PN-ABK-244
73229

WORKING PAPER SERIES

FISHERIES STOCK ASSESSMENT

TITLE XII
Collaborative Research Support Program

Fisheries Stock Assessment CRSP Management Office
International Programs, College of Agriculture
The University of Maryland, College Park, Maryland 20742

In cooperation with the United States Agency for International Development (Grant No. DAN-4146-G-SS-5071-00) the Fisheries Stock Assessment CRSP involves the following participating institutions:

The University of Maryland—Center for Environmental and Estuarine Studies
The University of Rhode Island—International Center for Marine Resource Development
The University of Washington—Center for Quantitative Sciences
The University of Costa Rica—Centro de Investigación en Ciencias del Mar y Limnología
The University of the Philippines—Marine Science Institute (Diliman)—College of Fisheries (Visayas)

In collaboration with The University of Delaware; The University of Maryland—College of Business and Management; The University of Miami; and The International Center for Living Aquatic Resources Management (ICLARM).

WORKING PAPER SERIES

WORKING PAPER NO. 55

DEVELOPMENT OF A SAMPLING EXPERT SYSTEM: FISHMAP

by
B.J. Rothschild, C.T. DiNardo
and M. Bhandary

University of Maryland

June, 1989

Fisheries Stock Assessment
Title XII
Collaborative Research Support Program

The Fisheries Stock Assessment CRSP (sponsored in part by USAID Grant No. DAN-4146-G-SS-5071-00) is intended to support collaborative research between the U.S. and developing countries' universities and research institutions on fisheries stock assessment and management strategies.

This Working Paper was produced by The University of Maryland-Center for Environmental and Estuarine Studies and The University of Costa Rica-Centro de Investigacion en Ciencias del Mar y Limnologia (CIMAR) in association with The University of Delaware and The University of Miami. Additional copies are available from the CRSP Management Office and from:

1454

Chesapeake Biological Laboratory
Center for Environmental and Estuarine Studies
The University of Maryland
P.O. Box 38
Solomons, Maryland 20688

University of Maryland
Center for Environmental and Estuarine Studies
Chesapeake Biological Laboratory
Solomons, MD 20688 USA

MD/SACRSP Working Paper 88/10

Development of a Sampling Expert System: FISHMAP

by

Brian J. Rothschild[1]
Gerard T. DiNardo[1]
Madhusudan Bhandary[1]

PREFACE

The Stock Assessment Collaborative Research Support Program is intended to support collaborative research between U.S. universities and institutions in the developing countries on stock assessment. The research is intended to build institutions and to improve capabilities in stock assessment, a set of techniques which will contribute to improving the economic and social efficiency of fish harvest. Three U.S. universities, the University of Maryland, the University of Rhode Island, and the University of Washington have primary responsibilities for working together with the University of Costa Rica and the University of the Philippines. The University of Maryland and the University of Washington are working with the University of Costa Rica. This document is an informal report, one of a series produced by a team of researchers at the University of Maryland in collaboration with the University of Costa Rica, the University of Miami, and the University of Delaware.

[1]University of Maryland, Center for Environmental and Estuarine Studies, Solomons, MD 20688 USA

SUMMARY

This report describes a baywide fishery independent sampling system for fish and shellfish. The system (FISHMAP) will have generalized applicability but will be targeted and operational on Chesapeake Bay waters of both Maryland and Virginia. This project will serve as a major component of the CHESFISH program in which we are presently engaged.

The primary rationale for developing the FISHMAP system is to provide a framework for identifying optimal sampling methodologies. Specifically, the system will provide a framework for identifying optimal trawl sampling procedures by 1) providing a strategy to determine the dimensions of the sampling units and their number, 2) providing strategies to efficiently allocate trawling effort among sampling units, 3) providing real-time methodologies to pilot the shipboard sampling process, and 4) providing a strategy to determine what types or classes of data should be sampled. In particular, the system will have the capability to define the temporal and spatial characteristics of the sampling process; where, when, and how often to sample. Also, once engaged in the field component of the sampling process, the system will be capable of determining where and when to fish the trawl based on the stream of real-time information being observed aboard the research vessel.

The system will be comprised of several components, data input, map model, decision-making meetings, operations research, and system coordination. The data input component houses pertinent data that will be used towards the development of the FISHMAP system. Three data sets reside in this component; historical trawling data (1960-1975), recent trawling data (1988), and real-time current data. The historical and recent trawl data sets comprise area and species specific estimates of CPUE and physicochemical data. The real-time current data will include data observed on board the research vessel during the actual sampling process (*i.e.*, hydroacoustic and physicochemical data). The data housed in the data input component will be used to construct n-dimensional maps of fish and shellfish distribution. This information will also be used to develop strategical and tactical sampling models. The strategic models will define the spatiotemporal characteristics of sampling; where and when to sample. Once on board the research vessel the tactical model will define when and where to deploy the sampling gear. Next, a roundtable management meeting will ensue to evaluate the program and to set the specifics of sampling; gear types, target-species, types of data to collect, the temporal and spatiotemporal characteristics of sampling, and sampling platforms. The statistical tradeoffs associated with the decisions will be evaluated. Once the specifics of the sampling process have been defined, the operations research component will assess the costs and benefits associated with collecting the data and outline optimal procedures to acquire the data (*i.e.*, vessel-routing, number of vessels to employ in the sampling and where to deploy the vessels). The coordination of all of the components will be facilitated through the use of an expert system.

The fundamental premise of the FISHMAP system is that the utilization of techniques in the data input, map models, decision making, and operations research components in the sampling of fish populations, integrated in the setting of an expert system represents a significant advance in the

application of these techniques to stock assessment. Integration of the system is made tractable through the use of expert system technology. Improvement in the application of these techniques is gained through the expert system facility for interpreting very large databases in approximately real-time.

INTRODUCTION - THE NEED FOR A COMPREHENSIVE SAMPLING SYSTEM

This report describes a baywide fishery-independent sampling system for fish and shellfish. The system (FISHMAP) is targeted to involve Chesapeake Bay waters of both Maryland and Virginia and to have general applicability in other areas. FISHMAP is a major component of the CHESFISH program[1].

The FISHMAP system will comprise an optimal, cost-effective, adaptive, real-time, sampling methodology to report on the dynamics of fish and shellfish in Chesapeake Bay. FISHMAP will provide basic data to evaluate the effects of the environment, anthropogenic activity and fishing, and to trigger appropriate management actions regarding these effects.

The foundation of any comprehensive management regime is an accurate assessment of population abundance (Gulland, 1983). The need for improved sampling procedures, and accompanying decision-making approaches in fishery management, has greatly increased in recent years. Part of this need arises from the socioeconomic importance of fisheries required for management planning. Part arises from the inadequacy of contemporary fishery models and sampling programs to generate accurate information for effective decision-making, particularly for the complex of problems in Chesapeake Bay. For instance, inferences concerning the status of stocks and the influence of the environment on Chesapeake Bay living resources are difficult to measure using

---

[1]CHESFISH is intended to be conducted cooperatively with the Virginia Institute of Marine Science (VIMS) to develop a baywide assessment of fish stocks. During the first year of CHESFISH the Maryland component completed the field phase of a localized pilot study to 1) identify efficient trawling gears, 2) identify sources of variability associated with trawling, and 3) begin construction of spatiotemporal maps of fish distribution. In the forthcoming and second year the prototype system will be made operational in sections of the Bay proximal to Solomons. In the third year the prototype system will be extended to the entire Bay with the idea of making the system operational in years 4 and 5.

only commercial fishery statistics. The problem is that the large number of fishermen in Chesapeake Bay who contribute to indices of even nominal effort challenges the feasibility of obtaining adequate estimates of stock abundance using the well known procedures. Furthermore, a large portion of the fish harvested in Chesapeake Bay is taken by recreational fishermen and statistics from this component of fishing mortality are either virtually impossible to obtain in a timely and cost effective fashion or contain such high sampling variance and biases as to make the estimate difficult to use. In addition, by sampling only that portion of the population recruited to the fishery, estimates of juvenile abundance cannot be obtained. This information is critical for making timely management decisions, especially when the population consists of dominant year classes.

Adding to the complexity of the sampling process is the fact that more than one species is caught simultaneously by a particular fishing gear. Very few fisheries are based solely on a single species. In practice, however, the interpretation of catch and effort data concerning a single species rarely accounts for effects resulting from interactions with other species. Also, the nature of many fisheries is such that catches of fish and shellfish are landed at a multitude of sites, both discrete and indiscrete. If catch and effort sampling does not account for all landing sites there is the potential for interpretations of these data to be biased. One obvious approach to solving these problems would be to increase the scope of sampling. However, with an increase in sampling comes an increase in costs.

Efforts to improve existing sampling methods, to estimate precise measures of population abundance, have lagged even in spite of many recent efforts concerned with modifying existing fishery models and developing new

modeling approaches (Rothschild, 1986; Golden *et al.*, 1987; Stagg, 1987; Walters, 1986; Mangel, 1986). In many areas where fish abundance has been measured, the estimates contain high sampling variance and biases, making them difficult to use with any confidence. In many cases, there is no consideration of optimal and most-cost effective sample design; sampling strategies are often chosen for either their consistency over the years regardless of immediate optimality, whether the estimates are biased, or imprecise, even though funds are available to generate a much more efficient sampling program. In resource-limited situations (constrained by budget), it is important to allocate sampling effort to maximize the return from programs designed to collect appropriate data. For example, decisions about where, when and how often to sample will determine how precise our estimates of abundance are. If these decisions are suboptimal then the resulting estimates are also suboptimal.

This report describes work plans, objectives, tasks, and anticipated products of the FISHMAP system.

WORK PLAN - THE FISHMAP SYSTEM

FISHMAP will provide a framework for identifying optimal trawl sampling procedures by 1) providing a strategy to determine the dimensions of the sampling unit and their number, 2) providing strategies to efficiently allocate trawling effort among sampling units, 3) providing real-time methodologies to pilot the shipboard sampling process, and 4) providing a strategy to determine what types or classes of data should be sampled. In particular, the system will have the capability to define the temporal and spatial characteristics of the sampling process; where, when, and how often to sample. Also, once engaged in the field component of the sampling process,

the system will be capable of determining where and when to fish the trawl based on the stream of real-time information being observed aboard the vessel (*i.e.*, hydroacoustic information, physicochemical data, catch in previous trawl, and distance from last trawl). In addition, the system will be capable of identifying which data needs to be collected to arrive at a more precise estimate of standing stock.

The design of the FISHMAP system is presented in Figure 1. The system is comprised of several components, data input, map model, decision-making meetings, operations research, and system coordination. These work as follows. The system coordination component, the "brain" of FISHMAP, is an expert system which integrates the other components. The expert system will be capable of making decisions concerning where, when, how, and what to sample based on information residing in each of the other components. Procedures to formulate the decisions will be in the form of rules, IF-THEN statements.

Data residing in the data input component are used as input into the map model component. This component houses both strategic and tactical models to identify where and when to sample, prior to actual field sampling, and where and when to fish the trawl when in the field. If, for example, a fishery manager wants to conduct field sampling next month, the strategic model is used to define the spatiotemporal characteristics of the sampling process based on prior knowledge of the distribution of the targeted species. Put another way, the strategic model will indicate where (location) sampling should be conducted. Procedures in the operation research component will then be used to map the route of the vessel in the most-cost effective manner. In addition, the optimal number of vessels to use in the sampling process will be identified based on vessel operation costs. Once in the field at a particular sampling site, the tactical model will be employed to pilot the shipboard

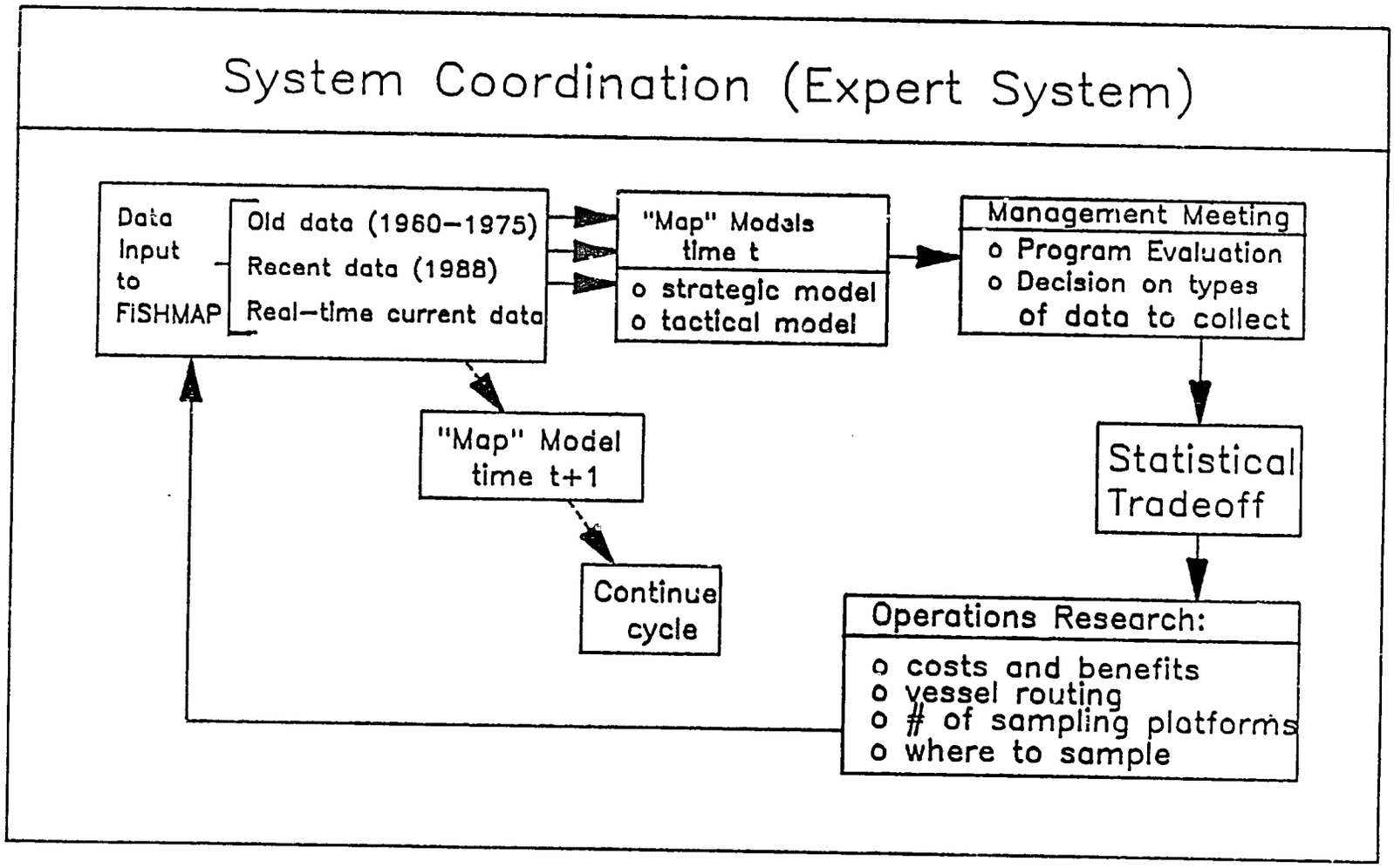


Figure 1. Components of the FISHMAP system.

sampling process. This model will be used on board the vessel to decide when to fish and trawl based on the stream of real-time data being collected. Put another way, this model will predict the catch of species i at a subsequent multidimensional time, T+1, based on hydroacoustic, physicochemical, and catch information observed at the multidimensional time T. If the prediction is "good" the trawl is not fished and biomass is measured via hydroacoustics. If the prediction is "poor" the trawl is fished. For example, if the vessel is steaming along a transect, data being collected is fed directly into the shipboard portable computer and real-time catch predictions generated. If the catch prediction for the targeted species is good then biomass estimates via hydroacoustics are generated. If the prediction is poor the trawl is fished.

It may be found that some of the data being collected on board the vessel is too "expensive" to collect or that some locations just cannot be sampled. In these cases, the decision-making component will define the statistical tradeoffs associated with changes in our sampling and allow for discussions regarding these changes.

The fundamental premise of the FISHMAP system is that the utilization of techniques in the data input, map models, decision-making, and operations research components in the sampling of fish populations, integrated in the setting of an expert system represents a significant advance in the application of these techniques to stock assessment. Integration of the system is made tractable through the use of expert system technology. Improvement in the application of these techniques is gained through the expert system facility for interpreting very large databases in approximately real-time.

An operational prototype of the FISHMAP system will have been developed by November 30, 1989. A detailed description of each component of the system

and anticipated products follows. In addition, preliminary results of year 1 will be discussed as they pertain to components of the work plan.

Data Inputs to FISHMAP

The data input component houses pertinent data that will be used towards the development of the FISHMAP system. Three data sets reside in this component; historical trawling data (1960-1975), recent trawl data (1988), and real-time current data. In order to facilitate analysis of this data, a database will be compiled. The following is a description of the data and the proposed database.

Data Analysis

Three data sets comprise the data input component; the historical trawl data (1960-1975), the recent trawl data (1988), and the real-time current data. Data include trawl generated area and species specific estimates of CPUE and physicochemical data from the historical and recent trawl data sets. The real-time current data will include data observed on board the research vessel during actual sampling (*i.e.*, hydroacoustic signals and physicochemical information).

Because there was no consistent set of trawl stations sampled, historical trawl stations have been grouped into eleven areas based on physical proximity (Figure 2). Each area was further subdivided into 3 depth zones, 0-30 ft., 31-50 ft., and $\geq$ 51 ft. For each area and depth zone combination the monthly number of trawls fished has been tallied. For example, the monthly tallies for Area 6 are shown in Figure 3. In addition, for each area, month, depth zone, and species combination relative abundance indices and mean length of catch has been estimated. Figure 4 shows relative abundance for striped bass in Area 6 and depth zone 0-30 ft.

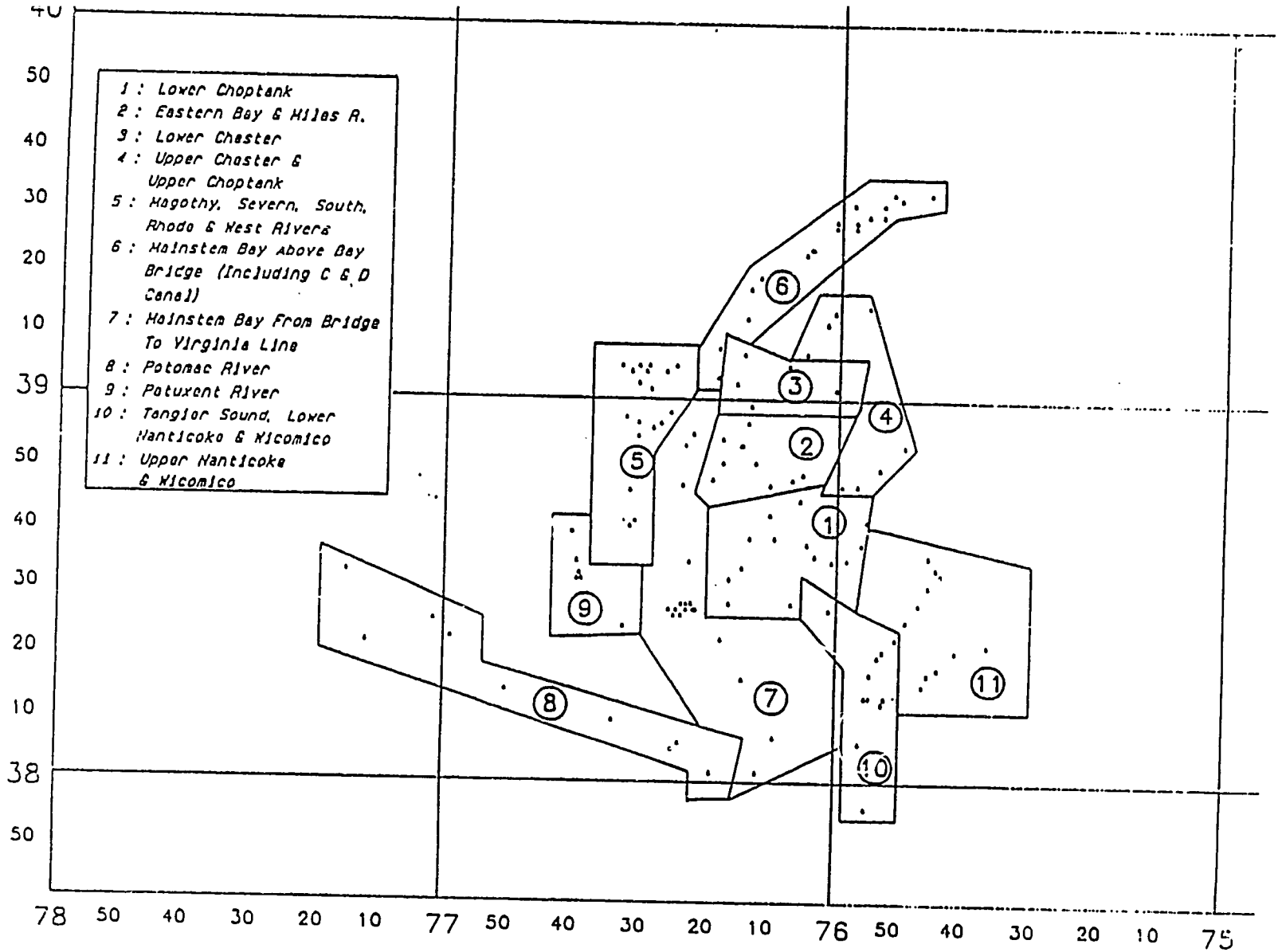


Figure 2. Location of the historical trawl stations and zones.

SAMPLING (ZONE 6) : 1965–1975
(DEPTH : 0–30 FEET)

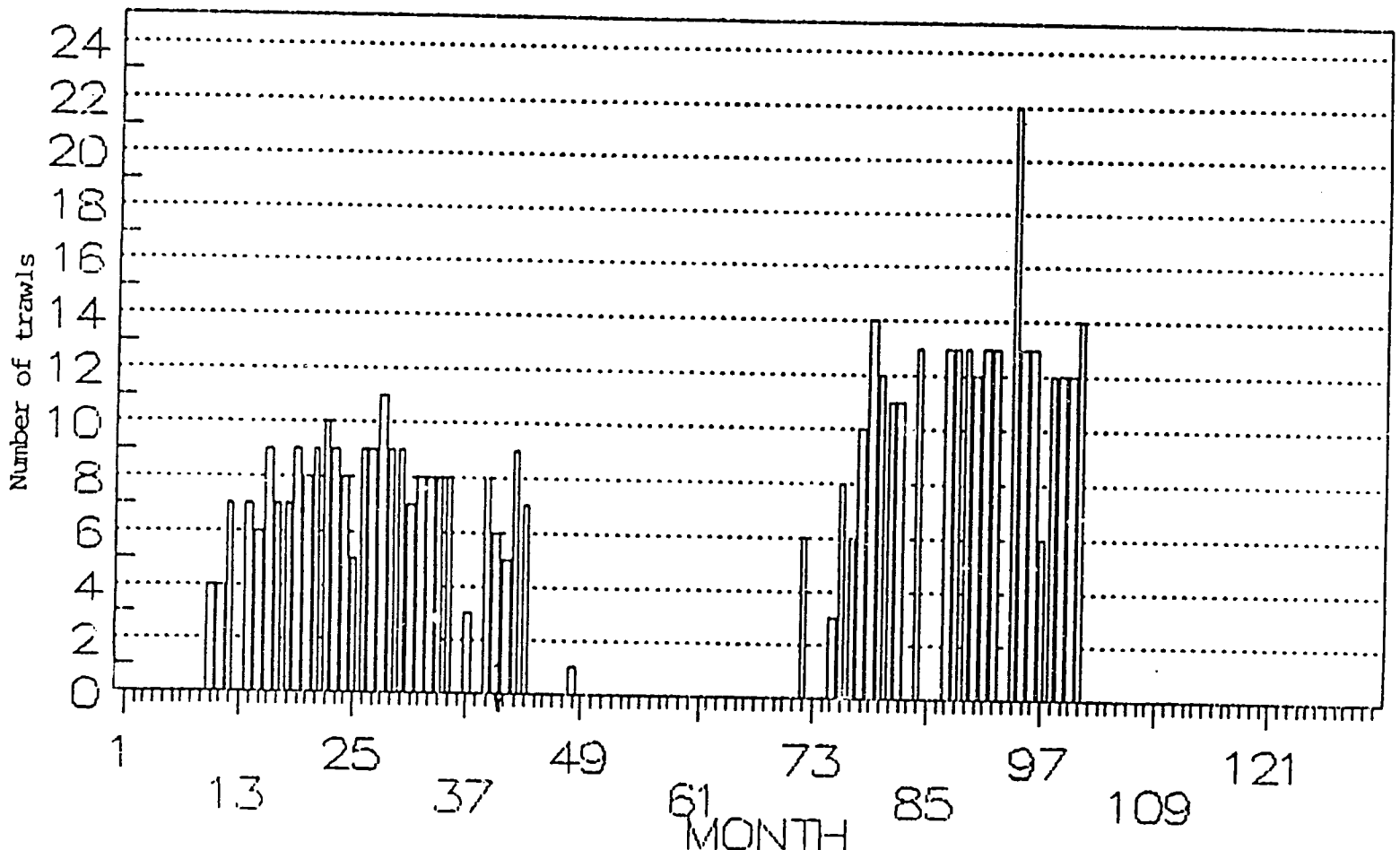


Figure 3 . Number of trawls fished per month in Zone 6 at a water depth of 0-30 feet between 1965 and 1975. (month 1 = January 1965, month 132 = December 1975).

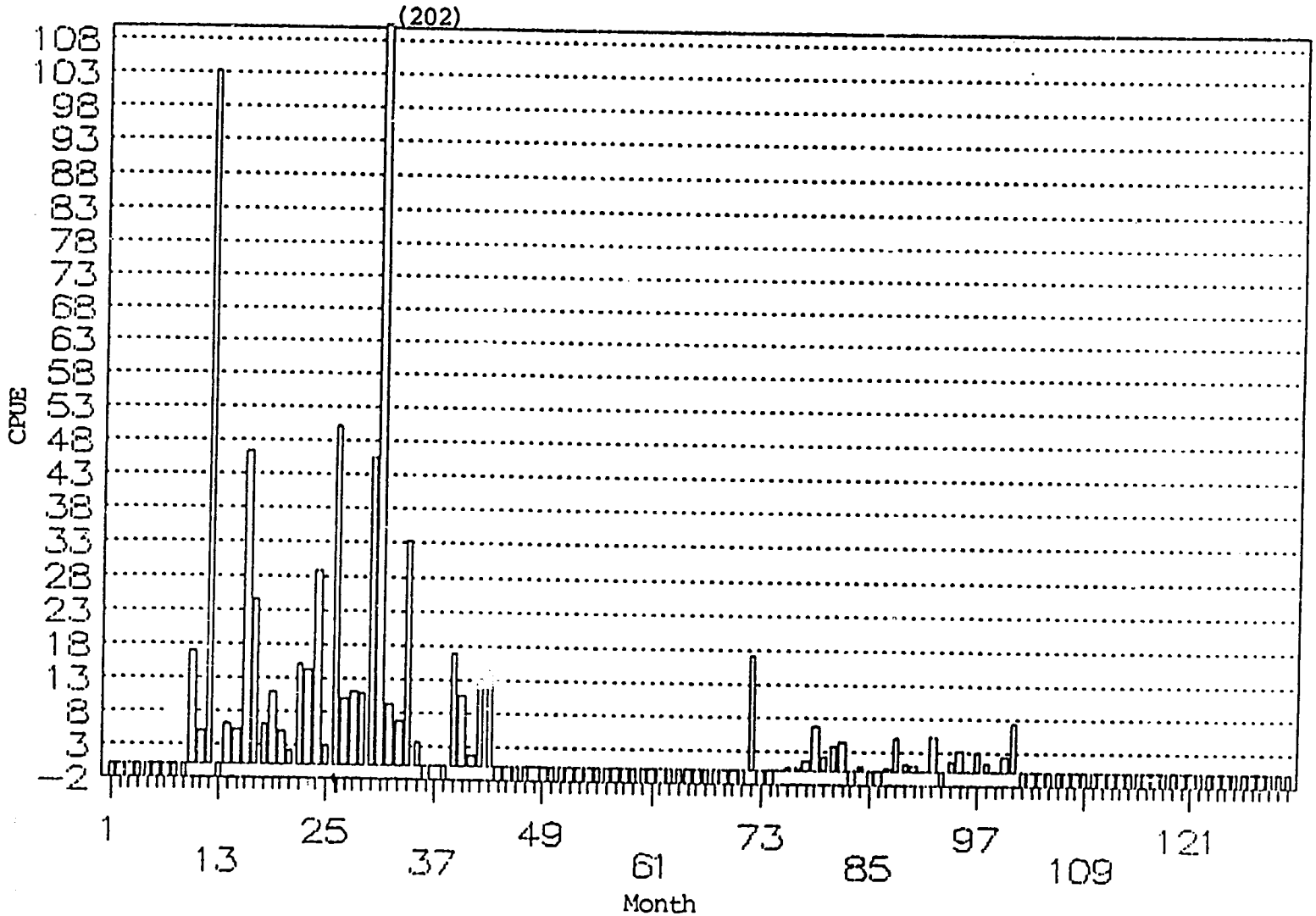


Figure 4. Monthly average CPUE (number caught/min. of trawling) for striped bass in Zone 6 from trawls fished at a water depth of 0-30 ft. between 1965 and 1975 (month 1 = January 1965, month 132 = December 1975). Bars below zero line on y-axis indicate months when no trawls were fished. Bars equal to zero indicate that trawling was conducted but no striped bass were caught.

Recent trawl data (1988) includes information on species density in relation to location, time of year and environmental parameters. Monthly trawling was conducted in the Patuxent River and adjacent Chesapeake Bay transects in the vicinity of Solomons, Maryland. This data provides preliminary distributional information needed to define the temporal and spatial characteristics of the sampling process.

Since the data describe distribution in space and time, the data also describe population dynamics in space and time. This will enable us to study how environmental variables, Chesapeake Bay "clean-up" measures, or stock enhancement measures have affected the population of a specific species in the Chesapeake Bay.

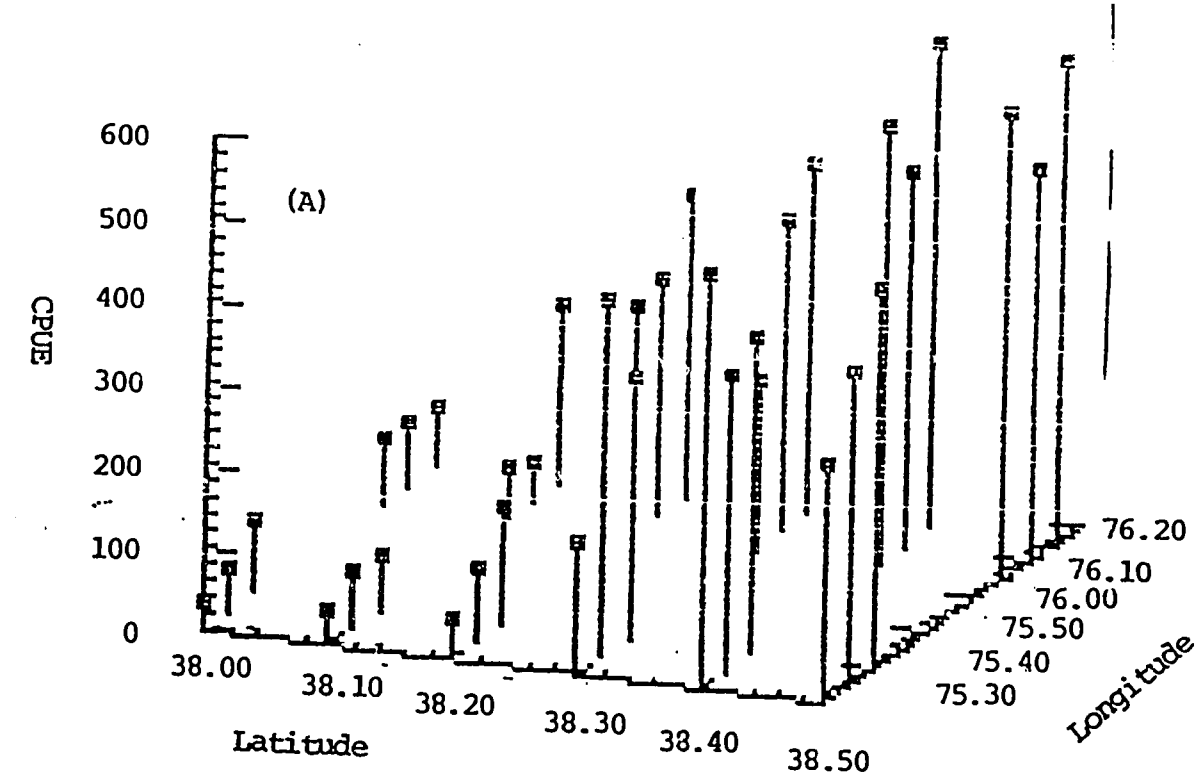
Database Management

The literally vast amount of raw data anticipated in this program necessitates the use of software which will allow us the fastest, most efficient access possible. D-Base IV (Ashton-Tate, 1988) was the chosen software for this database for two reasons. First, it is highly compatible with various software packages currently being used to load the raw data, such as Quattro (Borland, 1987) and Lotus 1-2-3 (Lotus, 1988). This factor will make conversion of the current database structure to a new, more efficient structure relatively uncomplicated. Second, D-Base IV is a true relational commercial database. This is crucial because we intend to build the new database in a fully relational format. A relational format is important because it allows for data to be accessed, condensed, aggregated, and otherwise manipulated across multiple files through the use of various data attributes which are common to these files. This allows large data sets to be broken into smaller, more manageable files which can then be handled

efficiently. These attributes are known as keys, and are classified as primary, secondary, etc. The nature of the ten years of compiled historical data and the data we are presently compiling lends itself to this specific type of database structure; for example, each trawl has a unique haul number which appears in every record and geographic regions are identified by unique station numbers which are also in every record. Documentation pertaining to the databases architecture and characteristics of the data files will be generated. This will make the data accessible to a wide range of users.

Map Models

Historical and recent trawl information will be used to generate n-dimensional "maps". These maps will show the relationship between species density and location, time of year, and environmental variables for a particular species. An example of 3-dimensional maps for spot, generated from the recent data is shown in Figure 5. The maps indicate the occurrence and density of spot in June and July at specific locations in Chesapeake Bay. This information is crucial in defining future temporal and spatial characteristics of the sampling process. For example, if we were interested in sampling for spot in August, we could use these maps to identify where we should sample. Thus we can use these maps to forecast future spatiotemporal sampling characteristics. It should be noted, however, that the n-dimensional maps generated from historical trawl data (1960-1975) will probably be highly inefficient for forecasting in terms of statistics. This is because there was no consistent trawling protocol followed nor was there any consistent set of trawl stations sampled. Thus, to study the dynamics of fish stocks trawling stations were pooled.



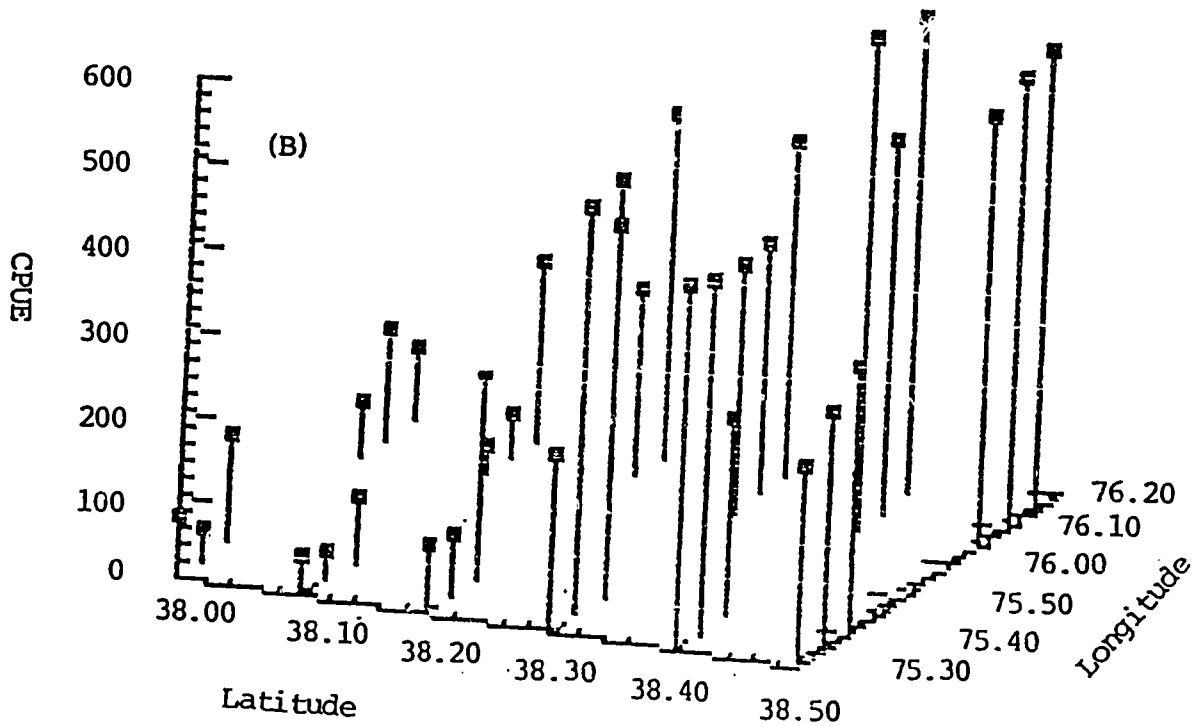


Figure 5. Hypothetical 3-dimensional map for spot. (A) - June, (B) - July.

Two sampling models will be developed, a strategic and a tactical model. These models will aid in the identification of optimal sampling strategies. Specifically, the strategic model will be used prior to actual field sampling (*i.e.*, in the office) to define the spatial and temporal characteristics of the sampling process; where and when to sample. The tactical model will be used on board the research vessel to decide when to fish the trawl based on characteristics of the stream of real-time data being observed. Data input for the strategic model will include the historical and recent n-dimensional map data. Data for tactical model will comprise the real-time current data.

Strategic Model

The strategic model will be used prior to actual field sampling to identify where to sample for a particular species. For example, based on prior spatiotemporal trawl data and physicochemical data, the strategic model will predict where we are likely to catch species i, and the biomass of the catch, at some future multidimensional time T+1. If the expected catch at the multidimensional time T+1 is zero, we do not sample; if it is greater than zero it is a potential sampling area. Put another way, based on past data for a particular species this model will indicate where to sample and at what month to sample.

In situations where sampling is constrained by monetary factors, knowing which areas are void of certain species, perhaps due to unfavorable environmental conditions, is important if we are concerned with optimality. In other words, it is inefficient to sample in areas where fish and shellfish are not found. For example, preliminary results from the 1988 trawl survey indicate that during the summer months fish and shellfish were absent from those portions of the water column deeper than 40 feet. Concurrent with this observation were dissolved oxygen values of 1.0 ppm or less in this same

portion of the water column. Since one of our primary gears is a bottom trawl, sampling should be conducted in areas where the water depth is less than 40 feet.

As a first step towards the development of the strategic model a multiple regression model has been formulated to predict future spatiotemporal sampling characteristics. Let the multidimensional time $T=(t, s_1, s_2, i)$, where t=time (i.e. day, month, etc.), s_1=latitude, s_2=longitude and i=species. Then, we can write the predictive catch model as,

$$W_T = W_{t-1} + B_1 X_{T_1} + B_2 X_{T_2} + \ldots + B_p X_{T_p} + E_T$$

Where W_T = total biomass at multidimensional time T;

W_{T-1} = biomass at previous multidimensional time T-1;

$B_1, B_2, \ldots, B_p$ are regression parameters to be estimated from past data;

$X_{T_1}, X_{T_2}, \ldots, X_{T_p}$ are depth, temperature, salinity, dissolved oxygen, etc. at multidimensional time T; E_T is random error at T.

By using regression techniques, it is easy to estimate the parameters.

Tactical Model

To pilot the shipboard sampling process, a tactical model has been formulated. This model will be used on board the research vessel to decide when to fish the trawl based on characteristics of the stream of real-time data being observed. Put another way, the model will predict the catch of species i at a subsequent multidimensional time, T+1, based on hydroacoustic, physicochemical and catch information observed at multidimensional time T. If the prediction is "good" the trawl is not fished and biomass is measured via

hydroacoustics. If, however, the prediction is "poor", the trawl is fished.

It is intended that this type of shipboard sampling approach will decrease sampling time. For example, if predictions are good for striped bass, it is conceivable that a baywide assessment of standing stock could be completed in less than one week.

The statistical formulation of the problem is as follows.

Let $T_j^{(i)} = (t, s_1, s_2)_j$ denote the multidimensional time indexed for species i, where t = time, s_1 = latitude, s_2 = longitude, j=1,2,...,n. Then biomass caught for species i at the spatiotemporal time j, $W_{T_j^{(i)}}$, can be expressed as a function of physicochemical parameters (*i.e.*, water temperature, salinity, and dissolved oxygen), hydrography, and previous catch. We can think of this data vector, $X_{T_j^{(i)}} = \begin{pmatrix} X_{1T}(i) \\ X_{2T}(i) \\ \vdots \\ X_{pT}(i) \end{pmatrix}$ as a set of observable signals where a subset of the signals are correlated in time. For example, catch of species i at the spatiotemporal time $T_j^{(i)}$ will most likely be correlated with the catch of species i observed at the spatiotemporal time $T_{j-1}^{(i)}$. We can express this model as:

$$W_{T_j^{(i)}} = \beta X_{T_j^{(i)}} + E_{T_j^{(i)}}$$

where $\beta = \begin{matrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{matrix}$ are regression parameters

and $E_{T_j^{(i)}}$ = random error for the model which is assumed to be autocorrelated. We can express the random error as the first order autoregressive error model

$$E_{T_j^{(i)}} = \phi E_{T_{j-1}^{(i)}} + N_{T_j^{(i)}}$$

where $|\phi| < 1$ and the $N_T(i)$ are independent $(N(0,\sigma^2)$.

We can estimate the regression parameters β and ϕ, by maximum likelihood procedures introduced by Cochran and Orcutt (1949). The solution procedure is as follows:

Solution Procedure

The log-likelihood conditional on E_{T_1}'s is of the form (E_{T_1} is the initial value)

$$\log L = -\frac{(n-1)}{2}\log\sigma_n^2 - \frac{1}{2\sigma_n^2}\sum_{j=2}^{n} n_{T_j}^2(i)$$

$$= -\frac{(n-1)}{2}\log\sigma_n^2 - \frac{1}{2\sigma_n^2}\sum_{j=2}^{n}(E_{T_j}(i) - \phi\, E_{T_{j-1}}(i))^2 \tag{3}$$

$$= -\frac{(n-1)}{2}\log\log\sigma_n^2 - \frac{1}{2\sigma_n^2}\sum_{j=2}^{n}[(W_{T_j}(i) - \beta' X_{T_j}(i)) \tag{4}$$

$$\phi(W_{T_{j-1}}(i) - \beta' W_{T_{j-1}}(i))]^2$$

The Cochran-Orcutt method proposes that, start with the ordinary least square estimator of β

$$i.e.\quad \hat{\beta}^o = \left(\sum_{j=2}^{n} \underline{X}_{T_j}(i)\, \underline{X}'_{T_j}(i)\right)^{-1}\left(\sum_{j=2}^{n} \underline{X}_{T_j}(i)\, W_{T_j}(i)\right) \tag{5}$$

Maximizing (3) w.r.t. ϕ, we get the estimator of ϕ

$$\text{as } \hat{\phi}^o = \frac{\sum_{j=2}^{n} E^o_{T_j}(i)\, E^o_{T_{j-1}}(i)}{\sum_{j=2}^{n}(E^o_{T_j}(i))^2} \tag{6}$$

$$\text{where } E^o_{T_j}(i) = W_{T_j}(i) - \hat{\beta}^{o\prime}\, \underline{X}_{T_j}(i) \tag{7}$$

The second improved esitmate of β is

$$\hat{\beta}(1) = \left(\sum_{j=2}^{n} \underline{X}^o_{T_j}(i)\, \underline{X}^{o\prime}_{T_j}(i)\right)^{-1}\left(\sum_{j=2}^{n} \underline{X}^o_{T_j}(i)\, W^o_{T_j}(i)\right) \tag{8}$$

$$\text{where } W^o_{T_j}(i) = W_{T_j}(i) - \hat{\phi}^o\, W_{T_{j-1}}(i) \tag{9}$$

$$\text{and}\quad \underline{X}^o_{T_j}(i) = \underline{X}_{T_j}(i) - \hat{\phi}^o\, \underline{X}_{T_{j-1}}(i) \tag{10}$$

and the second estimate of ϕ is

$$\hat{\phi}(1) = \frac{\sum_{j=2}^{n} E^{(1)}_{T_j}(i)\, E^{(1)}_{T_{j-1}}(i)}{\sum_{j=2}^{n}(E^{(1)}_{T_j}(i))^2} \tag{11}$$

where $E^{(1)}_{T_j}(i) = W_{T_j}(i) - \hat{\beta}(1)' \, \underline{X}_{T_j}(i)$ (12)

The procedure is continued until the log-likelihood in (3) or (4) attains its maximum.

In practice, we can go on for 10 to 15 iterations. After final iteration, let $\hat{\phi}$ and $\hat{\beta}$ be the final estimate of ϕ and β, respectively.

We can now estimate the predicted catch as

$$
\begin{aligned}
W_{T_{j+1}}(i) &= \hat{\beta}' \, \underline{X}_{T_{j+1}}(i) + E_{T_{j+1}}(i) \\
&= \hat{\beta}' \, \underline{X}_{T_{j+1}}(i) + \hat{\phi} \, E_{T_j}(i) \\
&= \hat{\beta}' \, \underline{X}_{T_{j+1}}(i) + \hat{\phi} \, (W_{T_j}(i) - \hat{\beta}' \, \underline{X}_{T_j}(i)) \\
&= \hat{\phi} \, W_{T_j}(i) + \hat{\beta}' \, (\underline{X}_{T_{j+1}}(i) - \underline{X}_{T_j}(i)) \qquad (13)
\end{aligned}
$$

where W_{T_j} is the previous catch and $\underline{X}_{T_j}$ and $\underline{X}_{T_{j+1}}$ are known observations and the estimates $\hat{\phi}$ and $\hat{\beta}$ are estimates we have obtained.

It should be noted that these models are preliminary and should be considered points of departure in the map modeling phase of FISHMAP. Therefore, efforts will center around the testing of these models using the historical trawl data, the 1988 trawl data and the 1989 trawl data.

Decision-Making Meetings

As with all sampling programs, budget constraints dictate the types and amount of data that can be collected. Thus, decisions must be made concerning the characteristics of the data. Periodic meetings will be held to evaluate the program and to decide on specifics of the sampling process; gear types, types of data to collect, the spatiotemporal characteristics of sampling, sampling platforms, and target species. In addition, the statistical tradeoffs associated with these decisions will also be discussed (see Appendix 1 for a discussion on statistical tradeoffs). Workshop participants will include Maryland Department of Natural Resources, Virginia Institute of Marine Science, and Chesapeake Biological Laboratory personnel.

To assist in this process, the Analytic Hierarchy Process (AHP), a decision analysis tool, will be used. First developed by Saaty in the early 1970's, this technique has gained wide acceptance in a number of diverse disciplines, especially when the decision problem involves the ranking of decision alternatives for evaluation, selection, or prediction (see Saaty (1982) and Zahedi (1986) for examples). A special feature of the AHP is its ability to combine judgements with quantitative measurements.

The AHP is a procedure that seeks to structure a complex decision problem hierarchically. Briefly, elements at each level of the hierarchy are compared to each other with respect to an element at the next higher level (ranked as to their importance). Using the "eigenvalue" method to estimate the relative

weights of the comparisons and then aggregating these relative weights, a set of scores for the decision alternatives is determined. The alternative with the highest score is the one chosen. DiNardo *et al.* (1988) demonstrated the applicability of the AHP to decisions facing fishery managers. The specific example considered involves Maryland's river herring fishery.

Operations Research

The actual construction of abundance maps based on longitude and latitude coordinates possess a host of interesting statistical issues. Once a map is assembled, decisions need to be made regarding where to sample in order to update and revise the most recent abundance map. These decisions are of the sort commonly studied by operations researchers. Budgetary and time constraints preclude us completing a sampling exercise as one would ideally like.

In deciding where to sample, several key variables need to be considered. The cost of sampling is one major factor. What resources are required to sample area a_i in time period t? Does abundance vary widely at area a_i over time? If yes, then sampling should be encouraged. If not, then the information gained from sampling at a_i may be negligible. If abundance at a_i and a_j are highly correlated, then it may not pay to sample both areas in time period t. We anticipate that the Analytic Hierarchy Process will be used to help select sampling areas.

Once the sampling areas are chosen, boats need to be scheduled. Each schedule specifies a set of areas to be visited by a particular vessel and the associated resources required for sampling. In general, the construction of schedules is a complicated task due to the size and mix of the fleet, the number of areas to be sampled, the periodic nature of sampling requirements,

etc. This scheduling activity can be aided via the use of computerized routing and scheduling techniques in order to reduce costs.

Important attributes of the operations research component will be the formulation of data collection stopping rules, *i.e.* when do we stop sampling. In contrast to the classical methods of statistics/sampling, according to which the number of observations is fixed in advance (according to a predetermined detection rate and significance level), an essential feature of the data collection stopping rules will be that the number of observations depends on the outcome of the observations and is therefore not predetermined, but a random variable. Samples are drawn sequentially until a predetermined degree of precision is attained, at which point sampling terminates.

These and related problems have been studied by Cochran (1977), Seber (1982), and Cassel *et al.* (1977). Francis (1984) looked at the problem of effort allocation in stratified random trawl surveys and suggested an adaptive strategy based on prior information. Schweigert *et al.* (1985) have suggested an optimal two-stage sampling design for herring spawn surveys, as has Chester and Walters (1986) for sampling for age distribution in Atlantic menhaden.

Wald (1947) discussed the advantages of sequential decision methods in the problem of testing two hypotheses. He proved that such methods yield on the average a smaller number of observations than any other method using fixed sample size. Furthermore, Wald described an optimal sequential test procedure, the sequential probability rates test, for testing H_0 against H_1. Similarly, Kuno (1969) described a sequential sampling procedure which is designed so that "sampling is terminated when a defined level of precision is reached" with precision measured by the ratio of the standard error to the mean (equivalent to estimating the mean within plus or minus some specified percentage of the mean). Green (1979) discussed the application of sequential

analysis procedures to environmental impact assessment and described how this procedure could be used to estimate sample size and number in various sampling designs. Jackson and Resh (1988) tested the efficacy of using sequential decision plans to monitor benthic macroinvertebrates and found that sequential decision plans could provide correct descriptions of environmental conditions based on fewer sample units than fixed sample-size procedures.

System Coordination

The development of a system to identify and update an optimal cost-effective sampling methodology to report on the dynamics of fish and shellfish in Chesapeake Bay is the primary goal of this project. The components outlined above are all essential to the development of such a system. The user of such a system, however, would have to be proficient in a number of diverse disciplines. In addition, someone proficient in these disciplines would need to be on board the vessel during each cruise to evaluate incoming data and make decisions as to whether or not to fish the trawl. In many cases, however, the user is a "nonexpert." Thus, a system that allows the user to imitate the decisions of an "expert" is needed.

The problems detailed above can be largely overcome by integrating the disciplines through the use of an expert system. The fundamental premise of the FISHMAP system is that the utilization of techniques in the data input, map models, decision-making workshops, and operations research components in the sampling of fish populations, integrated in the setting of an expert system represents a significant advance in the application of these techniques to stock assessment. Integration of the system is made tractable through the use of expert system technology. Improvement in the application of these techniques is gained through the expert system facility for interpreting very

large databases in approximately real-time.

Specific objectives of FISHMAP are twofold.

Objective 1: The identification of the most efficient sampling design to estimate standing stock.

If distributional information pertaining to a particular species is good, the system will identify the most efficient sampling design using the strategic model; where to sample and when to sample. If, however, no information exists, or the extant data is poor, the system will still identify a sampling design (not necessarily optimal) but will also identify what information needs to be collected in order to identify an optimal sampling design. For example, suppose we are interested in estimating the standing stock of spot in the northern Chesapeake Bay. The system will query the user about the types of information that are currently available for spot (this information is housed in the data input component). Based on the user's response to the queries, the system will determine which information state exists (data-rich or data-poor). If it is determined that a data-rich state exists, an optimal sampling design will be identified for estimating spot standing stock via the strategic model. If, however, a data-poor information state exists, the system will identify a sampling design (not necessarily optimal) and which information needs to be collected in order to define an optimal design. Say, for instance, that spatiotemporal distributional data for spot is nonexistent. In order to allocate sampling effort in an efficient manner, distributional information is required. Thus the system will identify a sampling design and will also indicate that this information should be collected to produce an optimal sampling design.

Based on the results of the strategic model that tell us where and when

to sample, the operations research component will be used to optimize the sampling process. In particular the number of research vessels to use in sampling will be identified as well as vessel-routes.

Objective 2: The development of an information interpretation subsystem that will aid in the identification of an efficient sampling protocol; where and when to fish the trawl and how often to trawl.

In many instances, prior to the commencement of trawling in a particular area and at a specific time (say multidimensional time T), information describing the physicochemical characteristics, commercial and recreational catches, and distributional characteristics of the species is available. In addition, there may also be hydroacoustic information available, both in terms of species specific densities and relative abundance of different size classes. If each piece of data was assessed individually, our expectation as to whether or not we could expect to catch species i at the next multidimensional time T+1 would be less than if we were to incorporate all of the information into our assessment. The information interpretation subsystem will be a modular system that allows for an assessment based on all available information. Each module will be information specific. The tactical model will be the tool used to interpret the stream of real-time data originating from each module. The model will be used on board the research vessel to decide when to fish the trawl based on characteristics of the stream of real-time data being observed. Put another way, the model will predict the catch of species i of a subsequent multidimensional time, T+1, based on hydroacoustic, physicochemical, and catch information observed at multidimensional time T. If the prediction is "good" the trawl is not fished and biomass is measured via hydroacoustics. If, however, the prediction is poor, the trawl is fished. One feature of each module will be a decision

criteria model that can be used to measure our uncertainty associated with the stream of information.

REFERENCES

Cassel, C.-M., C.-E. Sarndal, and J.H. Wretman. 1977. *Foundations of inference in survey sampling*. John Wiley & Sons, New York.

Chester, A.J. and J.R. Waters. 1985. Two-stage sampling for age distribution in the Atlantic menhaden fishery, with comments on optimal survey design. *No. Amer. J. Fish. Mgmt.* 5:449-456.

Cochran, W.G. 1977. *Sampling Techniques*. John Wiley & Sons, New York.

Cochran, D. and G.H. Orcutt. 1949. Applications of least squares regression to relationships containing autocorrelated errors. *J. Amer. Stat. Assoc.* 44:32-61.

D-Base IV Version 1. 1988. Ashton-Tate.

DiNardo, G.T., D.E. Levy, and B.L. Golden. 1988. Using decision analysis to manage Maryland's river herring fishery: An application of AHP. *J. Environ. Mgmt.* In Press.

Francis, R.I.C.C. 1984. An adaptive strategy for stratified random trawl surveys. *New Zealand J. Marine and Freshwater Res.* 18:59-71.

Golden, B.L., B. Rothschild, and A. Assad. 1987. A microcomputer-based decision support system for multi-species fishery management. pp. 333-339 *In* Lev, B., J.A. Bloom, A.S. Gleit, F.H. Murphy, and C. Shoemaker, *Strategic Planning in Energy and Natural Resources*, Elsevier Science Publishers, 339pp.

Green, R.H. 1979. *Sampling Design and Statistical Methods for Environmental Biologists*. John Wiley & Sons, New York, NY. 257pp.

Gulland, J.A. 1983. *Fish Stock Assessment: A Manual of Basic Methods*. John Wiley and Sons, Inc., New York, 223pp.

Jackson, J.K. and V.H. Resh. 1988. Sequential decision plans in monitoring benthic macroinvertebrates: Cost savings, classification accuracy, and development of plans. *Can. J. Fish. Aquat. Sci.* 45:280-286.

Kuno, E. 1969. A new method of sequential sampling to obtain the population estimates with a fixed level of precision. *Res. Pop. Ecol.* 11:127-136.

Lotus 123. v. 2.01. 1988. Cambridge, MA.

Mangel, M. 1986. Abundance estimates based on encounter rates. Report 86-4 on Sea Grant Project 6-F-N "New Methods in Stock Abundance Estimation." 33pp.

Quattro Version 1. 1987. Scotts Valley, CA: Borland

Rothschild, B.J. 1986. *Dynamics of Marine Fish Populations*. Harvard University Press, Cambridge, MA, 277pp.

Saaty, T. 1982. *Decision Making for Leaders*. Lifetime Learning Publications, 291p.

Schweigert, J.F., C.W. Haegele, and M. Stocker. 1985. Optimizing sampling design for herring spawn surveys in the Strait of Georgia, B.C. *Can. J. Fish. Aquat. Sci.* 42:1806-1814.

Seber, G.A.F. 1982. *The Estimation of Animal Abundance*. MacMillan Publishing Co., New York.

Stagg, C.M. 1987. The design and development of a fishery management expert system. Ph.D. Thesis, University of Maryland.

Wald, A. 1947. *Sequential Sampling*. John Wiley & Sons, New York, NY. xii + 212p.

Walters, C. 1986. *Adaptive Management of Renewable Resources*. MacMillan Publishing Company, New York, 374pp.

Zahedi, F. 1986. The analytic hierarchy process - a survey of the methods and its applications. *Interfaces* 16:96-108.

APPENDIX 1

One way to think about statistical tradeoffs is to consider the "value" of information. For example, let the expected value of sample information (EVSI) be equal to the ratio of the estimated parameter over the true parameter, the expected value of perfect information (EVPI) (we are assuming that we know the true value). If we were interested in getting as precise an estimate as possible, the optimal sample size would be that which results in a ratio value of unity. Figure A.1 shows the relationship between the values of information and sample size. In this graph the EVSI approaches the EVPI quickly and then levels off. The leveling off reflects the fact that as the sample size increases, the additional value of each trial decreases (*i.e.*, diminishing returns), and EVSI approaches EVPI. Choosing a large sample size, say n", will result in a better estimate then a sample size of n', but the increase in precision may not be significant enough to warrant the extra effort. Thus, what are the statistical tradeoffs associated with choosing n' over n"? These are the types of problems that the FISHMAP system will address.

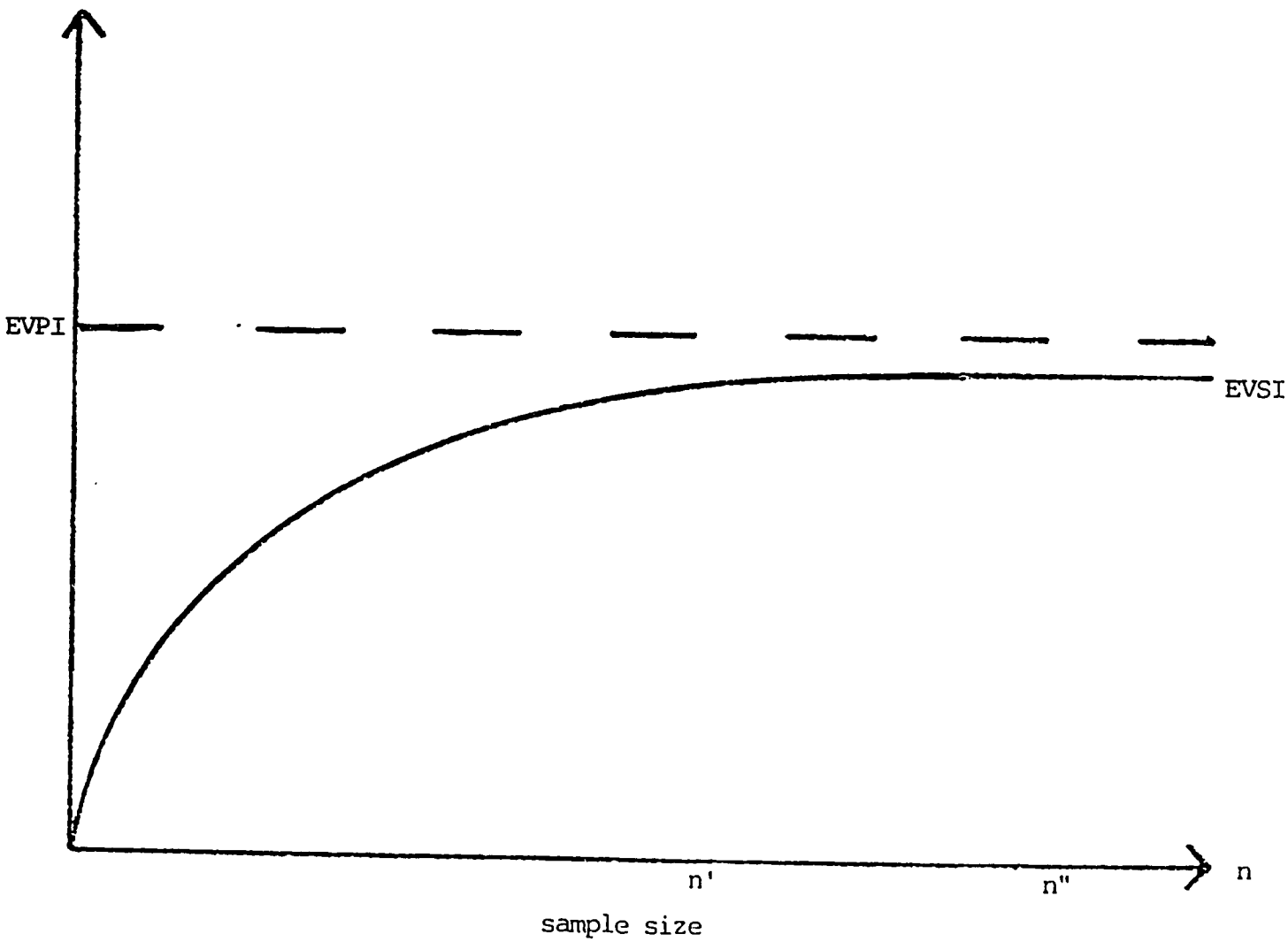


Figure A1. Relationship between the expected value of perfect information (EVPI), the expected value of sample information (EVSI), and sample size (n).