years of the survey. We assessed statistical significance of the Mantel test statistic for each time lag based on a Monte Carlo randomization procedure with 1000 permutations.

Second, we used redundancy analysis (RDA) to examine fish assemblage abundance profiles in relation to environmental and climate variables. RDA is a constrained ordination technique that seeks to find the major axes of variation in the dependent variable set (fish data set) that can be explained by linear constraints in the independent variable set (environmental data set) (Legendre and Legendre 1998). We sought to determine how much of the variance in the fish assemblage data set could be explained by variation in the climate indices (AMO and NAO), freshwater flow and water temperature. We judged the efficacy of the ordination by the percent of inertia in the overall fish assemblage and in each species-stage explained by the environmental constraints. To gain insight into the species-environment relationships expressed by the RDA, we scaled the plotting symbol for each year by the abundance of a selected fish.
species-stage and used generalized additive models (GAMs) to fit surfaces of the explanatory variables to the ordination plot for the selected species-stage.

Lastly, we used variance partitioning to decompose the constrained ordination into the independent and joint effects of local hydrology (freshwater flow and water temperature) and regional climate (AMO and NAO indices) (Legendre and Legendre 1998). Specifically, we used variance partitioning to determine how much of the variation in the fish assemblage over time could be explained by (1) local hydrology alone, after accounting for regional climate effects, (2) regional climate alone, after accounting for local hydrology effects, and (3) the joint or confounded effects of local hydrology and regional climate.

All statistical analyses were conducted using R (R Development Core Team 2008), including functions in the vegan library and programs written by the authors.

3.4 Results

The HRE fish assemblage has changed gradually from 1974 to 2005, as inferred by the Mantel correlogram (Figure 3.2). The pattern of steadily decreasing correlation (Mantel r) with increasing distance in time (i.e., number of years between surveys) indicates a long-term trend in the fish assemblage structure.

Variation in the HRE fish community over time was significantly explained by the linear environmental constraints of freshwater flow, water temperature and AMO based on RDA. Specifically, over 46% of the total inertia in the fish assemblage was explained by the environmental constraints. Three species-stages were fit well by the ordination based on the percentage of their inertia explained by the environmental constraints, including post yolk sac larval (PYSL) striped bass (73%), juvenile American
shad (69%) and yolk sac larval (YSL) striped bass (49%). The ordination triplot in figure 3.3 shows the first two constrained ordination axes, both of which explained a significant (p<.01) amount of variation in the fish assemblage data. The triplot depicts the species-stages and environmental variables as vectors representing the magnitude (vector length) and direction (vector angle) of greatest linear rate of change in the corresponding variables. For example, larval striped bass shows a strong positive relationship to high freshwater flows, indicating an increase in relative abundance during years with high freshwater flows. Similarly, juvenile shad shows a strong inverse relationship to AMO, indicating that juvenile shad increased in relative abundance during years in the extreme positive phase of AMO. The strong and linear relationship between juvenile shad and AMO was even more evident in the enhanced ordination plot in figure 3.4. In contrast, nonlinear configurations in the water temperature and freshwater flow GAM surfaces (not shown) indicate a more complex relationship with juvenile shad.

Partitioning of the explained variance in the constrained ordination indicated that both local hydrology (freshwater flow and water temperature) and regional climate (AMO and NAO) had strong and independent abilities to explain variation in the fish community (Figure 3.5). Of the 46.4% of the inertia in the fish community explained by the environmental constraints, all but 2% was explained uniquely by either local hydrology or regional climate. Local hydrology independently (i.e., after accounting for that due to regional climate) explained 21.1% of the variation in the fish community; regional climate independently explained 23.5% of the species inertia.
3.5 Discussion

There was a significant long-term trend in the selected species-stage abundances from 1974 to 2005 indicating that the HRE fish assemblage has changed significantly over this period. This finding is consistent with similar findings from other estuaries. For example, Collie et al. (2008) observed a similar shift in assemblage structure in Narragansett Bay and adjacent Rhode Island Sound from 1959 to 2005, where they documented a decrease in demersal species and an increase in warm-water fish and invertebrates. Within the HRE, Strayer et al. (2004) discovered a shift in the fish assemblage caused by the invasion of the zebra mussel during the early 1990’s. Specifically, they found a decrease in openwater fish (e.g. *Alosa*) and an increase in littoral fish (e.g. centrarchids). Daniels et al (2005) also documented temporal shifts in the HRE fish assemblage; however, they did not quantitatively assess the factors associated with the longterm trend, they only speculated on the importance of biotic and abiotic factors.

We found a remarkably strong relationship between abiotic factors and the shift in fish assemblage abundance profiles over time. Specifically, freshwater flow, water temperature and AMO explained 46% of the total inertia or variance in the species-stage abundance patterns. Moreover, these abiotic factors were mostly independent (i.e., not confounded) in their ability to explain the species inertia. Of the 20 species-stages we analyzed, only juvenile American shad and larval striped bass were well fit by the constrained ordination, suggesting that they are the only species/stages that are highly influenced by local hydrology and regional climate in the same manner as the entire assemblage. Additional individual species/stage statistical models would be required to
examine whether other species/stages are uniquely influenced by these abiotic factors.

Our results however provide different conclusions about the relative importance of local and regional climate on the two most abundant anadromous species in the HRE. Juvenile shad abundances are affected by regional climate as indicated by their strong negative correlation with the AMO (Figure 3.3 and Figure 3.4), whereas larval striped bass respond to local hydrology as indicated by their strong positive correlation with freshwater flow (Figure 3.3). Currently, American shad abundances are decreasing and striped bass abundances are increasing in the HRE (ASMFC 2007). The differential population trajectories for these two species may be a function of their differing life histories.

Both American shad and striped bass are anadromous, with the adults entering the estuary in early spring to spawn (Waldman 2006). Adult shad and striped bass potentially overlap spatially and temporally in their spawning patterns, causing competition and predation among larval and juvenile stages (Bilovic et al. 2002). However, juvenile striped bass spend at least two years in the HRE, whereas juvenile shad leave the estuary during the fall of their first year. Although both ocean-phase shad and striped bass can undergo extensive, seasonal, ocean migrations from Nova Scotia to North Carolina (Dadswell et al. 1987, Waldman et al. 1990), HRE adult striped bass are primarily residents of the estuary, New York Harbor, and western Long Island Sound (Secor 1999). In general, striped bass are inshore fish and do not usually occur more than 6 to 8 km offshore (Collette and Klein-MacPhee 2002). In contrast, in late fall and winter, shad often move to deeper waters and have been caught up to 175 km offshore (Dadswell et al. 1987). Shad seasonal migrations also include both short coastal
migrations and more extensive offshore migrations along the continental shelf at depths up to 200 meters (Neves and Dupres 1979). Unfortunately, most of the current data available on shad movements are fishery-dependent and because the fishery is conducted inshore, catch data are of limited use in determining offshore migration patterns (Neves and Depres 1979).

Because the AMO index calculation is based upon North Atlantic SST values, our results suggest that increases in juvenile shad abundances are negatively correlated with North Atlantic SSTs. During a negative (positive) AMO phase shad populations are high (low) because of cooler (warmer) SSTs. The migrating shad are likely directly influenced by SSTs because they are in direct contact with the North Atlantic during their oceanic migrations. Changes in juvenile abundances are due to direct mortality (Limburg 2001) and are also probably due to oceanic phase shad being influenced by climate-related factors experienced during their ocean migration. Climate change, and temperature in particular, can have both direct and indirect impacts on fish. Direct effects act on physiology and behavior and alter growth, development, reproductive capacity, mortality and distribution. Indirect effects alter the productivity, structure and composition of the ecosystem (Brander 2007). Although shad move seasonally within preferred isotherms and their migration patterns are controlled by temperature (Leggett and Whitney 1972; Neves and Depres 1979), they also cross these thermal barriers when they migrate inshore (Dadswell et al. 1987). Hobson et al. (2008) found that summer isotherms are now reaching their northern limit more frequently in the North Atlantic. Warm SSTs are thus occurring further north more often in summer months implying that shad are also experiencing warmer temperatures more often. Similarly, Condron et al. (2005)
demonstrated this causal link spatially and graphically with North American Atlantic salmon, the AMO and North Atlantic SSTs. They found that salmon abundances fluctuate with the AMO and were able to link survival with temperature changes in the Grand Banks, a known overwintering area for salmon and concluded that continued warming near the Grand Banks will further reduce stock numbers. Future work regarding American shad should further investigate the effects of the AMO and North Atlantic SSTs on migrating oceanic phase shad.

Local hydrology also affects the HRE fish community. In particular, our results show that larval striped bass abundances are positively correlated with freshwater flow, indicating an increase in relative abundance during years with high freshwater flows. In contrast, Pace et al. (1993b) found that variability in water temperature and freshwater flow in the HRE explained little of the interannual variation of larval striped bass abundances from 1974 to 1990. However, in a more recent analysis, Dunning et al. (2006) showed that the abundance of PYSL striped bass in the Battery region from 1999-2002 was higher when freshwater flow in the estuary during May and early June was higher. They concluded that freshwater flow strongly affects the abundance of striped bass larvae in the lower HRE, a finding similar to ours. This result has also been documented in the upper Chesapeake Bay estuary. While studying striped bass recruitment, North and Houde (2003) found that abundances of striped bass PYSL in the upper Chesapeake Bay estuary are lower during low freshwater flow years. They concluded that annual changes in freshwater flow could control larval survival and recruitment by modifying the physical and biological characteristics of the estuarine turbidity maximum (ETM) region.
We did not include biotic factors in this study although Limburg (2001) suggests predation and hypoxia negatively affect the juvenile shad migrating from the Hudson River and thus shape year class strength. Similarly, Strayer et al. (2004), while studying the effects of zebra mussels in the HRE, found that populations of open-water species (e.g., *Alosa* spp.) had shifted downriver away from the zebra mussel populations and decreased in abundance. Although within-river biotic and abiotic conditions clearly affect juvenile shad abundance, our results indicate that ocean climate or North Atlantic SST also strongly affects juvenile shad abundances within the river. Thus, even without considering biotic factors, robust climate patterns are evident, further highlighting the importance of regional climate on HRE juvenile shad abundances. Furthermore, the relative importance of local hydrology and regional climate in driving changes in fish abundance patterns in the HRE from 1974 to 2005 behave independently in this system (Figure 3.5) and can influence the fish community in different ways.

The effects of climate change are a central contemporary issue that is at the forefront of many ecological research endeavors (Walther et al. 2002). Addressing these effects will provide guidance on how we manage our resources presently and in the future. We are currently experiencing the warmest global temperatures in recorded history. Changes due to warming cause variations in AMO or SST patterns (Enfield et al. 2001) and may affect HRE fish abundances and distributions. Quantifying the effects of fluctuation in these patterns with regards to fish population dynamics could help Hudson River fisheries managers better manage the resource in light of future climate scenarios. Currently, climate change is not listed as a potential cause of Hudson River estuary American shad decline in the latest American shad assessment and recovery plan.
(ASFMC 2007, NYSDEC 2008). Our results provide clear evidence for climate-related variability in abundances of HRE juvenile shad populations.
Table 3.1: Fish Abundance Indices from 1974 to 2005 used in the multivariate analyses.

<table>
<thead>
<tr>
<th>Species</th>
<th>Species Code</th>
<th>Life History Stage</th>
<th>Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Shad</td>
<td>Alosa sapidissima</td>
<td>egg, ysl, pysl</td>
<td>LRS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>juv</td>
<td>BSS</td>
</tr>
<tr>
<td>Striped Bass</td>
<td>Morone saxatilis</td>
<td>egg, ysl, pysl</td>
<td>LRS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>juv</td>
<td>BSS</td>
</tr>
<tr>
<td>White Perch</td>
<td>Morone americana</td>
<td>egg, ysl, pysl</td>
<td>BSS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>juv, yearling</td>
<td>LRS</td>
</tr>
<tr>
<td>Bluefish</td>
<td>Pomatomus saltatrix</td>
<td>egg, ysl, pysl</td>
<td>BSS</td>
</tr>
<tr>
<td>White Catfish</td>
<td>Ameiurus catus</td>
<td>egg, ysl, pysl</td>
<td>BSS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>juv, yearling</td>
<td>LRS</td>
</tr>
<tr>
<td>Hogchoker</td>
<td>Trinectes maculatus</td>
<td>egg, ysl, pysl</td>
<td>BSS</td>
</tr>
<tr>
<td>Rainbow Smelt</td>
<td>Osmerus mordax</td>
<td>egg, ysl, pysl</td>
<td>LRS</td>
</tr>
<tr>
<td>Atlantic Tomcod</td>
<td>Microgadus tomcod</td>
<td>egg, ysl, pysl</td>
<td>BSS</td>
</tr>
<tr>
<td>Spottail Shiner</td>
<td>Notropus hudsonius</td>
<td>egg, ysl, pysl</td>
<td>LRS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>juv</td>
<td>BSS</td>
</tr>
</tbody>
</table>

**Notes:** ysl- Denotes yolk-sac life history stage; pysl- Denotes post yolk-sac life history stage; Juv- Denotes juvenile life history stage.
Figure 3.1: Map of the study area, Hudson River estuary (New York, USA). Study area, Hudson River estuary from the Battery to Albany.
Figure 3.2: Mantel correlogram showing the correlation in the fish community structure in relation to the time (years) between observations. Each point shows the correlation between the fish community at different time intervals (represented by the distance class); solid circles represent significant correlations. The monotonically decreasing correlation over time indicates a long-term trend; i.e., fish community similarity decreases steadily over time.
Figure 3.3: Constrained ordination triplot based on redundancy analysis depicting the first and second constrained axes (p<0.01). The length and direction of the vectors reveals the strength and direction, respectively, of maximum rate of change in the corresponding variable. Environmental variables are depicted with dashed vectors and species-stage variables are depicted with solid vectors. Species codes are listed in Table 1. AMOextr and NAO extr denote the respective extreme yearly climate index. And FlowMax and TempMax denote maximum yearly values of freshwater flow and water temperature in the HRE.
Figure 3.4: Enhanced ordination plot depicting the relative abundance of juvenile shad each year from 1974-2005. The solid circles represent years; the size of the circle is proportional to the relative abundance of juvenile shad. Superimposed on the ordination plot is a fitted surface of the Atlantic Multidecadal Oscillation (AMO) based on a generalized additive model. The regular and straight contours indicate that a linear model was optimal. The contour labels represent the predicted values of AMO.
Figure 3.5: Venn diagram showing the variance partitioning of the constrained ordination. The non-overlapping portion of each circle represents the percentage of “variance” in the fish community explained independently by the corresponding constraining variables (e.g., hydrology) after accounting the other constraining variables. The overlapping portion of the circles represents the confounded portion of the explained variance; i.e., the portion that cannot be distinguished between the two sets of constraining variables. The “residual” is the unconstrained inertia or the total “variance” in the species data set left unexplained or unaccounted for by the environmental constraints; it is equal to the sum of the unconstrained eigenvalues.
CHAPTER 4

DISTRIBUTION AND HABITAT UTILIZATION OF JUVENILE AMERICAN SHAD AND STRIPED BASS IN THE HUDSON RIVER ESTUARY

4.1 Abstract

We present species distribution models (2000-2005) based on generalized linear models of the fall distribution or presence/absence of juvenile American shad (*Alosa sapidissima*) and juvenile striped bass (*Morone saxatilis*) in the Hudson River estuary (HRE) based on data from a fishery-independent survey. The distribution of both species were modeled as a function of dissolved oxygen, salinity, water temperature, distance along the HRE denoted as river mile, time or Julian day, distance from submerged aquatic vegetation (SAV), and sediment characteristics for the 6 yrs of data chosen (2000-2005). Salinity, river mile, and Julian day were found to be the most important environmental determinants of juvenile American shad presence. And sediment type, salinity, river mile, and Julian day were found to be the most important environmental determinants of juvenile striped bass presence. Calibration plots show a high level of agreement between predictions generated by each global model and actual observations of presence/absence of juvenile American shad and juvenile striped bass. Based on this result we mapped predictions from the models for both species and found the highest predicted probabilities of juvenile American shad presence are found in the Upper HRE. Conversely, highest predicted probabilities of juvenile striped bass presence are found in the Lower HRE. Habitat partitioning between these two species is present during the fall in this system but the mechanism is unclear.
4.2 Introduction

Identifying important fish habitat has become a primary focus in fisheries management because the Magnuson-Stevens Fishery Conservation and Management Act of 1996 mandated the National Marine Fisheries Service (NMFS) and the eight regional fishery management councils to identify essential fish habitat (EFH) for all managed fish species. EFH is defined as “those waters and substrate necessary for spawning, breeding, feeding, or growth to maturity” (Rosenberg et al. 2000). This policy has been indirectly applied to non-federally managed fishes through regional (e.g., Atlantic States Marine Fisheries Commission or ASMFC) and state (e.g., New York State Department of Environmental Conservation or NYSDEC) agencies. For example, the goals of the Hudson River Action Agenda for 2005 to 2009 are “to restore the signature fisheries to their full potential” and to conserve and protect critical river habitat to assure that these characteristic species are supported throughout their life stages. The Hudson River Estuary Management Plan (HREMP) states that currently declining species including American shad (*Alosa sapidissima*), Atlantic sturgeon (*Acipenser oxyrhynchus*), river herring (*Alosa pseudoharengus* and *Alosa aestivalis*), American eel (*Anguilla rostrata*) and largemouth bass (*Micropterus salmoides*), and the recovering striped bass fishery (*Morone saxatilis*) need to be managed through an ecosystem approach. Evaluating fish habitat suitability for all life stages is a goal of the HREMP; however no procedures were described as to how this goal was to be achieved (NYSDEC 2005).

Juvenile American shad and striped bass are two anadromous species that utilize the Hudson River estuary (HRE) and are managed by the ASMFC, but currently have very different population statuses. The HRE commercial and recreational fishery for
American shad has been closed recently because abundances have been decreasing steadily since the late 1980’s and have shown little evidence of a recovery. The HRE also supports a fully recovered striped bass population demonstrating dramatic increasing abundances from the mid 1990’s to the present (Limburg et al. 2006). Both American shad and striped bass are well-represented (or sampled) in the Utilities’ surveys, a fisheries independent survey. But, we know little about their habitat requirements during the juvenile stage. Our objectives are to examine whether factors such as sediment type, water characteristics and distance to nearest submerged aquatic vegetation (SAV) beds affect the occurrence or presence/absence of these species.

Our specific research question is, do factors such as sediment type or grain size, distance to nearest SAV, specifically water celery bed (*Vallisneria Americana*) and nearest water chestnut bed (*Trapa natans*), salinity, water temperature and dissolved oxygen influence the distribution of American shad and striped bass in the Hudson River estuary? To address this question, we utilize a hierarchical approach and apply generalized linear mixed models (GLMMs) to describe habitat suitability for American shad and striped bass. This approach has been used to identify EFH (see Valavanis et al. 2008 for a general example). Our results include maps depicting probability of occurrence (or presence) for both species throughout the HRE. Although the approach could be applied and be useful for all signature species, these two species were chosen because they can be compared directly due to their similar life histories but differing abundance trends and management status. This study represents a thorough first step in identifying essential/critical fish habitat for two important HRE species.
4.3 Methods

We modeled the distribution or probability of presence for juvenile American shad and striped bass in the HRE. Juveniles were chosen as the focus of this study because they are well sampled in the utilities’ survey (described below) and because these species are listed as ‘signature fisheries’ in the HREMP. Signature fisheries are those fisheries that support a commercial and/or recreational fishery in the HRE, and identifying critical habitat for all their life stages is a goal of the HREMP (NYSDEC 2005).

4.3.1 Study Area

After examining studies concerning spatio-temporal fish abundance trends in the Hudson River Estuary (Pace et al. 1993, Daniels 1995, ASA 2004, Hurst et al. 2004, Strayer et al. 2004) and evaluating existing Hudson River geographic information systems (GIS), including sediment type (Nitsche et al. 2005), submerged aquatic vegetation characteristics (Nieder et al. 2004), and water quality parameters (ASA 2004), we decided to focus the research on the entire Hudson River estuary. The study area or extent is specifically defined as the Battery (km 0) north to Albany (km 250) (Figure 4.1).

4.3.2 Fish Data

In the HRE, there has been an annual intensive multi-species survey since 1974 when year class reports were first prepared for the electric utility generating stations located on the HRE. The goal of these reports is to determine the direct impact that these utility companies have on the HRE fish populations. The utilities’ survey consists of three annual surveys, the Long River Survey (LRS), designed to assess egg and larval
densities; the Fall Juvenile Survey (FJS), to assess juvenile densities offshore; and the Beach Seine Survey (BSS), to assess nearshore fish communities and abundance. The survey area begins at the Battery and ends in Albany (ASA 2007). Over thirty years of spatio-temporal fish abundance data exist due to this survey. The utilities’ fish survey data have been subjected to the rigors of complete QA/QC protocol and have been used in numerous analyses published in peer-reviewed journals (Pace et al. 1993, Strayer et al. 2004, Dunning et al. 2006). Approximately 275 fish sample locations are visited each year following a stratified random sampling design (Figure 4.1). For this study we used presence or absence of young-of-the-year or juvenile American shad and striped bass throughout the HRE for the fall season for each year, 2000 to 2005. We only looked at the fall season because juvenile shad presence was very low through the estuary in the summer during the time period 2000 to 2005 and we were primarily interested in presence of juvenile American shad throughout the HRE. The time period, 2000 to 2005 was chosen because the sediment and submerged aquatic vegetation environmental data (see below) were collected during this time period only.

4.3.3. Environmental Data

The environmental data used were a collection of biotic and abiotic factors from two data sources, the New York State Department of Environmental Conservation (NYSDEC) and the utilities’ survey. The Euclidean distance, measured in meters, from the nearest submerged aquatic vegetation (SAV) bed to a fish sampling location were calculated using a map of the Hudson River estuary SAV distributions derived from aerial photography conducted by the New York State Department of Environmental Conservation (NYSDEC) in 2002. The SAV distribution map was classified as water
celery and water chestnut vegetation and exists as vector/polygon feature types.

The NYSDEC also mapped sediment type or grain size based primarily on sidescan sonar, subbottom profiling, and sediment grab or core samples collected throughout the HRE from 1998 through 2003. Sediment type and bathymetry data were obtained from Cornell University Geospatial Information Repository (CUGIR) (http://cugir.mannlib.cornell.edu/index.jsp) (Nitsche et al. 2004, Bell et al. 2006). The sediment types derived from this study were used as an environmental variable in our models. There were six sediment types identified in the HRE: mud, sandy mud, muddy sand, sand, gravelly sand and gravel (http://cugir.mannlib.cornell.edu/bucketinfo.jsp?id=7884).

Three additional environmental variables, water temperature, salinity and dissolved oxygen were chosen for consideration in the analysis based on availability and known or plausible roles influencing fish distributions. These variables, obtained from the utilities’ survey, were not collected on-station or co-located with the fish samples and were not subsequently georeferenced. However, spatial location, including the river mile, water depth and strata location (eg. beach, channel or east or west of the channel) was recorded for each water sampling site. We created a thalweg for the HRE (according to Merwade et al. 2005) to measure north-south distance within the river. The thalweg in conjunction with the aforementioned spatial location information allowed us to assign approximate position coordinates to the water and fish sample locations (see Figure 4.1).

Because the water samples were not co-located with the fish samples, we took a seasonal approach and interpolated fall bottom temperature, salinity and dissolved oxygen values for each year (2000-2005) using Universal Kriging. Universal Kriging is
an interpolation procedure that uses a regression model as part of the kriging process, modeling the unknown local mean values as having a local linear or quadratic trend. We modeled a linear longitudinal or north-south trend and a quadratic east-west trend for all water characteristic variables. Necessary variogram modeling and subsequent universal kriging procedures were conducted with the automap package in R. Interpolated bottom temperature, salinity and dissolved oxygen were mapped in R and corresponding values were extracted for each fish sampling site.

All continuous explanatory variables (water temperature, salinity, dissolved oxygen, distance to nearest water celery and water chestnut bed, river mile and Julian day), values were standardized by scaling each within a column mean of zero. This allows for a direct comparison of the relative importance of the explanatory variables (Gelman & Hill 2006).

4.3.4 Analytical Approach

We used GLMMs with a binomial error distribution to model juvenile shad and juvenile striped bass occurrence over the time period 2000 to 2005 during the fall season. We used the environmental variables described above as fixed effects, as well as river mile to denote spatial dependence, and Julian date to denote specific time during the fall season. Our model also included 2 random effects, year and site location, because we chose to build a global model (2000-2005) and subsequently created a repeated measures sampling design. We modeled the response variable as a binomial variable only to denote presence/absence for that particular species and estimated the probability of juvenile shad or juvenile striped bass occurrence throughout the HRE. The final
products are probability of presence maps for the global models for juvenile shad and juvenile striped bass.

We evaluated the models using calibration plots that show the agreement between predictions generated by a model and actual observations. This is also considered the “goodness-of-fit” of the model. The calibration plots were produced by breaking the predicted probability range up into bins, and plotting the proportion of sampling sites that are recorded present within each of these bins against the mean or median predicted value of each bin. Because the models will be used to produce probability of presence maps for each species, it is desirable that the predicted probability is well calibrated to the true probability.

All statistical analyses were conducted using R (R Development Core Team 2008), and programs written by the authors. Final predicted probability of occurrence maps for juvenile shad and juvenile striped bass were projected in ArcMap 9.3.

4.4 Results

Strong correlations were present between many pairs of explanatory variables (Table 4.1). Most notably, there were strong negative correlations between river mile and salinity \((r = -0.65)\) and river mile and distance to nearest water chestnut bed \((r = -0.74)\). Other strong correlations occurred between salinity and distance to nearest SAV bed \((r = 0.68)\) and between salinity and distance to nearest water chestnut bed \((r = 0.76)\). However, collinearity among the explanatory variables was not deemed sufficient to drop variables from the models, but will be considered in the interpretation of the model results.
Based on our global GLMMs, the primary drivers for probability of juvenile American shad presence were Julian day, river mile, and salinity (Table 4.2). We found a significant non-linear relationship between Julian day and American shad presence (Figure 4.2). Probability of presence is 0.4 in early fall and only approaches 0.5 before sharply declining to near zero in late Fall. A significant non-linear relationship also exists between probability of juvenile American shad presence and spatial dependence or river mile. American shad presence is positively correlated with lower river miles until mid-estuary (river mile 75) where it becomes slightly negatively correlated (Figure 4.3). Probability of juvenile American shad presence was also appropriately modeled as a non-linear relationship with salinity. There is a significant negative correlation (p≤0.01) between American shad presence and salinity at low salinity values (Figure 4.4). As salinity increases, probability of American shad presence decreases sharply and then asymptotes at salinities around 5 which are usually located mid-estuary during the fall season. Overall, the probability of juvenile American shad presence increases as river mile increases and the probability of juvenile American shad presence decreases as Julian day increases and as salinity increases.

The primary drivers for juvenile striped bass presence were multifaceted and included many significant correlations (p<0.05) with various explanatory variables (Table 4.2). A weak significant negative correlation (p<0.05) with Julian day exists but probability of juvenile striped bass presence remain high (0.8) throughout the fall season (Figure 4.2). A significant negative correlation (p≤0.01) between the probability of juvenile striped bass presence and river mile means that as river mile increases probability of striped bass presence decreases (Figure 4.3). Probability of juvenile striped
bass presence was appropriately modeled as a non-linear relationship with salinity. There is a weak significant positive correlation ($p<0.05$) between striped bass presence and salinity (Figure 4.4). As salinity increases, striped bass presence increases sharply and then asymptotes at salinities around 5 which are usually located mid-estuary during the fall season. A non-linear significant correlation ($p \leq 0.01$) with water temperature shows that juvenile striped bass presence is positively driven by increasing temperature although a threshold might exist (Figure 4.5). There are also strong positive significant correlations ($p \leq 0.01$) between probability of striped bass presence and fine grained sediments. Overall, the probability of juvenile striped bass presence increases with increasing salinity, water temperature and finer grained sediments and decreases with increasing river mile. Also, common significant correlations between juvenile American shad and striped bass include Julian day, river mile and salinity but the direction of the correlations are very different and in most cases are opposite as seen in Figures 4.2 through 4.4.

Calibration or ‘goodness of fit’ plots (Figures 4.6 and 4.7) show a high level of agreement between predictions generated by each global model and actual observations of presence/absence of juvenile American shad and juvenile striped bass. Based on this result we mapped predictions from the models for both species (Figure 4.8 and 4.9). Highest predicted probabilities of juvenile American shad presence are found in the Upper HRE. Conversely, highest predicted probabilities of juvenile striped bass presence are found in the Lower HRE.
4.5 Discussion

In addition to meeting our primary goal identifying essential or critical habitat for juvenile American shad and striped bass during the fall (2000 to 2005) throughout the HRE (Figure 4.8 and Figure 4.9), we were able to outline a method that can be used for all signature species and life stages in the HRE. By applying GLMMs and interpreting the resulting parameter estimates, we are also able to describe the relationship between juvenile American shad and juvenile striped bass and the environmental variables. Identifying habitat requirements for all life stages and seasons over varying time periods could be useful in management decisions and to promote additional research needs.

The spatial patterns of juvenile American shad and juvenile striped bass probability of presence during the fall in the HRE were significantly related to several explanatory variables over the time period 2000 to 2005. Julian day, spatial dependence or river mile and salinity were the foremost variables for predicting presence/absence of juvenile American shad. Based on the magnitude of significant (p<0.05) parameter estimates for the juvenile striped bass GLMM (Table 4.2), we conclude that Julian day, river mile, salinity and water temperature are the major drivers for predicting presence/absence of juvenile striped bass although many significant (p<0.05) drivers exist (eg. sediment type, distance to nearest SAV and distance to nearest water chestnut bed). This discussion will primarily focus on the common drivers found between the two species, Julian day, river mile and salinity after discussing visible collinearity among certain variables (Table 4.1).

First, salinity and river mile have a strong negative correlation (-0.65) because salinity decreases as river mile increases. The salt water/freshwater interface vary with
season and can occur anywhere from rkm 25 in the spring to rkm 100 in the fall. But generally the HRE becomes a tidal freshwater environment mid-estuary, near Kingston, New York (rkm 125). Salinity is also highly correlated with both distance to nearest SAV bed (0.68) and distance to water chestnut bed (0.76). Water chestnut beds are found only in the freshwater region of the Upper HRE and water celery beds are found through the HRE in all salinity environments. But from this collection of intercorrelated variables, it was possible to discern a general pattern of juvenile American shad and juvenile striped bass presence/absence for the entire HRE and accomplish our goal to map EFH for both species during the fall.

The relationship between juvenile American shad and juvenile striped bass can be viewed as a negative one when considering Julian day, river mile and salinity as depicted in Figures 4.2 through 4.4. As time progresses throughout the fall as depicted (Figure 4.2) striped bass probability of presence remains unchanged while juvenile American shad decreases. Striped bass probability presence remains constant over the fall suggesting that they remain in the HRE during this season. From the probability of presence map for juvenile American shad (Figure 4.8) we see that they are probably not leaving the estuary, based on their low probability of presence in the lower HRE, but perhaps they are experiencing mortality which we discuss in further detail below. There is also a strong negative spatial relationship during the fall between these species as seen in Figure 4.3 and in probability of presence maps for both species (Figure 4.8 and Figure 4.9). Juvenile American shad utilize the upper HRE and juvenile striped bass primarily utilized the lower HRE. Bilkovic et al. (2002b) found that peak striped bass spawning activity in the York River system took place below areas of peak American shad.
spawning, with little spatial overlap. We conclude that this spatial partitioning may occur in juvenile American shad and juvenile striped bass also. Finally, we suspect that the significant relationship with salinity for both species is probably a confounding effect of spatial dependence due to the strong north south salinity gradient in the HRE (Fig. 4.4) as denoted by the spatial partitioning seen in Figure 4.8 and in Figure 4.9.

We also found a significant ($p<0.05$) positive relationship between both finer grained sediments probability of juvenile striped bass presence. Kennish et al. (2004) in Mullica Bay-Great Bay and Strayer et al. (2004) in the Hudson River found that sediment type or grain size and bottom structure strongly influences benthic invertebrate assemblages. Howe et al. (2008) found that age-0 striped bass collected in the lower HRE (rkm 35 to rkm 75) selectively feed on invertebrates, specifically, gammarid amphipods, Crangon spp. shrimp, and chironomid larvae. They also found higher numbers of invertebrates and higher numbers of age-0 striped bass at sights with finer sediments (Howe et al. 2008). Also, Overton et al. (2009) while studying diets of striped bass in the Chesapeake Bay found small sized (150 to 300 mm) striped bass primarily prey on invertebrates.

American shad are anadromous, spending most of their adult life in salt water and entering the Hudson River estuary to spawn in the freshwaters of the upper estuary from April to June. Their migration from the sea begins when the estuary reaches 12° C (Werner 1980). A female shad can produce 116,000 to 468,000 semibuoyant, non-adhesive eggs per spawning season. Historically, shad eggs, yolk sac larvae (YSL) and post yolk sac larvae (PYSL) are found mainly in the upper estuary. As young-of-the-year (YOY) they are found in the upper, middle and lower estuary and emigrate from the
Hudson River at the end of the summer. This emigration is probably triggered by declining water temperature and increased fish size (ASA 2007). However, YOY American shad were usually absent through the HRE during the summer over the time period 2000 to 2005 per the utilities’ survey. This could be due to the species being already in a decline. Adult and juvenile American shad abundances are low in the HRE and have been declining since the early 1990’s (ASMFC 2007).

The movement of juvenile American shad has been well studied in the HRE. Limburg (1996) conducted a fine-scale demographic study in the HRE of the 1990 year class of YOY American shad using otolith microanalysis and found that migration was both size- and age-biased. Cohorts moved down the river as early as June, older cohorts tending to move earlier than younger ones. She found that by September most size classes of fish utilize the middle estuary, and by late October, fish move to the lower estuary (even in the face of higher predation risk), due to a combination of lower food resources and thermal risks in the upper and middle estuary. Limburg (2001) confirmed this bimodal first emigration pattern with respect to age and size for the 1989 and 1990 cohorts also through otolith microanalysis but was unable to validate this finding with actual YOY data collected in 1989 and 1990 thus suggesting differential mortality due to predation and hypoxia experienced in the lower HRE (Limburg 2001). Our results show that YOY or juvenile American shad probability of presence is very low in the lower HRE and higher in the upper HRE during the fall (Figure 4.8) and therefore does not illustrate successful migration. Also, American shad probability of presence decreases through the season (Fig. 4.2) suggesting that they are experiencing mortality. Based on
the spatial partitioning we found between juvenile American shad and juvenile striped bass, differential mortality due to predation could explain our findings.

Striped bass are also anadromous and spawn between mid-May and mid-June in the middle of the HRE. Peak spawning occurs when water temperatures are 15.5 °C to 18 °C in the freshwater areas where the currents are moderate to swift. Depending on their age and size, female striped bass can produce up to several million semibuoyant eggs that are suspended by currents. Historically, striped bass eggs and yolk-sac larvae (YSL) are found mainly in the mid HRE but YSL are sometimes found further upriver. Striped bass post yolk-sac larvae (PYSL) are most abundant in the mid to lower estuary but can be found throughout the HRE. As YOY they are found primarily in the lower estuary but at the end of their first summer they can also be found in New York Harbor, western Long Island Sound, and the south shore of Long Island. Striped bass usually emigrate from Atlantic coast estuaries at age-2 or 3 (ASA 2007).

Juvenile American shad are common prey for striped bass in other systems but have not been consistently observed in the HRE. Gardinier and Hoff (1982) reported finding American shad in the stomachs of striped bass collected over the period April through November. However, Dunning et al. (1997) and Kahnle and Hattala (2007) examined stomachs of adult striped bass collected in the winter and spring, respectively, and did not find American shad as a prey item suggesting that season may influence diet. Jordan et al. (2003) found that among-year differences and spatial differences were substantially associated with age-0 or juvenile striped bass diet in the lower HRE in late July through November from 1994 to 1997. These diets primarily consisted of amphipods. However, HRE striped bass feeding habits in the mid-HRE has not been well
studied and this is where we might expect the highest levels of predation on juvenile American shad to occur (Figure 4.8) if this relationship exists. Also, investigating age-1 or older striped bass diets in the HRE would also be valuable in evaluating this potential relationship. For example, Tuomikoski et al. (2008) found that age-1 striped bass in the western Albemarle Sound, North Carolina preyed upon several commercially important species including juvenile American shad.

In our attempt to identify essential or critical fish habitat for juvenile American shad and juvenile striped bass we found distinct fall presence/absence patterns for both species. Based on the probability of presence maps we see a clear north-south partitioning of the estuary by juvenile American shad and striped bass (Figure 4.8 and Figure 4.9) that could be due to habitat partitioning or a predator-prey relationship. A strong predator prey linkage between age-1 striped bass and juvenile American shad was found in the western Albemarle Sound, North Carolina (Tuomikoski et al. 2008). Limburg’s (2001) suggestion of differential mortality due to predation supports our idea of a possible predator prey linkage between striped bass or other predatory fish (drum, spotted sea trout, predatory flounder etc.) and juvenile American shad in the mid and lower HRE.
Figure 4.1: Figure of the approximate fish and water sampling locations for Electric Utilities’ Hudson River estuary survey (2000-2005).
Table 4.1: Pearson correlation coefficients between all pairs of environmental variables.

<table>
<thead>
<tr>
<th></th>
<th>Julian day</th>
<th>River mile</th>
<th>Salinity</th>
<th>Water temperature</th>
<th>Dissolved Oxygen</th>
<th>Distance to SAV*</th>
<th>Distance to WCh**</th>
<th>Sediment Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julian day</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>River mile</td>
<td>-0.16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salinity</td>
<td>0.25</td>
<td>-0.65</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water temperature</td>
<td>-0.27</td>
<td>-0.25</td>
<td>0.14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissolved Oxygen</td>
<td>0.06</td>
<td>0.36</td>
<td>-0.37</td>
<td>-0.48</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to SAV*</td>
<td>0.32</td>
<td>-0.38</td>
<td>0.68</td>
<td>-0.14</td>
<td>-0.16</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to WCh**</td>
<td>0.19</td>
<td>-0.74</td>
<td>0.76</td>
<td>0.17</td>
<td>-0.24</td>
<td>0.58</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Sediment Type</td>
<td>-0.02</td>
<td>0.62</td>
<td>-0.30</td>
<td>0.29</td>
<td>0.37</td>
<td>-0.12</td>
<td>-0.32</td>
<td>1</td>
</tr>
</tbody>
</table>

* SAV=Submerged aquatic vegetation
** WCh=Water chestnut
Table 4.2: Parameter estimates and standard error from the global generalized mixed models (GLMM).

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>American shad Estimates</th>
<th>striped bass Estimates</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.36</td>
<td>0.23</td>
<td>0.61*</td>
</tr>
<tr>
<td>Julian Day</td>
<td>9.34***</td>
<td>1.03</td>
<td>-1.71*</td>
</tr>
<tr>
<td>Julian Day²</td>
<td>-9.66***</td>
<td>1.05</td>
<td>1.28</td>
</tr>
<tr>
<td>Salinity</td>
<td>-1.44***</td>
<td>0.21</td>
<td>0.46*</td>
</tr>
<tr>
<td>Salinity²</td>
<td>1.14***</td>
<td>0.20</td>
<td>-0.84***</td>
</tr>
<tr>
<td>River mile</td>
<td>1.25***</td>
<td>0.44</td>
<td>-1.13**</td>
</tr>
<tr>
<td>River mile²</td>
<td>-0.83**</td>
<td>0.36</td>
<td>-0.08</td>
</tr>
<tr>
<td>Water Temperature</td>
<td>-0.27</td>
<td>0.47</td>
<td>1.35**</td>
</tr>
<tr>
<td>Water Temperature²</td>
<td>0.58</td>
<td>0.48</td>
<td>-1.54***</td>
</tr>
<tr>
<td>Dissolved Oxygen</td>
<td>0.05</td>
<td>0.39</td>
<td>-0.12</td>
</tr>
<tr>
<td>Dissolved Oxygen²</td>
<td>-0.09</td>
<td>0.39</td>
<td>-0.09</td>
</tr>
<tr>
<td>Distance to SAV</td>
<td>0.18</td>
<td>0.11</td>
<td>0.25*</td>
</tr>
<tr>
<td>Distance to WCh</td>
<td>-0.08</td>
<td>0.11</td>
<td>-0.26*</td>
</tr>
</tbody>
</table>

**Sediment Type**

<table>
<thead>
<tr>
<th>Sediment Type</th>
<th>American shad Estimates</th>
<th>striped bass Estimates</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mud</td>
<td>0.07</td>
<td>0.11</td>
<td>0.33***</td>
</tr>
<tr>
<td>Sandy Mud</td>
<td>-0.05</td>
<td>0.19</td>
<td>0.72***</td>
</tr>
<tr>
<td>Muddy Sand</td>
<td>0.06</td>
<td>0.23</td>
<td>0.86***</td>
</tr>
<tr>
<td>Sand</td>
<td>-0.49</td>
<td>0.40</td>
<td>0.10</td>
</tr>
<tr>
<td>Gravelly Sand</td>
<td>1.00</td>
<td>0.68</td>
<td>0.36</td>
</tr>
<tr>
<td>Gravel</td>
<td>-0.48</td>
<td>1.31</td>
<td>-1.58</td>
</tr>
</tbody>
</table>

**P < 0.01**

**P=0.01**

*P<0.05
Figure 4.2: Plot of the probability of juvenile American shad and juvenile striped bass presence versus Julian day in the HRE based on the global GLMMs.
Figure 4.3: Plot of the probability of juvenile American shad and juvenile striped bass presence versus river mile in the HRE based on the global GLMMs.
Figure 4.4: Plot of the probability of juvenile American shad and juvenile striped bass presence versus salinity in the HRE based on the global GLMMs.
Figure 4.5: Plot of the probability of juvenile striped bass presence versus water temperature in the HRE based on the global GLMMs.
Figure 4.6: Calibration plot for juvenile American shad global GLMM.
Figure 4.7: Calibration plot for striped bass global GLMM.
Figure 4.8: Map of the probability of juvenile American shad presence in the HRE based on the global GLMM (2000-2005). Higher probabilities are depicted in red and lower probabilities are depicted in blue.
Figure 4.9: Map of the probability of juvenile striped bass presence in the HRE based on the global GLMM (2000-2005). Higher probabilities are depicted in red and lower probabilities are depicted in blue.
CHAPTER 5
A CRITICAL EXAMINATION OF FINDINGS, AND PROPOSED FUTURE DIRECTIONS

5.1 Climate and Hudson River American shad

We found that the Hudson River estuary (HRE) fish community is influenced by regional climate and local hydrology. Specifically, we confirmed that larval striped bass have a strong positive relationship with freshwater flow and we also found that juvenile American shad have a strong negative relationship with the Atlantic Multidecadal Oscillation (AMO). Next, we used cross correlation functions (CCF), a time series approach (Crawley 2007), to examine how the AMO affected juvenile American shad. Specifically, we looked at whether changes in juvenile shad from 1974-2005 lagged AMO changes. First, we used CCFs to analyze two stationary time series, juvenile shad abundance indices and the AMO from 1974 to 2005, to indicate any direct or delayed dependence between them. We found a significant negative correlation between the AMO and HRE juvenile shad abundances at a 3 year positive lag (Figure 5.1). Future studies should examine this lag further because it may reveal a mechanistic link between ocean climate and HRE juvenile American shad populations.

We propose that additional work should be conducted to determine if the significant 3 year positive lag between the AMO and juvenile American shad is ‘real’. This could be tested by analyzing secondary data using a correlational approach. American shad abundance or presence data could be obtained from the fishery-independent trawl surveys conducted by the Northeast Fisheries Science Center, Woods
Hole MA. The SST data could originate from the 5X5 Kaplan data set version 2; and the AMO climate index in this study could be an unsmoothed time series of the basin wide average of these SST data in the North Atlantic basin (www.cdc.noaa.gov). The first step would be to demonstrate the probable correlation between the AMO and SST data and then to examine possible correlations between American shad abundance or presence and the AMO and SSTs. The results could be represented spatially on a map to aid in future conceptual modeling.

If the aforementioned lag is validated, future studies should focus on the effects of climate, specifically SSTs, on the physiology or reproductive biology/characteristics of ocean migrating shad. Developing a conceptual model of American shad reproductive strategies (Figure 5.2) is an important first step to elucidating possible mechanistic links between ocean climate and HRE American shad populations. One facet of a possible causal link could involve geographic variations in the life history of American shad and climate change as demonstrated in Figure 5.2 based on Leggett and Carscadden’s (1978) research. They found a correlation between latitude and population variation in American shad reproductive characteristics or strategies. American shad populations are semelparous south of Cape Hatteras, North Carolina and they are iteroparous north of Cape Hatteras. Leggett and Carscadden (1978) also found that the degree of iteroparity or percent repeat spawning increases in American shad populations from North Carolina to Canada. This gradient demonstrates an adaptive strategy driven by the thermal regime of their natal rivers (Leggett and Carscadden 1978). Proposed additional work includes re-examining differences in the proportion of repeat spawners among the iteroparous
American shad populations in light of changing thermal regimes in many riverine systems.

5.2 Relationship between Juvenile American shad and striped bass in the Hudson River Estuary

While studying habitat suitability for juvenile American shad and juvenile striped bass, we found a possible predator prey relationship between the two species, at this life stage, based upon an obvious north-south spatial partition (Figure 4.8 and Figure 4.9). This relationship has not been confirmed nor invalidated in the HRE. Season specific, especially fall, diets of HRE striped bass should be examined in the mid and lower HRE to determine if they are preying on juvenile American shad. If this is occurring, we could quantify the amount of American shad removed from the system due to predation and get a more accurate measurement of mortality.

We also propose a telemetry study because information on the movement of juvenile American shad could be useful in determining migration patterns from the upper HRE to the lower HRE in the summer and fall. The north-south partition between juvenile American shad and juvenile striped bass is still evident in the summer based on preliminary results (Figure 5.3 and Figure 5.4) but there are some patchy near-shore areas of higher probability of presence for juvenile American shad in the lower HRE. In figure 4.8, these areas of higher probability of presence do not exist during the fall and only occur in the upper HRE. This suggests that juvenile American shad start to migrate in the summer but do not continue to migrate during the fall. This result could be validated or invalidated through the proposed telemetry study.
Lastly, other spatial habitat variables or possible habitat requirements should be included in the global models such as, land use or contaminant levels or biotic variables including prey items and competing invasive species. These types of stressors could negatively affect the overall health of American shad during all life stages. For example, the HRE has been affected by an invasive species, the zebra mussel since the early 1990’s. Strayer et al. (2006) found larval American shad were shifted down river because of limited prey due zebra mussel colonization in the upper HRE. This result suggests that the larvae are in poor condition because of reduced food availability and are now forced to migrate to the mid and lower HRE where they may face continued starvation or predation. This type of stressor could reduce recruitment to the already closed American shad fishery.
Figure 5.1: Cross-correlation function (CCF) plot for HRE juvenile American shad abundance indices and the AMO over the time period, 1974 to 2005.
Figure 5.2: Conceptual model of population-level variation in reproductive characteristics for adult American shad under the influence of increased water temperatures.
Figure 5.3: Map of the probability of juvenile American shad presence in the HRE during the summer based on the global GLMM (2000-2005). Higher probabilities are depicted in red and lower probabilities are depicted in blue.
Figure 5.4: Map of the probability of juvenile striped bass presence in the HRE during the summer based on the global GLMM (2000-2005). Higher probabilities are depicted in red and lower probabilities are depicted in blue.
LITERATURE CITED


Delworth, T.L., and Mann, M.E. 2000. Observed and simulated multidecadal variability in the Northern Hemisphere. Climate Dynamics. 16: 661-676


ESRI. Environmental Science Research Institute, Inc. 2004. Arc/Info 9.0 software.

Everly, A. W., and J. Boreman 1996. Habitat requirements of important fish species inhabiting the Hudson river estuary. NMFS, Office of Habitat Protection.


Hoffman, J.C., K. E. Limburg, D.A. Bronk, and J. E. Olney


