Special Report No. 48

of the

ATLANTIC STATES MARINE FISHERIES COMMISSION

Report of the Juvenile Abundances Indices Workshop

DECEMBER 1995

Report of the

Juvenile Abundance Indices Workshop

by

Paul J. Rago
National Marine Fisheries Service
Woods Hole, MA

C. Dianne Stephan
Atlantic States Marine Fisheries Commission
Washington, DC

and

Herbert M. Austin
Virginia Institute of Marine Science
Gloucester Point, VA

Convened by:

Atlantic States Marine Fisheries Commission
January 21-22, 1992
Grasonville, MD

December 1995

Acknowledgements

Funding for this workshop was provided by the National Marine Fisheries Service, U.S. Fish and Wildlife Service, Environmental Protection Agency, and Chesapeake Bay Stock Assessment Committee.

Many thanks to Steering Committee members Harry Hornick and Herb Austin, the workshop participants, and Paul Perra, Anne Richards, and Steve Jordan.

Printing of this report was supported by the U.S. Fish and Wildlife Service Sportfish Restoration Program.

Table of Contents

List of Tables

List of Figures

Executive Summary

Juvenile abundance indices (JAIs) are frequently a critical component of interstate fisheries management programs. Such indices can provide a measure of annual recruitment success, a predictor of potential fishery yields, and a trigger for either relaxing or restricting fisheries. A workshop was convened on January 20-21, 1992 to address the development and use of JAIs and to provide guidance for future work.

Participants included representatives from state and federal agencies, universities and consulting firms. This report provides a summary of the key findings and recommendations of the workshop participants. Striped bass was used as a model species, though findings should be applicable to similar surveys for other species. Summaries of existing state striped bass and/or alosid juvenile sampling programs and data are included. Technical issues are addressed in detail in appendices prepared by independent consultants. These include short introductions to key problems and a guide to relevant literature.

Key workshop recommendations are summarized below:

Sample Design

- Replicate seine hauls are not independent samples, and should be discontinued. Historical estimates should be recomputed using catches from first replicate haul.

- Surveys based on the concept of partial replacement designs--a mix of fixed and random sites, may be advantageous. Effort now devoted to auxiliary sites (or other effort that can be redirected) should be devoted toward sampling of randomly selected beaches, so that studies can be conducted to determine the most appropriate mix of fixed and random beaches.

- Development of appropriate weights for comparing results from several systems is critical to proper interpretation of the data. The approach of developing weights based on the extent of habitat in each systems should be continued.

- A list of seinable beaches, and habitat in general, should be developed for each system to monitor habitat change and to provide an expanded sampling frame for random samples.

- Efforts should be made to identify shorter time windows for sampling and reallocating effort toward greater spatial coverage through selection of random sites.

- Revisions to long term programs must allow comparability with historical data.

Estimation and Analyses

- Interannual changes in fixed station indices can be used to detect qualitative changes in abundance. In contrast to the problematic nature of fixed station designs, variances of station differences may satisfy conditions necessary of statistical inference.

■ Nonparametric and graphical methods should be used to explore the qualitative aspects of data. An index based on ranks should be developed to supplement information obtained from catches per haul.

■ Adjustment of catches with ancillary information (e.g., area swept by gear) is not likely to be useful. Thus, efforts should be made to use "standard" hauls as much as possible. Relating individual catches (seine hauls or trawls) to environmental variables will be difficult.

■ Transformations of data to achieve normality are possible but data should be stratified appropriately. Log transformation of the (catch +1) per tow data appears to satisfy the normality assumption but this becomes problematic when large numbers of zeros are present.

Validation

■ Correlations of juvenile indices with commercial catches appears to be a viable method of validation. Resampling methods (e.g., bootstrap, cross-validation, jackknife) should be done to verify derived relationships.

■ Correlations with later life stages are the most viable means of validating more recent time series. Statistical techniques which incorporate the errors in both independent and dependent variables should be used.

■ Utility of marked hatchery fish to estimate population abundance has been demonstrated in the Delaware and Patuxent rivers. More detailed examination of data from these experiments might assist in design of validation studies.

Many of the current management regulations for striped bass are based on the Maryland juvenile index. This index played a central role in the development of management measures in the early 1980's that lead to the recovery of striped bass. Today the Maryland juvenile index continues to be an important component of quota formulation and the estimation of stock status. As the precision and accuracy of indices from other states improve and are validated with independent data, coast-wide management will rely more heavily on these measures of recruitment. The importance of developing and agreeing upon a methodology for combining multiple indices into an overall estimate of juvenile abundance cannot be overstated.

Several states have implemented one or more of the workshop recommendations. Random stations have been added to survey programs, replicate samples have been dropped and additional sampling sites have been identified. The Striped Bass Technical Committee has re-evaluated the Maryland juvenile index and recommended the use of the geometric mean as the best index of abundance. The Technical Committee also recommended that the index be weighted by the area of the juvenile nursery habitat in each river system. The former recommendation has been endorsed by the Striped Bass Management Board of ASMFC, while the latter is still under consideration.

LIST OF PARTICIPANTS

NAME	AGENCY	ADDRESS
Lynn Henry	North Carolina Division of Marine Fisheries	1367 US 17 South Elizabeth City, NC 27909
Kim McKown	New York Department of Environmental Conservation	Bldg #40, SUNY Stony Brook, NY 11790-2356
Dianne Stephan	Atlantic States Marine Fisheries Commission	1776 Massachusetts Ave., NW Washington, DC 20036
Thomas Squires	Maine Department of Marine Resources	State House Station #21 Augusta, ME 04333
Gary Shepard Anne Richards	National Marine Fisheries Service	166 Water Street Woods Hole, MA 02543
Ginny Fay	National Marine Fisheries Service	One Blackburn Drive Gloucester, MA 01930
Andy Kahnle	New York Department of Environmental Conservation	21 South Putt Corners Road New Paltz, NY 12561
John M. Hoenig William G. Warren	Independent Consultants	Department of Fisheries & Oceans P.O. Box 5667 St. John's, CN A1C 5X1
Erik Barth	Virginia Marine Resource Commission	2600 Washington Avenue Newport News, VA 23607
Mark Gibson	Rhode Island Division of Fish & Wildlife	P.O. Box 218 West Kingston, RI 02892
Tom Mosca III Jim Colvocoresses Herb Austin	Virginia Institute of Marine Sciences	Gloucester Point, VA 23062
Harry T. Hornick Jim Uphoff Don Cosden Steve Jordan Carol McCollough	Maryland Department of Natural Resources	Tidewater Administration C-2 Tawes State Office Bldg Annapolis, MD 21401
Raj Tilak	D.C. Fisheries Management Board	2100 Martin Luther King Ave., SE Washington, DC 20020

NAME	AGENCY	ADDRESS
Steve Weisberg	Versar	9200 Rumsey Road Columbia, MD 21045
Hal Wilson	Coastal Environmental Services, Inc.	1099 Winterson Road Linthcum, MD 21090
Peter Himchak	New Jersey Bureau of Marine Fisheries	P.O. Box 418 Port Republic, NJ 08241
Paul Rago Bob Dorazio	U.S. Fish & Wildlife Service	P.O. Box 700, Leetown Kearneysville, WV 25430
Vic Crecco	Connecticut Department of Environmental Protection	P.O. Box 719 Old Lyme, CT 06371
Jim Francesconi	North Carolina Division of Marine Fisheries	P.O. Box 769 Morehead City, NC 28557

Introduction

The collection of juvenile abundance data to predict future harvestable stock size was given early consideration by Hjort (1914) when he proposed annual monitoring of juvenile distribution and abundance. Since that time, many researchers have implemented juvenile abundance monitoring programs from which indices of finfish recruitment have been developed. These recruitment indices are important components of the information base used by fishery managers to manage fish stocks.

Two of the primary issues associated with the use of juvenile abundance indices (JAIs) in fisheries management are: 1) the accuracy with which the indices represent year class strength, and 2) whether the indices are sufficiently precise to allow detection of differences among year-classes of different strengths.

Use of JAIs to emulate recruitment assumes that the indices accurately reflect year class strength. This assumption can be tested by comparing JAIs with future stock abundance (Goodyear 1985, McKown 1992b). However, the imprecise nature of JAIs also requires careful consideration when integrating this information into management programs.

The lack of precision in JAIs stems from several sources, including patchy distribution of schools, shifts in general distribution patterns, variation in production among different areas within the survey area, and shifts in distribution during the time periods of the survey. Some of the effects of these factors can be reduced during index estimation. For example, stratification by area and sampling period reduced the coefficient of variation of geometric means calculated from a striped bass juvenile abundance survey from 103% to just 30% (Richards 1992).

Another issue relevant to the management of interstate fisheries is the development of coastwide JAIs from surveys located in different geographical regions. Usually monitoring surveys performed by different agencies employ different survey methodologies. In order to manage a species on a coastwide basis, it may be desirable to develop a method for deriving a coastwide index based upon the data from several different regional surveys.

In order to address these and other issues relating to juvenile abundance survey design and data analysis, the Juvenile Abundance Indices Workshop was convened on January 21-22, 1992. The workshop was sponsored by the Atlantic States Marine Fisheries Commission, U.S. Fish and Wildlife Service, and Maryland Department of Natural Resources, and supported by the Emergency Striped Bass Study of both the National Marine Fisheries Service and U.S. Fish and Wildlife Service, the Chesapeake Bay Program of the Environmental Protection Agency, and the Chesapeake Bay Stock Assessment Committee.

Workshop discussions were based on sampling methodology and data analysis for Atlantic coast state striped bass and/or alosid juvenile abundance surveys. Striped bass was used as a model species because of the many rigorous monitoring programs in place and the availability of long-term databases for this species. Recommendations developed based on these databases should be applicable to other similarly conducted surveys. An ancillary workshop immediately following the Juvenile Abundance Indices Workshop discussed such broader application.

Workshop participants made short presentations about their monitoring programs or special research efforts. Representatives from each state provided catch data from each monitoring program in a standard format for use during data analysis sessions. Central issues included sampling strategies, historical comparability of data, index validation, and development of composite indices.

Independent consultants John M. Hoenig and William G. Warren participated in the workshop, and provided written reports which are appended. These reports discuss the technical issues in detail, and include a guide to relevant literature.

This workshop proceedings summarizes the state monitoring programs that were presented, and discusses the central issues raised regarding survey design, and index estimation, analysis, and validation. Recommendations which resulted from workshop discussions are included under each discussion.

State Monitoring Programs

Kennebec River, Maine

Improvements in water quality and fish passage initiated in the early 1970s have sought to restore the habitat of the Kennebec River estuary and dependent stocks of anadromous fish. The Kennebec River Juvenile Survey (KRJS) was initiated in 1979 in order to establish a time series on juvenile alosid recruitment which would track the recovery of these stocks (Squires 1992).

The survey area totals 45.1 square kilometers of tidal riverine habitat, which consist of the tributaries and main branch of the Kennebec River, including the Androscoggin, Eastern, Cathance, and Abagadasset Rivers and Merrymeeting Bay (Figures 1 and 2). Four sites are sampled biweekly on the mainstem of the Kennebec River from Merrymeeting Bay to Edwards Dam, from midJuly through October. Merrymeeting Bay is sampled biweekly at four sites, and the Androscoggin River is sampled biweekly at three sites. The Eastern, Abagadasset, and Cathance rivers are sampled monthly at one site. All sampling sites are located at fixed stations. . . . .

The current sampling gear is a beach seine (17 m x 1.8 m, 6 mm stretched mesh) with a 1.8 m by 1.8 m bag. The seine is set from a boat perpendicular to shore. The inshore end is held stationary at the edge of the water while the offshore end is towed in an upcurrent arc toward shore and then hauled in. One haul is made at each site, and covers approximately 220 m^2. A summary of the current sampling design is included in Table 1.

A smaller seine (9.1 m x 1.8 m, 6 mm stretched mesh) was used when the survey was initiated in 1979. Seine poles were held approximately 6 m apart and towed 30 m, parallel to the shore, which swept an area of approximately 214 m^2. From 1979 to 1982, the Kennebec River and its tributaries were sampled at over 35 sites from Popham upriver to Augusta. In 1983 the current gear was employed. Comparison between the past and present seines found no significant difference in the catch of alewife (Alosa pseudoharengus), which is the most abundant species present.

Juvenile indices from the KRJS are calculated as the arithmetic mean of individuals captured per seine haul. Annual indices for striped bass are given in Table 2. Indices have not been validated, and are used only as an indication of relative abundance of spawning stock.

Hudson River, New York

The New York Department of Environmental Conservation conducts two surveys in the Hudson River and surrounding estuary, primarily targeting juvenile striped bass (McKown 1992a). The beach seine survey samples the Tappan Zee to Haverstraw Bay area of the

Hudson River, from river mile 23 through 40 (Figure 3). The trawl survey samples river mile 24 through 43. In addition, a beach seine survey to sample juvenile American shad (Alosa sapidissima) is conducted upriver of the striped bass surveys.

Beach Seine Survey - The gear used is a 61 m x 3 m beach seine with a 5 mm stretched mesh bag and 6 mm stretched mesh wing, and is set by boat and retrieved by hand. A total of 33 fixed stations are currently designated. Station selection is based on prevailing wind and tide conditions. Twenty-five stations are sampled biweekly from mid-July through early November. One haul is made per station, with varying area coverage based on station and tide conditions. The first round of sampling is scheduled to have a midday slack tide in order to sample all tide stages per season, and the same lunar cycle each year. Current sampling design is summarized in Table 1.

The survey has evolved since its inception in 1976. Forty-one stations were designated from 1976 through 1985; in 1985 eight stations were dropped because of obstructions. Beginning in 1980 the survey ran for six weeks from late August through early November. In 1985 survey duration was increased to nine weeks.

Trawl Survey - A Carolina wing bottom trawl with a 8.5 m headrope and stretched mesh sizes 38 mm for the body, 32 mm cod end, and 13 mm cod end liner is towed for five minutes at each station at a speed of 3.9 m/s. Twenty-six fixed stations are located on offshore shoals in water ranging from 2 to 10 m in depth. All stations are sampled biweekly, from mid-July through early November. Sampling dates for this survey are also chosen to be on the same lunar cycle each month and to sample all tidal stages over the season. Current sampling design is summarized in Table 1.

A number of changes in sampling design have occurred over the life of the survey. From initiation in 1981 until 1983, river miles 56 through 62 were also sampled, but were dropped because of low striped bass catches. The survey was performed for six weeks from early September through early November in 1981. From 1982 through 1984, the survey was expanded to eight weeks and lasted from early August through early November. The survey began in late June or early July and lasted for ten weeks in 1985, 1987, and 1988. In 1985 the survey was expanded to its present duration. The survey was not performed in 1991 because of vessel repairs.

In both the beach seine and trawl surveys, the arithmetic mean is used as the relative index of abundance, with each station and survey having equal weight. Annual striped bass indices and statistical properties are given in Table 2. Because of a seasonal onshore - offshore distribution pattern, an index combining these surveys has also been generated (McKown 1992b).

In order to use a juvenile index as an indicator for recruitment in ASMFC's management program, the index must be validated by comparing it to future landings or abundance of the year class later in life (ASMFC 1989). McKown (1992b) compared the beach seine, trawl and combined indices with indices of abundance generated from a DEC survey of yearling striped bass in the bays of western Long Island (Figure 3). All three indices showed good correlation. In addition, the indices were compared with yearling indices of abundance generated by a New York Power Authority (NYPA) survey. Yearling abundance from the NYPA survey correlated with the beach seine survey index and the combined index.

Delaware River, New Jersey

The New Jersey Department of Environmental Protection initiated a survey in 1980 to provide an annual relative index of abundance for young-of-the-year striped bass in the Delaware River (Himchak 1992). The survey area extends from Trenton, NJ south to the salt/freshwater interface south of the Delaware Memorial Bridge (Figure 4), and is divided into three regions. Region I is brackish tidal water, and extends from the southernmost portion of the survey area north to the Delaware Memorial Bridge. Region II is brackish to fresh tidal water and extends from the Delaware Memorial Bridge to the Schuylkill River at the Philadelphia Naval Yard. Region III is a tidal freshwater area which extends from Philadelphia north to Trenton. Regions I and II represent the historical striped bass spawning grounds in the Delaware River.

A bagged beach seine (30.5 m x 1.8 m, 10 mm mesh) is used at both fixed and random stations, which are sampled biweekly from August through October. A total of 256 samples are taken, with 50% of the effort concentrated in Region II. Current sampling design is summarized in Table 1.

The survey has undergone a number of changes to sampling design since its inception in 1980. During 1980 and 1981, a 30.5 m x 3.1 m x 6 mm mesh beach seine was used at randomly selected stations sampled at varying frequency from August through November. Two hauls were made at each station, with the second considered a replicate. A total of 30 samples were completed in 1980 and 25 in 1981. In 1982, the current gear was employed and other aspects of sampling design remained the same, with 47 samples collected in 1982, 43 in 1983, and 76 in 1984. For 1985 and 1986, fixed stations were sampled monthly, for a total of 136 samples in 1985 and 107 in 1986. From 1987 through 1990, 16 fixed stations were sampled biweekly from midJuly through midNovember, for a total of 256 samples annually (250 in 1989).

After ten years of the survey were completed, the data were analyzed to determine whether improvements in the sampling design were warranted. Changes which were implemented based on this analysis include narrowing of the sampling season to August through October, use of fixed and random stations, elimination of replicate hauls, and concentration of sampling effort in Region II (Baum 1992).

Relative abundance indices for striped bass are calculated as the mean number of young-of-year captured per seine haul. Estimates are reported for each region and pooled for an annual river index. Annual indices are reported in Table 2.

Chesapeake Bay, Maryland

The Maryland striped bass juvenile survey samples nearshore estuarine finfish communities, although the primary focus is on striped bass (Maryland Department of Natural Resources, 1992). The survey area is divided into four regions, including the Upper Bay, and Choptank, Nanticoke, and Potomac Rivers (Figure 5). The number of stations per region is based on region size, with seven stations in each of the Upper Bay and Potomac River, and four stations in the Nanticoke and Choptank Rivers. Stations are fixed, and sampled once during each monthly round performed during July, August, and September.

A bagless beach seine (30.5 m x 1.2 m, 13 mm stretched mesh) is set by hand, with one end fixed on the beach and the other fully extended, perpendicular to the beach. The seine is swept with the current, and covers approximately 729 m^2. At stations where depth or tide prevents full deployment, the area covered is recorded. Two hauls are taken at each site, with 30 minutes allowed between each haul. The second haul is considered to be a replicate. Current sampling design is summarized in Table 1.

The survey was initiated in 1954, at which time it consisted of sampling rounds in the summer and late fall, and included stations scattered throughout the Chesapeake and its tributaries. Annual indices were calculated using only the summer data. The annual number of samples taken ranged from 34 to 46 until sampling was standardized in 1962. In 1962 a second summer sampling round was added, the fall round was dropped and 22 stations were standardized. The total number of annual samples was 88 through 1966. A third round was added in 1966, raising the total number of annual samples to the current level of 132.

Station locations have been constant wherever possible; however, some have been relocated because of habitat alteration. There have been a total of eight site relocations, with at least one in each system. Auxiliary sites were added in 1974 in response to a downward trend in the juvenile striped bass abundance index. These sites were chosen to bracket permanent sites within a system or fill gaps in geographic coverage. Since site location and number have varied, they are not used to compute indices of abundance. Presently there are 18 auxiliary stations.

Abundance indices are computed as the number of species captured per haul. Striped bass juvenile abundance indices and associated statistics are given in Table 2. Goodyear (1985) validated this index by comparison with commercial landings.

Chesapeake Bay, Virginia

The juvenile striped bass sampling program was initiated by Virginia Institute of Marine Science (VIMS) in 1967, and continued annually until 1973, when it was discontinued due to the termination of federal funding. It was reinstituted with state funds in 1980, and then switched back to federal support for 1981 to the present. The three major Chesapeake Bay tributary rivers of Virginia are sampled, including the James, York system, and Rappahannock (Figure 6). A trawl survey which also collects juvenile striped bass has been extant since 1955; however it is not used to calculate an annual striped bass index.

The current sampling gear for the striped bass juvenile survey is a 30.5 m x 1.2 m, 6 mm bar mesh flat minnow seine, identical to that used by Maryland Department of Natural Resources. Virginia employed this gear in 1986. During 1967 through 1973, a 30.5 m 1.8 m, 6 mm bar mesh bagged seine was used. A 30.5 m x 1.8 m bagged seine set from a boat was used from 1981-1986. The current seining operation is to deploy the net perpendicular to the beach out to the full extension of the net or maximal effective depth for a manual haul, then sweep the offshore end downcurrent and back to shore while keeping the onshore end in a fixed position. Two sets are made at each index station. Sampling design is summarized in Table 1.

There are 18 index stations which are sampled five times a year on a biweekly basis from mid-July through September (Figure 6). Twenty auxiliary stations provide geographically expanded coverage during years of unusual precipitation or drought when the normal index stations do not yield samples. Stations are fixed and 3.2 to 16.1 km apart, with closest spacing in the primary nursery areas.

All finfish species are identified and counted. Lengths are measured for all species. Catch data are transformed using the natural log prior to analysis. Indices are calculated as an adjusted geometric mean (Colvocoresses 1984). The geometric mean is adjusted to an arithmetic scale by multiplying the geometric:arithmetic mean ratio during the first ten years of sampling (2.28). This scaling permits better comparison with the Maryland arithmetic index. To date, the second haul from each station has been treated as a replicate. Only data from index stations are used in comparing relative juvenile abundance. Striped bass abundance indices and associated statistics are given in Table 2.

No validation methodology has been developed, but since the field sampling is identical to that of Maryland Department of Natural Resources, and Maryland's methodology was validated by Goodyear (1985), it is assumed that the application is the same.

The trawl survey uses a 9.5 m semi-balloon trawl, with 19 mm bar mesh in the body and a 6 mm mesh cod-end liner, and samples the tributary rivers at mid-channel and the bay. Tow speed is 2.5 knots and tow time is five minutes. River stations are occupied from river mouth at approximately 8.0 km intervals up river to the freshwater interface. Seven stations are sampled in the York system, six in the James, and seven in the Rappahannock. Additional stations are added upstream during the winter in order to more adequately sample anadromous species. Bay stations are random stratified, with stratification by depth and region. During the winter (December-April), 33 stations are sampled per month, while 39 stations are sampled during the summer (May-November). Sampling occurs monthly, with the exception of January-February and March-April. During each of these two month segments, sampling occurs once. Generally, the trawl survey does not adequately sample the young-of-the-year striped bass, however they are taken during the colder months (December-February).

An index combining the Maryland and Virginia striped bass juvenile abundance indices into a Chesapeake Bay-wide index was prepared by VIMS (Austin et al. 1992) Each index incorporated into the combined index is geometric, scaled, and weighted by the area of the spawning ground in order to maintain the twenty year long term average. Validation of the index using catches of three year old striped bass in gill nets is being attempted. The index was submitted to the Chesapeake Bay Stock Assessment Committee for use in management the Summer 1993.

Potomac and Anacostia Rivers, District of Columbia

In 1986 a survey of District of Columbia waters was initiated to determine relative abundance of juvenile striped bass (Tilak and Palmer 1992). Four sites on the Potomac River and two sites on the Anacostia River are sampled annually (Figure 7). Gear was chosen based on the neighboring programs of Maryland and Virginia, and includes a bagless beach seine (30.5 m x 1.2 m) of 6 mm mesh. Each site is sampled once per month from May through August, with 2 hauls per site. The second haul is considered a replicate. A summary of current sampling design is included in Table 1. There have not been any substantial changes to this program since its inception.

The annual index is calculated using the mean number of fish captured for the largest of the two replicates per station. Annual abundance indices are given in Table 2.

Albemarle Sound, North Carolina

The North Carolina Division of Marine Fisheries (NCDMF) began conducting the juvenile striped bass survey in 1982. Previously (1955-1987) the survey was conducted by North Carolina State University (NCSU). Phalen (1988) found statistically significant differences during two of the six years that the surveys overlapped; however, Monroe (1991) suggested that since the overlap years were during a period of low indices, it was not possible to compare them.

A sampling strategy identical to that of NCSU is used by NCDMF. The trawl is a semi-balloon trawl, with a 5.5 m headrope, and 13 mm stretched mesh cod end. Tows are made for 15 minutes at 2.75 knots. Depth of tow ranges from 1.8 m to 3.0 m. For the NCDMF survey, samples are taken every two weeks from July through October, for a total of 56 samples per year. The NCDMF survey samples 12 fixed stations (Figure 8) while the NCSU survey sampled 7 fixed stations (Figure 9). A summary of the NCDMF sampling strategy is given in Table 1.

Indices are computed as the arithmetic mean of catch per unit effort at all stations over the sampling period. Annual indices for the NCDMF survey and associated statistics are given in Table 2.

Development and Application of Juvenile Abundance Indices: Central Issues

Background

The primary objective of juvenile finfish monitoring is to obtain a valid measure of relative abundance. Validity of the juvenile index relates to the usual statistical properties of central tendency and dispersion. Whether the abundance measures are based on sampling theory (e.g., stratified random design) or models (e.g., delta-distribution), it is also desirable that the estimate be unbiased and precise. The true accuracy and precision of juvenile abundance indices can never be known but certain steps can be taken to reduce bias and variability. In the following sections these steps are described and the summary recommendations from the Workshop are listed.

Before discussing the central issues, a general review of the biological and environmental factors affecting juvenile sampling is worthwhile. In a striped bass population with a broad age distribution of spawners, reproduction typically occurs over a several month period. The survival of cohorts produced over this extended period is highly variable. Catastrophic events such as sharp temperature drops (Dey 1981, Boreman 1983) or low Ph (Hall 1987) can eliminate entire cohorts. More subtle changes in environmental conditions, however, can produce equally variable survival rates (Houde 1989). In any given year, the juvenile year class can consist of multiple cohorts with different birth dates. The timing of these "windows" for successful reproduction varies with river system and across years. The windows are unpredictable but recent improvements in otolith analysis (Secor et al. 1989, Kline 1990) have allowed for retrospective examination of early life history dynamics.

Most surveys sample juvenile finfish during the limited period of time in which the fish congregate in regions where sampling gear can be effectively deployed. The residency period for juveniles in such areas is controlled by ontological changes in temperature preferences, predator avoidance, and availability of forage. As fish mature, their habitat requirements change and movement out of the areas that can be sampled is likely. Thus, sampling of juveniles reflects a snapshot of the dynamic balance of recruitment of cohorts to the sampling area, natural mortality, and emigration due to changing habitat preferences.

Superimposed on the dynamics of recruitment to the sampling area is the usual variation induced by schooling and gear selectivity. Both of these factors are influenced by prevailing environmental conditions. To compensate for these known sources of variation, most agencies have employed sampling designs that might be called "bet-hedging" strategies. The sampling period is extended to account for recruitment of all cohorts to the sampling area, and the sampling area is big enough to encompass even marginally suitable habitat. The "average" estimate of abundance derived under these conditions integrates all factors influencing abundance in the sampling area. Inasmuch as these factor vary from year to year, the high variability and potential bias of juvenile abundance indices is not surprising. The magnitude of this variability is important for drawing proper inferences about population status, and ultimately, for fishery management decisions.

Design

Standard methods for design of sampling surveys have been well described (e.g., Cochran 1977). The basis of sampling theory is founded on the concept of a random sample. If one wants to draw inferences about some attribute of a population of size N, then each element of the population has a finite probability (= 1/N) of being sampled. In a juvenile survey the population consists of all sampling units (i.e. area swept by the gear) in the nursery areas of a given water body. The attribute of interest is the density of juveniles present. When sampling units or stations are fixed, the probability of the fixed station being included in the sample is 1. For all other sampling units, the probability of inclusion is 0. Thus the scope of inference is restricted to the set of sites sampled (See p. 46, Appendix 1, for further discussion).

In a strict sense, the use of fixed stations precludes the computation of variance because the observations only include variation attributable to measurement error; the conditions necessary for estimation of sampling error are not met. In a random sampling design, observations include both measurement error and sampling error. Thus estimates of variance from fixed sampling design will, on the average, be lower than estimates from a comparable random sampling design. Fixed stations can provide excellent estimates of relative change in abundance as several analyses of striped bass juvenile indices have shown (Goodyear 1985, McKown 1992b). The validity of juvenile indices is addressed in a later section of this report.

The relationship between fixed and random samples in surveys is an active area of investigation in statistics, particularly in aquatic systems (Steinarsson and Stefansson 1986, Nicholson et al. 1991). More recently, Warren (1994) investigated the relative ability of fixed and random stations to detect trends in relative abundance. A hybrid design that includes both fixed and random stations is known as sampling with partial replacement (SPR). Appendix 1 provides an excellent introduction to SPR designs. Warren (1994) derived conditions under which fixed station designs will be more accurate than SPR designs. He also formulated conditions under which SPR designs will be more accurate than designs based on random station selection.

Unlike sampling designs applied to human populations, the population and sampling frame for fish populations are unknown. The sampling design must not only provide an estimate of relative abundance, but also determine when and where the fish are present. Moreover, the relative vulnerability of fish to capture changes with fish size, and weather and habitat conditions. As these unknown factors place additional burdens on the design, most sampling programs for juvenile fish populations might be called "bet-hedging" designs. Appendix 2 provides additional insights into the problems of statistical design and guidance on survey design.

Workshop Recommendations: Design

1. Replicate seine hauls are not independent samples, and should be discontinued. Historical estimates should be recomputed using catches from first replicate haul.

2. The inferential aspects of variance estimates computed from fixed sample designs are unknown. A survey based on the concept of partial replacement--a mix of fixed and random sites, may be advantageous. Effort now devoted to auxiliary sites (or other effort that can be redirected) should be devoted toward sampling of randomly selected beaches, so that studies can be conducted to determine the most appropriate mix of fixed and random beaches.

3. The inconsistent response in catch among systems makes development of an appropriate system weighting mechanism critical to proper interpretation of the data. The approach of developing weights based on the extent of habitat in each systems should be continued. Those weights, when developed, should be tested using a validation technique, such as that used by Goodyear.

4. As an alternative approach, empirically-derived weights for individual stations or groups of stations (e.g., river systems) might be developed by applying the Goodyear validation approach. The weights would be proportional to the R-squared of the multiple linear regression. Such weights should be corroborated with additional evidence before being accepted as meaningful.

5. A list of seinable beaches, and habitat in general, should be developed for each system to monitor habitat change and to provide an expanded sampling frame for random samples.

6. Efforts should be made to identify shorter time windows for sampling and reallocating effort toward greater spatial coverage through selection of random sites.

7. Catch data should be analyzed in context of a sampling design. Two designs have been proposed, but both are flawed. The stations can be treated as fixed and subjected to a paired analysis, but this method assumes that stations stay fixed in space and in quality, which they have not. An alternative is to treat them as random sites within a two-stage stratified design. Implementing recommendation 2 under this section will develop the data that would allow testing of the representativeness of the presently sampled sites.

8. Revisions to long term programs must allow comparability with historical data.

Estimation and Analysis

Estimation of juvenile striped bass abundance is complicated by biological processes regulating the survival and growth of multiple cohorts produced over an extended spawning season. Movements of juvenile fish into and out of the sampling region, variation in weather and tide condition, and gear selectivity collectively result in high sampling variability. Such variability tends to increase with abundance, a phenomenon known to fisheries science since at least the early 1950's from the work of Taylor (1953). Log transformation of the variable is the usual method recommended for "stabilizing" the variance (e.g., Sokal and Rohlf 1969, Zar 1974, Neter et al. 1990). Development of efficient estimators however, continues to be a source of controversy in the fisheries literature. Smith (1990) showed the sample mean and variance estimates to be more robust than lognormal-based estimators, particularly when the assumed lognormal distribution is incorrect. Myers and Pepin (1990) examined a number of data sets for their underlying statistical properties and concluded that the efficiency of the delta distribution estimators relies heavily upon the validity of the assumed lognormal distribution for the non-zero values.

From a theoretical standpoint, the most appropriate statistical distribution for modeling numbers per tow should be some form of discrete distribution such as the negative binomial (Taylor 1953, Bannerot and Austin 1983), gamma (Myers and Pepin 1990) or binomial (Sampson 1988) distribution. In large measure, the discrete nature of numbers per tow has been approximated with continuous distributions, most notably the lognormal distribution. For example, most models of population dynamics models assume a lognormal error structure (e.g., Deriso et al. 1985, Collie and Sissenwine 1983, Fournier and Archibald 1982). The delta distribution is simply a mixture of proportion of zeros and a lognormal distribution for the nonzero values (Aitchison 1955). Pennington (1983) first introduced the delta distribution to fisheries and his work has been widely applied in fisheries surveys.

A full treatment of the relevant theory is beyond the scope of this report, but several practical considerations are of interest. When the frequency of zero observations is low the delta distribution is often treated as a lognormal distribution by adding a small constant c (=1 since $\log_e(1)=0$) to the numbers per tow. If the random variable Y is distributed lognormally, $\log_e(Y)$ is distributed normally with parameters μ, σ^2. A simple geometric mean if often used to approximate the expected value of Y and is computed as either

$$\overline{Y_g} = \exp\left(\frac{\sum_{i=1}^{n} \log_e(Y_i + c)}{n}\right) \qquad (1)$$

or

$$\overline{Y_g} = \exp\left(\frac{\sum_{i=1}^{n} \log_e(Y_i + c)}{n}\right) - c \tag{2}$$

Neither Eq. 1 or 2 properly estimates the expected value of a lognormally distributed variable which is

$$E[Y] = \exp^{\left(\mu + \frac{\sigma^2}{2}\right)} \tag{3}$$

Substitution of the sample estimates for μ and σ^2 gives

$$\overline{Y_g} = \exp^{\left(\overline{Y} + \frac{s^2}{2}\right)} \tag{4}$$

where

$$\overline{Y} = \frac{\sum_{i=1}^{n} \log_e(Y_i + c)}{n} \tag{5}$$

Bradu and Mundlak (1970) demonstrated that Eq. 4 is an unbiased estimator of the mean when Y is lognormally distributed; Eq 4 is sometimes termed the bias-corrected geometric mean. The treatment of the constant c in the backtransformation (Eq. 4) does not appear to have been addressed in the fisheries literature.

The simple geometric mean (Eq. 1) and bias corrected geometric mean (Eq. 4) were computed for the New York, Maryland, Virginia and North Carolina data sets (Table 2). Simple geometric means are always less than equal to the arithmetic mean and the magnitude of the difference is related to the degree of skewness in the original observations. By the same token, increased skewness in the original data increases the variance such that the bias corrected geometric means often exceed the arithmetic means. Estimates of variance incorporated in the bias corrected geometric means in Table 2 do not incorporate survey design stratification and are therefore probably too high. More refined estimates of the mean square error of the survey should be incorporated into the estimates of bias corrected geometric means. Note that the bias corrected geometric means are similar in magnitude to the arithmetic mean. Adhoc re-scaling procedures designed to achieve comparability among the arithmetic and geometric series would not be necessary.

Comparisons across programs of the relative variability of annual abundance estimates suggest that most indices have comparable properties. The coefficient of variation of the arithmetic indices ranges between 70 and 92% for states which have had programs in place for more than 12 years (i.e., NY, MD, VA, Table 3). The CV of the annual standard deviation ranges from 66 to 91% and the coefficients of variability of the within year variation (i.e., SD_y/AVE_y) range over a narrow interval. The ratio of the maximum value to the mean is a measure of the skewness and influence of large catches on the overall mean. The relative variability of this measure exhibited surprisingly good agreement across state programs (Table 3). Relative interannual variability of simple geometric means was less than the arithmetic means but the variability of bias-adjusted geometric means was comparable to the arithmetic means. Overall, the results in Table 3 suggest that all existing striped bass juvenile abundance indices have similar patterns of variation.

The log transformation is a special case of general class of transformations known as the Box-Cox transformation (see Sakia 1992 for a recent review). The Box-Cox technique results in a transformation that best approximates a normal distribution. Most applications of the Box-Cox transformation have been economics and agriculture (Sakia 1992); actual applications in fisheries assessment have not been extensively reported. Additional comments on the Box-Cox technique and an example are summarized in Appendix 1.

Appendices 1 and 2 contain additional information on estimation and analyses. Bayesian methods rest on the assumption that the parameter to be estimated is a random variable from a known distribution whose properties can be specified in advance (i.e., the prior distribution). Pages 76-82 (Appendix 2) give an introduction to the theory and a worked example. Kalman filters (See pp 57-58, Appendix 1) are a particular type of empirical Bayes analysis that have been useful in fisheries assessment. The concept of persistence is addressed rigorously by Warren (1994). Finally, methods for combining the results of several studies have been called "meta analysis". A brief introduction to these approaches is on pp 49-50, Appendix 1.

<u>Workshop Recommendations: Estimation and Analyses</u>

1. Fixed sites should be analyzed with respect to their "persistence". Persistence is a measure of the consistency of the rank of catches at a single site over time. For example, if a particular station has the third highest catch in one year, then it should have a very high rank (around the third highest) in most other years. If there is persistence of rank order over years, then the fixed station design can provide a sensitive indicator of trends. Note that the longer the period of time over which persistence occurs, the more sensitive the index.

2. Interannual changes in fixed station indices can be used to detect qualitative changes in abundance. In contrast to the problematic nature of fixed station designs, variances of station differences may satisfy conditions necessary of statistical inference.

3. Because fish move, even fixed stations contain a random component, the magnitude of which is presently unknown. Analyses of dispersal of tagged fish might quantify the degree of randomness present in fixed stations. This random component would be small in a highly persistent system.

4. Nonparametric and graphical methods should be used to explore the qualitative aspects of data. An index based on ranks should be developed to supplement information obtained from catches per haul.

5. Annual estimates should not be interpreted in isolation from what has gone before. Empirical Bayesian methods, including the Kalman filter (when covariates are available), provide a mechanism by which such prior data may be incorporated. In general empirical Bayesian methods yield estimates with smaller mean-squared error.

6. Adjustment of catches with ancillary information (e.g., area swept by gear) is not likely to be useful. Thus, efforts should be made to use "standard" hauls as much as possible. Relating individual catches (seine hauls or trawls) to environmental variables will be difficult.

7. Transformations of data to achieve normality are possible but data should be stratified appropriately. Recent advances in delta distribution models (i.e., partitioning of zero catches into "expected" and "true" categories) might be useful.

Validation

Validation of juvenile indices is not a rigorously defined concept. In the most general sense, validation reflects general acceptance of an index by the scientific and management community. The basis for acceptance rests on one or more statistical procedures, but probability levels for Type I and II statistical errors are not specified. Validation techniques can be classified into two types--indirect and direct.

Indirect techniques depend on correlation between the juvenile index and some other estimate of population abundance or yield. Goodyear (1985) first established the strong statistical association of the Maryland Juvenile index and subsequent commercial landings in Maryland's portion of Chesapeake Bay. The nature of the fishery at that time (small minimum size limit and high rate of fishing mortality) provided ideal conditions for demonstrating a correlation between recruitment and yield. The lag between the survey estimate and entry to the fishery was less than two years and a number of strong year classes, especially 1964 and 1970, provided a strong "signal" that provided significant contrast between average and exceptional year classes.

The highly regulated nature of contemporary fisheries (e.g., quotas, high size limits) diminishes the utility of Goodyear's technique for other state indices. Additional information on fishing mortality rates, growth rates and migration can be incorporated into a projection model to link juvenile indices with yields. This methodology was used in the formulation of the Amendment 4 of the ASMFC Management Plan for Striped Bass and for implementation of the Harvest Control Model (Rugolo and Jones 1989). A similar analysis has been conducted by the New York Power Authority for striped bass in the Hudson River.

In the absence of a long juvenile index time series and ancillary biological information, validation will usually depend on correlations between the juvenile index and another life stage. For example correlations between age 0 fish and age 1 fish in the following year in the same survey demonstrate consistency of the sampling design (Mosca et al. 1994).

Correlations between juvenile indices and later life stage indices from other surveys can also help to validate a juvenile survey. McKown (1992b) correlated Hudson River beach seine survey indices of age 0 abundance with abundance of age 1 striped bass caught by beach seines in western Long Island Sound the following year. Correlations between juvenile indices and age 2 and older age classes should be attempted but variable fishing mortality, migration, and maturation rates can obscure underlying relationships.

Another indirect validation approach might be termed "validation by analogy". For example it might be argued that a survey employing the same methodology and design as a "validated" survey is also valid. In general this approach has not been accepted because of potential habitat differences among surveys.

Direct validation approaches are those in which the juvenile index is shown to be correlated with an absolute estimate of abundance. For example a juvenile abundance index could be compared with the results of a mark-recapture experiment that provides an estimate of true abundance. While this approach is conceptually simple, the logistics of carrying out a synoptic study over several years would be daunting. The use of hatchery fish for abundance estimation has been demonstrated by Dorazio et al. (1991) but specific studies to validate juvenile surveys have not been attempted.

Workshop Recommendations: Validation

1. Correlations of juvenile indices with commercial catches appears to be a viable method of validation. Resampling methods (e.g., bootstrap, cross-validation, jackknife) should be done to verify derived relationships.

2. Correlations with later life stages are the most viable means of validating more recent

time series. Statistical techniques which incorporate the errors in both independent and dependent variables should be used.

3. Utility of marked hatchery fish to estimate population abundance has been demonstrated in the Delaware and Patuxent rivers. More detailed examination of data from these experiments might assist in design of validation studies.

Future Work

Young of the year indices play an important part in the monitoring and management of Atlantic striped bass populations. The Maryland juvenile index played a central role in the development of management measures in the early 1980's that lead to the recovery of striped bass. Today the Maryland juvenile index continues to be an important component for the formulation of quotas and estimation of stock status. As the precision and accuracy of indices from other states improve and are validated with independent data, coast-wide management will rely more heavily on these measures of recruitment. The importance of developing, and agreeing upon, a methodology for combining multiple indices into an overall estimate of juvenile abundance cannot be overstated.

Since the date of this workshop, several states have implemented recommendations. Random stations have been added to survey programs, replicate samples have been dropped and additional sampling sites have been identified. The Striped Bass Technical Committee has re-evaluated the Maryland juvenile index and recommended the use of the geometric mean as the best index of abundance. The Technical Committee also recommended that the index be weighted by the area of the juvenile nursery habitat in each river system. The former recommendation has been endorsed by the Striped Bass Management Board of ASMFC, while the latter is still under consideration.

References and Literature Cited

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

ASMFC. 1989. Supplement to the Striped Bass FMP - Amendment #4. Fisheries Management Report No. 15 of the Atlantic States Marine Fisheries Commission. Washington, DC.

Austin, H.M. and C.R. Hickey. 1977. Predicting abundance of striped bass, Morone saxatilis, in New York waters from modal lengths. Fish. Bull. 76(2): 467-473.

Austin, H., J. Colvocoresses, and T. Mosca, III. 1992. Develop a bay-wide young-of-the-year striped bass index. Final Report to the Chesapeake Bay Stock Assessment Committee, Cooperative Agreement No. NA16F00393-053. Virginia Institute of Marine Science, Gloucester Point, VA. 57 p.

Bannerot, S. P. and C. B. Austin. 1983. Using frequency distributions of catch per unit effort to measure fish stock abundance. Trans. Am. Fish. Soc. 112:608-617.

Baum, T. 1992. Delaware River striped bass young-of-year recruitment survey. Abstract in Abstracts: Emergency Striped Bass Study Annual Workshop 1992, National Marine Fisheries Service, Silver Spring, MD, 20910.

Boreman, J. 1983. Simulation of striped bass egg and larva development based on temperature. Transactions of the American Fisheries Society 114(1): 3-7.

Bradu, D. and Y. Mundlak. 1970. Estimation in lognormal linear models. J. Amer. Stat. Assoc. 65:198-211.

Cochran, W.G. 1977. Sampling Techniques. John Wiley and Sons, New York, NY.

Collie, J. S. and M. P. Sissenwine. 1983. Estimating population size from relative abundance data measured with error. Can. J. Fish. Aquat. Sci. 40:1871-1879.

Colvocoresses, J.A. 1984. Striped bass research, Virginia. Part I: Juvenile striped bass seining program. Annual Report, 1984, Virginia Institute of Marine Science, Gloucester Point, Virginia. 64 p.

Deriso, R. B., T. J. Quinn and P. R. Neale. 1985. Catch-at-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42:815-824.

Dey, W.P. 1981. Mortality and growth of young-of-the-year striped bass in the Hudson

River estuary. Transactions of the American Fisheries Society 110(1): 151-157.

Dorazio, R. M., B.M. Florence, and C. M. Wooley. 1991. Stocking of hatchery-reared striped bass in the Patuxent River, Maryland: survival, relative abundance, and cost-effectiveness. North American Journal of Fisheries Management 11:435-442.

Fournier, D. and C.P. Archibald. 1982. A general theory for analyzing catch at age data. Can. J. Fish. and Aquati. Sci. 39:1195-1207.

Goodyear, C.P. 1985. Relationship between reported commercial landings and abundance of young striped bass in Chesapeake Bay, Maryland. Trans. Amer. Fish. Soc. 114(1): 92-96.

Hall, L.W. 1987. Acidification effects on larval striped bass, Morone saxatilis, in Chesapeake Bay tributaries--a review. Water, Air and Soil Pollution 35: 87-96.

Himchak, P.J. 1992. Delaware River striped bass young-of-year recruitment survey. Report submitted to the Juvenile Abundance Indices Workshop, Grasonville, MD, January 21-23,1992.

Hjort, J. 1914. Fluctuations in the great fisheries of northern Europe, viewed in the light of biological research. Rapp. et Proces-Verbaux, Vol XX. 228 p.

Houde, E.D. 1989. Subtleties and episodes in the early life of fishes. Journal of Fish Biology 35(Supplement A): 29-38.

Kline, L.L. 1990. Population dynamics of young-of-the-year striped bass, Morone saxatilis, populations based on daily otolith increments. Ph.D. Thesis, Virginia Institute of Marine Science, College of William and Mary, Gloucester Point, VA. 253 p.

Maryland Department of Natural Resources (DNR). 1992. Maryland striped bass juvenile survey. Report submitted to the Juvenile Abundance Index Workshop, Grasonville, MD, january 21-23, 1992. 16 p.

McKown, K. 1992a. New York Department of Environmental Conservation Hudson River young-of-the-year striped bass surveys. Report submitted to the Juvenile Abundance Indices Workshop, Grasonville, MD, January 21-23, 1992. 5 p.

McKown, K. 1992b. Validation of the Hudson River young-of-the-year striped bass indices. New York Department of Environmental Conservation, Division of Marine Resources, Stony Brook, NY. 10 p.

Mosca, T.C., H.M. Austin, and D.M. Plotner. 1994. Evidence of a relation between a white perch young of the year index and indices of later life stages *in* The Chesapeake Experience, Toward a Sustainable Coastal Watershed, CRC Publ. 149, pp. 58-62.

Monroe, R.J. 1991. Reconciliation of the striped bass juvenile abundance indices by NCSU and NCDMF for years 1982-1987 *in* NOAA Tech. Mem. NMFS-SEFC-291.

Myers, R. and P. Pepin. 1990. The robustness of lognormal-based estimators of abundance. Biometrics 46:1185-1192.

Neter, J., W. Wasserman, and M. H. Kutner. 1990. Applied linear statistical model. 3rd ed. Irwin Publ., Homewood, IL.

Nicholson, M. D., T.K. Stokes, and A. B. Thompson. 1991. The interaction between fish distribution, survey design, and analysis. ICES CM 1991/D:11, 9 p.

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

Phalen, P.S. 1988. Evaluation of juvenile striped bass surveys in Albemarle Sound, NC. North Carolina Department of Natural Resources and Community Development, Division of Marine Fisheries, Morehead City, NC. 8 p.

Raney, E.C. 1952. The life history of striped bass Roccus saxatilis (Walbaum), Bull Bingham Oceanog. Coll., Vol. XIV, Art. 1. pp. 5-97.

Richards, R.A. 1992. Incorporating precision into a management trigger based on Maryland's juvenile index. Report to the Atlantic States Marine Fisheries Commission, March 18, 1992.

Rugolo, L. J. and P. W. Jones. 1989. A recruitment-based interseason harvest control model for Chesapeake Bay striped bass. Maryland Department of Natural Resources Chesapeake Bay Research and Monitoring Program.

Sakia, R. M. 1992. The Box-Cox transformation technique: a review. The Statistician 41:169-178.

Sampson, D.B. 1988. Fish capture as a stochastic process. J. Cons. Int. Explor. Mer. 45:39-60.

Secor, D.H., J.M. Dean and R.B. Baldevarona. 1989. Comparison of otolith growth and somatic growth in larval and juvenile fishes based on otolith length/fish length relationships. Rapp. P-V. Reun. Cons. Int. Explor. Mer. 191: 431-438.

Smith, S. J. 1990. Use of statistical models for the estimation of abundance from trawl survey data. Can. J. Fish. Aquat. Sci. 47:894-903.

Sokal, R. R. and F. J. Rohlf. 1969. Biometry. Freeman Co., San Francisco.

Squires, T.J. 1992. State of Maine Kennebec River juvenile survey. Report submitted to the Juvenile Abundance Indices Workshop, Grasonville, MD, January 21-23, 1992. 11 p.

Steinarsson, B. and G. Stefansson. 1986. Comparison of random and fixed trawl stations in Icelandic groundfish surveys and some computational considerations. C. M. 1986/D:13. Copenhagen. International Council for the Exploration of the Sea.

Taylor, C. C. 1953. Nature of variability in trawl catches. Fish. Bull. 54:145-166.

Tilak, R. and I. Palmer. 1992. Juvenile striped bass monitoring in District of Columbia during 1991. Report submitted to the Juvenile Abundance Workshop, Grasonville, MD, January 21-23, 1992. 12 p.

Warren, W. G. 1994. The potential of sampling with partial replacement for fisheries surveys. ICES Journal of Marine Science 51:315-324.

Zar, J. H. 1974. Biostatistical analysis. Prentice-Hall, Inc. Englewood Cliffs, NJ

Table 1. State Striped bass and/or alosid juvenile abundance survey sampling strategies for Maine, New York, New Jersey, Maryland, District of Columbia, Virginia, and North Carolina

State	Water Body	Duration	Season	Gear	Fishing Method	Stations	Geographic Stratification	Sampling Frequency	Annual N	Comments
ME	Kennebec River and Tributaries	1979 - Present	Mid-July-October	Beach seine: 17m x 1.8m 6mm stretched mesh 220m 2 swept	Set by boat, against current, stationary inshore end, sampled at changing tide.	12 fixed	Upper Kennebec- 4 sites Androscoggin- 3 sites Merrymeeting Bay- 4 sites Eastern- 1 site Cathance- 1 site Abagadasset- 1 site	every 2 weeks monthly	74	
NY	Hudson River	1980-Present	Mid-July-early November	Beach seine: 61m x 3m 5mm stretched mesh bag 6mm stretched mesh wing	Set by boat, all tide stages sampled per season.	33 fixed	none	25 stations every 2 weeks	225	same lunar cycle each year
NY	Hudon River	1991-Present	Mid-July-early November	Trawl: 8.5m headrope 38mm stretched mesh body 32mm mesh cod end 13mm stretched mesh cod end linar	Tow time: 5 minutes speed: 1.3m/sec	26 fixed	none	every 2 weeks	234	same lunar cycle each year

Table 1. Continued

State	Water Body	Duration	Season	Gear	Fishing Method	Stations	Geographic Stratification	Sampling Frequency	Annual N	Comments
NJ	Delaware River	1980-Present	August-October	Beach seine: 30.5 x 1.8m 10mm mesh	Set by hand and fished with current, stationary inshore end.	Variable no., fixed and random	50% effort in Region II	every 2 weeks	256	Survey site selected wrt salinity, tidal stage not considered
MD	Chesapeake Bay and tribtaries	1954-Present	July-September	Beach seine: 30.5m x 1.2m 6mm mesh bagless 729 m2 swept	Set by hand and fished with current, stationary inshore end.	22 fixed	Potomac- 7 sites Upper Bay- 7 sites Choptank- 4 sites Nanticoke- 4 sites	monthly	132	2 hauls are made at each station & used as replicates.
DC	Potomac & Anacostia Rivers	1986-Present	late May-early August	Beach seine: 30.5m x 1.2m 6mm mesh	Set by hand with current stationary inshore end.	6 fixed	Potomac- 4 sites Anacostia- 2 sites	monthly	48	2 hauls are made at each station and used as replicates

Table 1. Continued

State	Water Body	Duration	Season	Gear	Fishing Method	Stations	Geographic Stratification	Sampling Frequency	Annual N	Comments
VA	Chesapeake Bay tributaries (James, York, Pamunkey, Mattaponi & Rappahanock Rivers	1967-1973 1980-Present	mid-July-September	Beach seine: 30.5m x 1.2m 6mm mesh	Set by hand with current, stationary inshore end.	18 fixed index 20 fixed auxiliary	none	every two weeks	180	Upriver stations are closer together. Use of auxiliary in addition to index stations is weather dependent.
VA	Chesapeake Bay and Tributaries	1955-Present	All year Winter: Dec-April Summer: May-Nov.	Trawl: 9.5m headrope 19mm mesh body 6mm mesh cod-end linar	Tow time: 5 minutes Tow speed: 2.5 knots	Random, 53+ per month-winter, 59 per month-summer	stratified by region and depth in bay fixed stations at 5 mi intervals in rivers	monthly	572+	A variable number (+) of river stations are added upriver during winters.
NC	Albemarle Sound	1955-Present	July-October	Trawl: 5.5m headrope 13mm stretched mesh cod end	Tow time: 15 minutes Tow speed: 2.8 knots	7 fixed	none	every 2 weeks	56	

Table 2. Summary of annual values and selected statistics for striped bass juvenile abundance surveys. Geometric means and bias-adjusted geometric means are based on analyses of unstratified data sets.

State	Year	Number of Samples	Arithmetic Mean	Standard Deviation	Coefficient of Variation	Maximum Value	Percent Zero	Geometric Mean	Bias Adjusted Geometric Mean
NY Beach Seine	1979	117	5.00	7.98	1.60	43	33.3	3.16	5.72
	1980	149	24.00	57.80	2.41	547	22.1		
	1981	131	21.55	42.53	1.97	346	6.9	9.87	21.79
	1982	143	30.50	47.98	1.57	285	5.6	15.17	33.04
	1983	148	48.05	110.71	2.30	1178	5.4	17.25	52.51
	1984	146	37.11	89.84	2.42	906	4.1	16.01	35.67
	1985	216	4.63	6.62	1.43	32	32.9	3.20	5.56
	1986	222	8.75	11.30	1.29	57	17.1	5.29	10.20
	1987	225	82.88	184.57	2.23	1432	5.8	26.13	99.45
	1988	220	70.40	85.38	1.21	869	0.9	43.16	78.78
	1989	225	59.55	86.16	1.45	642	1.8	29.43	69.50
	1990	217	58.00	65.71	1.13	473	0.8	35.87	77.00
	1991	215	15.23	22.57	1.48	160	14.9	7.56	17.51
New York Trawl	1981	78	9.78	12.68	1.30	68	12.8	6.22	11.15

Table 2. Continued

State	Year	Number of Samples	Arithmetic Mean	Standard Deviation	Coefficient of Variation	Maximum Value	Percent Zero	Geometric Mean	Bias Adjusted Geometric Mean
(No samples in 1991)									
	1982	119	18.39	27.41	1.49	127	26.1	7.06	22.91
	1983	103	26.41	50.72	1.92	311	24.3	8.53	31.29
	1984	101	9.95	13.78	1.38	65	24.8	5.19	11.74
	1985	187	3.86	8.84	2.29	60	49.7	2.27	3.97
	1986	177	15.97	27.22	1.70	219	24.9	6.55	18.72
	1987	200	53.57	86.97	1.62	733	8.5	20.03	68.36
	1988	199	115.25	176.58	1.53	1220	6	41.64	176.32
	1989	180	35.70	50.44	1.41	419	14.4	14.13	49.84
	1990	179	62.72	118.30	1.89	1155	15.6	18.88	96.78
	1991								
Maryland	1963	88	4.03	7.37	1.83	42	44.3	2.61	4.64
	1964	88	23.49	38.13	1.62	220	10.2	10.04	26.47
	1965	88	7.42	18.38	2.48	109	53.4	2.56	6.09
	1966	132	16.68	30.84	1.85	216	14.4	7.24	17.50
	1967	132	7.85	20.71	2.64	129	31.1	3.28	6.69

Table 2 Continued

State	Year	Number of Samples	Arithmetic Mean	Standard Deviation	Coefficient of Variation	Maximum Value	Percent Zero	Geometric Mean	Bias Adjusted Geometric Mean
	1968	132	7.17	12.56	1.75	83	34.9	3.70	7.87
	1969	132	10.52	23.56	2.24	189	37.9	3.81	10.11
	1970	132	30.43	47.94	1.58	371	5.3	13.48	34.54
	1971	132	11.77	22.01	1.87	121	18.9	5.02	11.65
	1972	132	11.01	27.61	2.51	204	28	4.26	10.27
	1973	132	8.87	20.45	2.31	156	39.4	3.29	8.04
	1974	132	10.12	26.51	2.62	168	34.6	3.61	8.60
	1975	132	6.69	10.18	1.52	49	26.5	3.81	7.48
	1976	132	4.89	13.72	2.81	121	40.9	2.55	4.58
	1977	132	4.85	11.46	2.36	75	37.9	2.60	4.68
	1978	132	8.45	12.29	1.45	81	23.5	4.75	9.80
	1979	132	4.00	7.27	1.82	43	34.1	2.73	4.53
	1980	132	1.96	3.42	1.74	20	46.2	2.01	2.76
	1981	132	1.21	2.77	2.29	20	61.4	1.59	2.02
	1982	132	8.45	13.52	1.60	76	24.2	4.54	9.42
	1983	132	1.36	3.68	2.71	35	62.9	1.60	2.09
	1984	132	4.21	9.60	2.28	62	34.9	2.64	4.37
	1985	132	2.93	7.42	2.53	47	58.3	1.91	3.10

Table 2 Continued

State	Year	Number of Samples	Arithmetic Mean	Standard Deviation	Coefficient of Variation	Maximum Value	Percent Zero	Geometric Mean	Bias Adjusted Geometric Mean
	1986	132	4.14	11.28	2.72	80	4.7	2.34	3.98
	1987	132	4.80	12.59	2.62	98	49.2	2.46	4.60
	1988	132	2.67	8.32	3.12	67	61.4	1.74	2.68
	1989	132	25.17	77.93	3.10	725	26.5	5.78	21.08
	1990	132	2.13	3.75	1.76	20	50.8	2.02	2.91
	1991	132	4.44	9.05	2.04	53	48.5	2.52	4.69
Virginia	1967	53	4.13	6.77	1.64	40	34	3.02	4.95
	1968	66	3.30	6.76	2.05	50	34.9	2.62	3.96
	1969	77	2.84	5.87	2.07	40	41.6	2.28	3.44
	1970	77	6.09	8.79	1.44	43	27.3	3.82	7.10
	1971	80	2.13	3.40	1.60	16	42.5	2.24	3.21
	1972	116	0.89	1.67	1.88	10	59.5	1.52	1.78
	1973	84	1.66	4.09	2.46	30	59.6	1.70	2.32
	1980	89	2.57	5.46	2.12	38	48.3	2.11	3.17
	1981	116	1.42	4.20	2.96	42	53.3	1.69	2.14
	1982	106	3.06	6.62	2.16	38	49.1	2.19	3.50

Table 2 Continued

State	Year	Number of Samples	Arithmetic Mean	Standard Deviation	Coefficient of Variation	Maximum Value	Percent Zero	Geometric Mean	Bias Adjusted Geometric Mean
	1983	102	2.94	5.81	1.98	43	28.4	2.52	3.57
	1984	106	4.38	8.63	1.97	65	32.1	2.91	4.85
	1985	142	2.27	4.19	1.85	23	44.8	2.06	2.98
	1986	144	4.67	8.02	1.72	59	31.9	3.08	5.28
	1987	144	15.22	21.21	1.39	107	9.7	7.91	16.81
	1988	180	7.49	11.71	1.56	78	18.9	4.35	8.21
	1989	180	11.01	22.99	2.09	243	11.1	5.92	11.08
	1990	180	6.93	10.71	1.55	58	21.7	4.22	7.69
	1991	180	3.71	7.88	2.12	76	33.3	2.66	4.18
North Carolina	1982	72	0.74	1.67	2.26	9	68.1	1.39	1.64
	1983	99	0.74	1.62	2.19	8	74.8	1.37	1.64
	1984	131	0.26	1.02	3.92	7	90.1	1.12	1.20
	1985	147	0.17	1.03	6.06	10	95.2	1.06	1.12
	1986	119	0.06	0.35	5.83	3	96.6	1.03	1.05
	1987	178	0.27	1.14	4.22	12	87.6	1.13	1.21
	1988	140	1.93	6.15	3.19	53	72.9	1.55	2.24

Table 2 Continued

State	Year	Number of Samples	Arithmetic Mean	Standard Deviation	Coefficient of Variation	Maximum Value	Percent Zero	Geometric Mean	Bias Adjusted Geometric Mean
	1989	147	3.88	6.80	1.75	32	51.7	2.43	4.46
	1990	139	0.83	2.94	3.54	28	74.8	1.34	1.60
	1991	140	0.76	1.61	2.12	8	71.4	1.40	1.66
Maine	1979-86		0.00						
	1987	74	0.35						
	1988	68	0.04						
	1989	68	0.01						
	1990	68	0.06						
	1991	63	0.25						
New Jersey	1980	30	0.07						
	1981	25	0.00						
	1982	47	0.17						
	1983	43	0.05						
	1984	76	0.04						
	1985	136	0.04						

Table 2 Continued

State	Year	Number of Samples	Arithmetic Mean	Standard Deviation	Coefficient of Variation	Maximum Value	Percent Zero	Geometric Mean	Bias Adjusted Geometric Mean
	1986	107	0.51						
	1987	256	1.11						
	1988	256	0.58						
	1989	250	2.71						
	1990	256	2.06						
	1991	256	1.05						
District of Columbia	1986	48	3.60						
	1987	48	1.70						
	1988	48	4.96						
	1989	48	1.16						
	1990	48	1.40						
	1991	48	1.33						

Table 3. Relative variability of annual summary statistics for striped bass juvenile abundance indices. Table entries represent the coefficients of variation of the annual statistics presented in Table 2.

State	Period	Coefficient of Variation						
		Arithmetic Mean					Geometric Mean Indices	
		Mean	SD	CV	Ratio Max: Mean	% Zeros	Mean	Bias Ad-justed
NY Seine	1979-91	69.5	76.2	26.2	46.6	94.8	71.9	73.4
NY Trawl	1981-90	92.2	90.6	17.4	32.4	57.3	84.8	103.1
Maryland	1963-91	85.3	87.6	21.8	36.2	44.7	67.2	87.3
Virginia	1967-73							
	1980-91	76.1	65.9	19.2	40.9	39.8	50.2	67.7
North Carolina	1982-91	113.7	87.2	41.3	61.0	16.9	27.9	53.6
Maine	1987-91	94.2						
New Jersey	1980-91	121.7						
District of Columbia	1986-91	60.4						

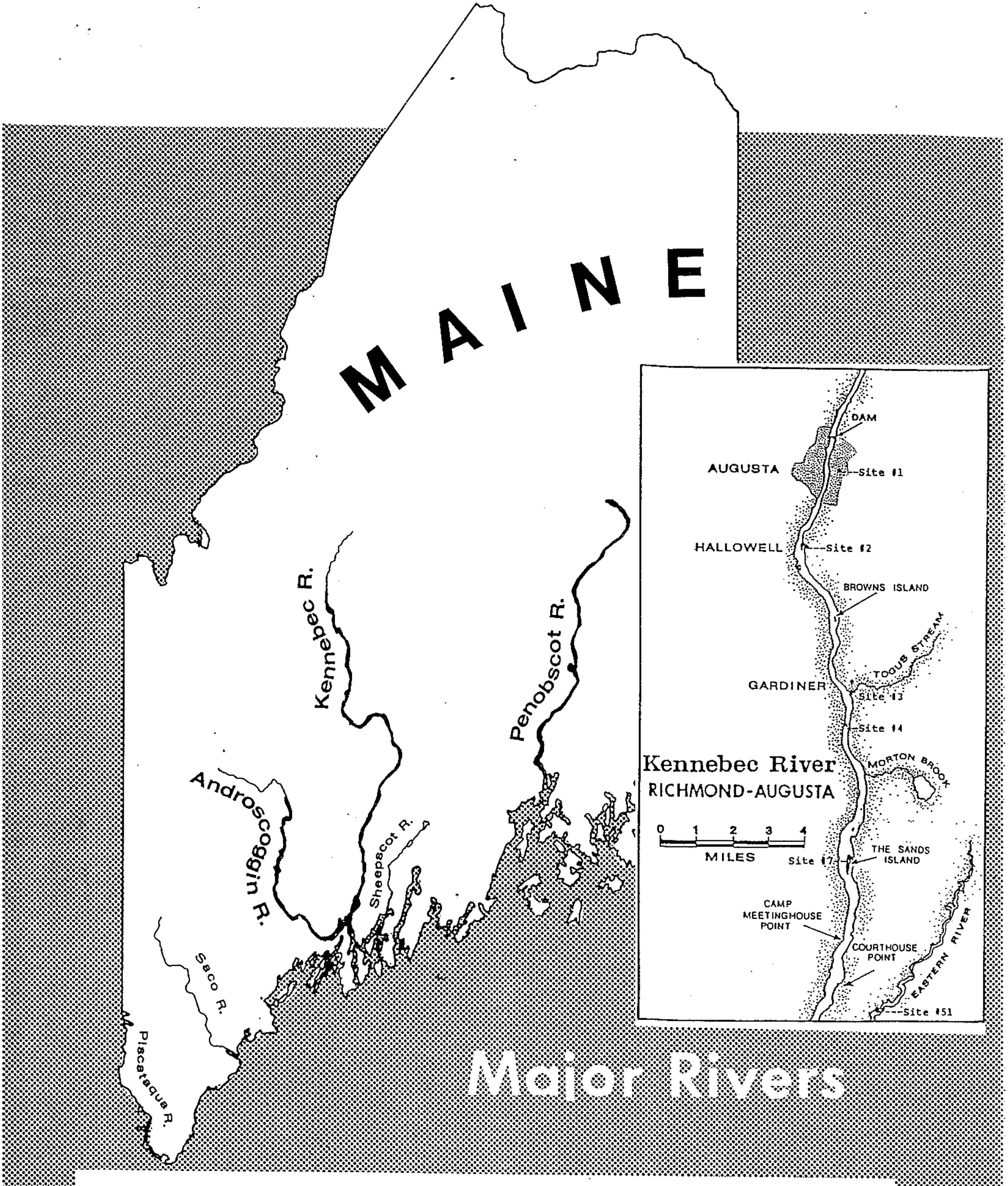


Figure 1. Major rivers in Maine and location of Maine Department of Marine Resources alosid juvenile abundance survey stations in the upper Kennebec and Eastern Rivers.

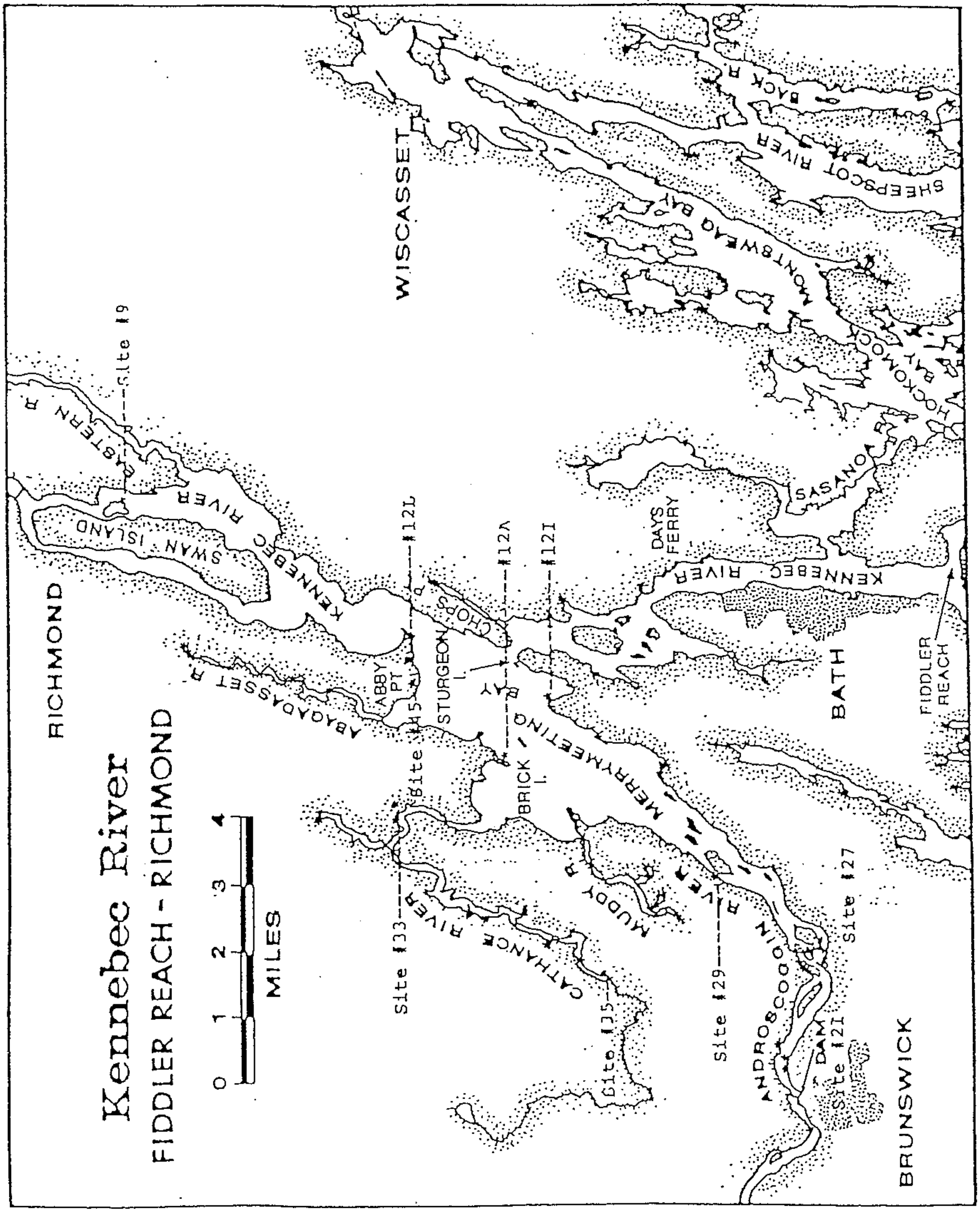
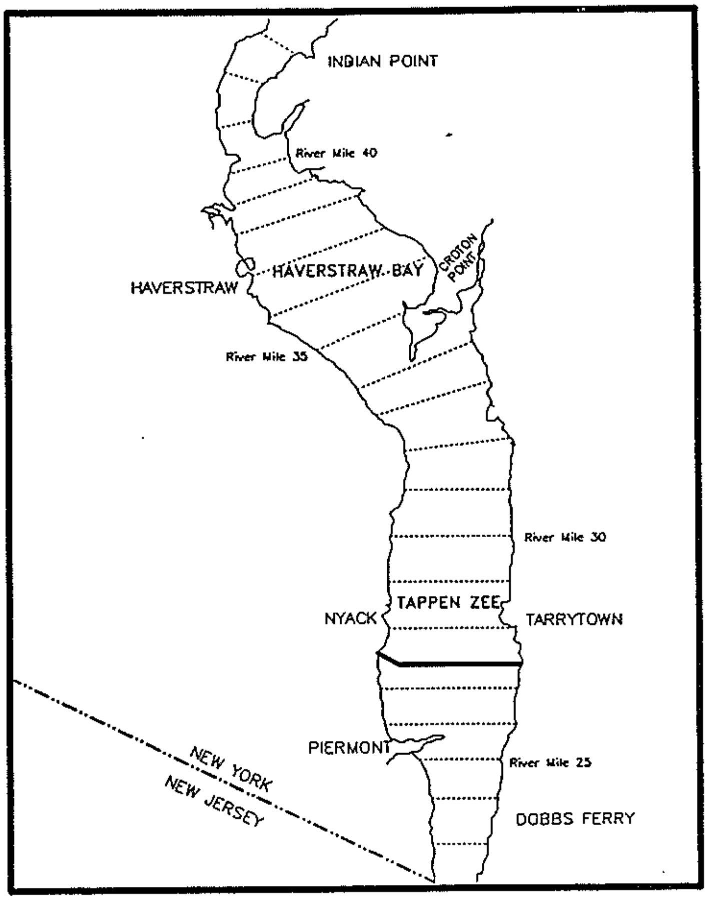
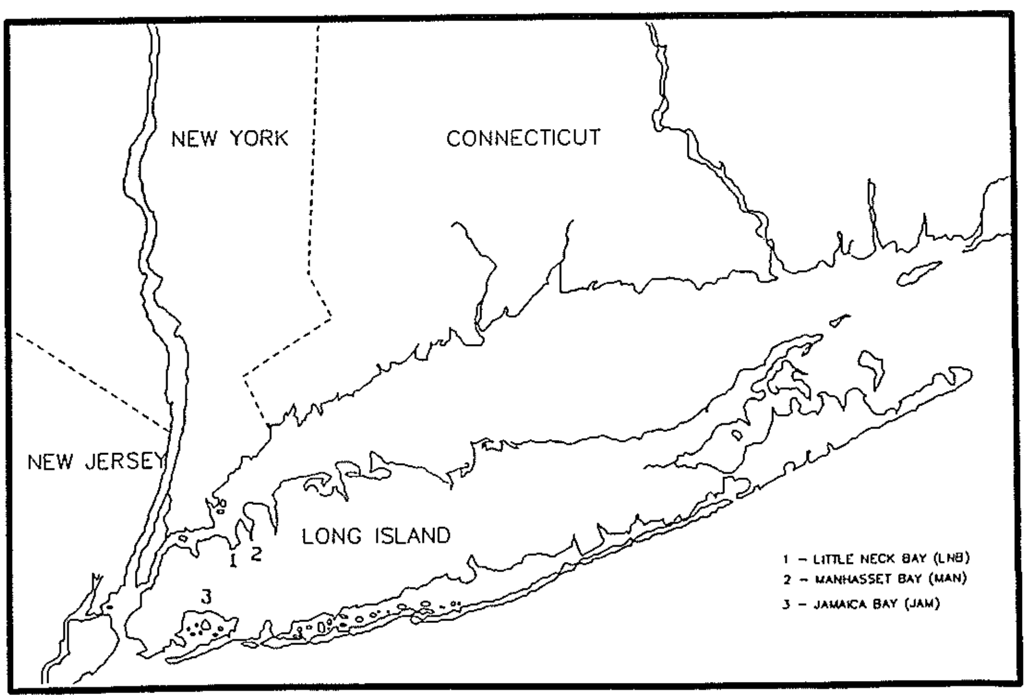


Figure 2. Location of Maine Department of Marine Resources alosid juvenile abundance survey stations in the Kennebec, Androscoggin, Abagadasset and Cathance Rivers and Merrymeeting Bay.



HUDSON RIVER YOY

STRIPED BASS SAMPLING AREA



LONG ISLAND BEACH SEINE SURVEY AREAS

Figure 3. Location of the New York Department of Environmental Conservation Hudson River beach seine and trawl striped bass juvenile abundance surveys.

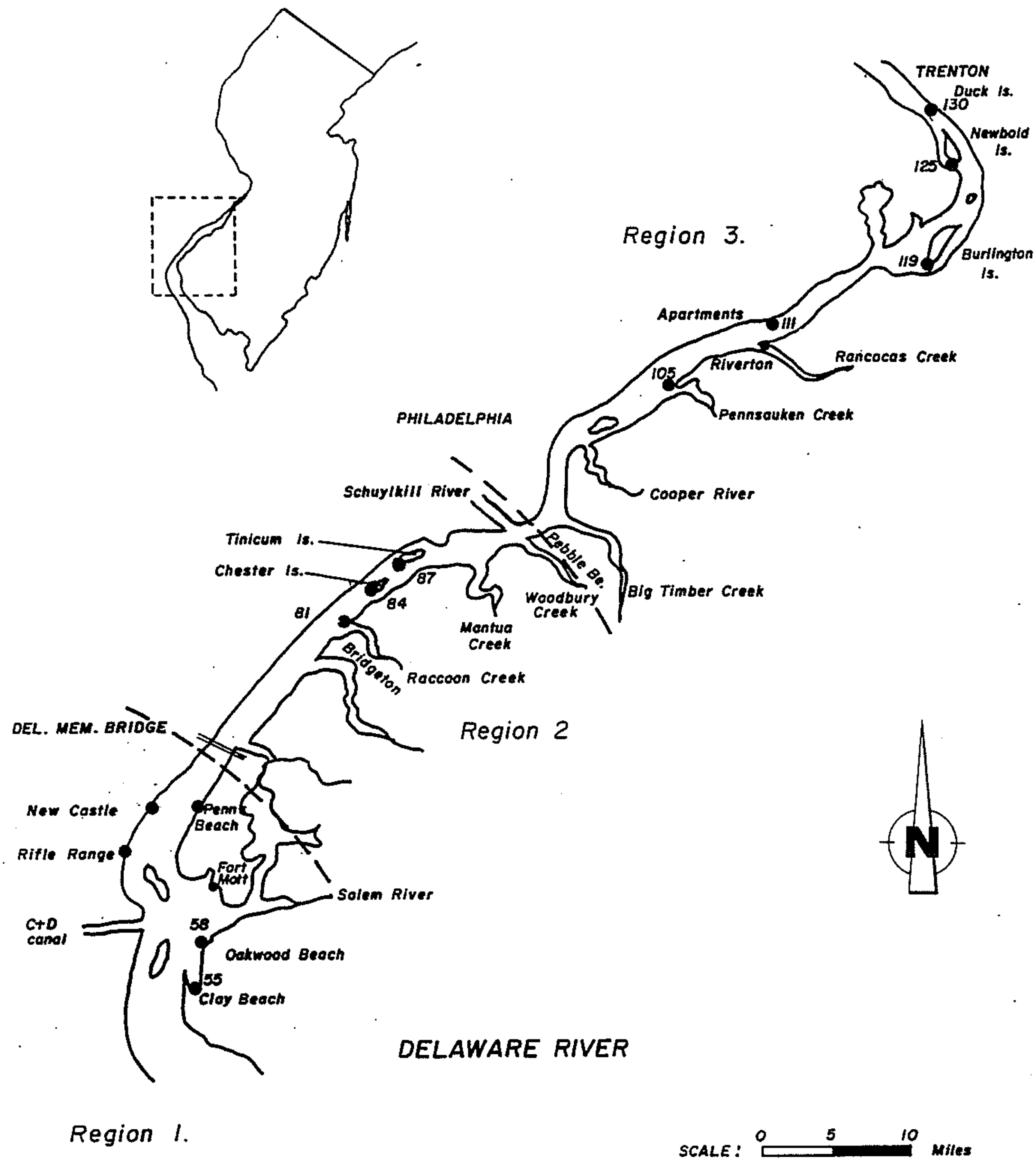


Figure 4. Location of New Jersey Bureau of Marine Fisheries striped bass juvenile abundance survey stations in the Delaware River.

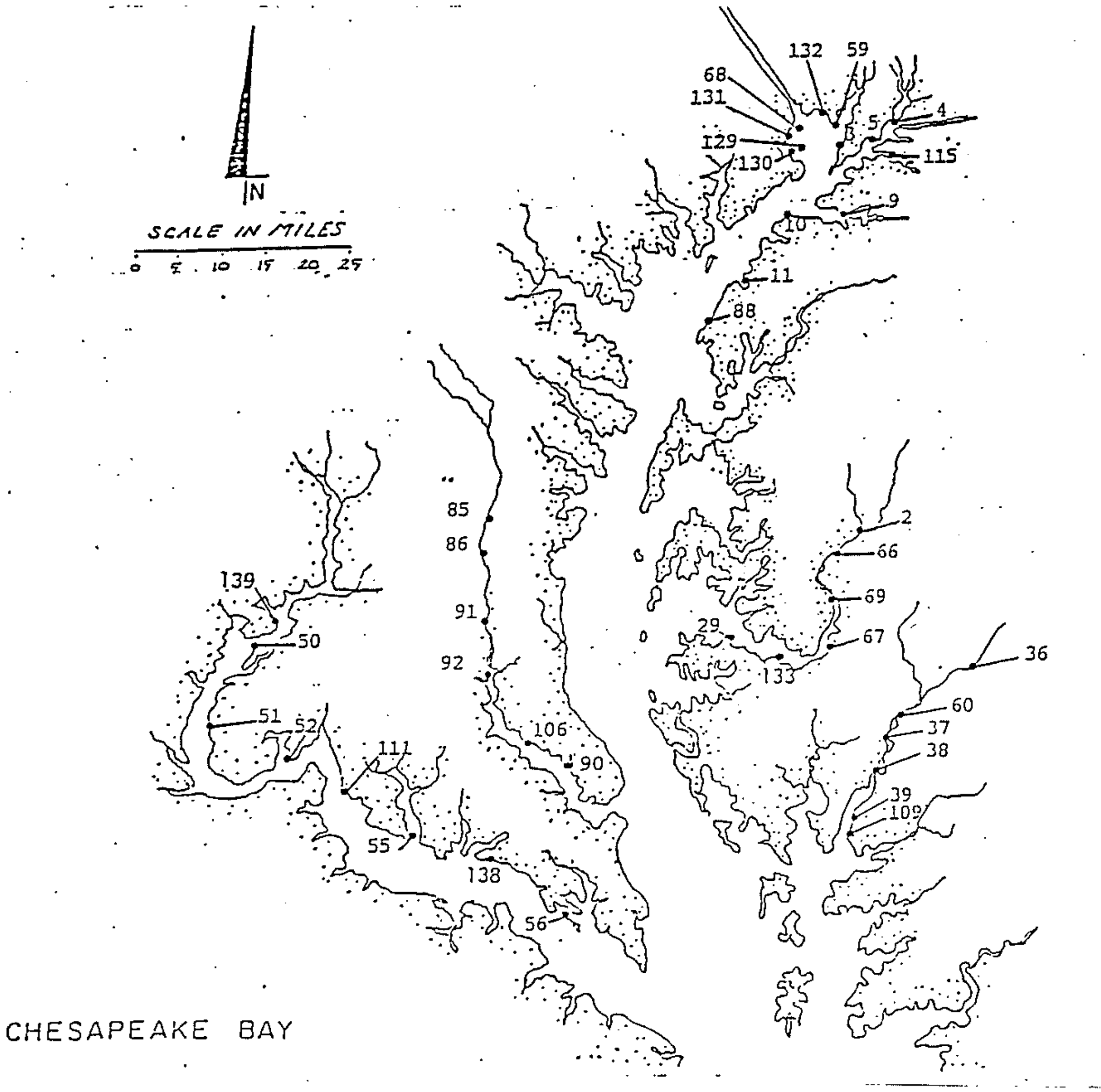


Figure 5. Location of Maryland Department of Natural Resources striped bass juvenile abundance survey stations in Chesapeake Bay and tributaries.

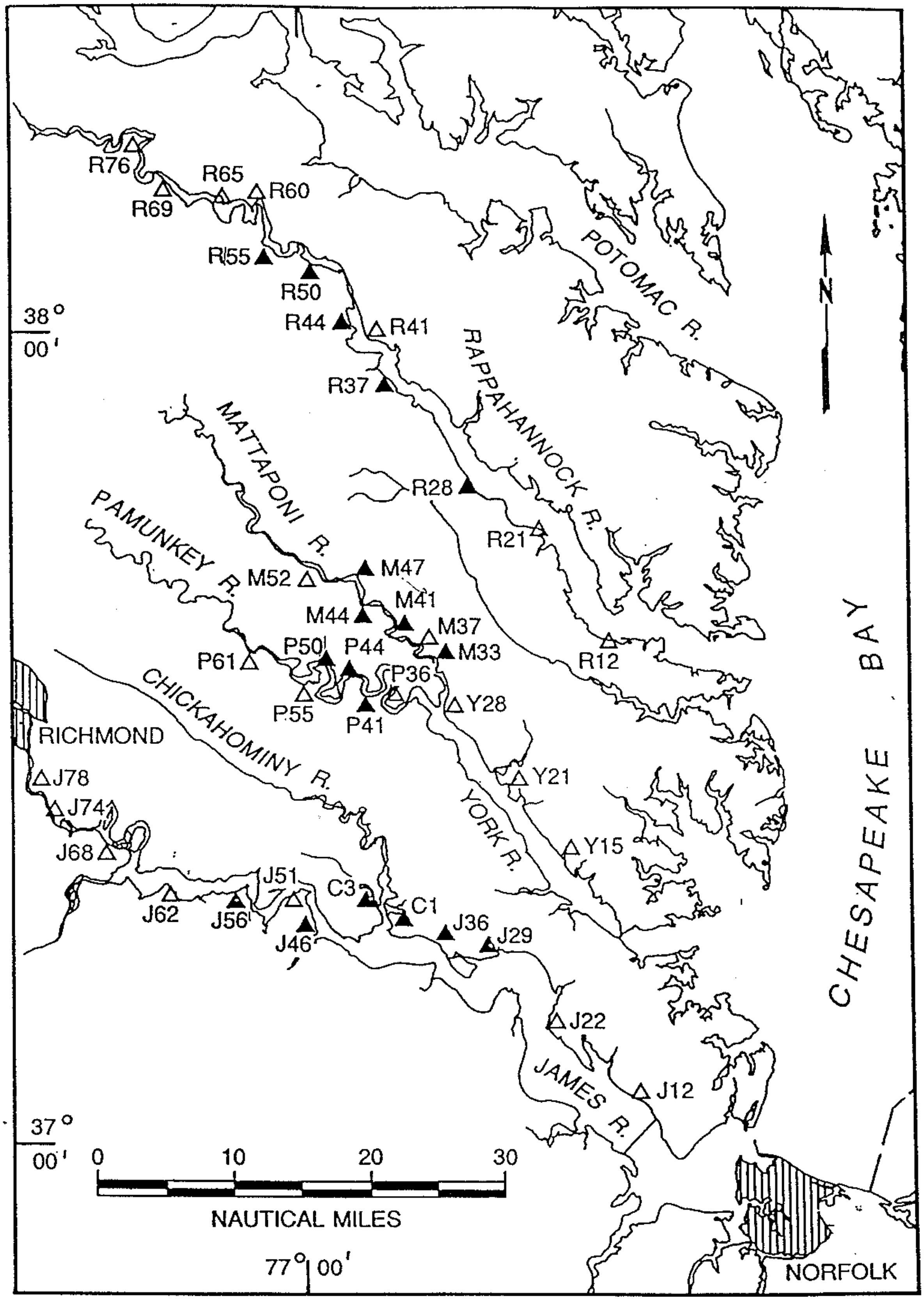


Figure 6. Location of Virginia Institute of Marine Sciences striped bass juvenile abundance survey stations in Virginia tributaries of Chesapeake Bay.

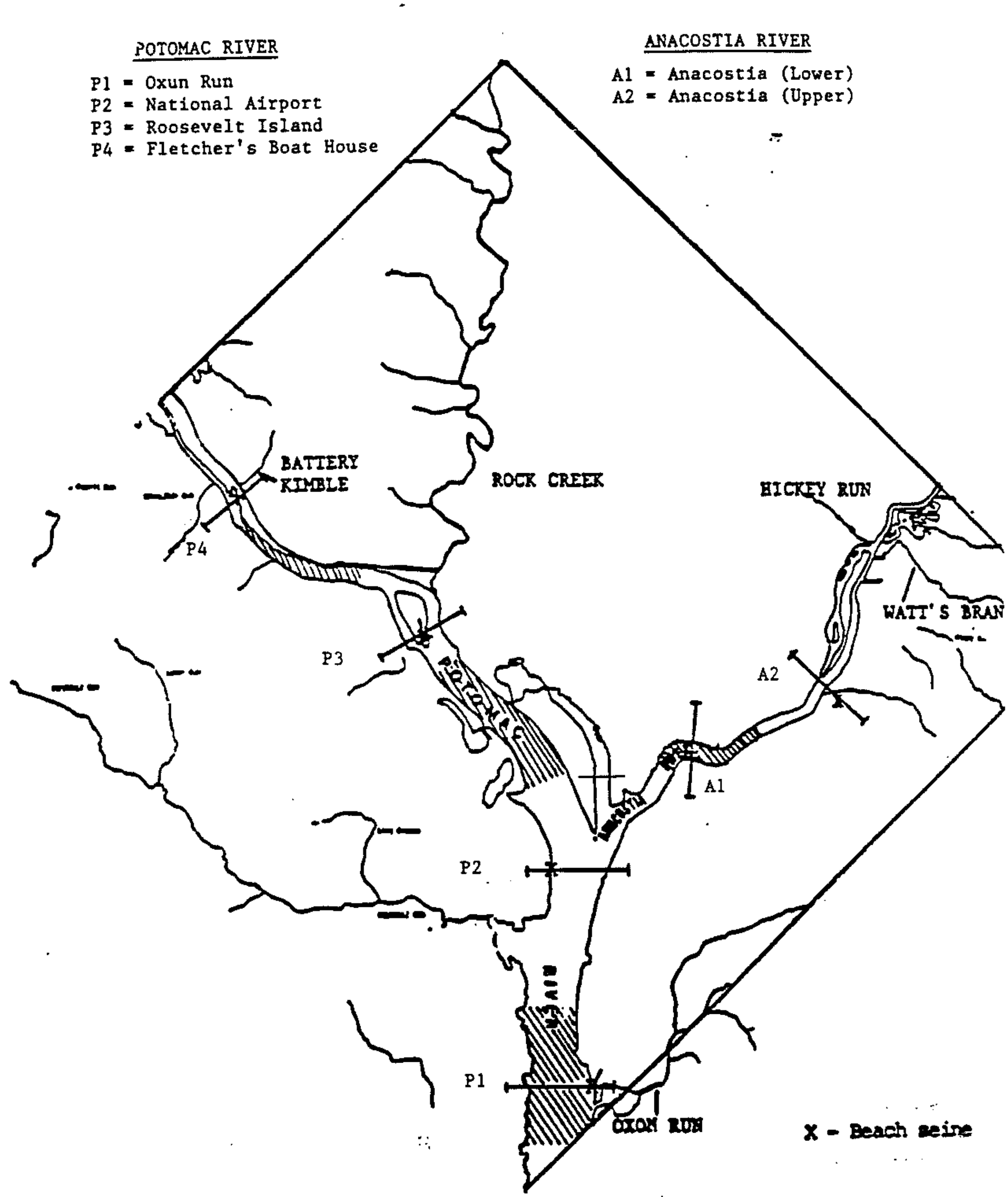


Figure 7. Location of District of Columbia Fisheries Program juvenile abundance survey stations in the Potomac and Anacostia Rivers.

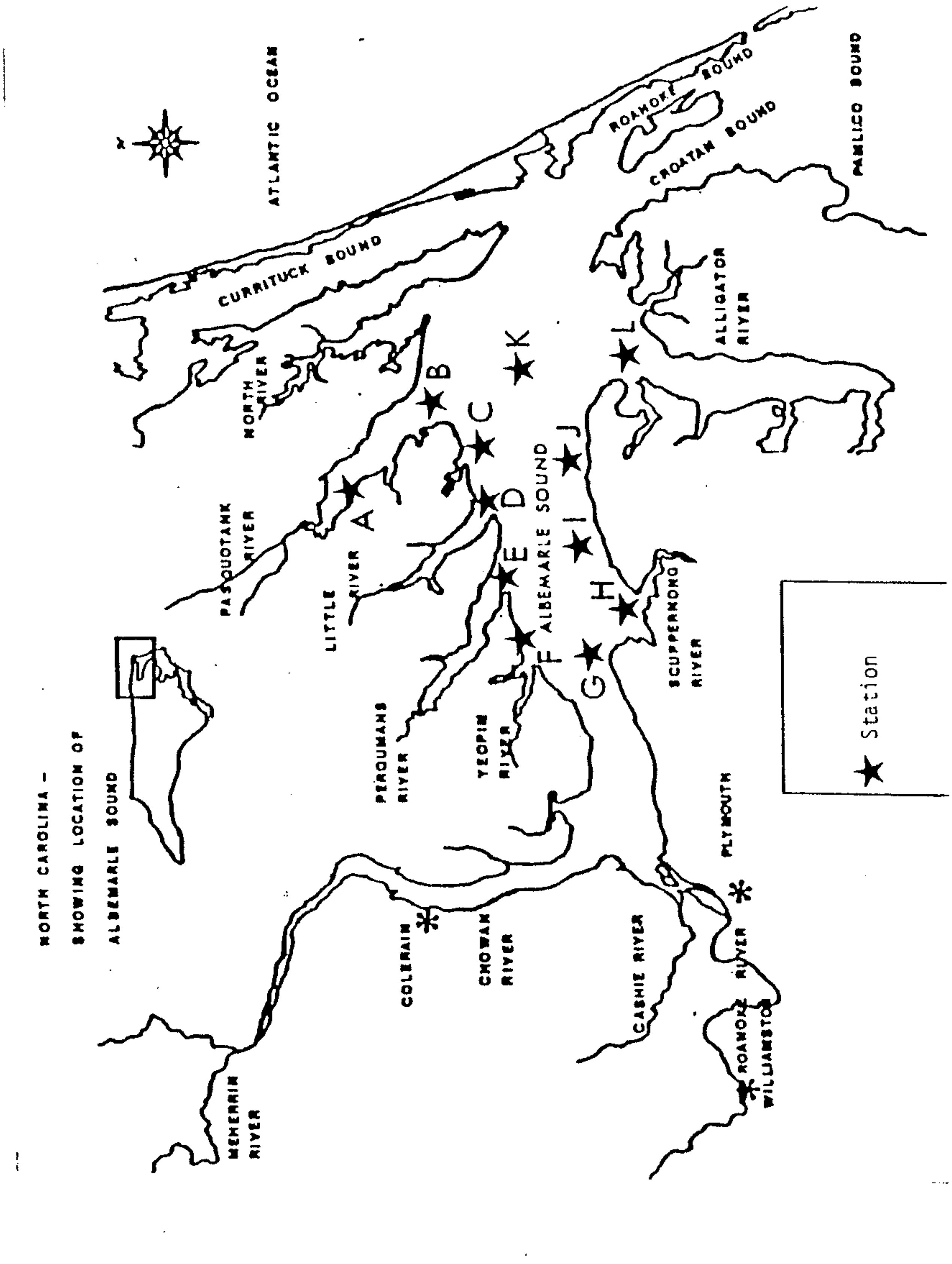


Figure 8. Location of North Carolina Division of Marine Fisheries striped bass juvenile abundance survey stations in Albemarle Sound.

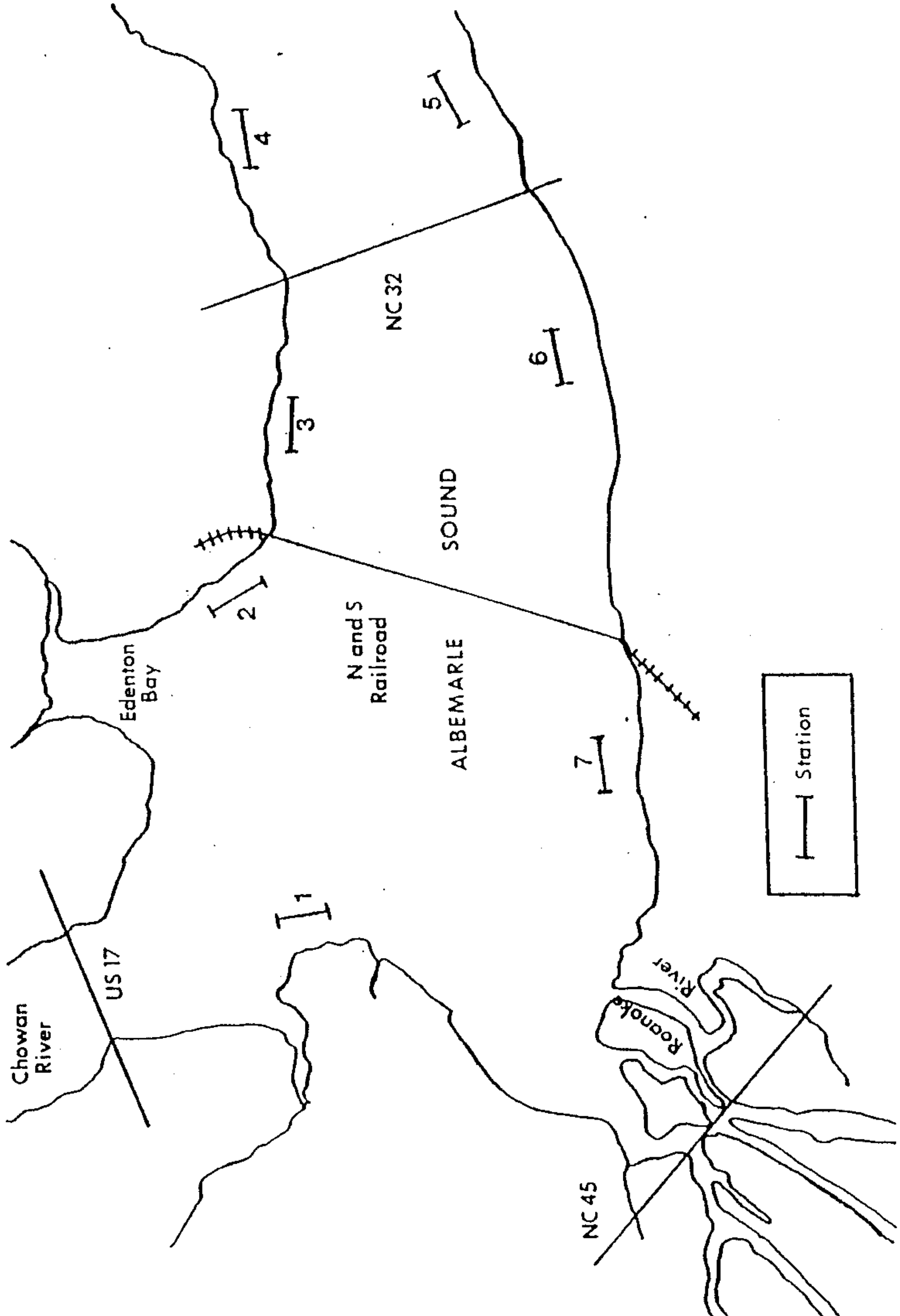


Figure 9. Location of historical North Carolina State University striped bass juvenile abundance survey stations in Albemarle Sound.

Appendix 1

JUVENILE ABUNDANCE INDEX WORKSHOP

Grasonville, MD
January 21-23, 1992

Consultant's Report

William G. Warren, M.Sc., Ph.D.
26 Virginia Place
St. John's, NF
Canada A1A 3G6

JUVENILE ABUNDANCE INDEX WORKSHOP
Consultant's Report

Survey Design.

The reports on the surveys striped bass juveniles presented by the several participants contained one notable feature in common. All rely on samples obtained over a set of fixed stations. The fact that, for various reasons, some stations have at times had to be replaced by others does not alter in any fundamental way the fixed nature of these sampling sites. Estimates of precision (confidence limits) of annual abundance of juveniles have, however, been calculated as if these stations provided a random sample of the population. In this respect, the draft manuscript on sampling with partial replacement, appended to this report, is highly relevant.

Fixed stations and random stations, the latter chosen independently each year, provide the two extreme cases of sampling with partial replacement. In fact, for simplicity, the appended ms. develops the results for these special cases before attempting to treat the general, and therefore more complex, case. Basically, the variance of an estimate based on fixed samples contains only a component stemming from what we shall call "measurement" error. It contains no sampling error *per se*. On the other hand, the variance of an estimate based on random sampling contains both measurement and sampling error components. Fixed stations will, therefore, commonly generate estimates that are, in the strict sense, more precise than those obtained via random sampling of stations.

On the other hand, random stations always provide an unbiased estimator of the population mean, whereas fixed stations yield, in general, a biased estimator, although, depending on the choice of the fixed stations, the magnitude of the bias could be anywhere from negligible to substantial. The problem is that there is no way of determining the extent of the bias from the data themselves.

Accordingly, if the mean squared error (defined as the sum of the bias squared and the variance of the estimate) is used as a measure of accuracy, an estimate based on fixed samples can be more accurate than one based on random samples, but it is also possible for it to be less accurate; without some additional assumptions or information, it is not possible to determine which situation has arisen in any particular case.

It is shown in the Appendix that, depending on circumstances, it is possible for sampling with partial replacement (SPR), i.e. working with a combination of fixed stations coupled with additional stations that are independently and randomly selected each year, to yield a more accurate estimate than either purely fixed or purely random samples.

With respect to an index of abundance, bias in the estimate in any one year may be of little concern if changes in abundance between years is accurately reflected by changes in the index. The same arguments apply to estimates of change; fixed samples will, in general, yield biased but relatively precise estimates, whereas random samples will yield unbiased but less precise estimates with the former sometimes being more accurate than the latter, and vice versa. Again, in terms of accuracy, SPR may be superior to either.

The property that determines the bias, and thus which strategy provides the more accurate estimates, is "persistence". Persistence can be looked on as analagous to interaction (or the lack of it) in a two-way classification, here thought of as stations and years. No interaction, i.e. the station differences maintaining themselves from year to year (or, conversely, inter-year differences being the same from station to station) corresponds to a system that is fully (or completely) persistent. In a fully persistent situation, changes in abundance estimated from fixed stations will be unbiased (and hence, more accurate than changes estimated from random samples). As the interaction increases (or persistence decreases) the scales tip more and more in favor of random samples, with possible advantage to SPR with intermediate persistence. Station-by-station time series of catch data provide a means for determining the degree of persistence in, at least, relative terms.

What then is the role of the between-station variance calculated over fixed stations in any one year? This is certainly the variance of the population of those fixed stations. It is not, however, a valid measure of the variance of the population of interest which comprises the fixed stations along with all other habitable areas of the river or estuary, etc, or as a measure of the precision of sample mean as an estimate of the population mean. (It may not be possible to sample all habitable areas. In this circumstance, one should perhaps define the population in terms of the stations that it would be possible to sample and assume that the proportion of the population in the remaining regions remains constant over time. It is doubtful that this would be the case but it is difficult to see what else could be done. Violation will contribute to the error, or noise, in forecasts of fish populations generated by juveniles in the system as a whole). This comes about from the probability of a station being included in the sample being either 1 (for fixed stations) or 0 (for other than fixed stations) instead of $1/N$ (for all stations) under random sampling, where N is the number of stations that can be sampled.

Fixed samples, thus, do *not* provide a frame for making inferences about the population as a whole (either over years or for any one year). This does not necessarily mean that an index of abundance derived from fixed samples is useless. Depending in the persistence property and the choice of stations, it could provide an excellent measure of, in particular, relative change in abundance. The drawback is that the fixed-station data do not provide a valid estimate of the precision or accuracy of that index.

Reference has been made to measurement error. By this is meant the difference that would arise if more than one sample could be taken at the same station at the same time. Of course, we have no direct measure of this, but replicate sampling of stations, after a suitable time lag, has been a common practice. Apparently, because of the disturbance caused by the first sample, the catch in the second is on the average, and fairly consistently, less than the first. Suppose that, on the average, the ratio of the second catch (y) to the first (x) is p. Then a plot of y against x would yield points scattered about the line $y = px$. The residuals of the points about that line would, seemingly, provide a reasonable estimate of the measurement error.

If, as appears to be the case, the replicate sample is well correlated with the first it is adding relatively little information, and certainly not in proportion to its cost. It appears that one would be getting better value by sampling at additional stations. Further, averaging catches of the first and replicate samples will negatively bias the estimates relative to estimates obtained from a single sample. Since replicate sampling has not always been carried out, unless appropriate adjustment is made, time series would have to be based solely on first sample estimates. Also, variances calculated from individual catches would be inflated by the (systematic) difference between the first and second samples at a station.

It may be argued that, since fish move, fixed stations could be treated as *de facto* random. Fish movement is, clearly, a contributor to any lack of persistence. Persistence is here viewed as a between-survey, i.e. between-year, characteristic. It is fish movement within a survey that might possibly justify the treating of fixed stations as random. It would, in effect require the dispersal of fish to be random throughout the region surveyed. Habitat preference and the distance between fixed stations would seem to work against this being a viable assumption. However, sampling has generally been carried out at several times in a season. One could then consider persistence within a season. A lack of persistence within seasons would not necessarily imply random movement but might relate to a systematic shift of habitat as the season progressed. Similar shifts due to changing habitat quality may well occur over longer time periods. Neither of these consititute random dispersal. Treating fixed stations as random on the basis of fish movement should, therefore, be viewed with caution.

If within-season persistence exists, one could consider reducing the number of sampling occasions and redirect the effort to sampling more stations.

Transformation of Data.

Some indices of striped bass abundance have been based on logarithmically transformed data. The geometric mean is, of course, the backtransformation of the arithmetic mean of data so transformed and is a biased estimate of the population mean. This does not necessarily imply that, as an index, it would be inferior to the arithmetic mean of the untransformed data; it depends very much on how it is used. Certainly, transformation will generally be necessary for the construction of reasonable confidence limits from skewed data (although this does not necessarily imply that the same transformation has to be used to obtain a point estimate). Further, there is nothing sacrosanct about the logarithmic transformation. It is, in fact, a special case of the Box-Cox family of transformations, namely

$$z = (x^{\lambda} - 1)/\lambda, \ \lambda \neq 0$$

$$z = \log(x), \ \lambda = 0$$

where λ is chosen to maximize the likelihood under the assumption that z is normally distributed. This does not necessarily mean that z is normally distributed, but the λ so chosen will result in z being as close to normal as possible for this family of transformations. An example is appended.

The maxim that the author has worked by is "transformations if necessary, but not necessarily transformations". Since with transformation one is working in an artifical scale, interpretaion is sometimes dificult. Accordingly the writer generally avoids transforming data unless he can see some clear benefits.

Temporal Analysis.

The annual abundance indices consititute a time series. Annual estimates should not be interpreted in isolation from what has gone before. Empirical Bayesian methods provide a mechanism for incorporating prior information to improve the accuracy of an estimate based on current information alone. Emprical Bayes procedures come in various stripes. In particular, one may use prior information, i.e. the time series up to, but not including, the current year to predict the index for the current year. Since this estimate is based on the data of several years, it should be relatively precise. However, as an estimate of the current year's index, it it potentially biased. On the other hand, an estimate based on a random sample for the current year will be unbiased but, since based on a smaller amount of data, less precise. There are ways of combining these estimates to produce another estimate that, although possibly biased, has smaller mean squared error than either and, in that sense, is more accurate.

The Kalman filter is a particular type of empirical Bayes procedure. The relationship between the vector of observations, Y_t, at time t and a vector of state parameters, α_t, is assumed to be

$$Y_t = F_t \alpha_t + v_t$$

where F_t is a matrix of known constants and v_t a vector of residuals with zero expectation and variance matrix V_t. To put this in more familiar terms, the above could be an ordinary linear regression equation, i.e.

$$Y_t = \alpha_{1t} + \alpha_{3t} x_t + v_t$$

The difference here is that the state parameters are assumed to evolve over time (hence the subscript t on the α) according to

$$\alpha_t = G_t \alpha_{t-1} + w_t$$

where G_t is a matrix n known constants and w_t a vector of residuals with zero expectation and variance matrix W_t.

The writer has used this formulation, with apparent success, to model the temporal trend of pollutant concentration as a function of fish length. Specifically, Y were the concentrations of mercury in fish muscle and x the fish lengths (treated as known constants). It was assumed that slope of the regression, α_{3t}, fluctuated randomly from year to year, i.e.

$$\alpha_{3t} = \alpha_{3t-1} + \epsilon_{3t}$$

but that there was a temporal trend in the intercept, α_{1t}, modeled by

$$\alpha_{1t} = \alpha_{2t} + \epsilon_{1t}$$

$$\alpha_{2t} = \alpha_{2t-1} + \epsilon_{2t}$$

i.e.

$$G_t = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In this application, it was assumed that the variance matrices, V_t and W_t, were constant over time, i.e. the subscript t could be dropped from them. Note that, in general, it is not necessary to assume that the components of the residual vectors, v_t and w_t, are independent, only that the v_t are independent of the w_t.

Given V_t, W_t and estimates at time zero, one may then estimate successively $\alpha_1, \alpha_2, \ldots \alpha_T$. We denote these estimates by $a_{t/t-1}$ i.e.

$$a_{1/0}, a_{2/1}, a_{3/2}, \ldots a_{T/T-1}$$

where $a_{3/2}$ depends on the data of year 1 as well as year 2, etc. When the data of year t becomes available, $a_{t/t-1}$ may be updated to a_t.

If we wish, we may use the data of all years to revise the estimate of a previous year (referred to as smoothing). We denote these estimates by $a_{t/T}$ where, of course, $a_{T/T} = a_T$.

In parctice, V_t and W_t will generally be unkown but can be estimated by maximizing the log likelihood. The choice of the initial values (the values at time zero) is arbitrary. This, in general, is of little consequence since the data themselves soon take over. One could, of course use the conventional estimate based on the first year's data in which case $a_1 = a_{1/0}$. Full details are given in a working paper that may be obtained from the writer on request.

In the application referred to above, the Kalman filter model revealed a trend that was consistent with what would be expected if pollution abatement actions were being effective, but was not obvious from conventional year-by-year regression analysis. This was largely due to the very different character of the length distribution of the fish sampled in the various years. In some years there was a very compact cluster of fish lengths which provided little information as to the parameters of the regression for those years. In other years there was a compact cluster coupled with one or two extreme values which then dictated the slopes of those regressions. In other years, the fish lengths were fairly uniformly distributed over a relatively wide range. In such years, the Kalman filter had negligible effect on the parameter estimates, but in years with compact clusters with, perhaps, one or two extremes, the Kalman filter generated estimates more in keeping with the overall picture. Based on a least-squares criterion, the individual least squares fits must, of course, be best; the increase in the residual mean square using the Kalman filter was modest and, in all years, the filter estimates just as credible as the individual least squares.

Various Kalman filter formulations are possible. One does not need to have a regression. For example, Aoki (1987), citing Harvey (1984), defines a random trend model by

$$y_t = \mu_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$

$$\beta_t = \beta_{t-1} + \zeta_t$$

where the term μ_t can be though of as a random trend because it reduces to a constant plus a linear term in time when the additive noises are absent. Autoregressive (AR) and moving average (MA) models, and their combination (ARMA) can be given a similar state-space representation. These models are fairly flexible and may well provide a good representation of abundance index time series.

Whether one prefers estimates based on each year's data individually or estimates that incorporate prior data along the lines indicated above, depends on how much one believes that there is some structure underlying the series in contrast to it being a succession of haphazard and unrelated events.

Meta-Analysis.

Somewhat related to the above is the concept of meta-analysis. Although the term "meta-analysis" is of relatively recent origin, the concept has been around for a long time. Hedges and Oklin (1985) define meta-analysis as "the rubric used to describe quantitative methods for combining evidence across studies". They note that "the earliest reference to a statistical procedure for combining significance tests appears to be in a book by L.H.C. Tippett published in 1931". The more commonly used procedure is, however, that first given by Fisher (1932).

Specifically, let p be the probability that the observed value of some test statistics would be exceeded under the null hypothesis, e.g. $p = P(t > T)$ where T is the observed value of the test statistic. Since T is a function of the observations, regarded as random variables, it is also a random variable, as is p since it, in turn, is a function of a random variable. It, therefore, has a distribution which, if one is dealing with continuous random variables, is uniform between 0 and 1. It then follows that, if we have several, say n, tests of the same hypothesis, the quantity $-2\sum \log(p_i)$ will be distributed as a chi-squared on $2n$ degrees of freedom. One does not have to use the the same test statistic in each study. (Mathematically, one does not even have to have the same null hypothesis, although it would make little sense to be combining tests of different hypotheses).

This approach is most useful if the individual studies are too small to detect, with reasonably high probability, meaningful departures from the null hypothesis, i.e. have low power that is a low probability of correctly rejecting the null hypothesis. A well publicized recent case concerned the beneficial effect of aspirin on heart attack. A relatively large number of studies had been carried out at various research institutions throughout the world. Because of cost, these studies were generally small scale and all found the null hypothesis of no benefit to be acceptable. Some researchers at Oxford, however, collected all the published results and, after putting them together and doing an appropriate meta-analysis, found very stong evidence of a beneficial effect. Fortunately, seemingly negative results had been widely reported. There is, however, a tendency to report only positive results; studies with negative results often remain unpublished. Meta-analysis requires access the the results of *all* relevant studies. Analysis of a subset may well bias the inference. Acquistion of all relevent studies is often difficult. This would seem less of a problem for the type of application envisioned here; i.e. the combining of information from different river systems feeding into a common fishery, since there seems to be a strong degree of cooperation between the interested parties.

Often one may be interested not so much in testing an hypothesis but in estimating the magnitude of an effect from the collection of relevant studies. This, in general, presents more of a challenge since effect magnitudes are likely to be less robust against differences in the measurement process than are hypothesis tests. A good account of available methodologies is given in Hedges and Olkin (1985). The recent text by Hunter and Schmidt (1990) seems more philosophical than operational. It may be of interest to note, in passing, that many of the early applications of meta-analysis (before the word was coined) occurred in the biological sciences. Most papers on meta-analysis *per se* have appeared, however, in the psychological and educational literature. More recently, there seems to have been some recognition of the part that the more sophisticated methods developed for these areas can play in enviromental studies.

Spatial Analysis.

One topic that it may be worth touching on briefly is the geostatistical analysis of spatial data. Geostatistics attempts to utilize spatial structure to interpolate abundances at unsampled locations. The basic asumption is that abundances (or densities) at geographically close locations will be more similar than abundances (or densities) at spatially more distant locations. Habitat prefence could upset a purely spatial structure but may, perhaps, be used as a covariate. The potential for a geostatistical approach in the analysis of juvenile striped bass survey data cannot be evaluated from the data presented.

References

Aoki, M. 1987. State Space Modeling of Time Series. Springer-Verlag, Berlin.

Fisher, R.A. 1932. Statistical Methods for Research Workers (4th. ed.) Oliver and Boyd, London U.K.

Harvey, A.C. 1984. Trends and cycles in macroeconomic time series. London School of Economics Working Paper. London U.K.

Hedges, L.V. and Olkin. I. 1985. Statistical Methods for Meta-Analysis. Academic Press, San Diego CA.

Hunter, J.E. and Schmidt. 1990. Methods of Meta-Analysis. Sage, Newbury Park, CA.

EXAMPLES OF BOX-COX TRANSFORMATION

The Box-Cox transformation is designed to transform data to a new variable having an approximately normal distribution. If x_i denotes the data, $i = 1, 2, \ldots, n$ the transformation is

$$y_i = (x_i^{\lambda} - 1)/\lambda, \ \lambda \neq 0$$
$$= \log(x_i), \ \lambda = 0$$

where λ is chosen to maximize the likelihood under the assumption that the y_i are normally distributed. This is equivalent to maximizing

$$L = -(n/2)\log(\hat{\sigma}^2) + (\lambda - 1)\sum \log(x_i)$$

where

$$\hat{\sigma}^2 = \sum (y_i - \bar{y})^2/n$$
$$\bar{y} = \sum y_i/n$$

Consider the following data set ($n = 20$).
$\underline{x}' =$ [6 46 50 71 129 190 303 329 362 408 446 519 553 834 837 870 965 1193 1228 1254]. We obtain the following values of L for selected λ.

λ	L
1.0	-120.1718
0.9	-119.1891
0.8	-118.3512
0.7	-117.6793
0.6	-117.1986
0.5	-116.9393
0.4	-116.9373
0.3	-117.2349
0.2	-117.8809
0.1	-118.9289
0.05	-119.6208
0	-120.4336

L is maximized with λ approximately 0.45 (= 1/2.2) (Fig. A.1). This is comforting since the data were obtained by squaring a random sample of normally distributed variables. In practice one would likely apply a square-root transformation.

Consider now the following data set ($n = 20$).
$\underline{x}' =$ [13 20 20 23 31 40 57 61 67 75 83 98 105 180 181 191 224 316 333 345]. Again we obtain values of L for selected λ, thus:

λ	L
1.0	-93.3557
0.9	-92.1981
0.8	-91.1354
0.7	-90.1868
0.6	-89.3607
0.5	-88.6657
0.4	-88.1099
0.3	-87.7011
0.2	-87.4460
0.1	-87.3504
0.05	-87.3637
0	-87.4184

L is maximized with λ approximately 0.1 (=1/10) (Fig. A.2). In practice one would likely take $\lambda = 0$ i.e. $y = \log(x)$. (The data were here obtained by exponentiating a normal random variable).

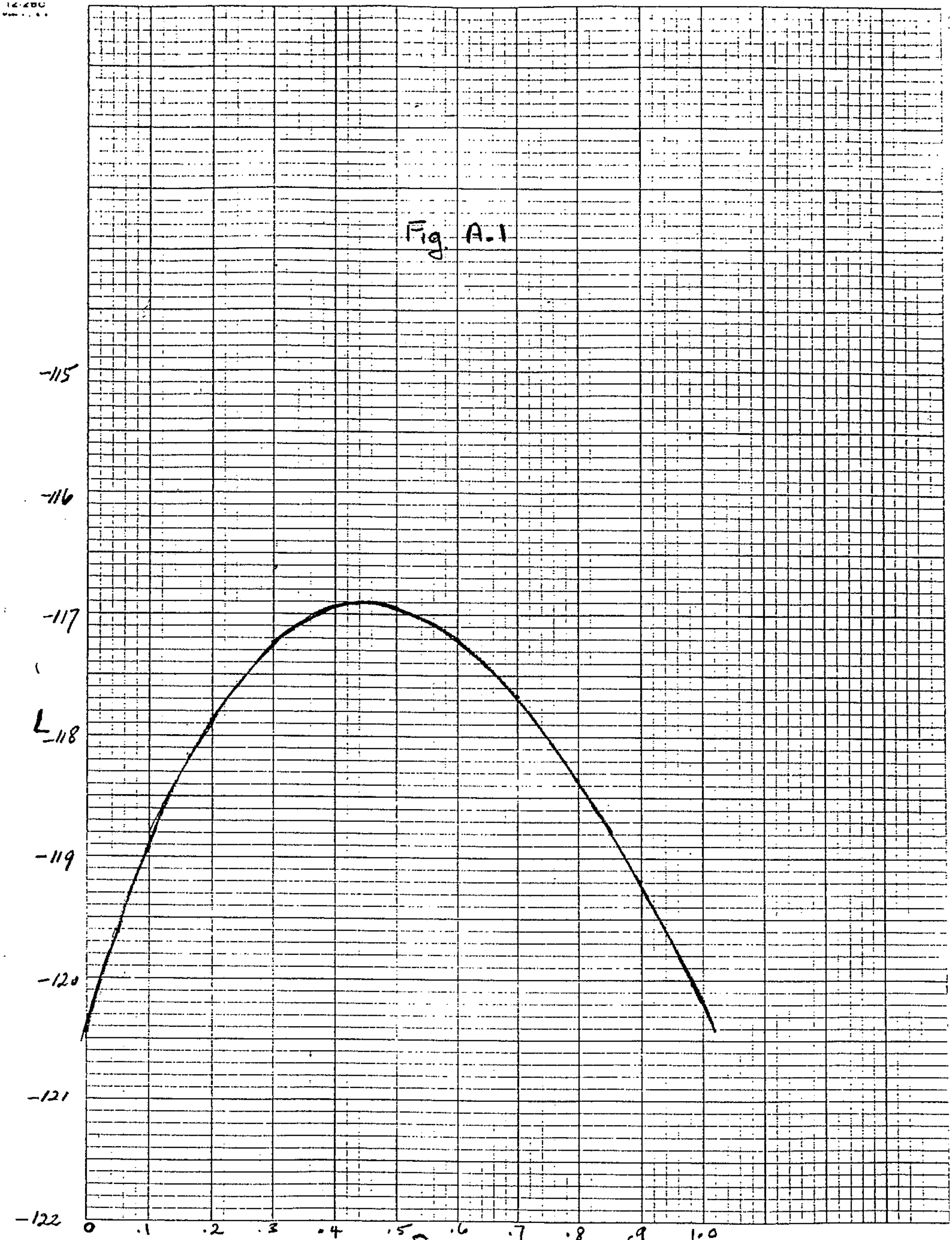


Fig. A.1
-115
-116
-117
L
-118
-119
-120
-121
-122
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0

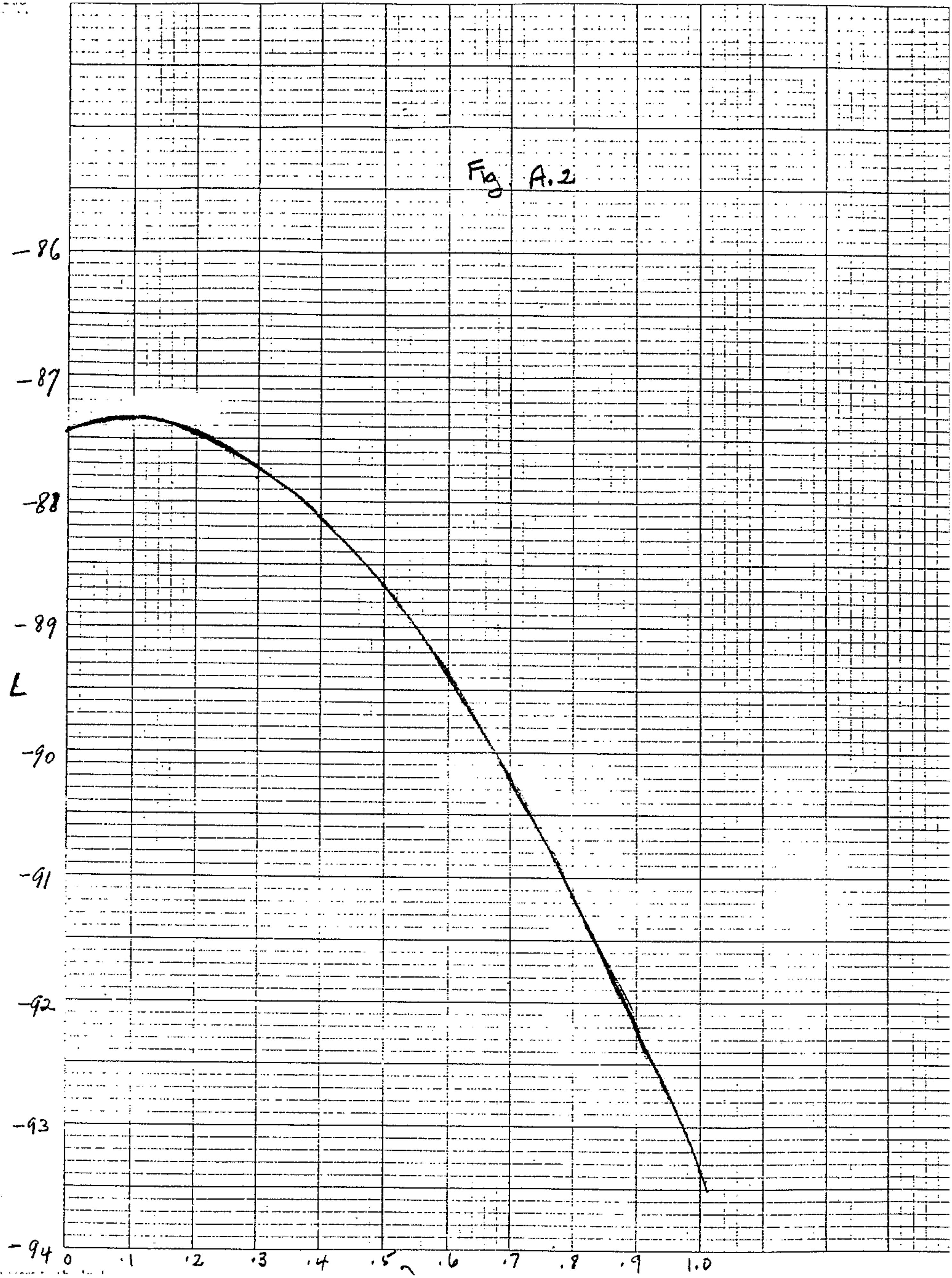


Fig. A.2
-86
-87
-88
-89
L
-90
-91
-92
-93
-94
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0

SAMPLING WITH PARTIAL REPLACEMENT
(Not to be cited without reference to the author)

Following Nicholson et al. (1991) we let μ_{iy} denote the (true) index of abundance at the i^{th} station in year y, and suppose that the total number of (possible) stations in the area is N. Then the (true) mean index of abundance in the area in year y is given by

$$\bar{\mu}_y = \sum_{i=1}^{N} \mu_{iy}/N$$

Let x_{iy}, $i = 1, 2, \ldots n$, be the observations made at n $(< N)$ sample stations in year y. Further, we suppose that the sampling variance of x_{iy} at the i^{th} station (in year y) is σ_{iy}^2. This variation somes about from replicate observations at the same location not yielding exactly the same observations and the station being an area, rather than a point, so that repeated measurements within the same station are not necessarily made at exactly the same point. (This is akin to the nugget effect in kriging).

Firstly, suppose that the n sample stations are selected at random form the N available. Then, as is well known, the sample mean, $\bar{x}_y = \sum_{i=1}^{n} x_{iy}/n$, is an unbiased estimator of $\bar{\mu}_y$. It follows that the variance of $\bar{x}_y$ is sum of two components, one stemming from the usual sampling variance, i.e.

$$\sigma_\mu^2 = \sum_{i=1}^{N} (\mu_{iy} - \bar{\mu}_y)^2/N$$

and the the other from the "measurement" error, i.e.

$$\sum_{i=1}^{N} \sigma_{iy}^2/nN$$

An unbiased estimator of $Var(\bar{x}_y)$ is

$$\sum_{i=1}^{n} (x_{iy} - \bar{x}_y)^2/n(n-1) = s^2/n, \text{ say}$$

If the sample stations are fixed, the data refer specifically to those stations and

$$E(\bar{x}_y) = \sum_{i=1}^{n} \mu_{iy}/n \neq \bar{\mu}_y, \text{in general}$$

Further

$$Var(\bar{x}_y) = \sum_{i=1}^{n} \sigma_{iy}^2/n^2$$

Commonly, but not necessarily,

$$\sum_{i=1}^{n} \sigma_{iy}^2/n^2 < \sigma_\mu^2/n + \sum_{i=1}^{N} \sigma_{iy}^2/Nn$$

i.e. the variance of the fixed-station mean will be less than the variance of the random-station mean. In a sense this is not a fair comparison, since the former is a biased estimator of $\bar{\mu}_y$. Accuracy is, perhaps, best measured by the mean squared error which, for fixed stations is

$$(\bar{\mu}_y - \sum_{i=1}^{n} \mu_{iy}/n)^2 + \sum_{i=1}^{n} \sigma_{iy}^2/n^2$$

It is difficult to generalize how the magnitude of this quantity compares with the variance of the random-station mean, which is also its mean squared error. A fortuitous set of fixed stations may yield a highly accurate mean; unfortunately there is no way of determining which circumstance exists from the sample data *per se.*

Our primary interest here is, however, the change in abundance over time. Again, following Nicholson et al. (1991) let

$$\mu_{iy} = \mu + \phi_i + \psi_y + \xi_{iy}$$

where, for convenience, it is assumed that

$$\sum_{i=1}^{N} \phi_i = \sum_{y=1}^{Y} \psi_y = \sum_{i=1}^{N} \xi_{iy} = \sum_{y=1}^{Y} \xi_{iy} = 0$$

(No loss of generality is introduced by these contraints although they can be an inconvenience if the data set is augmented by an additional year). Only when the $\xi_{iy} = 0$ for all i and y will the difference between years at any specific station be the same as the difference between the overall means; this property has been described by Houghton (1987) as persistence.

If the sample stations are selected at random in each year we have

$$E(\bar{x}_2 - \bar{x}_1) = [\sum_{i=1}^{N}(\mu + \phi_i + \psi_2 + \xi_{i2}) - \sum_{i=1}^{N}(\mu + \phi_i + \psi_1 + \xi_{i1})]/N$$

$$= \psi_2 - \psi_1$$

i.e. $\bar{x}_2 - \bar{x}_1$ is an unbiased estimator of the difference between the year effects.

If, however, the same set of fixed stations is used in each year then

$$E(\bar{x}_2 - \bar{x}_1) = [\sum_{i=1}^{n}(\mu + \phi_i + \psi_2 + \xi_{i2}) - \sum_{i=1}^{n}(\mu + \phi_i + \psi_1 + \xi_{i1})]/n$$

$$= \psi_2 - \psi_1 + \sum_{i=1}^{n}(\xi_{i2} - \xi_{i1})/n$$

Thus, in general, unless one has the proprerty of persistence, the difference between the means of fixed stations is a biased estimator of the difference bewteen year effects.

We now consider $Var(\bar{x}_2 - \bar{x}_1)$ under independent random selection of stations in each year.

Firstly write $\bar{x}_2 - \bar{x}_1$ as

$$[(\Sigma_2\phi_i - \Sigma_1\phi_i) + n(\psi_2 - \psi_1) + (\Sigma_2\xi_{i2} - \Sigma_1\xi_{i1}) + (\Sigma_2\epsilon_{i2} - \Sigma_1\epsilon_{i1})]/n$$

where Σ_y denotes summation over the stations in year y and the ϵ_{iy} denote the "measurement" errors. For convenience, it is assumed that, in each year, there is one observation at each of the sampled stations. We require, therefore,

(1) $Var(\Sigma_2\phi_i - \Sigma_1\phi_i) = Var(\Sigma_2\phi_i) + Var(\Sigma_1\phi_i)$ - $2Cov(\Sigma_2\phi_i, \Sigma_1\phi_i)$. Now

$$Var(\Sigma_y\phi_i) = n\frac{N-n}{N}\frac{\sum_{i=1}^{N}\phi_i^2}{N-1}$$

and, as shown in Appendix I,

$$Cov(\Sigma_2\phi_i, \Sigma_1\phi_i) = 0$$

Define

$$\sigma_\phi^2 = \sum_{i=1}^{N} \phi_i^2/N$$

Then

$$Var(\Sigma_2\phi_i - \Sigma_1\phi_i) = 2n\frac{N-n}{N}\frac{N}{N-1}\sigma_\phi^2$$

(2) Likewise

$$Var(\Sigma_y\xi_{iy}) = n\frac{N-n}{N}\frac{\sum_{i=1}^{N}\xi_{iy}^2}{N-1}$$

and, clearly,

$$Cov(\Sigma_1\xi_{i1}, \Sigma_2\xi_{i2}) = 0$$

Define

$$\sigma_\xi^2 = \sum_{i=1}^{N}\sum_{y=1}^{2}\xi_{iy}^2/2N$$

i.e., for the time being, we take $Y = 2$. Then

$$Var(\Sigma_2\xi_{i2} - \Sigma_1\xi_{i1}) = 2n\frac{N-n}{N}\frac{N}{N-1}\sigma_\xi^2$$

(3) Next $Var(\Sigma_2\epsilon_{i2} - \Sigma_1\epsilon_{i1}) = \sum_{i=1}^{n}\sum_{y=1}^{2}\sigma_{iy}^2$. It may not be unreasonable to suppose $\sigma_{iy}^2 = \sigma_\epsilon^2$, for all i, y, in which case

$$Var(\Sigma_2\epsilon_{i2} - \Sigma_1\epsilon_{i1}) = 2n\sigma_\epsilon^2$$

(4) Clearly, $Var(\psi_2 - \psi_1) = 0$.

(5) Finally, all the covariances such as $Cov(\Sigma_2\phi_i - \Sigma_1\phi_i, \Sigma_2\xi_{i2} - \Sigma_1\xi_{i1})$ are zero.

Bringing the above together, we have that

$$Var(\bar{x}_2 - \bar{x}_1) = \frac{2}{n}\left[\sigma_\epsilon^2 + \frac{N-n}{N}\frac{N}{N-1}(\sigma_\phi^2 + \sigma_\xi^2)\right]$$

If, as is usually the case, $n << N$ then

$$Var(\bar{x}_2 - \bar{x}_1) \approx 2[\sigma_\phi^2 + \sigma_\xi^2 + \sigma_\epsilon^2]/n$$

Next, consider $Var(\bar{x}_2 - \bar{x}_1)$, under the assumption of fixed stations; it is assumed that these stations are purposively selected by some criterion. Then

$$\bar{x}_2 - \bar{x}_1 = [(\psi_2 - \psi_1) + \sum_{i=1}^{n}(\xi_{i2} - \xi_{i1}) + \sum_{i=1}^{n}(\epsilon_{i2} - \epsilon_{i1})]/n$$

The only random components here are the ϵ_{iy}. Thus

$$Var(\bar{x}_2 - \bar{x}_1) = 2\sigma_\epsilon^2/n$$

It is clear that the variance of the difference is less with fixed stations than with random sampling, but what about the mean squared error? With fixed stations this is

$$[2\sigma_\epsilon^2 + (\sum_{i=1}^{n}(\xi_{i2} - \xi_{i1}))^2/n]/n$$

Thus, fixed stations will be more accurate in estimating change if, for $n << N$,

$$[\sum_{i=1}^{n}(\xi_{i2} - \xi_{i1})]^2/n < 2(\sigma_\phi^2 + \sigma_\xi^2)$$

This will depend on what happens at the subset of fixed stations. Under persistence, $\xi_{iy} = 0$ for all i and y, so that, with fixed stations, the estimator will always be the more accurate (as well as unbiased). However, if there is a lack of persistence, the differences $\xi_{i2} - \xi_{i1}$ might well be large, yielding appreciable bias and, hence, a mean squared error greater than under independent random sampling.

Even if the fixed-station estimator is the more accurate we may still prefer independent random samples, if there is a cost differential in favour of the latter. For example, in forestry, permanent plots are considerably more expensive to establish than temporary plots. The plot boundaries have to be carefully marked to ensure that, in future surveys, exactly the same trees, apart from ingrowth and mortality, are measured. Such cost differential is unlikely to occur in fishery surveys, although there be some added navigational cost in returning to a relatively precise location if the station area is small. On the other hand, foresters are dealing with units that are fixed in space and have to suffer the environmental conditions imposed on them. Fish, however, are mobile and can react spatially to environmental changes. Accordingly, lack of persistence is likely to be a more serious problem with fishery than with forestry surveys.

Sampling with partial replacement, i.e. keeping some sampled units fixed but selecting the others at random on each occasion, has been used with some success in forestry and might well produce some gains in fishery surveys, albeit for a different reason (lack of persistence as opposed to a cost differential between fixed and random stations). We, therefore, now develop the properties of estimators based on sampling with partial replacement.

Let the sample be as before except that n_2 of the observations are from fixed stations and $n_1 = n - n_2$, independently selected at random from the remaining $N - n_2$ stations in each year. We now let Σ_y denote the some over the random stations in year y, and Σ_0 the sum over the fixed stations. We then have

$$\bar{x}_2 - \bar{x}_1 = [\Sigma_2\phi_i - \Sigma_1\phi_i + n(\psi_2 - \psi_1) + \Sigma_2\xi_{i2} + \Sigma_0\xi_{i2} - \Sigma_1\xi_{i1} - \Sigma_0\xi_{i1} + \sum_{i=1}^{n}\epsilon_{i2} - \sum_{i=1}^{n}\epsilon_{i1}]/n$$

Now

$$E(\Sigma_2\phi_i) = E(\Sigma_1\phi_i) = n_1\Sigma_0\phi_i/(N - n_2)$$

By the same token

$$E(\Sigma_2\xi_{iy}) = -n_1\Sigma_0\xi_{iy}/(N - n_2)$$

whence

$$E(\bar{x}_2 - \bar{x}_1) = \psi_2 - \psi_1 + \frac{N - n}{n(N - n_2)}\Sigma_0(\xi_{i2} - \xi_{i1})$$

The result for fixed stations ($n = n_2$, $n_1 = 0$) and independent random samples ($n_2 = 0$, $n_1 = n$) can, of course, be obtained from this general case.

Finally, we consider $Var(\bar{x}_2 - \bar{x}_1)$. Since, in effect, we are dealing with independent random samples of n_1 stations out of a possible $N - n_2$

$$Var(\Sigma_2\phi_i - \Sigma_1\phi_i) = n_1\frac{N - n}{N - n_2}\frac{\sum^{N-n_2}\phi_i^2}{N - n_2 - 1}$$

where $\sum^{N-n_2}$ denotes summation over the $N - n_2$ non-fixed stations. Now

$$\sum^{N-n_2}\phi_i^2 = \sum_{i=1}^{N}\phi_i^2 - \Sigma_0\phi_i^2$$

$$= N\sigma_\phi^2 - \Sigma_0\phi_i^2$$

Define

$$\sigma^2_{\phi,0} = \Sigma_0(\phi - \Sigma_0\phi_i/n_2)^2/n_2$$
$$= (\Sigma_0\phi_i^2 - n_2\bar{\phi}_0^2)/n_2, \text{ say}$$

whence

$$Var(\Sigma_2\phi_i - \Sigma_1\phi_i) = n_1\frac{N-n}{N-n_2}\frac{N(\sigma^2_\phi - n_2[\sigma^2_{\phi,0} + \bar{\phi}_0^2]/N)}{N-n_2-1}$$

Likewise, define

$$\bar{\xi}_0 = \Sigma_0\sum_{y=1}^{2}\xi_{iy}/2n_2$$

and

$$\sigma^2_{\xi,0} = \Sigma_0\sum_{y=1}^{2}\xi^2_{iy}/2n_2$$

Then

$$Var(\Sigma_2\xi_{i2} - \Sigma_1\xi_{i1}) = 2n_1\frac{N-n}{N-n_2}\frac{N(\sigma^2_\xi - n_2[\sigma^2_{\xi,0} + \bar{\xi}_0^2]/N)}{N-n_2-1}$$

Thus

$$Var(\bar{x}_2 - \bar{x}_1) = \frac{2\sigma^2_\epsilon}{n} + \frac{2n_1}{n^2}\frac{N-n}{N-n_2}\frac{N}{N-n_2-1}\left[\sigma^2_\phi + \sigma^2_\xi - \frac{n_2}{N}(\sigma^2_{\phi,0} + \sigma^2_{\xi,0} + \bar{\phi}_0^2 + \bar{\xi}_0^2)\right]$$

ILLUSTRATION

To obtain some idea of what might happen in practice, persistence, i.e. σ^2_ξ, was estimated from research trawl survey data for cod in NAFO Division 2J for two successive years (1988, 1989). These are stratified random surveys with approximately 150 stations in each year. For the present purpose, the stratification was ignored and station in one year within 2.5 nm of a station in the other year was regarded as being at the same location. (The 2.5 nm was somewhat arbitrary; it was chosen as the smallest distance from which a reasonable number of "fixed" stations could be generated from these data).

From the above we have, for a given year,

$$Var(x_{iy}) = \sigma^2_\phi + \sigma^2_\xi + \sigma^2_\epsilon$$

for which we may obtain a pooled estimate in the usual way as

$$s^2_x = [\sum_{y=1}^{2}\sum_{i=1}^{n_i}(x_{iy} - \bar{x}_y)^2]/(n_1 + n_2 - 2)$$

where, here, n_1 and n_2 denote the number of stations in the two years. Further

$$d_i = x_{i2} - x_{i1} = \psi_2 - \psi_1 + \xi_{i2} - \xi_{i1} + \epsilon_{i2} - \epsilon_{i1}$$
$$= \psi_2 - \psi_1 + 2\xi_{i2} + \epsilon_{i2} - \epsilon_{i1}$$

since $\xi_{i1} + \xi_{i2} = 0$. Thus

$$Var(d_i) = 4\sigma^2_\xi + 2\sigma^2_\epsilon$$

which can be estimated as

$$s^2_d = \sum_{i=1}^{n}(d_i - \bar{d})^2/(n-1)$$

where, here, n is the number of "fixed" stations.

We now have two equations from which to estimate the three unknowns, σ_ϕ^2, σ_ξ^2 and σ_ϵ^2. Let us suppose that the measurement is negligible, then, if $\sigma_\xi^2 = p\sigma_\phi^2$ p can be estimated as

$$p = \frac{s_d^2/4}{s_x^2 - s_d^2/4}$$

In this way σ_ξ^2 was estimated as approximately $0.8\sigma_\phi^2$.

It has been shown above that fixed stations will be more accurate in estimating change if

$$[\sum_{i=1}^{n}(\xi_{i2} - \xi_{i1})]^2/n < 2(\sigma_\phi^2 + \sigma_\xi^2)$$

Since we assume $\xi_{i1} + \xi_{i2} = 0$, the left hand side can be written

$$4[\sum_{i=1}^{n}\xi_{i1}]^2/n$$

Suppose that the ξ_{i1} are normally distributed (justifiable since N is assumed large) and that the "fixed" stations have been chosen at random. Then $[\sum_{i=1}^{n}\xi_{i1}]^2$ will be distributed as $n\sigma_\xi^2$ times a chi-squared on 1 degree of freedom. The condition thus becomes

$$\chi_1^2 < 1.125$$

In this sense, the probability that fixed samples yield a more accurate estimate of change than random independent random samples is computed as

$$Prob(\chi_1^2 < 1.125) = 71.1\%$$

Clearly, if the persistence is greater ($\sigma_\xi^2 < 0.8\sigma_\phi^2$) the probability will increase, and tends to unity as σ_ξ^2 tends to zero. The smaller the persistence (the greater σ_ξ^2 relative to σ_ϕ^2) the smaller this probability, with a limit of

$$Prob(\chi_1^2 < 0.5) = 52.0\%$$

We have here assumed equal costs for fixed and random stations.

It is, perhaps, more germane to determine the chance of fixed samples yielding a substantially more accurate estimate of change than independent random samples. Recall that to halve a confidence interval one must quarter the variance. Hence we consider $Prob(\chi_1^2 < 1.125\alpha^2)$.

α	0.95	0.9	0.8	0.7	0.6	0.5
Probability (%)	68.6	66.0	60.4	54.2	47.5	40.4

In the case of sampling with partial replacement, if $N >> n$, the bias is approximately $\Sigma_0(\xi_{i2} - \xi_{i1})/n$ and $Var(\bar{x}_2 - \bar{x}_1)$ approximately

$$\frac{2\sigma_\epsilon^2}{n} + \frac{2n_1}{n^2}[\sigma_\phi^2 + \sigma_\xi^2]$$

Thus, the mean squared error is, approximately,

$$\frac{[\Sigma_0(\xi_{i2} - \xi_{i1})]^2}{n^2} + \frac{2\sigma_\epsilon^2}{n} + \frac{2n_1}{n^2}[\sigma_\phi^2 + \sigma_\xi^2]$$

How does this relate to the special cases of all samples fixed ($n_1 = 0$, $n_2 = n$) and independent random samples ($n_1 = n$, $n_2 = 0$)? If the common term, $2\sigma_\epsilon^2/n$ is omitted, this means comparing

$$\frac{[\sum_{i=1}^{n}(\xi_{i2}-\xi_{i1})]^2}{n^2}$$

$$\frac{[\Sigma_0(\xi_{i2}-\xi_{i1})]^2}{n^2}+\frac{2n_1}{n^2}[\sigma_\phi^2+\sigma_\xi^2]$$

and

$$2[\sigma_\phi^2+\sigma_\xi^2]/n$$

Consider the probability that sampling with partial replacement will have a smaller mean squared error than independent random sampling under the assumption that the "fixed" stations are selected at random. Then

$$\frac{[\Sigma_0(\xi_{i2}-\xi_{i1})]^2}{n^2}+\frac{2n_1}{n^2}[\sigma_\phi^2+\sigma_\xi^2]<2[\sigma_\phi^2+\sigma_\xi^2]/n$$

is equivalent to

$$\frac{4n_2\sigma_\xi^2\chi_1^2}{n}<2(\sigma_\phi^2+\sigma_\xi^2)[1-n_1/n]$$

which reduces to

$$\chi_1^2<\frac{1}{2}[1+\sigma_\phi^2/\sigma_\xi^2]$$

that is, the same condition, for fixed samples to be more accurate in estimating change than independent random samples. The difference is that, under partial replacement, (n_1 strictly less than n) the variance component is reduced; this may or may not be counterbalanced by the introduction of bias.

The condition for sampling with partial replacement to yield smaller mean squared error than fixed samples is somewhat more complicated. Consider

$$\frac{[\Sigma_0(\xi_{i2}-\xi_{i1})]^2}{n^2}+\frac{2n_1}{n^2}[\sigma_\phi^2+\sigma_\xi^2]<\frac{[\sum_{i=1}^{n}(\xi_{i2}-\xi_{i1})]^2}{n^2}$$

This is equivalent to

$$4n_2\sigma_\xi^2\chi_{1(P)}^2+2n_1(\sigma_\phi^2+\sigma_\xi^2)<4n\sigma_\xi^2\chi_{1(F)}^2$$

where $\chi_{1(P)}^2$ and $\chi_{1(F)}^2$ denote two separate chi-squared variables. (If we assume that the fixed stations in these two situations are independently chosen [at random] then these chi-squared variables may be regarded as independent). Let p denote the proportion of fixed stations ($= n_2/n$). The condition can then be written

$$p\chi_{1(P)}^2-\chi_{1(F)}^2<\frac{1-p}{2}[1+\sigma_\phi^2/\sigma_\xi^2]$$

Since, in general, a linear combination of chi-squared variables is *not* itself a chi-squared variable - the exception is a sum of chi-squared variables - the right hand side would have to be evaluated by double integration. This is beyond the scope of the present note. However the bias component will be reduced under partial replacement if

$$p\chi_{1(P)}^2<\chi_{1(F)}^2$$

or

$$F_{1,1}<1/p$$

where $F_{1,1}$ is distributed as Snedecor's F on 1 and 1 degrees of freedom. We obtain the following (values in %)

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Probability (%)	80.5	73.2	68.1	64.1	60.8	58.0	55.6	53.5	51.7

where the limits as p tends to 0 and 1.0 are 100% and 50%, respectively. In other words, partial replacement is likely to reduce the (absolute) bias, but this may or may not be counterbalanced by the introduced variance.

It must be emphasized that, in these illustrations, it has been assumed that the "fixed" stations have been randomly selected. The inferences do not, therefore, necessarily apply to purposive selection of the fixed stations. With sufficient prior knowledge it may be possible to select the fixed stations with small or negligible bias in the estimate of change. On the other hand, we have not here considered stratification to reduce the inter-station component of variation. Also we have not considered specifically and cost differential in sampling fixed versus random stations which, again, could alter the balance. Finally, the focus here has been on the estimation of change. The balance could again be changed if, in addition to change, estimates of the actual abundance are required in one or both years and how much emphasis is given to the latter.

APPENDIX I

We here show that $Cov(\Sigma_1\phi_1, \Sigma_2\phi_i) = 0$.

Since the sample stations in each year are selected independently and at random, there is a possibility that one or more stations will be common to both samples; indeed, the number of common stations can be 0, 1, 2, ... n.

The total number of possible sample combinations is $\binom{N}{n}^2$, and the number of these for which there will be no station in comon is $\binom{N}{n}\binom{N-n}{n}$.

The number of sample combinations that will contain exactly one station in common is $\binom{N}{1}\binom{N-1}{n-1}\binom{N-n}{n-1}$, the number with exactly two stations in common is $\binom{N}{2}\binom{N-2}{n-2}\binom{N-n}{n-2}$ and, in general, the number with exactly j stations in common is $\binom{N}{j}\binom{N-j}{n-j}\binom{N-n}{n-j}$. Accordingly

$$\binom{N}{n}^2 = \sum_{j=1}^{n}\binom{N}{j}\binom{N-j}{n-j}\binom{N-n}{n-j}$$

For simplicity we assume $N > 2n$ which, in practice, will always be the case.

If there is no station in common then $\Sigma_1\phi_i\Sigma_2\phi_i$ will contain terms $\phi_i\phi_j$, $i < j$, and there will be n^2 such terms. On taking expectations every possible $\phi_i\phi_j$ $(i < j)$ must occur an equal number of times. Thus, the number of time any particular $\phi_i\phi_j$, with $i \neq j$, will occur is

$$n^2\binom{N}{0}\binom{N}{n}\binom{N-n}{n}/\binom{N}{2}$$

Next, suppose that the samples have exactly one station in common. Then, $\Sigma_1\phi_i\Sigma_2\phi_i$ will contain *one* term of the form ϕ_i^2 and $n^2 - 1$ terms such as $\phi_i\phi_j$, $i < j$. Again, on taking expectations, each ϕ_i^2 must occur an equal number of times, as must each $\phi_i\phi_j$ with $i < j$. The number of occurrences of each ϕ_i^2 is, therefore

$$1\binom{N}{1}\binom{N-1}{n-1}\binom{N-n}{n-1}/\binom{N}{1}$$

and the number of occurrences of each $\phi_i\phi_j$, $i < j$

$$(n^2-1)\binom{N}{1}\binom{N-1}{n-1}\binom{N-n}{n-1}/\binom{N}{2}$$

More generally, suppose that the samples have exactly j stations in common. Then, $\Sigma_1\phi_i\Sigma_2\phi_i$ will contain j terms of the form ϕ_i^2 and $n^2 - j$ terms like $\phi_i\phi_j$, $i < j$. Then, on taking expectations, the number of occurrences of any particular ϕ_i^2 will be

$$j\binom{N}{j}\binom{N-j}{n-j}\binom{N-n}{n-j}/\binom{N}{1}$$

and the number of occurrences of each $\phi_i\phi_j$, $i < j$

$$(n^2-j)\binom{N}{j}\binom{N-j}{n-j}\binom{N-n}{n-j}/\binom{N}{2}$$

Thus, overall, the number of occurrences of each ϕ_i^2 will be

$$\sum_{j=0}^{n} j \binom{N}{j}\binom{N-j}{n-j}\binom{N-n}{n-j} / \binom{N}{1}$$

and the number of occurrences of each $\phi_i \phi_j$, $i < j$

$$\sum_{j=0}^{n} (n^2 - j) \binom{N}{j}\binom{N-j}{n-j}\binom{N-n}{n-j} / \binom{N}{2}$$

Now

$$\sum_{j=0}^{n} n^2 \binom{N}{j}\binom{N-j}{n-j}\binom{N-n}{n-j} / \binom{N}{2} = n^2 \binom{N}{n}^2 / \binom{N}{2}$$

We therefore focus on

$$\sum_{j=0}^{n} j \binom{N}{j}\binom{N-j}{n-j}\binom{N-n}{n-j}$$

$$= \sum_{j=1}^{n} N \binom{N-1}{j-1}\binom{N-j}{j-1}\binom{N-n}{n-j}$$

$$= \sum_{j=0}^{n-1} N \binom{N-1-j}{j}\binom{N-1-j}{n-1-j}\binom{N-n}{n-1-j} = N \binom{N-1}{n-1}^2$$

Thus, each ϕ_i^2 occurs

$$N \binom{N-1}{n-1}^2 / \binom{N}{1} = \binom{N-1}{n-1}^2$$

times and each $\phi_i \phi_j$ occurs

$$\left[n^2 \binom{N}{n}^2 - N \binom{N-1}{n-1}^2 \right] / \binom{N}{2} = 2 \binom{N-1}{n-1}^2$$

times, Hence

$$Cov(\Sigma_1 \phi_i \Sigma_2 \phi_i) = (n/N)[\sum_{i=1}^{N} \phi_i^2 + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \phi_i \phi_j]$$

$$= (n/N)[\sum_{i=1}^{N} \phi_i]^2 = 0$$

Appendix 2

Striped Bass Young-of-the-Year Indices of Abundance: Comments by John M. Hoenig on a Workshop held in Grasonville, Maryland, January 21 to 23, 1992

prepared by

John M. Hoenig
Department of Fisheries and Oceans[1]
P.O. Box 5667 (1 White Hills Road)
St. John's, Nfld A1C 5X1, CANADA
tel. (709) 772-2802

1) address is for identification purposes only; the opinions expressed in this report are those of the author and do not necessarily represent the views of the Department of Fisheries and Oceans of Canada.

Contents

Executive Summary

1) Most of the striped bass index of abundance programs use fixed stations which are selected from a portion of the total habitat occupied by the young-of-the-year (yoy) fish. Therefore, standard statistical methods for treating the data do not seem appropriate. In particular, it is questionable whether any of the estimated variances are appropriate.

2) I believe that conclusions and decisions concerning the status of the striped bass populations must be made more on the basis of biological arguments than on the basis of statistical reasoning. For example, if an index in the current year is higher than in the previous year, and if catch per seine haul is generally higher everywhere but especially so in one region, then this would support the idea that yoy abundance is indeed higher this year. In contrast, if abundance in the current year is sharply up in a few scattered locations and more or less normal elsewhere, this is not strong evidence for a rise in abundance. Furthermore, if examination of past data shows that yoy striped bass have a patchy distribution, so that in any year one could observe a few high catches at scattered locations (but the location of the high catches are not consistent from year to year), then there is even less reason to believe the index in the current year is higher than in the previous year.

3) Point 2 - the advisability of making conclusions on biological grounds - implies that the data should be looked at in disaggregated form, i.e., that summarizing the data for each year as a single value of an index results in the loss of ability to evaluate the biological significance of the sampling data.

4) Other reasons for viewing the data in disaggregated form are: a) one can study spatial patterns, e.g., where are the highest catches made, and what environmental factors correlate with catch level? b) one can deal with the problem of changes in the number and location of the sampling stations over time by looking at trends within individual stations.

5) It remains somewhat unresolved how an index of abundance should be constructed from the survey data. Some possibilities are: use the highest catch rate observed during the sampling season, use the average catch rate observed during a fixed window of time, use the average catch rate observed in a year-specific window of time. This cannot be resolved on purely theoretical grounds. However, there is empirical evidence from the Hudson River which sheds some light on the question. In the absence of empirical validation, one should probably compute indices in more than one way to see if the conclusions depend critically on the particular form of the index.

6) The Workshop participants discussed the use of a "trigger" for opening a fishery. I believe it is inadvisable to base such a decision on a single criterion (such as the mean of the index over a period of three consecutive years exceeding a proscribed value). Rather, it would be better to have a set of criteria that reflect biological realism, such as a) the value of the index exceeds some threshold value two years out of three and is not below another critical value in the third year, b) there is some spatial cohesiveness to the observed increase in abundance, and c) the results of (a) and (b) do not depend critically on which of several possible indices of abundance is used. Also, as noted at the workshop, most of the indices are geared towards predicting future recruitment to the fishery rather than estimating the spawning stock biomass. Since the relationship between estimated spawning stock biomass and subsequent estimated recruitment is notoriously weak, it is hard to conceive of a young-of-the-year striped bass index being useful for both predicting spawning biomass and future recruitment. Therefore, the logical basis for using an index program to trigger the opening of a striped bass fishery ought to be spelled out clearly, even if only in qualitative terms. Finally, if a trigger for opening a fishery is initiated, there should also be one or more mechanisms in place for closing the fishery if there are indications of a population decline.

I) Introduction

The U.S. Fish and Wildlife Service, Atlantic States Marine Fisheries Commission, Maryland Department of Natural Resources, and U.S. Environmental Protection Agency sponsored an international workshop on evaluation of indices of young-of-the-year (yoy) striped bass abundance. The workshop was held from 21 to 23 January, 1992, in Grasonville, Maryland. Representatives from various states along the Atlantic seaboard and the District of Columbia presented summaries of their index programs. Drs. William G. Warren and John M. Hoenig were invited as outside experts to critique the programs on statistical and biological grounds. This report summarizes Dr. Hoenig's views on the striped bass index programs.

It was noted at the Workshop that there can be two separate reasons for sampling bass in their first year of life. One possible goal is to obtain an index of spawning stock biomass. The other goal is to obtain an index of future recruitment to the fishery from the cohort or cohorts already in existence. Because the relationship between estimated spawning stock biomass and subsequent estimated recruitment is notoriously weak, it is hard to conceive of a young-of-the-year index program achieving both goals. Technical problems in sampling aside, the first goal is best achieved by sampling early in the season before most of the first-year mortality occurs; the second goal is best achieved by sampling late in the season after year-class strength is largely determined. Most of the programs discussed at the Workshop were oriented towards predicting future recruitment (at least in relative terms).

Most of the programs involve beach seining at non-randomly selected locations which are nominally fixed over time. That is, the same stations are supposed to be used year after year. This approach has several problems.

1) The seineable sites represent only a portion of the total habitat occupied by yoy striped bass so that, strictly speaking, the conclusions drawn from data obtained from seining are valid only for that portion of the population in seinable areas. In order to make conclusions about the entire year class of striped bass, it is necessary to assume something about the relationship between the fish in the seined or seinable areas and the rest of the population. Typically, it is assumed that a fixed proportion of the population is in the sampleable or sampled area each year so that trends of abundance in the data reflect trends in the population.

2) Statistical inference using standard statistical methods is not possible because of the fixed station design and because only a portion of the population (i.e. those fish in sampleable areas) are subject to study. If it were possible to randomly sample all of the habitat of yoy striped bass, then classical methods described in any text on sampling theory could be used to draw valid conclusions about trends in abundance. These methods are "design-based" meaning the validity of the conclusions is assured by the randomness introduced into the sampling design. Since it is not possible to sample all of the habitat and because fixed stations are used, any conclusions drawn must be conditional on what is assumed. Thus, the methods are "model-based" and the validity of the conclusions depend on the validity of the model.

3) Over the years, some sampling stations have been dropped and additional ones added in most of the programs. This means that, if stations are "persistent" over time (i.e., have consistently high or consistently low catches), then the observed abundances will depend on both the actual abundance of fish and the particular stations included in the sample. For example, if stations with consistently poor catches are dropped and replaced with stations with high catches then this would tend to cause the index to go

up over time (all other things being equal). Also, if stations are dropped because of habitat degradation, then there may be real loss of productivity and this loss is not being measured because the stations are dropped from consideration.

In what follows, I discuss: how the catches at a station can be summarized in the form of an index, why data should be viewed on a station by station basis rather than being oversummarized in a single index, and I provide a bound to the error caused by some areas being unsampleable.

II) Construction of an index: summarizing the season's catches

Over the course of a summer season, the catch rate of yoy striped bass should rise to a peak and then decline. The increase in catch rate is due to a combination of fish growing to a size where they are catchable and fish moving into shallow water where they can be seined. Similarly, catch rate declinces after reaching a peak due to a combination of fish reaching a size where they can avoid the seine, fish leaving the shallow areas, and reduction in abundance through mortality.

One might suppose that the peak catch rate would provide the best index of abundance. This supposes that, when the catch rate is at its maximum, all of the cohort - or a *fixed* proportion of the cohort - is present in the seineable areas. Also, the population present when the maximum occurs must be indicative of the eventual recruitment arising from the yoy population. If the peak occurs on July 15 one year and on August 15 the next year, one would have to assume that the effect of natural mortality (after the peak occurs) on year class strength is the same in the two years.

Another possibility is that the maximum proportion present in the seineable areas (i.e., peak catch rate over the season) is variable from year to year but that the average time (per fish) spent in seineable areas is constant from year to year. That is, the seasonal pattern of catch rates varies but the total usage of the seineable habitat (in fish-days) is proportional to the population size. In this case, the appropriate index of populaton size would be the area under the curve relating catch rate to time of season or, equivalently, the average catch rate over the season. This scenario might not seem very intuitive but might arise, for example, as follows: suppose that as autumn arises and fish begin to reach a certain size, the fish begin a downstream migration through shallow water; in some years many fish might be moving downstream at the same time whereas in other years, the migration might be spread out over a prolonged period of time. In this case, the maximum proportion found in the seined sites might vary from year to year whereas the individual amount of time spent in the seined areas might not vary much from year to year. It should be noted that in some years there is a clearly defined peak in the graph of catch rate over time but in other years the curve can be very flat-topped or even bimodal. This variability in the shape of the curve casts doubt on the validity of the assumption of a constant proportion of the population in the seined areas when the peak catch rate occurs.

Both of these approaches to constructing an index present operational difficulties. For the peak abundance method it may be necessary to conduct a great deal of sampling simply to define the peak and, except for the week when the peak occurs, the sampling effort will be ignored. The peak may be subject to a great deal of sampling error, and the peak might not be well defined (the curve could be bimodal). For the areal method, a great deal of sampling effort will be necessary throughout the season if you want to define when the fish first come inshore and when they leave. It should be asked if it is really worth studying the fish when they are not very susceptible to the sampling gear and when they are not using the habitat much.

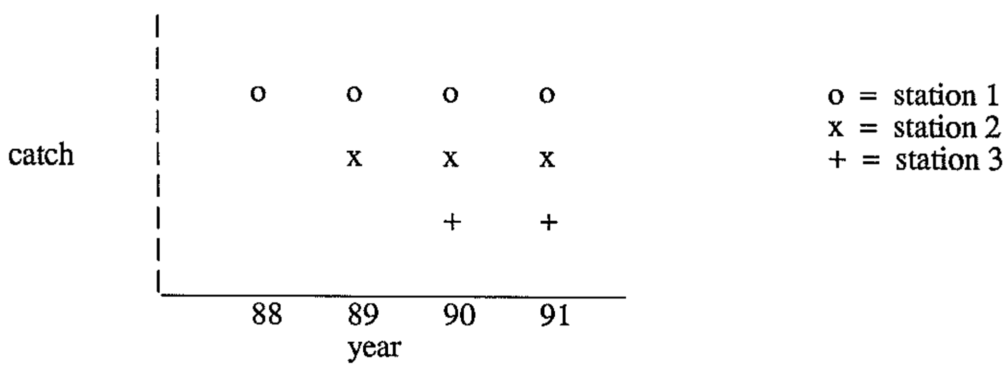
In practice, people have effectuated compromises which, in effect, combine the two approaches. That is, they use a window of time. This can be viewed as smoothing the peak or as looking at the bulk of the habitat usage. The window of time can be fixed *a priori* (e.g., the month of August every year) or it can be variable over the years and determined by conditions in each year.

There is no way to tell on theoretical grounds which is the best approach. However, it should be noted that there has been empirical validation (for Hudson River striped bass) which consisted of computing various kinds of indices and seeing which ones correlate the best with subsequent measures of year class strength. This approach should be continued. Also, the New York Department of Environmental Conservation has been looking at deeper water usage by yoy bass at the same time they conduct the beach seining. This can be used to test the peak abundance model. According to K. McKown (NYSDEC, pers. comm.), the catch rates from trawls and beach seines show an inverse relationship suggesting that there are net movements between shallow and deep water. This suggests the possibility of using the change-in-ratio estimator of Heimbuch and Hoenig (1989) or Hoenig (1990) to estimate the proportion of the population in each habitat.

Recommendation. Since it is not clear which formulation for the index of abundance is best, one should look at various possible indices to see if apparent trends depend critically on which index is considered.

III) Disaggregation of indices: how and why

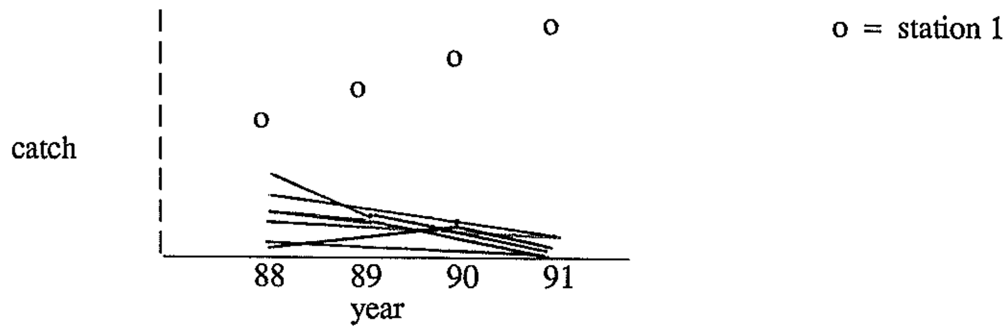
Theoretical considerations. In this section, I discuss the importance of looking at the catch data on a station by station basis. Consider the data plotted below:



Here, station 1 has been sampled each year for the last four years while station 2 has only been sampled for the last three years and station 3 has been sampled for the last 2 years. For each station, there is no indication of a trend in the catch per seine haul over time. However, if we were to compute the average catch (over all stations) for each year and use this as our index of abundance, we would conclude that there had been a sharp and steady decline in abundance. It is clear in the example above that there is a strong "station effect". That is, some stations have consistently high catches while other stations have consistently low catches. Changing the sampling locations over time adds variability to the data when there are strong station effects. If sampling stations are chosen randomly from a list of possible sites, then there is no bias associated with changing sampling sites each year -

only an increase in variance occurs. However, if stations are replaced in a non-random fashion then bias is likely to occur.

Now consider the graph below:



One station (station 1) seems to indicate a trend that is opposite to the trend indicated by the other stations. Which interpretation is the correct one - has the abundance gone up or down? Clearly, it is not possible to tell for sure from the information given because the stations are not a representative sample of the available habitat for yoy striped bass. However, one might suspect that the trend is more likely in the downward direction than upward because the catch has gone down at almost all of the stations. If one were to simply average all of the catches in each year, one might obtain an index which suggests just the opposite. In effect, one would be saying that one trusts station 1 more than the other stations combined. (Note that if one were to use the median catch in each year as the index of abundance, then one would conclude that there had been a downward trend and one would not accept the information from station 1 as being the correct interpretation.)

Finally, consider the case where some of the stations (say half) indicate that the abundance has gone up since last year and the rest of the stations indicate that abundance has declined. Suppose the mean and the median both indicate no change since last year. Do we really want to take as our estimate that there has been no change? One thing that could be done is to look at the geographical location of the stations indicating a positive change versus the location of those stations indicating a decline. If the two kinds of stations appear to be independently distributed over space then perhaps there has been no overall change. However, one might notice that the stations indicating a positive change occur in one place and those indicating a negative change occur in a different location. This would suggest that further analysis is necessary. For example, it might be that one nursery area was extremely productive while conditions in the other area were poor. (One should also check if the results could be an artifact of sampling.) If we were to conclude that different nursery areas had different levels of production, then it would be necessary to have some idea of the relative sizes of the different nursery areas in order to interpret the catch data. (In this regard, it is worth noting that efforts are under way to measure and map the nursery areas in Virginia).

The key points are that:

1) When there are strong station effects, simple averages will tend to give the most weight to those stations with the highest catches whereas medians, trimmed means, etc. will tend to reflect the trend of the majority of the stations; it is not clear which is the correct approach.

Note that if more than 50% of the stations regularly have null catches then the median catch will be 0 much or all of the time. Thus, the median may not be a very informative index. Again, the best approach is not clear. One could look at another percentile, e.g., the 60th percentile, instead of the median. Alternatively, if the percentage of stations with null catch remains constant than one can look at just the nonzero catches; otherwise, if the number of stations with null catches changes appreciably then this might be important information which should not be ignored in constructing an index. The spatial distribution of null catches might also be examined.

2) If there are strong station effects and the list of sampled stations changes from year to year, then changes in the index over time will reflect both changes in the abundance and changes in the list of stations. Given that the replacement of stations is not random, biases will result but the direction of the bias cannot be stated with any certainty.

3) If data from different regions show different trends, then determining the overall change in the population requires information on the amount of habitat in the different regions.

Recommendations. The above discussion leads me to make three recommendations for further research:

1) One could examine the existing data to see how strong the station effects might be. For example, one might rank the stations in each year by the catch at the station and then examine if some stations have consistently low ranks and if some have consistently high ranks.

2) Furthermore, one could examine the geographic locations where high and low ranking stations occur. (For the Hudson River, one could plot the average rank for a station versus the river mile where the station is located). One could also try to relate rank to physical factors associated with the stations. This would not only provide interesting ecological information (locations of apparently good and apparently bad nursery areas and factors associated with such areas) but would also help to interpret conflicting information. For example, if some stations indicate an increase in abundance and some indicate a decrease one might be able to determine that the good catches are clustered in one area and the bad catches are in another area. This might suggest that conditions were good in one nursery area but poor in another area.

3) One could compute more than one index each year (e.g., mean catch and median catch) and see if the results are in agreement. If the agreement is poor then it is necessary to decide which approach is more likely to provide the right conclusion.

Real-world situations. In many of the monitoring programs some sites have been removed from the sampling list because of habitat change (e.g., construction projects, arrival of *Hydrilla*), because some sites proved to be consistently unproductive, and because occurrence of juvenile striped bass appeared to be further upstream than previously believed (in the Delaware Bay system). These observations support the idea that there may be persistent station effects for at least some of the stations. Thus, some bias may have been introduced by adding and dropping stations.

In addition, some of the States are collecting data at "auxiliary stations" for various purposes, e.g., to better define the range of yoy striped bass and check for range extensions. It was recognized by Workshop participants that adding these stations to the list of stations used to compute the abundance index could cause bias.

Some possible methodology. The suggestions which follow are based on the idea of looking at changes within a station over time. I assume that, for each station, the season's catch can be summarized as a station-specific index. Then, some options are:

option 1: for each station, compute a series of annual differences in the index value as

$$\text{difference}_{t+1} = \text{index}_{t+1} - \text{index}_t$$

where the subscript refers to the year. Then plot the series of differences versus year for all of the stations. By plotting the year-to-year differences, instead of the year-to-year index values, we reduce the range of values that must be accommodated in the y-axis. This is very important if one station has much higher catches than the bulk of the others because, in this case, the bulk of the plotted data will hug the x-axis and trends will be hard to see. Also, the human eye is not good at judging slopes (changes) so it is advantageous to plot the slopes (i.e., the differenced data) rather than having the viewer judge slopes from the time series of index values.

option 2: construct a histogram of all of the station differences for the most recent pair of years.

option 3: designate the first year as the base year and compute differences between the station-specific index in each year and the value in the first year, i.e., compute

$$\text{base difference}_{t+1} = \text{index}_{t+1} - \text{index}_1 \; .$$

Then, for each station, plot the series of base differences versus year. (Alternatively, designate the mean of the first k years as the base or reference level, and compute the differences from this reference level).

option 4: for each year, compute the proportion of stations showing a positive change over the previous year. Also, compute the mean and median change.

option 5: compute the proportion of stations with zero catches in each year. This might be useful for monitoring how the range expands and contracts over time and this, in turn, might provide insights into whether a particular year is likely to produce a strong or weak year class. Interpretation of such information would be enhanced if it could be displayed in such a way as to reflect spatial patterns. One way to do this would be to mark an x at each station with a nonzero catch and an o at each station with a null catch. Another way to display the information would be with an expanding bubble plot in which each station is marked with a circle the size of which is proportional to the number of fish caught at the station.

IV) Effect of unsamplable areas on estimates

In this section, I assume that there is a portion of the total habitat occupied by yoy striped bass that is potentially sampleable (e.g., water within reach of the beach seine). The usual situation is to assume that the population in the sampleable areas is a fixed (but unknown) percentage, P, of the total population. If this is false, and the actual percentage of the population in the sampleable area in year t is P_t, then the apparent abundance in year t will be biased by a factor I_t

$$I_t = \frac{P_t}{P}$$

There are two extremes to consider. Suppose first that, in a given year, all of the population is in the seineable area instead of the usual P%. Then, by sampling only with the beach seines, the population will appear large relative to a normal year because the catch rate is inflated by the abnormal occurrence of 100% of the fish in the seineable area (instead of P%). In fact, the index is inflated by the factor

$$\text{inflation factor} = \frac{100}{P}\ \% \ .$$

For example, if normally 25% of the population is in the seineable areas but this year 100% is in the sampleable area then the index will be inflated by a factor of 4.

Clearly, if we can increase P, the scope for overestimation decreases. For example, if we could use a different sampling gear so that 50% of the population is normally subject to sampling, then at most we could bias our estimate of relative abundance by the factor 2.

At the other extreme, suppose in a given year 0% occurs in the sampleable area. Then, clearly, our estimate of relative abundance is 0, i.e., a 100% underestimate.

Since the worst thing we can do, from a conservation point of view, is to overestimate the relative abundance of the fish, it is of interest to find a sampling scheme that maximizes the proportion of the population normally subject to sampling.

V) Other issues

Other indices of year-class strength. In addition to looking at catch per seine haul (an index of abundance) as a predictor of year-class strength, it seems reasonable to look at other variables that might be correlated with ultimate year class strength. Thus, it might be worthwhile to see if mean length or mean weight at a certain point in the first year of life is strongly correlated with ultimate year-class strength. I did not hear this discussed at the Workshop.

Habitat degradation. During the Workshop, I heard a few people mention that over the years a number of stations had to be dropped because of environmental change associated with construction projects, arrival of *Hydrilla*, and movement of sandbars. Such losses may be important because there could be long-term loss of habitat that is not being quantified and which is, in fact, being disregarded by the process of dropping stations. One way to assess the effects of dropping stations is to plot the catch at each station versus year. If there are strong "station effects" (i.e., if stations tend to be consistently good or bad - see figure on page 7), then the value of the overall index will depend on the particular set of stations considered, and dropping stations will lead to potentially misleading results.

I suggest that it would be worthwhile to make an inventory of seinable beach sites and, more generally, of total striped bass yoy habitat. This need not be done in a single year. For example, each year a few segments of the River or Bay could be selected and surveyed by looking at maps, making aerial photos, and/or surveying the coastal access points by car or by boat. The potential benefits are numerous: learning more about striped bass distribution, monitoring for long-term environmental degradation, compiling data useful for assessing damage caused by pollution incidents (oil spill, toxic discharges, etc.).

Triggers: Recognizing that an objective criterion or set of criteria is required to "trigger" the opening of a fishery, I would suggest that a compound trigger would be more appropriate than a single criterion. Thus, I would trigger the opening of the fishery if both the mean and median change in catch is "up" (in some sense) and the increase shows geographic cohesiveness (e.g., a certain percentage of the stations in a large region show increases). This helps guard against the possibility of a few "freak" observations triggering a management action of biological significance when the biological basis for the decision is weak. Furthermore, the conclusions drawn should not depend critically on how the index was formlulated. This implies that several indices should be computed and the indices should be in general agreement in order to trigger the opening of a fishery. I would also say that a trigger to open the fishery should only be considered when there is a well specified list of conditions each of which would trigger the closing of the fishery.

The issue of spatial cohesiveness of an apparent increase in abundance is an interesting question from a statistical point of view. It would be worthwhile to develop analytical approaches to study spatial patterns of change in abundance. I suggest one approach here. Since the sampling programs are shore-based, it often will be reasonable to think of the sampling stations as occurring along a one-dimensional axis rather than occurring in two-dimensional space. For the one-dimensional case, one can assign a score of "+" to those stations showing an increase over the previous year and a "-" to those showing a decrease. The null hypothesis of no cohesiveness can be expressed as: H_0: pluses occur at random along the shoreline versus H_1: pluses occur in clumps. This can be tested with a one-tailed runs test.

Improving precision. The ability to make correct management decisions is hampered by the imprecision of the available information. A number of statistical approaches are available to reduce or deal with uncertainty. In Appendix 1 an empirical Bayes approach is described in which the estimate of the index for the current year is combined with estimates from previous years as a weighted mean. If the current estimate is good (has low variance) and is within the realm of experience, the current estimate is given a high weight and the estimate is changed only slightly. However, if the current estimate is outside the realm of experience and has poor precision, then this estimate is not very believable and one would do better to shrink the estimate closer to the historical mean. This procedure can result in an estimator with greatly improved mean squared error. Other techniques that might be tried involve using model-based sampling (e.g., Hoenig et al. 1989), using regression estimators (Hoenig and Heywood 1991), and using outlier analysis.

Appendix 1: an empirical Bayes approach to a time series of indices

Abstract

In repeated, large-scale fish surveys, such as the striped bass young of the year index programs, there may be a great deal of auxiliary information available which can be used to obtain improved estimates. One approach is to use empirical Bayes estimators in which estimates from previous years or from surrounding regions are assumed to represent a random sample from the prior distribution (i.e., a random sample from the statistical distribution from which the current estimates were drawn). We describe and illustrate the use of an estimator which requires minimal assumptions. The low reliance on assumptions should overcome the objections of some researchers to the use of a Bayesian type of estimator.

Introduction

For many large-scale fish surveys, such as the striped bass young of the year index programs, there may be a great deal of auxiliary information available which can be used to improve the estimates of abundance. Suppose, for example, that for the past six years survey estimates of the average catch of a young of the year (yoy) striped bass have ranged from 100 to 150 fish per seine haul, and in the current year the estimate is 300 fish/haul. Should we accept the current estimate? If not, how can we use the auxiliary information to "improve" (in some sense) the current estimate. If the current survey was performed properly and the sample size was large, then we could easily accept the current estimate. But, if the current estimate is based on a very small sample size (say, for reasons external to the fishery, e.g., because of an across the board budget cut), then one would be tempted to use the information from previous years to improve or adjust the current estimates. This paper considers one approach to the problem of combining current information with auxiliary or "prior" information in order to obtain improved estimates. It should be noted that the term prior information is used in the statistical sense and thus does not have to come from earlier periods of time. For example, information from geographically similar areas could be considered as prior information.

The approach we will use is a type of empirical Bayes estimator (see Johnson [1989] for a very readable introduction to the empirical Bayes approach; see Efron and Morris [1972 and 1975] and Morris [1983] for basic theory.) We will begin by reviewing classical Bayesian estimation and then show how the subjectivity associated with Bayesian estimation can be overcome by the empirical Bayes approach. We also show that the empirical Bayes estimator can be written in the form of a weighted mean of the current estimate and the prior estimates. This makes the estimator more intuitive. We illustrate the method by applying the estimator to some yoy striped bass data from Maryland.

Classical Bayesian Estimation

In classical Bayesian estimation, the parameter to be estimated, θ, is viewed as being a random variable from some statistical distribution. For example, θ could be the catch rate and the value θ takes on might depend on the weather during the year. The researcher specifies a prior distribution for θ, then collects some current information about θ and uses the current information to update the prior distribution. The result is called the posterior distribution of θ. The updating is accomplished by way of Bayes Theorem. More formally, let the prior density of θ be denoted by $h(\theta)$, the posterior density of θ be denoted as $\psi(\theta|X)$, and the conditional density of the data given θ (or the likelihood of the data) be denoted by $f(X|\theta)$. (Here, X is a sufficent statistic of the sample observations such as the sample mean.) Then, by Bayes Theorem,

$$\psi(\theta|X) = \frac{f(X|\theta)\, h(\theta)}{g(X)} \qquad [1]$$

where $g(X)$ is the marginal density of the statistic X,

$$g(X) = \int f(X|\theta)\, h(\theta)\, d\theta \ .$$

Note that $g(X)$ does not depend on θ and, hence, is simply a constant that depends only on the sample observations. Once the posterior distribution of θ is found, there are a variety of ways to estimate θ. A common estimator, $\hat{\theta}_B$, is the mean of the posterior distribution. Thus,

$$\hat{\theta}_B = \frac{\int \theta\, f(X|\theta)\, h(\theta)\, d\theta}{\int f(X|\theta)\, h(\theta)\, d\theta} \ . \qquad [2]$$

It can be shown that this estimator minimizes the expected value of the square of the error $(E(\hat{\theta}_B - \theta)^2)$.

Many people are uncomfortable with having to specify a prior distribution for θ. Some people feel that the subjectivity involved with specifying a prior makes the approach inappropriate for scientific research. It is possible, however, to estimate the prior distribution of θ from prior data. This approach should be more acceptable to those who worry about subjectivity in the analysis of survey data.

An Approach to Empirical Bayes Estimation

The method presented here was described by Krutchkoff (1972). It is not the most modern approach nor is it necessarily the most efficient estimator. However, the estimator is rather easy to derive and it avoids not only having to specify the values of the parameters of the prior distribution, but also it frees the researcher from having to specify the <u>form</u> of the prior distribution.

Let us denote the estimate of catch rate in the current year, based on just the current year's data, by $\hat{\theta}_1$. The estimate from the previous year is denoted by $\hat{\theta}_2$, from the year before that by $\hat{\theta}_3$, and so on. Each $\hat{\theta}_i$ is an estimate of the actual catch rate in year i, θ_i. We consider the k values of θ_i as a random sample from the prior distribution of θ. Consequently, the k values of θ_i could be used to obtain an empirical characterization of the prior distribution of θ if the θ_i were known. We will simply estimate the prior distribution of θ by using the values of the estimates $\hat{\theta}_i$.

Thus, the prior density of θ, $h(\theta)$, is taken to be the k values of θ_i, each occurring with a probability mass of 1/k. The empirical Bayes estimator for the catch in the current year, $\hat{\theta}_{1EB}$, can then be obtained from equation [2] as

$$\hat{\theta}_{1EB} = \frac{\sum_{i=1}^{k} \hat{\theta}_i f(\hat{\theta}_1 | \hat{\theta}_i) \frac{1}{k}}{\sum_{i=1}^{k} f(\hat{\theta}_1 | \hat{\theta}_i) \frac{1}{k}}$$

$$= \frac{\sum_{i=1}^{k} \hat{\theta}_i f(\hat{\theta}_1 | \hat{\theta}_i)}{\sum_{i=1}^{k} f(\hat{\theta}_1 | \hat{\theta}_i)} \quad . \qquad [3]$$

Similarly, the empirical Bayes estimator for the previous year is the same as equation [3] with a 2 replacing each occurrence of 1 in the subscripts, and so on for each of the preceding years.

It remains only to specify the form of the conditional density f. This density essentially says, if θ_1 is actually $\hat{\theta}_i$, how likely would it be to obtain the value $\hat{\theta}_1$ as our estimate. We will assume that estimates of θ are normally distributed, i.e. that f is a normal density function. In the case of a beach-seine survey, the statistic of interest may be a sample mean. Then, by the central limit theorem, the sample mean will tend to a normal distribution as the sample size becomes large. This justifies assuming f is normal. It is also necessary to obtain a value for the variance of the normal density f. We use the sample estimate of the variance of $\hat{\theta}_i$, S_i^2, for this.

Example

We consider data from 1976 to 1991 from the Maryland juvenile striped bass survey. The goal is to obtain improved estimates of the index for each year. Here, we will compute only the estimate for 1989 (year 3) for illustrative purposes. The index in 1989 was much higher than in any other year (i.e., outside the realm of experience) so one might ask "should we accept such a high estimate as our best estimate for 1989?".

The necessary data for the estimator [3] are simply the annual estimates of the means and standard errors of the means (Table 1, cols. 2 and 3). From this, one can calculate the conditional densities $f(\hat{\theta}_3|\hat{\theta}_i)$ (col. 4), where

$$f(\hat{\theta}_3|\hat{\theta}_i) = \frac{1}{\sqrt{2\pi}\, S_i} \exp\left[- \frac{(\hat{\theta}_3 - \hat{\theta}_i)^2}{2S_i^2} \right] .$$

The marginal density of the data is calculated by summing the entries in column 4 and dividing by k = 16 years. (However, we need not actually divide by k since the k's cancel in equation 3.) The sum of the entries in column 5 is .05883. Next, we calculate the weights for each year by dividing the entries in column 4 by the sum of the entries in column 4. Finally, the estimator for mean catch per haul in 1989 is simply the weighted sum of the estimates for each year, i.e., the sum of the products of the entries in columns 2 and 5. The empirical Bayes estimate for mean catch per haul in 1989 is thus 25.174.

In this example, the weights assigned to each year other than 1989 are close to zero. In essence, the situation in 1989 is clearly so different from anything experienced before (or since) that the prior information is of virtually no value in determining what is happening in 1989. The estimate for 1989 remains unchanged. This result is due to the fact that the estimated standard errors are small so that 1989 stands out as clearly different from the other years.

I also computed the empirical Bayes estimates for the first three years and these show slight changes over the original estimates as shown below:

year	original estimate	empirical Bayes estimate
1976	4.89	4.52
1977	4.85	4.54
1978	8.45	8.48

This example did not turn out as expected - we didn't see much shrinkage towards the mean when the data are uncertain. However, it does point out that if the variance estimators are not appropriate then the technique is not meaningful.

Discussion

The Empirical Bayes Estimator as a Weighted Mean

Equation [3] can be written as

$$\hat{\theta}_{1EB} = \sum_{i=1}^{k} W_i \, \hat{\theta}_i$$

where

$$W_i = \frac{f(\hat{\theta}_1 | \hat{\theta}_i)}{\sum_{i=1}^{k} f(\hat{\theta}_1 | \hat{\theta}_i)} .$$

This shows that the empirical Bayes estimator is a weighted mean of the estimate for the current year and the estimates from previous years. Recall that

$$f(\hat{\theta}_1 | \hat{\theta}_i) = \frac{1}{\sqrt{2\pi}\, S_i} \exp\left[- \frac{(\hat{\theta}_1 - \hat{\theta}_i)^2}{2S_i^2} \right] .$$

From this it can be seen that W_i will be large when S_i^2 is small and when the difference between the current estimate and the ith estimate $(\hat{\theta}_1 - \hat{\theta}_i)$ is small. Thus, if the variance in the current year, S_1^2, is small (indicating there is a great deal of information about the current year), the weight given to the current estimate is likely to be the largest weight. Prior years get high weights if they both agree closely with the current estimate and have low estimated variance.

What if Prior Estimates Are Not a Random Sample?

It may happen that the previous observations (estimates) do not constitute a good description of the prior distribution from which the current value of θ arose. For example, suppose there is a strong increasing trend over time in the 6 previous estimates of mean catch per haul. One would not expect to draw a random sample of size 6 from a distribution and have the observations appear in perfect ascending order. Thus, one would have serious doubts about the validity of the empirical prior.

It is worth considering how the estimator behaves when there is a trend in the previous estimates. Suppose the variance is constant over all years, i.e., $S_i^2 = S^2$, all i. Then, clearly, year 1 agrees with year 1 the best and gets the highest weight; year 2 agrees with year 1 the second best and gets the second hightest weight, etc. Thus, as you go further back in time, the estimates have less and less relevance to the current situation and they receive progressively less weight. The old estimates will pull the current estimate down a bit. However, since they don't agree well with the current data they receive low weights thus affording a measure of protection against the failure of the assumption.

Krutchkoff (1972) points out that there is a law of diminishing returns in that, as you add more and more years of prior information, the improvement in the mean square error diminishes. He suggests as a rule of thumb that there may not be much point in using more than 15 years of data. By limiting the number of years of prior information, one also lessens the chances that an underlying trend in the annual values of θ will cause problems.

Potential Application to Striped Bass YOY Indices

The method above pertains to the case where there exists a series of estimates of an index of abundance along with valid estimates of the precision (variance) of the estimates. Thus, this method does not appear to be appropriate for the fixed station design because of the lack of suitable estimates of the variance. Appropriate variance estimates might be obtainable if one can make certain assumptions about the sampling process. In this case, the approach may be very useful for reducing "noise" or uncertainty about the index estimates. I computed empirical Bayes estimates for a portion of the Maryland data to show estimates are pulled in towards the mean when they are outside the realm of experience and/or they have high uncertainty (variance). Unfortunately, the example wasn't very impressive because the estimates hardly changed at all. This is due to the fact that the nominal estimates of variance were not very large. Thus, although the estimate for 1989 was outside the realm of experience (higher than in any other year before or since 1989) the standard error was small enought that we must accept that the index was so high. However, as noted elsewhere, it is not clear that these estimates of variance are appropriate.

Table 1. Example of the use of the empirical Bayes estimation scheme to estimate mean catch per seine haul in year 3 (1989). Data are from the Maryland young-of-the-year striped bass index of abundance program. Year 1 is 1991, year 2 is 1990, and so on.

year, i	estimate, $\hat{\theta}_i$	std error, S_i	$f(\hat{\theta}_3 \mid \hat{\theta}_i)$*	W_i**
1 (1991)	4.439	0.787	$< 10^{-43}$	$< 10^{-43}$
2	2.130	0.328	$< 10^{-43}$	$< 10^{-43}$
3	25.174	6.783	0.05883	1.0000
4	2.667	0.724	$< 10^{-43}$	$< 10^{-43}$
5	4.803	1.096	$< 10^{-43}$	$< 10^{-43}$
6	4.144	0.982	$< 10^{-43}$	$< 10^{-43}$
7	2.932	0.646	$< 10^{-43}$	$< 10^{-43}$
8	4.205	0.835	$< 10^{-43}$	$< 10^{-43}$
9	1.364	0.320	$< 10^{-43}$	$< 10^{-43}$
10	8.447	1.177	$< 10^{-43}$	$< 10^{-43}$
11	1.212	0.241	$< 10^{-43}$	$< 10^{-43}$
12	1.955	0.297	$< 10^{-43}$	$< 10^{-43}$
13	3.992	0.633	$< 10^{-43}$	$< 10^{-43}$
14	8.447	1.070	$< 10^{-43}$	$< 10^{-43}$
15	4.848	0.997	$< 10^{-43}$	$< 10^{-43}$
16 (1976)	4.894	1.194	$\leq 10^{-43}$	$\leq 10^{-43}$
		sum	0.05883	sum 1.0000

*) $$f(\hat{\theta}_3 \mid \hat{\theta}_i) = \frac{1}{\sqrt{2\pi}\, S_i} \exp\left[- \frac{(\hat{\theta}_3 - \hat{\theta}_i)^2}{2S_i^2} \right] .$$

**) $$W_i = \frac{f(\hat{\theta}_3 \mid \hat{\theta}_i)}{\sum_{i=1}^{k} f(\hat{\theta}_3 \mid \hat{\theta}_i)} .$$

Appendix 2: Literature cited

Efron, B. and C. Morris. 1972. Limiting the risk of Bayes and empirical Bayes estimators - Part II: the empirical Bayes case. Journal of the American Statistical Association 67:130-139.

Efron, B. and C. Morris. 1975. Data analysis using Stein's estimator and its generalizations. Journal of the American Statistical Association 70:311-319.

Heimbuch, D.G. and J.M. Hoenig. 1989. Change-in-ratio Estimators for Habitat Usage and Relative Population Size. Biometrics 45:439-451.

Hoenig, J.M. 1990. Estimating the Proportion of a Stock in each Area (Region) when the Regional Catchabilities are Unknown but the Stock Redistributes in Space over Time: the Case where Mortality Occurs. pp. 83-85 *in* Report of the Working Group on Methods of Fish Stock Assessments, Nantes, France, 10-17 November, 1989. Int. Coun. Explor. Sea C.M. 1990/Assess:15.

Hoenig, J.M. and C.M. Heywood. 1991. Using Model Predictions as an Auxiliary Variable to Reduce Variance in a Creel Survey. Amer. Fish. Soc. Symp. 12:292-297.

Hoenig, J.M., F.B. Martin and M. Heywood. 1989. Model-Based Sampling Methods for Effort and Catch Estimation. *In* E. Edwards and B. Megrey (eds.), Mathematical Analysis of Fish Stock Dynamics. Amer. Fish. Soc. Sympos. 6:181-189.

Johnson, D.H. 1989. An empirical Bayes approach to analyzing recurring animal surveys. Ecology 70:945-952.

Krutchkoff, R.G. 1972. Empirical Bayes estimation. American Statistician 26(5):14-16.

Morris, C. 1983. Parametric empirical Bayes inference: theory and applications (with discussion). Journal of the American Statistical Association 78:47-65.