

92919

# **The Truss: A Geometric and Statistical Approach to the Analysis of Form in Fishes**

J.M. McGlade and E.G. Boulding

Fisheries Research Branch  
Scotia-Fundy Region  
Marine Fish Division  
Bedford Institute of Oceanography  
P.O. Box 1006, Dartmouth  
Nova Scotia, B2Y 4A2

September 1986

**Canadian Technical Report of  
Fisheries and Aquatic Sciences  
No. 1457**



Fisheries  
and Oceans

Pêches  
et Océans

**Canada**

## **Canadian Technical Report of Fisheries and Aquatic Sciences**

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of the Department of Fisheries and Oceans, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in *Aquatic Sciences and Fisheries Abstracts* and indexed in the Department's annual index to scientific and technical publications.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715-924 were issued as Department of Fisheries and the Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page. Out-of-stock reports will be supplied for a fee by commercial agents.

## **Rapport technique canadien des sciences halieutiques et aquatiques**

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. Il n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques du ministère des Pêches et des Océans, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications complètes. Le titre exact paraît au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la revue *Résumés des sciences aquatiques et halieutiques*, et ils sont classés dans l'index annuel des publications scientifiques et techniques du Ministère.

Les numéros 1 à 456 de cette série ont été publiés à titre de rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre. Les rapports épuisés seront fournis contre rétribution par des agents commerciaux.

Canadian Technical Report of Fisheries and Aquatic Sciences No. 1457.

The Truss: A Geometric and Statistical Approach  
to the Analysis of Form in Fishes

by

J.M. McGlade and E.G. Boulding<sup>1</sup>

Marine Fish Division  
Bedford Institute of Oceanography  
P.O. Box 1006, Dartmouth  
Nova Scotia, B2Y 4A2

<sup>1</sup> Present Address: University of Washington, Department of Zoology NJ-15, Seattle, Washington, 98195, U.S.A.

Minister of Supply and Services Canada 1986

Cat. No. Fs 97-6/1457E

ISSN 0706-6457

Correct citation for this publication:

McGlade, J.M. and E.G. Boulding. 1986. The Truss: A Geometric and Statistical Approach to the Analysis of Form in Fishes. Can. Tech. Rept. Fish. Aquat. Sci. No. 1457.

## TABLE OF CONTENTS

	<u>Page</u>
ABSTRACT . . . . .	Iv
INTRODUCTION . . . . .	1
THE PROBLEMS OF STOCK DEFINITION . . . . .	1
PATTERNS OF GROWTH IN DIFFERENT FISH POPULATIONS . . . . .	2
THE QUANTIFICATION OF MORPHOLOGICAL DIFFERENCES . . . . .	2
METHODS . . . . .	3
SAMPLING PROBLEMS . . . . .	3
HOW TO SET UP THE TRUSSES . . . . .	4
DATA VERIFICATION . . . . .	4
A GEOMETRIC APPROACH: THE COMPOSITE FORM . . . . .	4
Theory and Algebra . . . . .	6
Calculation of X-Y Co-ordinates . . . . .	6
Calculation of "Average" Truss Measures . . . . .	7
Flattening the Truss . . . . .	8
Instructions . . . . .	9
A STATISTICAL APPROACH: MULTIVARIATE ANALYSIS OF FORM . . . . .	9
Multivariate Procedures: Pattern Recognition in Morphological Hyperspace . . . . .	10
The Problem of Size . . . . .	11
Instructions . . . . .	12
AN EXAMPLE OF RESULTS -- THE SCOTIAN SHELF GADOLDS . . . . .	13
GULF OF MAINE AND SCOTIAN SHELF POLLOCK . . . . .	13
GULF OF MAINE AND SOUTH-WEST NOVA SCOTIAN HADDOCK . . . . .	13
POLLOCK VERSUS HADDOCK . . . . .	21
CONCLUSIONS . . . . .	32
ACKNOWLEDGEMENTS . . . . .	32
REFERENCES . . . . .	32
APPENDIX A: Flow Chart of Procedures . . . . .	35
APPENDIX B: CYBER Procedures Using FORTRAN V Programs . . . . .	36
REGPROC (BMDP1D With TRANSF and TRUSRG) . . . . .	39
LEASUB (CTRUS) . . . . .	44
TRUSPLT (TRUSSD) . . . . .	52
APPENDIX C: CYBER Procedures using BMDP Statistical Programs . . . . .	62
Notes on Using BMDP Statistical Software . . . . .	62
PCAONE (BMDP4M) . . . . .	64
PCAPLOT (BMDP6D) . . . . .	66
PCAPLT2 (BMDP6D) . . . . .	66
DISCONE (BMDP7M) . . . . .	67

## ABSTRACT

McGlade, J.M. and E.G. Boulding. 1986. The Truss: A Geometric and Statistical Approach to the Analysis of Form in Fishes. Can. Tech. Rept. Fish. Aquat. Sci. No. 1457.

This report describes the application of and software to produce a truss (sensu Strauss and Bookstein 1982) in reconstructing the outline of a fish, based on a geometric approach, and in providing the data necessary for a multivariate statistical analysis of morphology in fishes. The use of the method in stock identification is described, and illustrated by a case study of pollock (Pollichius virens) and haddock (Melanogrammus aeglefinus) on the Scotian Shelf and in the Gulf of Maine. The truss is viewed as an effective alternative to more traditional measures of morphological description, as it relies upon a rigorous definition of homology between individuals.

## RÉSUMÉ

McGlade, J.M. and E.G. Boulding. 1986. The Truss: A Geometric and Statistical Approach to the Analysis of Form in Fishes. Can. Tech. Rept. Fish. Aquat. Sci. No. 1457.

Ce rapport décrit l'application d'un réseau (après Strauss et Bookstein 1982) dans la reconstruction d'un profil de poisson, basé sur une approche géométrique, et la construction des données nécessaires pour analyser les statistiques de variables multiples de la morphologie des poissons. On décrit l'application de cette méthode à l'identification des stocks. En suite on démontre la méthode en utilisant nos études des goberges (Pollichius virens) et des aiglefin (Melanogrammus aeglefinus) du Plateau Scotian et du Golfe Maine. On considère que le réseau est plus efficace que les mesures classiques pour décrire la morphologie, parce qu'il est basé sur une définition rigoureuse de l'homologie entre les individus.

## INTRODUCTION

Paradoxically, the discrimination of marine finfish stocks relies as much upon similarity as upon variability. The simple fact that we can detect differences between groups means that we can also detect similarities within them. Human beings are inveterate classifiers, and so almost every field of inquiry begins with some kind of taxonomy or natural classification: physicists classify elementary particles, chemists use a Periodic Table and of course biologists have developed the Linnean system.

At the root of biological classification is the tacit assumption that the members of a particular group share a similarity of structure or behaviour, or a commonality of descent, from which we see a scheme of relatedness that allows us to assemble individuals into a class or taxon. The process of classification is generally referred to as systematics, which Simpson (1961) defined as "the scientific study of the kinds and diversity of organisms and of any and all relationships among them". This definition is used in its widest sense, and subsumes the problems of taxonomy which pertain to the theoretical study of classification, including its principles, bases, procedures and rules (*ibid.*).

Current developments in systematics have included rapid changes in concepts and procedures, hastened no doubt by the wide availability of computers, and advances in genetics, cytology and molecular biology. And yet it is still fair to say that for fisheries, the subject of the structure of fish populations is "...a bewildering array of semantic problems because there is little agreement on the meaning of the words used to define groups in the hierarchy with the rank of subspecies and below..." (Royce 1972). True, the literature abounds with observations, opinions and statements on the nature of a fish stock, but there is a fundamental disparity which is likely to keep any authoritative resolution in abeyance -- that is the belief in a genetically determined or even phenotypically defined group versus a more pragmatic group defined in greater part by a management plan in ignorance of any genetic or phenetic integrity.

Attempting to define stocks, however, remains an imperative for fisheries management especially in the marine environment, because through the Law of the Sea Conference, more and more nations have taken on the responsibility for common fisheries resources within the 200 mile zone adjacent to their coastlines, and in so doing have developed a need to estimate the abundance of fish stocks in response to fishing pressure. Stock definition is thus at the core of fisheries, because implicit in management is the sentiment put forward at the beginning of the century by Theodore Roosevelt that the Nation behaves well if it treats the natural resources as assets which it must turn over to the next generation increased and not impaired in value.

In many countries the principle adopted for the most efficient way to control extraction of marine fishes is one of total allowable catch (TAC), allocation within which is subject to

negotiations. The TAC is the output from the analytical, but essentially deterministic models developed by Beverton and Holt (1957). Within this legacy, the population (or stock) is first defined, and then the principal parameters, such as growth, mortality, maturity and recruitment are determined. Thus the first step in any fisheries management plan is to define the structure of the fish resources within a stipulated area, and state which stocks are transboundary, because in reality a jurisdictional boundary may represent not only a legal limit to national control but also to divergent philosophies as to the ways in which common property resources are viewed, and hence exploited. The evidence for stock separation must thus be as unequivocal as possible, for it is clear that subject to the division of oceans, the theoretic bases of stock delineation will come under serious review and attack.

## THE PROBLEMS OF STOCK DEFINITION

What then is a stock? Traditionally it has been defined as "a population of organisms which, sharing a common gene pool is sufficiently discrete to warrant consideration as a self-perpetuating system which can be managed" (Larkin 1972). However, the caveat of manageability distorts the term by bet-hedging against the possibility of irreconcilable demands from biology and politics.

Indeed it has been stated that whilst "biological management of fisheries has been built around the concept of the unit stock... this apparently common sense notion may be an instance of misplaced concreteness which places artificial constraints on analyses or on management rules and procedures. In fact the stock is an abstract term applied to provide a rationale for a certain kind of aggregation of catch data. This is not to say that there may or may not be such a thing as a discrete group of fish that may constitute an effective breeding group or stock, but in many cases there is significant uncertainty about the identity of the group from which successive annual catches are made, so that the operational term does not unequivocally refer to an identifiable physical entity." (Dickie 1979).

How then do we recognize a stock? Implicit in most statements is the fact that a stock is a population which by inference maintains itself in Castle-Hardy-Weinberg equilibrium, but by default can be distinguished by its phenotype (Booke 1981, Ihssen *et al.* 1981). However, these two statements do not proceed *pari passu*, and in fact there is now a developing literature on epigenesis (the phenotypic translation of the genome under different exogenous and endogenous constraints) and the hierarchical organization of adaptive potentials that lead from DNA to the phenotype and behaviour of an organism, which would suggest that congruency of genetic and phenotypic groups is highly unlikely. This of course begs the question as to the role of population genetics in fisheries management, because there is some doubt as to how selection, in the form of fishing, is operating on the phenome and hence the genome. Something as fundamental as a cline (*viz.* a non-uniform spatial

distribution in the genetic composition of a population in equilibrium) may not coincide with a cline in a phenotypic character.

The susceptibility of a fish species to form genetic stocks is probably related to the degree of spatial or temporal separation of encounters. A species whose range is well separated by geographic barriers or whose life history attributes include multiple spawning periods, homing to a spawning area, longevity, and a bottom-oriented fry stage will thus be likely to consist of multiple genetic stocks (Spangler *et al.* 1981, McGlade 1981). The susceptibility of a fish species to form phenotypic stocks is probably related to its morphological plasticity or genome-phenome linkage, to the amount it moves and to the presence of consistent environmental differences among different parts of its habitat. For example, a species whose life history involves juveniles originating from the same spawning area diffusing out into different areas, and then returning to an area to spawn is more likely to consist of multiple phenotypic stocks.

Unfortunately, the variability observed in electrophoretic analyses of enzymes -- the technique most generally used to ascertain levels of genetic variability -- cannot be implicitly assumed to be representative of the variability of the genome as a whole; structural proteins appear to be much less variable than enzymes, and the variation itself may depend on the subunit molecular weights of the enzymes. Collectively, these results suggest that estimates of genetic variation based upon the standard electrophoretic techniques represent a highly biased sample of genes. Measures of genetic differentiation, whether among individuals within populations, or among populations, must thus be coupled to the collection of other information such as morphology, physiology and ecology. However the explicit relationship between the genome and the organism has not been extensively studied.

Thus, fisheries managers have largely based their decisions about stock structures on evidence from tagging studies, and more traditionally meristic and morphometric analyses. The basic assumption then is that the results do in fact reflect some genetic homogeneity, albeit influenced by the environment.

#### PATTERNS OF GROWTH IN DIFFERENT FISH POPULATIONS

Patterns of growth from juvenile stages to adult stages often differ in different fish populations. The growth rate of a fish will not be generally uniform but there will be periods (or growth stanzas) during which it can be considered approximately so (Cock 1966). The transition between one growth stanza and another often occurs as the fishes go from one habitat to another or from one maturity stage to another as ontogenetic potentials are crossed.

Differences in form among different fish stocks are probably most conveniently considered as differences in shape as a function of size. This is complicated because fish of the same size are not necessarily of the same age or at the same stage of development especially if their average adult size differs. The genesis of differences in shape can occur through:

- 1) differences in shape generated during the egg and larvae stages;
- 2) differences in the relative growth rates of different body parts during a given growth stanza;
- 3) differences in the size at which the transition from one set of growth rates to another occurs (Cock 1966).

All of the above mechanisms for generating shape differences can occur because of genetic causes or because of different environmental conditions experienced by the various fish populations. That genetic variability can be responsible is easily seen by comparing the growth patterns of two fish species which differ in shape. But variation in environmental variables such as food availability and temperature can also generate shape differences. During a phenocritical period during the egg stage, meristic characters, which in turn effect external morphology, are very sensitive to temperature (Tåning 1946, Tåning 1952), salinity,  $O_2$  and  $CO_2$  pressure (Heuts 1949). The relative growth rate of different body parts can certainly be influenced by starvation as exemplified by the presence of stunted fish. And although there may be a genetically-determined critical minimum length (Thorpe *et al.* 1980) that must be reached before transition from one growth stanza to another can occur, the time required to reach that length may in part depend on the environmental conditions. Indeed there is evidence that fish may postpone transition to the next growth stanza depending on their physical state; Eastern-Atlantic salmon (*Salmo salar*) smoltify at either one or two years depending on their size by the fall of their first year; those which smoltify after two years do not feed over their first winter (Thorpe *et al.* 1980).

#### THE QUANTIFICATION OF MORPHOLOGICAL DIFFERENCES

Biologists have long been interested in quantifying differences in form among organisms with some of the early workers taking a statistical (Castle 1914, Huxley 1924, Wright 1932) and some a geometric (Thompson 1917) approach. It was recognized early in the 1900s that body parts did not necessarily remain in proportion to each other during growth (Castle 1914) and this led to the development of a mathematical model (Huxley 1924) and a statistical method (Gould 1966) to test for what became known as allometry.

Long-standing interest in the genetic basis of morphological variation was and still is a driving force behind many attempts to quantify morphology. An early application of the multivariate procedure, principal components analysis, and path analysis to morphometric data from chickens was prompted by an interest in defining a general factor that influenced overall body size and additional special growth factors which differentially affected different body parts (Wright 1932). In the study of fishes, it was also discovered that an environmental factor, namely temperature, influenced the number of vertebrae present in trout (Schmidt 1921), and that during the phenocritical period in embryology

a difference of up to 5 vertebrae could be induced even between genetically very similar (full siblings) embryos (Tåning 1946). This discovery -- that environmental factors could affect meristic characters and in turn external morphology -- was important because it showed that the adult form of an organism resulted from an interaction between its genome and the environment, the details of which are still incompletely understood.

Most of the recent advances in morphometrics use new multivariate methods for separating differences in size from differences in shape (Mosimann and James 1979, Humphries *et al.* 1981, McGlade 1981, Thorpe and Leamy 1983) or new geometric methods for quantifying changes in shape from one form to another (Bookstein 1978). It is important to separate variation in shape from variation in size, so that the shape of an average form can be compared at a given size, ontogenetic stage or chronological age. Indeed if the ontogeny behind a given shape difference is of interest, some workers favour a geometric approach to shape analysis rather than the multivariate methods of data analysis we discuss in a later section. They criticize multivariate methods because:

- 1) morphometric data used in such analyses is typically first reduced to a correlation or covariance matrix which results in the loss of their spatial relationships which would otherwise enable a functional morphologist to interpret the changes in shape (Bookstein 1978);
- 2) the interpretation of allometry in a multivariate context is debatable (Sprent 1972); and
- 3) the results of such analyses are often difficult to interpret and difficult to explain to non-statisticians.

As a result many morphologists have taken a geometric approach in comparing life-history stages.

These comparisons are facilitated if Strauss and Bookstein's (1982) truss of morphometric measurements is used. The truss which consists of the distances between homologous landmarks on the outline of a two-dimensional projection of a form has many advantages over the traditional morphometric data sets:

- 1) it provides a geometric protocol for morphometric character selection;
- 2) it archives the configuration of the landmarks so that the form of an individual specimen can be reconstructed;
- 3) it makes it possible to take morphometric measurements with a digitizing board;
- 4) it enables construction of a composite, average form that represents a given population at a given age or size; and
- 5) it allows visualization of multivariate trends of growth and allometry within populations (Strauss and Bookstein 1982).

The truss method has been used for cottid sculpins (Strauss and Bookstein 1982), for juvenile chinook salmon (Winans 1984), and in this report we present its application to the gadoids of the Scotian Shelf.

In this technical report we describe how to use the truss method in selecting morphometric characters for identification of environmental stocks. We also describe how to construct a composite truss that provides a geometric representation of the ontogenetic growth patterns in a given population, and provide FORTRAN V programs and command language from the BMDP Statistical Package that do the calculations. Finally we describe a multivariate procedure for analyzing morphometric data using the BMDP statistical package. We hope our experience will benefit other workers interested in applying morphometrics in a fisheries context.

## METHODS

### SAMPLING PROBLEMS

An effective sampling design may be obvious where there are geographical barriers that prevent or discourage mixing of the hypothesized stocks, but will be less so where a species is continuously distributed over a large, homogeneous geographic region. If no previous information is available, it is wise to do a preliminary study, comparing fish from the extremes of the species distribution before investing additional sampling effort. If there are no differences among the extremes of a continuous distribution over a homogenous geographical area, there are unlikely to be differences among intermediate areas. It is best to sample when the geographical separation of the hypothesized stocks is at a maximum. This will usually be when the fish are aggregated for spawning but if environmental stocks are of primary interest it could be just before the young fish leave a nursery area. We will refer to these spawning aggregations or nursery areas as geographical areas.

The sampling problem then reduces to two aspects: 1) to obtain enough locations (or sets) to characterize each geographical area, and 2) to obtain and measure enough fish to estimate the variation within each location (Thorpe 1976). Single sets are not sufficient to characterize an area if fish of the same age tend to school together or if segregation within a habitat is dependent on the fish's age. Therefore even in a preliminary study there should be at least two locations (or sets) for each area and at least 50 or more fish obtained and measured from each location depending on the amount of morphological variation present. A more detailed study would require many more fish and an even distribution of sizes within the size range being characterized. With rare species or where sampling for morphometrics is only a component of a larger sampling program, it may be difficult to obtain 50-100 fish per location and there will be a temptation to pool all the fish from several locations. This should be done with caution as it may result in grouping together two genetically distinct populations or subdividing a cline (Thorpe 1976).

## HOW TO SET UP THE TRUSSES

The exact configuration of the truss will depend on the fish species under investigation. Central to the concept of the truss is the idea of homologous anatomical landmarks (Strauss and Bookstein 1982). In practice homologous characters can be difficult to choose if the fish are from different genera. If only external morphological criteria are used, a character may be considered homologous with the character with which it shares the greatest degree of similarity or correspondence (Sneath and Sokal 1973), although ideally this will be based on evidence of evolutionary relationships from the fossil record or from comparative embryology.

When choosing landmarks it is important to 1) choose points identified by some consistent feature of the local morphology such as the insertion of a fin (Fig. 1), 2) to comprehensively and evenly cover the entire body form, and 3) to choose the points so that the interlandmark distances are as short as possible as short measures contain more localized information about shape (Strauss and Bookstein 1982).

The truss we used for these gadoid fishes contains 7 cells and is derived from a lateral projection of the 3 dimensional fish onto the two dimensions of the paper (Fig. 2). A truss containing 4 cells from a lateral projection and an additional cell with two appended triangles from a dorsal projection has been used for comparing species of cottids as head shape was important (Strauss and Bookstein 1982). Certainly if prior information on the type of interstock variation is available it makes sense to increase the density of landmarks in that body region.

If preserved fish are to be used it is necessary to preserve them flat particularly if they are large. Freshly caught fish are ideal and because the archiving of landmarks is relatively rapid, it may even be possible to archive the landmarks of an anesthetized fish which is being repeatedly measured over time for a growth study.

To archive the landmarks, the fish is laid on its side on a piece of water-resistant paper and its fins are spread out. The landmark positions are marked with a pencil. Interior landmarks are extended to the closest point on the body outline on a line perpendicular to the longitudinal axis (see Fig. 1, landmarks 2, 3, and 5). The fish is then removed, the landmarks are circled and the paper labelled to identify the fish. If insufficient landmarks exist around the periphery of the fish, it is possible to establish points by X-raying each fish, enhancing the edges using iron filings in vaseline, and then projecting points out to the edges from the vertebral column (pers. comm., R.L. Stephenson, Marine Fish Division, Biological Station, St. Andrews, New Brunswick).

There are two ways to transfer the landmarks to the computer. The first, is to measure the distances between the landmarks with a ruler, and then keypunch them into the computer in order. Alternatively a digitizer can be used to digitize the position of the landmarks; then the interlandmark distances can be calculated by the computer

using Pythagoras' theorem (Winans 1984).

The use of the digitizer prevents errors from measuring and keypunching but necessitates writing at least one special computer program for each particular digitizer.

## DATA VERIFICATION

A data file composed of a large number of similar measurements for a sample of fish will inevitably contain a number of errors due to measurement, coding, and data entry. These must be detected and removed before further analysis is possible (see Appendix A for a flow chart of procedures).

Gross errors, (viz. a datum point more than two standard deviations from the mean and its character state), may result from misplacing a decimal point or taking a truss measure between the wrong landmarks. These can be detected by packages such as BMDP1D. An advantage of the truss method is that the archiving of the landmarks on tracing paper allows measurements to be checked months after the processing of the live fish has taken place.

Minor errors in multivariate data sets are normally difficult or impossible to detect. However where the total number of fish is relatively small these can be identified by attempting to reconstruct the trusses of individual fishes using the FORTRAN program TRUSSD (see Appendix B for FORTRAN listings). The appropriate option should be selected, so that the resulting plotfile can be drawn on a small flatbed plotter (e.g. a Tektronix) and examined for distortion. Dubious cases can be reconciled by comparing plots of comparable sized fish. Any gross errors that remain in the data set will cause an error message stating that the program is unable to compute the x-y coordinates for that particular fish.

## A GEOMETRIC APPROACH: THE COMPOSITE FORM

The originator of a geometric approach to shape analysis was the noted Scottish biologist and classical scholar D'Arcy Wentworth Thompson (1917, 1942). In his classic book On Growth and Form, he showed the transformation from one shape to another by drawing the distortion of a grid (Fig. 3). Attempts to quantify the distortion of Thompson-type grids in terms of growth gradients (Huxley 1932) were largely unsuccessful because the mathematics were intractable (Bookstein 1978). A promising new quantitative approach to measuring the shape change between any two forms is the method of biorthogonal grids (Bookstein 1978). Bookstein's method differs from that of Thompson in the orientation and structure of the grid. A mesh of points from the first form is mapped onto the second form by interpolating between homologous landmarks. A local grid is then computed at each of these points so that one of its axes is oriented along the direction of maximum or minimum local rate of change. A set of curving lines is then derived by integration from the principal directions of the axes. The elongations or contractions at any desired point

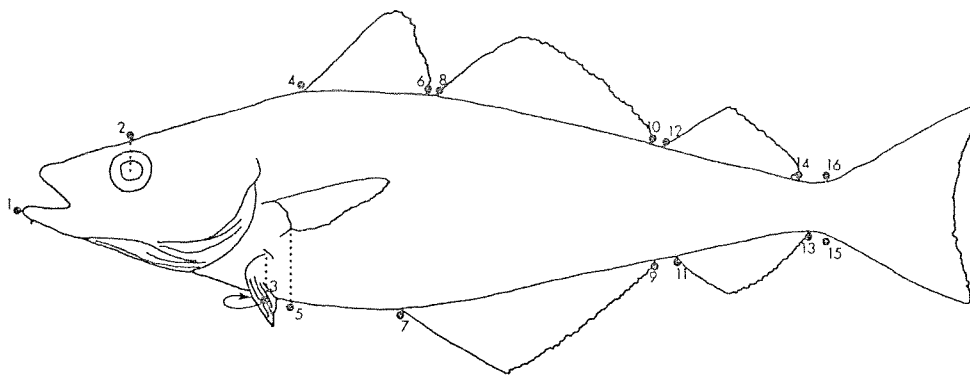
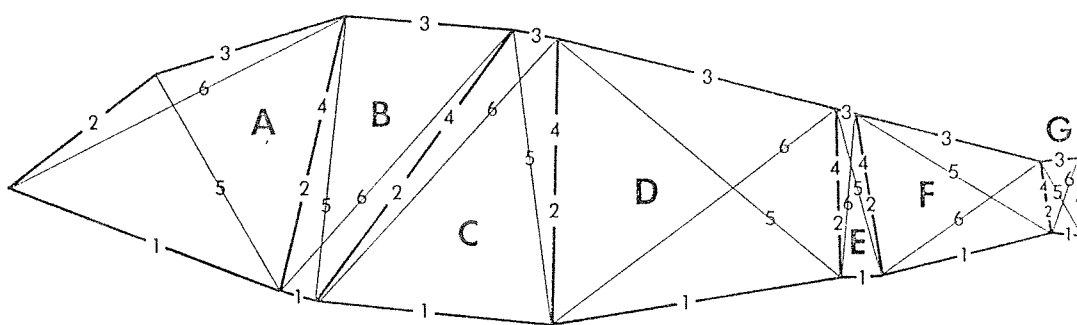


Figure 1. Positions of landmarks around a pollock (Pollachius virens).

V32805



TRUSS CONFIGURATION FOR SHAPE ANALYSIS

Figure 2. Trusses constructed from the landmarks indicated in Figure 1.

are thus obtained and can be depicted directly on the form (Bookstein 1978). While we do not include programs for biorthogonal analysis, we mention it here because the first step is the construction of a composite truss for each of the groups being compared (Strauss and Bookstein 1982).

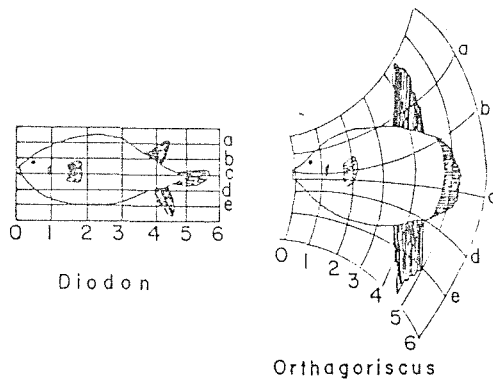


Fig. 3. Distortion of a grid giving the outlines of *Diodon* and *Orthogoriscus* (after D'Arcy Thompson, 1917).

The geometric technique we describe in detail in this report is the construction of a composite form (Strauss and Bookstein 1982) to represent a given fish population. The population characterized by a composite form is a function of the size-classes the sample consists of; problems will arise if extrapolation outside the range of the data is attempted. However, the strength of this technique is in its versatility -- one could characterize the entire ontogenetic growth pattern from juvenile to adult or alternatively all the two year olds by varying the data set used. The method for construction of the composite form is derived from the power function  $Y = bX^K$  described by Huxley (1924) to test for allometry where  $Y$  is a body part,  $X$  is a measure of total body length and  $b$  and  $K$  are the back-transformed intercept and slope, respectively from a linear least-squares regression on  $\log X$  and  $\log Y$  (Gould 1966). Thus if  $b$  and  $K$  are estimated for a given body part for a given population of fish, the size of that body part can be predicted for any desired size of fish. In the method we describe (a revised version of Strauss and Bookstein 1982), the scores on the first within-group principal component are used in place of total body length as the first within-group principal component best explains the patterns of covariance among the morphometric variables.

#### Theory and Algebra

##### Calculation of X-Y Co-ordinates:

The procedure for calculation of the x-y coordinates for each landmark is the same whether an individual truss is being reconstructed or a composite truss is being constructed from a

population. Calculation of the coordinates of the first two points is easier if both are arbitrarily placed on the y axis. Thus, if the first point is assigned the coordinates  $x = 0, y = 0$ , then the second point will be at  $x = 0, y = d_{12}$  where  $d_{12}$  is the distance between the first and second landmarks. The third point is at a distance  $d_{13}$  from the first point and  $d_{23}$  from the second point. If a circle with a radius of  $d_{13}$  is drawn, centered at the first point and another with a radius of  $d_{23}$  is drawn centered at the second point, the circles will intersect at two points, one to the right and one to the left of the y axis (Fig. 4).

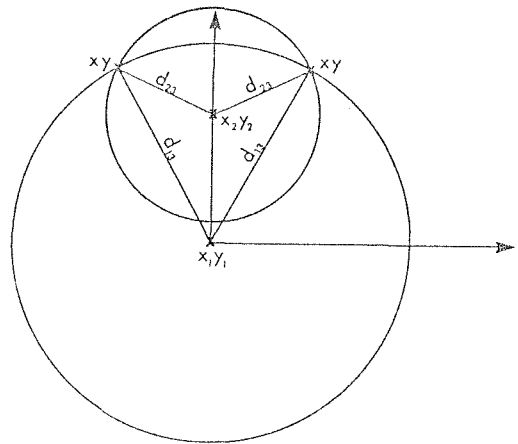


Fig. 4. Calculation of the X-Y coordinates for each landmark is based on the point of intersection of two circles with radii equal to the distance between the first two landmarks and an additional point.

We can determine the coordinates of these points algebraically:

Using Pythagoras' theorem twice:

$$(1a) \quad (x - x_1)^2 + (y - y_1)^2 = d_{13}^2$$

$$(1b) \quad (x - x_2)^2 + (y - y_2)^2 = d_{23}^2$$

where:  $x, y$  - are the coordinates of the new landmark (number 3)

$x_1, y_1$  - are the coordinates of the first point

$x_2, y_2$  - are the coordinates of the second point

$d_{12}$  - is the distance from the first to second landmark

$d_{23}$  - is the distance from the second to third landmark

Eliminating  $y$  from the simultaneous equations in 1a and 1b and solving for  $x$  using the quadratic formula gives us:

$$(2) \ x = \frac{-b + (b^2 - 4ac)^{\frac{1}{2}}}{2a} \text{ or } x = \frac{-b - (b^2 - 4ac)^{\frac{1}{2}}}{2a}$$

$$\text{where: } a = 1 + k_2^2$$

$$b = 2k_1k_2 - k_2y_1 - 2x_1$$

$$c = x_1^2 + k_1^2 + y_1^2 - 2k_1y_1 - d_{13}^2$$

$$\text{and: } k_1 = \frac{(x_1^2 + y_1^2 + d_{23}^2 - d_{13}^2 - x_2^2 - y_2^2)}{(y_1 - y_2)}$$

$$k_2 = \frac{(x_2 - x_1)}{(y_1 - y_2)}$$

Since the truss is to be built from left to right, the larger of the two values of  $x$  is chosen and substituted back into equations 1a or 1b to give:

$$(3) \ y = y_1 + (d_{13}^2 - (x_3 - x_1)^2)^{\frac{1}{2}}$$

where  $x_3$  is the larger value obtained in equation 2.

The coordinates of the fourth landmark can be obtained using  $d_{14}$  and  $d_{24}$ ,  $d_{14}$  and  $d_{34}$ , or  $d_{24}$  and  $d_{34}$ . Again two values for the  $x$  coordinate will be obtained. Since the definition of the truss is such that landmark four is above landmark three, and so  $d_{12}$  is on the  $y$  axis, the value of  $x$  chosen is that giving the largest value of  $y$ .

To calculate the coordinates of subsequent landmarks, it is important to rotate the truss to a standard orientation so that the correct value of  $x$  can be selected from the quadratic equation.

To carry out this rotation we first calculate the angle,  $\theta$ , that  $d_{13}$  makes with the  $x$  axis (Fig. 5a). Thus:

$$(4) \ \theta = \text{Arctan} \left( \frac{(y_3 - y_1)}{(x_3 - x_1)} \right)$$

and then apply an appropriate angular transformation to all the points calculated so far, so that  $d_{13}$  now lies on the  $x$  axis. Thus:

$$(5) \ \begin{aligned} x'_n &= x_n \cos \theta + y_n \sin \theta \\ y'_n &= y_n \cos \theta - x_n \sin \theta \end{aligned}$$

In a composite truss, the length of a shared side may have been adjusted to slightly different values in adjacent cells. To compensate for this, the position of the points in the previous cell are adjusted so that the shared side has the value of the next cell. This is done by lining up the segment's midpoints (Fig. 5b).

The above procedure is repeated for each cell until the coordinates of all the landmarks have been calculated. When the entire truss is completed, it is rotated, using a similar procedure to that shown above, until a

"pseudolateral" line from the "nose" to the "tail" is placed on the  $x$  axis.

Finally, to facilitate plotting, the truss is translated up the  $y$  axis until it is entirely within the first quadrant.

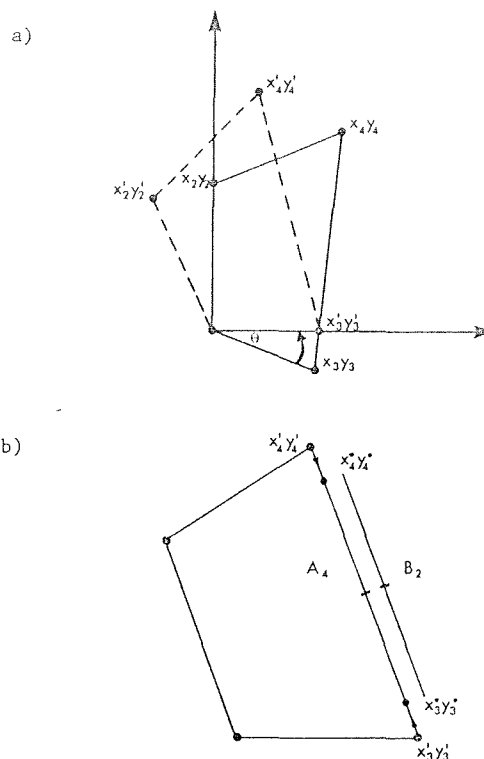


Fig. 5. Calculation of the X-Y coordinates of subsequent landmarks based on the rotation of the truss to a standard orientation by a) calculating  $\theta$  and applying the transformation to all points; b) for a composite truss the midpoints of the shared sides are lined up, as on  $A_4$  and  $B_2$ .

#### Calculation of "Average" Truss Measures:

A composite or "average" form can usually be constructed for any size that is within the size-range of fish found within the sample. Before this form can be plotted it is necessary to calculate the size of the truss measures comprising it. The method used is virtually identical to that described by Strauss and Bookstein (1982, p. 123) but we shall recapitulate it here briefly.

The measure of standard size used was the first within-group principal component scores, however the total length of the fish or some equivalent could be used. These scores were taken from a principal components analysis on the log-transformed truss measures, and should be done using the covariance matrix (Appendix C BMDP4M). Linear regressions between each truss element and the first within-group principal component scores as the independent variable were completed, as

described in the previous section. The coefficients from these regressions allowed the expected length of the truss measures for a fish of the desired standard size from the given population to be calculated (Fig. 6).

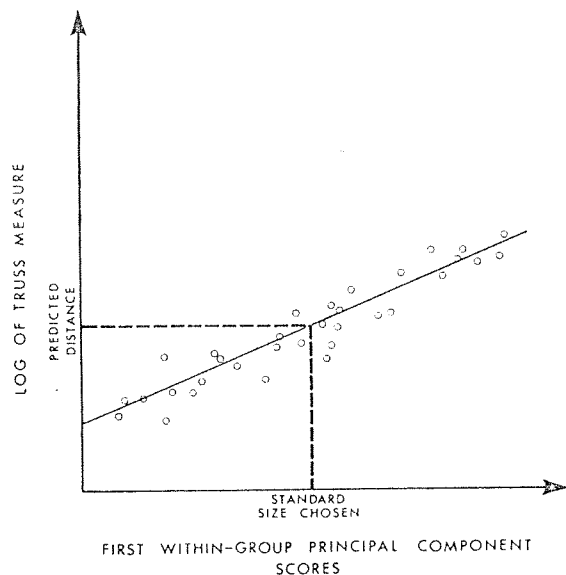


Fig. 6. Truss measurements for a fish of a desired standard size can be calculated from a regression of the first within-group principal component scores and the logarithm of each truss measure.

If  $k$  is the slope and  $b$  is the intercept from these regressions,  $d_{ij}$  is the distance between landmark  $i$  and landmark  $j$ , and  $S$  is the first within-group principal component score from the analysis on the  $\log(d_{ij})$ .

then:  $\log d_{ij} = \log b + k \cdot S$

which allows us to predict a  $d_{ij}$  for the desired  $S$ .

#### Flattening the Truss:

If each cell of the truss consisted of only five truss elements it would always be possible to form a planar truss. The presence of the sixth element provides redundancy. This facilitates the detection of errors by making it impossible to reconstruct a planar truss if one element contains a gross error. However even when the data set is composed of planar trusses from individual fishes, this does not guarantee that a planar composite truss can be constructed. By examining the simplest case where the population consists of only two individuals it can easily be shown that the population means for each type of truss element will not necessarily form a planar truss. Non-homogeneous variances among the different types of truss elements are the rule rather than

the exception. Even when care has been taken to delete all obvious errors from the raw data, some will inevitably remain. Thus a method of flattening the composite truss is required.

Usually only small adjustments to the truss elements making up each cell are needed to allow the cell to relax into a planar configuration. In order for these six distances (four edges and two diagonals) to be co-planar they must satisfy Salmon's (1914) criterion of planarity: i.e., the determinant  $V$  of a matrix must be equal to zero.

where:

$$|V| = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{12}^2 & d_{13}^2 & d_{14}^2 \\ 1 & d_{12}^2 & 0 & d_{23}^2 & d_{24}^2 \\ 1 & d_{13}^2 & d_{23}^2 & 0 & d_{34}^2 \\ 1 & d_{14}^2 & d_{24}^2 & d_{34}^2 & 0 \end{vmatrix} = 0$$

This determinant can be written algebraically as:

$$(6) \quad |V| = -2d_{13}^2 d_{24}^2 - 2d_{13}^2 d_{24}^2 - 2d_{12}^2 d_{34}^2 - 2d_{12}^2 d_{34}^2 \\ - 2d_{14}^2 d_{23}^2 - 2d_{14}^2 d_{23}^2 + 2d_{13}^2 d_{12}^2 d_{24}^2 + 2d_{13}^2 d_{12}^2 d_{34}^2 \\ - 2d_{13}^2 d_{12}^2 d_{14}^2 + 2d_{13}^2 d_{24}^2 d_{34}^2 + 2d_{13}^2 d_{24}^2 d_{14}^2 \\ + 2d_{13}^2 d_{24}^2 d_{23}^2 - 2d_{13}^2 d_{34}^2 d_{23}^2 + 2d_{13}^2 d_{14}^2 d_{23}^2 \\ + 2d_{12}^2 d_{24}^2 d_{34}^2 - 2d_{12}^2 d_{24}^2 d_{23}^2 + 2d_{12}^2 d_{34}^2 d_{14}^2 \\ + 2d_{12}^2 d_{34}^2 d_{23}^2 + 2d_{12}^2 d_{14}^2 d_{23}^2 - 2d_{24}^2 d_{34}^2 d_{14}^2 \\ + 2d_{24}^2 d_{14}^2 d_{23}^2 + 2d_{34}^2 d_{14}^2 d_{23}^2$$

This method of adjusting the truss elements until they are co-planar, differs from that of Strauss and Bookstein (1982) in that the determinant is used directly. In this way the magnitude of the correction to each truss element of the cell is minimized, while forcing the determinant  $V$ , as close to zero as possible.

Algebraically:

Let  $\underline{D}$  be the initial vector of the six truss elements for the cell being "flattened", then the corrected vector  $\underline{D}^* = \underline{D} + \underline{x}$  where  $\underline{x}$  is the vector of corrections.

We want:

$$|V((\underline{D} + \underline{x})^2)| = 0$$

where  $V$  is the determinant of the corrected truss elements for that cell, and the whole is termed the penalty function,

and:

$$\sum_{i=1}^6 x_i^2$$

to be minimized, where  $x_i$  is the difference between the initial and final cell vectors.

Therefore we want to minimize:

$$(7) F(\underline{x}) = \sum_{i=1}^6 x_i^2 + K |V(\underline{D} + \underline{x})^2|$$

where: K is a constant which the penalty function is multiplied by.

To do this we use the derivative-free version of the Levenberg-Marquandt algorithm. The FORTRAN program, TRUSS (Appendix B), calls the IMSL (International Mathematical and Statistical Library) (1984), subroutine ZXSSQ; LINPAC an equivalent program may have to be used if IMSL is not available. ZXSSQ is a subroutine which finds a local minimum of the sum of squares of  $m$  real functions in  $n$  real variables. The form of equation 7 used in the external subroutine, FUNC (Appendix B), which is called by ZXSSQ is:

```
DO 30 K=1,6
  F(K)=ABS(X(K))+(100,000*DETER)**2
30 CONTINUE
```

Where:

DETER is equal to the value of the determinant of the corrected truss elements. Each element of the vector  $\underline{F}$  is squared and added together by ZXSSQ.

The large value of the constant, K, means that the value of the determinant affects the value of  $\underline{F}$  more than the absolute value of  $\underline{X}$ . This means it is not technically a least squares problem but in practice the algorithm seems to work well.

When each cell of the composite truss has been "flattened" the adjusted truss elements are ready for point calculation and plotting by the FORTRAN program TRUSSD (Appendix B).

#### Instructions

Once the data have been verified and edited a composite form can be constructed for each group. Initially, we use the program BMDP4Ma (Appendix C) which transforms the data using  $\log_{10}$ , completes a principal components analysis using the covariance matrix option, and saves the data and principal component scores in a BMDP save file. Note that the USE statement in the TRAN paragraph can also be used to select a subset of the total number of cases so that only the fish from the desired location are present in the BMDP save file (Appendix C).

BMDP save files are an efficient method of data storage but are not easily read by non-BMDP programs. To get around this problem we use a BMDP program BMDP1DB, to read directly the BMDP save file then interface it with the FORTRAN subroutine TRANSF (Appendix B), which writes the data into a standard ASCII file. This subroutine precedes the FORTRAN program TRUSRG described below.

The FORTRAN programs TRUSRG and CTRUSS (Appendix B) are used to compute a series of univariate linear least squares regressions. The slope and intercept from one of these regressions

are used to predict the first principal component score of a fish of the desired size of the composite, and this is used with the other slopes and intercepts to predict its expected truss dimensions. The regressions are based on the log-transformed truss measurements plus the first principal component scores from the BMDP4M analysis. For each of the regressions used to predict the truss dimensions, the truss measure is the dependent variable and the first principal component score is the independent variable. For the additional regression the first principal component score is the dependent variable and the log-transformed total fork length is the independent variable. The coefficients from these regressions are saved in a standard file.

The next FORTRAN program, CTRUSS, uses the regression coefficients to calculate the expected truss dimensions for a composite fish of the desired size. These "average" truss measures must then be constructed into a composite truss. As previously discussed, although the "average" truss measures are derived from planar trusses, there is no guarantee that they will recombine to give a planar truss. Therefore the program adjusts the "average" truss measures, cell by cell, until they form a planar truss, by calling a subroutine from a non-linear least squares package (IMSL's ZXSSQ Appendix B). This subroutine minimizes the value of a determinant which is equal to zero when the truss is perfectly planar.

Non-linear optimization techniques are not foolproof, and the one used by ZXSSQ is no exception, but it works well in this application when the necessary corrections to the truss measures are small (see the IMSL manual, Appendix Z-2). It is important to check that the value of the variable Infer passed back from ZXSSQ is not equal to zero since a value of zero implies that convergence has failed. In practice this usually occurs because there are still uncorrected gross errors in the data set, or because the standard size chosen for the composite construction is well outside the size-range of fish in the data set. The program, CTRUSS, warns the user when this has occurred by printing a message at the bottom of the file assigned to unit 6, and called LDEBUG.

The adjusted "average" truss dimensions are then written into a standard file. They are then read by another FORTRAN program, TRUSSD (Appendix A), that calculates the x-y coordinates of landmarks corresponding to these truss measures. On the CYBER installation at the Bedford Institute of Oceanography this program interfaces with a graphics package "DISPLA" which generates a plot file. The plotfile can then be plotted on a Tektronix or any type of flatbed plotter.

#### A STATISTICAL APPROACH: MULTIVARIATE ANALYSIS OF FORM

The quantification of form has also been approached from a statistical point of view. As stated above, the presence of allometry can be tested by a simple power function, described by Huxley (1924) and Tessler (1936) as  $Y = BX^K$ , where Y is a body part, X a measure of body length, and b and K constants which can be estimated by a linear least-squares regression on a logarithmic transformation (Gould 1966).

However Jolicoeur (1963) subsequently realized that the allometric equation was too simplistic in its portrayal of growth, and attempted to derive a generalization to the multivariate case, using the direction cosines of the logarithmic covariance matrix. Unfortunately, the theoretical assumptions of this method are still not understood adequately to assess the effect on each variable (Sprent 1972).

Another approach to the problem, has been to use a number of multivariate techniques in conjunction with each other. Three methods commonly applied, although rarely together are principal component analysis (PCA), discriminant function analysis (DA), and multivariate analysis of variance (MANOVA). These techniques are described in various texts (e.g. Morrison 1967, 1976; Pimental 1979; Reyment et al. 1984), and will not be described in detail here. Suffice to say, however, that despite the widespread use of multivariate techniques in biological studies of growth and systematic variation, many of the statistical ramifications with regard to systematics have not been fully appreciated. Indeed, the multivariate approach to size and shape variation is still in an exploratory phase. With this in mind, only the important statistical operations performed by these methods will be stressed in this work.

#### Multivariate Procedures: Pattern Recognition in Morphological Hyperspace

Principal component analysis is an ordination technique, which can be used to simultaneously examine variation in a number of characters: there are two types -- R-mode and Q-mode. In an R-mode analysis where  $X$  is an  $n \times p$  data matrix, the  $p \times p$  minor product matrix is  $X'X$ , compared to a Q-mode analysis performed on the  $n \times n$  major product matrix,  $XX'$ . In R-mode analysis the relationship between variables is of interest, whereas in Q-mode the definition of inter-object similarity is of interest. Thus in an R-mode analysis, variation can be regarded as the differential response of characters along an ontogenetic trajectory. Each character will show a range in its response, which can be expressed statistically as its variance.

The main purpose of PCA is to describe parsimoniously the total variance for all characters, in as few dimensions as possible. The derived dimensions are linear combinations of the original variables, that successively account for the major patterns of variation. The relationship between each dimension is an expression of the correlation or covariance matrix of the original data set.

Geometrically, the relationship can be defined as the cosine of the angle between two vectors, which describe the equiprobability contours of two variables (Fig. 7a). If two variables are uncorrelated, then the vectors are orthogonal; whereas, if they are correlated, then the corresponding contours are elliptical. The first principal component lies along the long axis of the ellipse. The second component lies orthogonally to it, and is regarded to be independent of the first component (Fig. 7b).

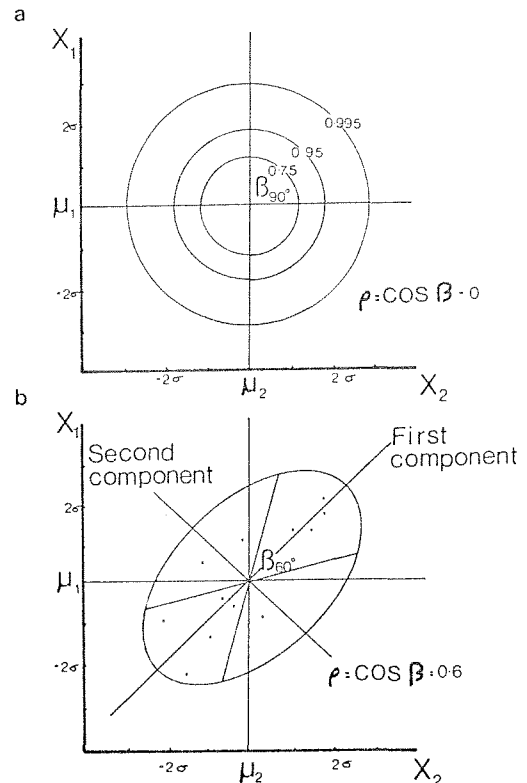


Fig. 7a. Equiprobability contours ( $p = 0.75, 0.95, \text{ and } 0.995$ ) for uncorrelated variables  $X_1$  and  $X_2$  (after Morrison, 1976).  $\mu_1$  and  $\mu_2$  refer to two vectors drawn at zero variance; these axes intersect at  $90^\circ$  thus the cosine of  $\beta = 0$ . Pearson's product moment correlation coefficient ( $\rho$ ) = 0.6.

7b. The lines of equiprobability are elliptic, defined by  $\beta = 60^\circ$ , and with  $\rho = 0.6$ .

Algebraically, the principal components can be obtained by solving for the eigenvalues and eigenvectors of the correlation or covariance matrix. The eigenvalues are derived from the characteristic equation of the matrix used:

$$D - \lambda I = 0$$

where  $D$  is a square matrix, of order  $p$ ,  $\lambda$  is the scalar (eigenvalue), and  $I$  is the identity matrix of order  $p$ . Each eigenvalue has an associated eigenvector, which satisfies the function:

$$(D - \lambda_i I) a_i = 0$$

where  $\lambda_i$  is one of the  $p$  eigenvalues, and  $a_i$  is one of the corresponding eigenvectors.

The eigenvalue indicates the length of each principal component. When these are standardized to represent the total variance, then their eigenvectors are equivalent to the principal components that they represent. In biological terms, the first principal component has been considered representative of overall size effects (Lee 1971, Kuhry and Marcus 1977), and the second component indicative of shape, an argument determined by the bipolarity of the loadings (Pimental 1979, Reyment *et al.* 1984). However, detailed work on this aspect of principal components analysis has yet to be done.

A second multivariate technique, discriminant function analysis (DA), is useful where a priori knowledge of the groupings is available. This method essentially assigns an individual to a group, based on a number of predesignated groups. The functions are derived from the inverse matrix of the within group variance-covariance matrix, and are given by:

$$Y = (\bar{x}_1 - \bar{x}_2)' S^{-1} X$$

where  $\bar{x}_1$  and  $\bar{x}_2$  are the mean vectors for the respective samples,  $S^{-1}$  is the inverse matrix, and  $X$  is a vector of variables. The details of this method are given in a number of statistical texts, such as Morrison (1976) and Reyment *et al.* (1984).

Biologists have used the technique in two ways, 1) the description of the differences between the groups on the basis of the sample data, and 2) the allocation of future elements, whose origins are not known with certainty (Habbema and Hermans 1977). Usually, the differences between groups are expressed in terms of a distance measure, such as the Mahalanobis' (1936) generalized distance,  $D^2$ , or Wilks' (1932) criterion. The former has a number of important properties, one of which is useful in taxonomic research, namely that the contribution of each character to the  $D^2$  can be ascertained (Rao 1952).

The linear discriminant function is connected with the Mahalanobis' generalized distance as follows:

$$D^2 = (\bar{X}_1 - \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2) = d'a$$

where vector  $d$  is the difference between the two sample mean vectors (Reyment *et al.* 1984). To test whether the centroids of two groups are significantly different, an F-test can be used, where:

$$F(p, n - g - p + 1) = \frac{D^2 \left( \frac{1}{n_A} + \frac{1}{n_B} \right)^{-1}}{\frac{(n - p - 1)}{p(n - g)}}$$

where  $n$  is the total number of observations, and  $g$  is the number of groups.

Some of the most important points to consider at the beginning of this analysis are 1) the costs of assignment to a particular group, 2) the a priori probability of belonging to one of the groups, and 3) the number of groups involved (Lachenbruch and Goldstein 1979). Other aspects that should be considered are 1) homogeneity of the within-group variation, 2) multi-normal distributional assumptions, 3) criteria for selection of variables, 4) the way in which selection is made during the computations, and 5) the derivation of the estimates of posterior probability. These points are discussed in most texts on multivariate techniques. Overall, this method provides a robust test of group association, and can be quite effective in an exploratory study of inter-group relationships.

The criteria for selection of variables is often the highest F-value. Generally, this method tends to distinguish well-separated groups further, instead of trying to separate poorly defined groups. A jackknife procedure is often applied (Jennrich and Sampson 1983) to reduce bias in the final computations of the discriminant functions. Each case is eliminated, in turn, from the computations of the group means and cross-products (Lachenbruch and Mickey 1975). Homogeneity of the within-group covariance matrices can be tested with a chi-square test (Kendall and Stuart 1966).

#### The Problem of Size

Much of the variation in many morphometric data sets, including our own, can be attributed to size. But defining what is meant by size is more difficult than might first appear. Early workers interested in allometry treated size as a linear distance measure from the extreme anterior to the extreme posterior of the organism (Huxley 1924, Gould 1966). This is still used since the multivariate portrayal of allometry (Jolliffe 1963) is not yet well developed (Sprent 1972). But most recent workers prefer a multivariate definition of size especially those who are interested in the genetic basis of form (eg. Thorpe and Leamy 1983): they consider size to be an unmeasured latent variable which explains the observed correlations among the morphometric variables best, arguing that this avoids confounding the variation of an explicitly measured size variable with that of the morphometric variables (Bookstein *et al.* 1985). The measure of size used for an individual specimen is often its score on the first principal component from a R-mode principal component using the covariance matrix (Humphries *et al.* 1981, Thorpe and Leamy 1983, see preceding section).

Differences in form among different groups of fish can be subdivided into differences in size and differences in shape. Differences in mean size at sexual maturity or in mean adult size are often important in characterizing different groups of fish (McGlade 1981). But differences in mean size among groups can also result from inadequate sampling. Thus it is often important to distinguish differences in shape from differences in size.

Three major classes of techniques have been used to separate size from shape: ratios, residuals from regressions against size, and

multivariate analyses. We will briefly discuss each of these three methods, but for a more thorough review see Humphries *et al.* (1981) or Reymont *et al.* (1984).

Ratios of morphometric variables over an explicit size measure such as total body length have been extensively used but have come under considerable criticism. A ratio will not be constant within a group for individuals showing differential growth, and the effect of allometry may be large compared to the differences in shape among groups (Reymont *et al.* 1984). The use of ratios assumes that a linear regression for a morphometric measure versus total length would give a good fit and would pass through the origin, something which may not be true even if the variables are first log-transformed (Thorpe and Leamy 1983). Finally the use of a ratio of a morphometric character over an explicitly measured variable such as total length confounds the variance in the numerator with that of the denominator (Bookstein *et al.* 1983). Moreover, the statistical properties of a ratio of two random variables can cause problems which may not be noticed by non-statisticians (Atchley *et al.* 1976). We think that all of these problems make it inadvisable to use ratios to correct for size -- but not everyone agrees with this conclusion (Mosimann and James 1979).

A second method of removing size is to undertake a statistical analysis of shape on the residuals from univariate regressions of the morphometric variables against size (Thorpe 1976). The patterns of growth and thus any allometry is likely to be different for groups from different environments. Therefore a separate univariate regression for each morphometric measure against size must be done separately for each group. The consequences of trying to fit regression lines to the pooled data from all the groups combined are shown in Fig. 8; if there are different patterns of growth in the different groups, the residuals will be non-random. One method of avoiding this is to use an average of the within-group slopes (Thorpe 1976), but this will not work if the within-group slopes are substantially different (Bookstein *et al.* 1985).

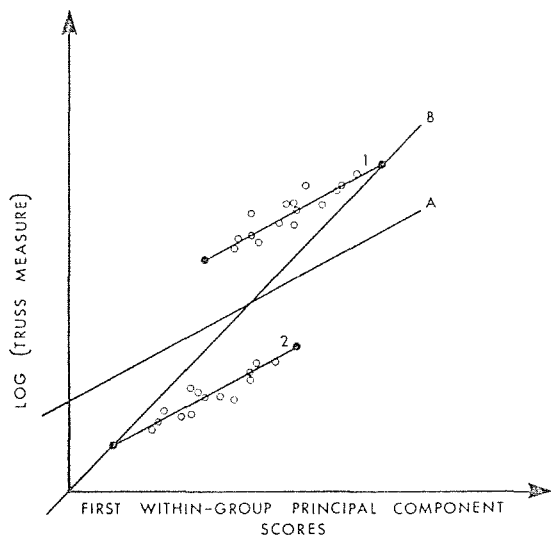


Fig. 8. Regression of two sets of data pooled to give a single regression line using Thorpe's (1976) method.

The third method of separating size from shape is with multivariate techniques, primarily principal components analysis. In practice the scores of the individual specimens on the first principal component are highly correlated with traditional measures of size such as total length. The scores on the second and subsequent principal components are then considered as measures of shape. The scores on the first principal component are plotted against those on the second principal component, and the plot examined to see if there is any separation of fish from different locations. If there is any differentiation along the axis of the first principal component then it is usually attributed to differences in size, and if there is separation along the axis of the second principal component then it is attributed to differences in shape. The procedure can be repeated using the third or subsequent principal component scores in place of the second if they account for a significant proportion of the variance. The strength of this method is that no a priori grouping of sample locations is required.

The major criticism of the technique however, is that the first principal component does not necessarily account for all the size variation in the data, so that the remaining components might contain a mixture of size and shape (Humphries *et al.* 1981; McGlade 1981). A method of shearing the scores on the second and subsequent principal components to remove size has been proposed (Humphries *et al.* 1981). This method has the disadvantages that it treats fish from different sample locations differently and that it is difficult to program on non-Michigan Terminal System (M.T.S) operating systems (although it has been done using IMSL subroutines (Winans 1984)). Indeed this shearing procedure is probably unnecessary unless a more aesthetic plot is desired. Any dependence of the second principal component scores on size can usually be seen in the original plot as the magnitude of scores on the second principal component will show a linear increase with those on the first (Humphries *et al.* 1981). This will most often occur when fish from different genera are being compared and not when differentiating stocks within a species where the patterns of growth are quite similar.

#### Instructions

Some parts of the procedure for analysing the truss measures will depend on the type of study. The first step, however, is always the same and consists of verification of the data-set (see previous description).

The next step depends on whether there is an a priori hypothesis about the subpopulation structure in the fish population. If there is no such hypothesis then a principal components analysis should be undertaken on the log-transformed truss measures using the covariance matrix. The resulting plots of the principal component scores can then be examined for clusters. Identification of the groups to which the points belong can be facilitated by saving the data and scores from this analysis in a BMDP save file and using BMDP6D (Appendix C) to produce a plot that is labelled with the group identity. If such plots reveal clustering by group then these groups can be considered subpopulations.

## AN EXAMPLE OF RESULTS -- THE SCOTIAN SHELVE GADOLDS

We will present three examples that use the techniques described in this report: 1) an investigation of stock structure in Scotian Shelf pollock (*Pollachius virens*), 2) an investigation of stock structure in Scotian Shelf haddock (*Melanogrammus aeglefinus*), and 3) a comparison of the morphology and of the growth patterns of pollock and haddock.

The fish were collected during three cruises on the *Lady Hammond* by Marine Fish Division personnel. These were cruise H088 on January 6-11, 1983 (Fig. 9), cruise H089 on January 12-18, 1983 (Fig. 9), and cruise H110 from November 28 to December 8, 1983 (Fig. 10). The sampling design was a compromise between what was desirable for stock identification purposes and what was desirable for other objectives of the Marine Fish Division.

A seven cell truss was used as a basis for collecting the morphometric data (Fig. 2). Meristic counts were also made on these fish and tissue samples were taken for electrophoresis: these data will be published elsewhere.

### GULF OF MAINE AND SCOTIAN SHELVE POLLOCK

The pollock were all collected during cruise H110 (Fig. 10). Fish were combined from sets 18 and 19 (hereafter set 18) which gave a total of 100 fish in set 3, 32 fish in set 12, and 100 fish in set 18.

We used BMDP4M (Appendix C) to do a principal components analysis on the correlation matrix of the log<sub>10</sub> - transformed truss measures. (We suggest that workers use the covariance matrix, but in this case, the results are very similar.) We then used BMDP6D (Appendix C) to produce labelled plots of the scores on the principal components.

The first principal component has high positive loadings from all of the truss measures, except for B1 (Table 1). The loadings suggest that the first principal component does in fact embody general size. Moreover, as the first principal component explains 89.6% of the variance, we can conclude that most of the variance in the pollock morphometric data can be attributed to variation in size. The only truss measure somewhat independent of size is B1 (Table 1), the distance between the attachment of the pectoral and the attachment of the pelvic fins (Figs. 1 and 2).

The second principal component has bipolar loadings for the variables (Table 1): the importance of a variable's contribution to a principal component score can be determined by comparing the magnitude of the absolute value of its loading coefficient to that of other variables. In this example, the second component shows a high positive loading for B1 and a moderate, negative loading for E3 (the distance between the second and third dorsal fins). This means a fish would have a high second principal component score if its value of B1 was large and

its value of E3 was small. This component accounts for an additional 2.6% of the variance.

The third principal component has a high positive loading from E3, moderate positive loadings from B1, C1, E1, E5, and G3, and small negative loadings from most of the rest of the variables (Table 1). It accounts for only 1.5% of the variance so will not be discussed further.

Before plotting the scores for a given component, the program standardizes them by subtracting their mean, dividing them by the standard deviation and then squaring them. The plot of the scores on the first principal component versus those on the second principal component shows two distinct clusters, one containing mostly fish from set 18 and one containing mostly fish from sets 3 and 12 (Fig. 11). The clusters are separated along the axis of the first principal component but not along that of the second suggesting that the separation is by size rather than by shape. Indeed the mean fork length, another measure of size, of fish from set 18 was only 37.0 cm compared to the mean of 67.9 cm for sets 3 and 12 combined.

We have constructed composite forms for average pollock of fork lengths of 25 cm, 50 cm, and 75 cm (Figs. 12a, b, c) using the regression coefficients for the data from all the sets combined (Table 2). Note that the distortion, the amount of correction needed to allow the average truss measurements to be constructed into an average form, is very small. The 50 cm form is a) proportionately shorter in cell A, in side B1, and in cell F, b) proportionately longer in cell D, and c) proportionately narrower in cell E compared to the 25 cm form. The same trends are noticeable when the 75 cm form is compared with the 50 cm form. Thus as a pollock grows from 25 to 75 cm, its head, the distance between the pectoral and the pelvic fins, and the body region below the 3rd dorsal fin become proportionately shorter, the body region below the 2nd dorsal fin becomes proportionately longer, and the body region between the 2nd and 3rd dorsal fins becomes proportionately narrower. Pollock thus, exhibits considerable differential growth of its various body regions.

We also constructed composite forms for the two groups separated in the principal component analysis: one for sets 3 and 12 and one for set 18. Both of these forms were constructed for an average fish of the size corresponding to the mean fork length for that group (Figs. 13a, b). The differences in shape between these two composite forms are what we would have expected for a 37.0 cm pollock and a 68.0 cm pollock on the basis of the changes we saw between the smaller and larger forms constructed from all the sets combined.

### GULF OF MAINE AND SOUTH-WEST NOVA SCOTIAN HADDOCK

The haddock were collected from sets 9, 10, 30, 31, 37, 41, 42, and 43 on cruise H088 (Fig. 9), from sets 3, 4, 6, 11, 28, 29, and 30 on cruise H089 (Fig. 9), and from sets 10 and 19 on H110. To avoid confusion, set 30 from H088 was recoded as set 88, set 30 from H089 was recoded as set 89, and sets 10 and 19 from H110 were recoded as sets 70 and 79 respectively.

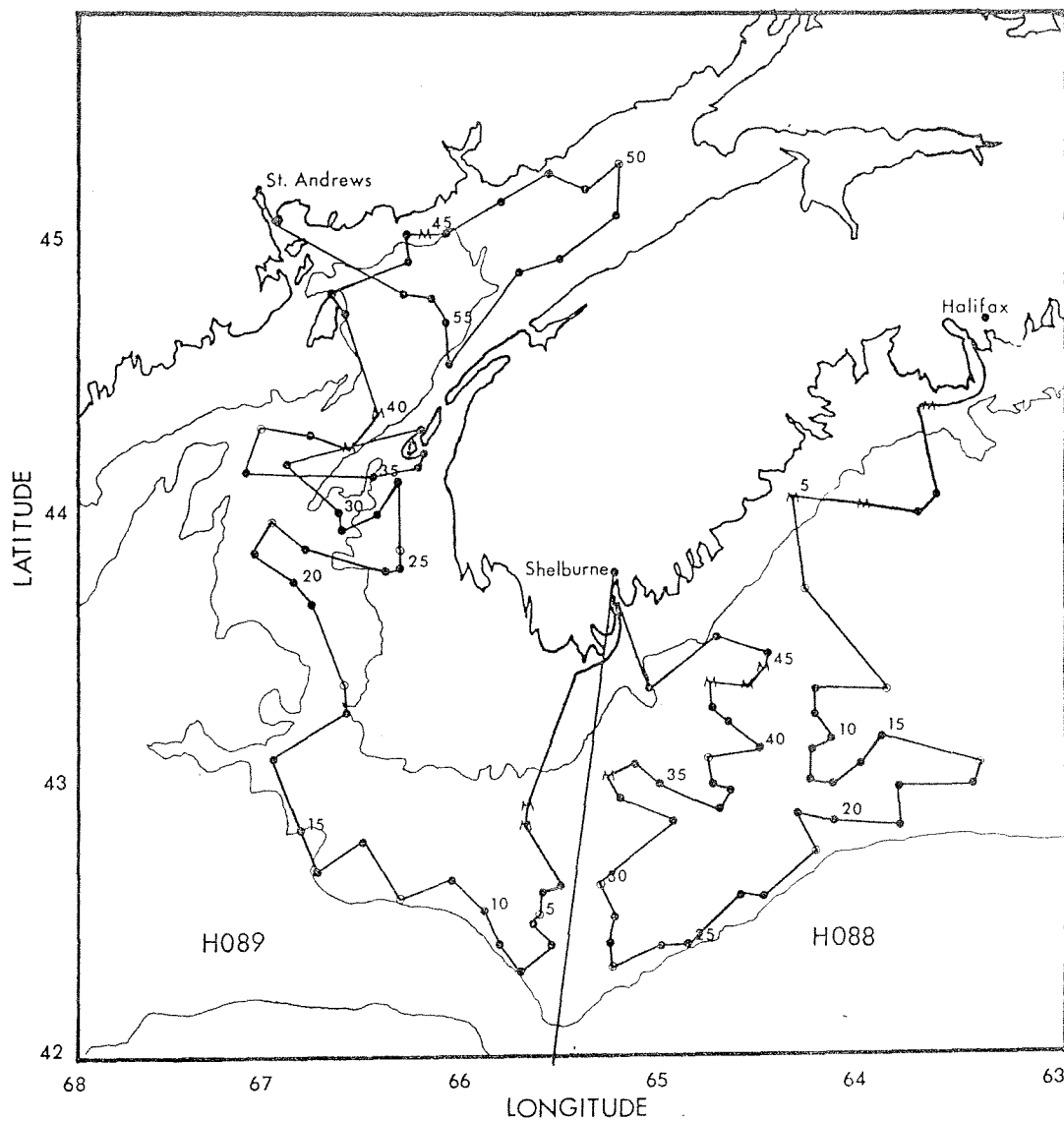


Figure 9. Cruise track off south-western Nova Scotia for the R.V. Lady Hammond, H088/89.

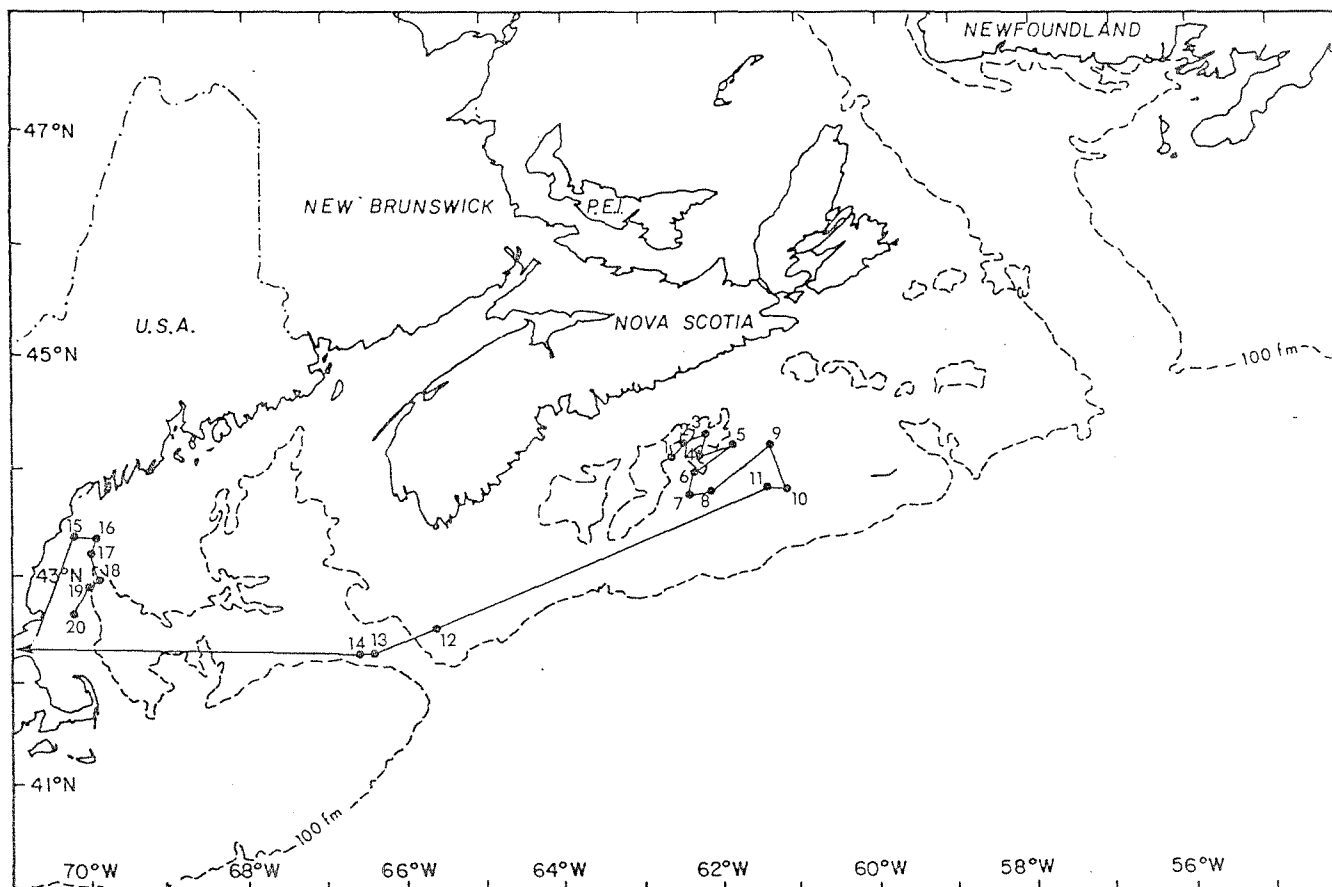


Figure 10. Cruise track for the R.V. Lady Hammond on the Scotian Shelf and Gulf of Maine, H110.

Table 1. Principal component scores for each truss measure of pollock, the variance explained and cumulative percentage for the first three unrotated components (PC1, PC2, PC3) (using the correlation matrix).

	PC1	PC2	PC3
A1	.991	-.021	-.021
A2	.981	-.017	-.037
A3	.990	-.054	-.007
A4	.996	-.006	-.023
A5	.991	-.012	-.026
A6	.990	-.026	-.023
B1	.365	.909	.176
B3	.984	-.000	-.075
B4	.995	-.030	-.034
B5	.993	-.005	-.027
B6	.996	-.006	-.028
C1	.973	-.125	-.016
C3	.767	-.020	.122
C4	.989	-.022	-.013
C5	.990	-.014	-.011
C6	.995	-.033	-.033
D1	.988	-.002	-.048
D3	.982	.006	-.065
D4	.990	.034	-.007
D5	.985	.017	-.049
D6	.909	-.017	-.063
E1	.787	-.068	.255
E3	.704	-.227	.605
E4	.993	.030	-.039
E5	.993	.019	.005
E6	.991	.017	-.013
F1	.982	-.017	-.064
F3	.985	.014	-.066
F4	.972	.021	-.064
F5	.930	.012	-.129
F6	.990	-.010	-.046
G1	.901	-.032	.099
G3	.918	-.026	.114
G4	.932	.121	-.011
G5	.934	.048	.017
G6	.964	.037	.014
Variance Explained	32.271	.930	.559
Cumulative %	89.6%	92.2%	93.7%

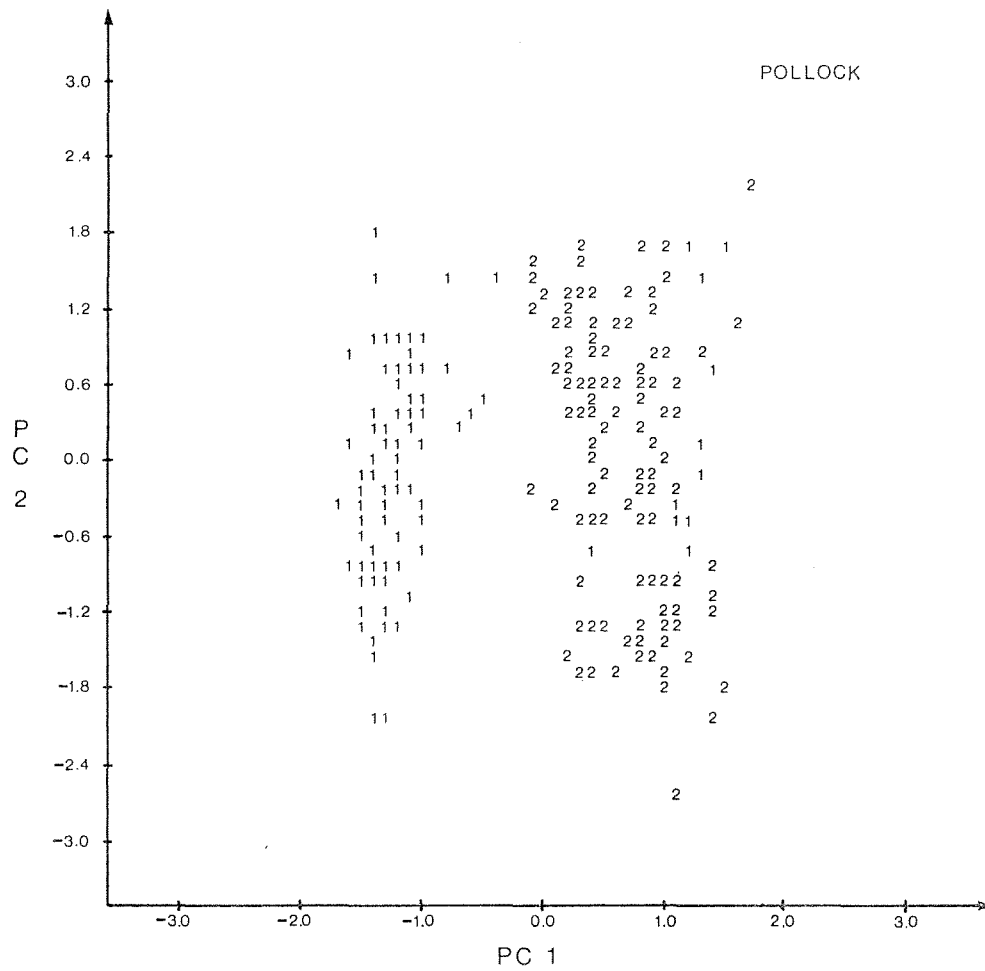


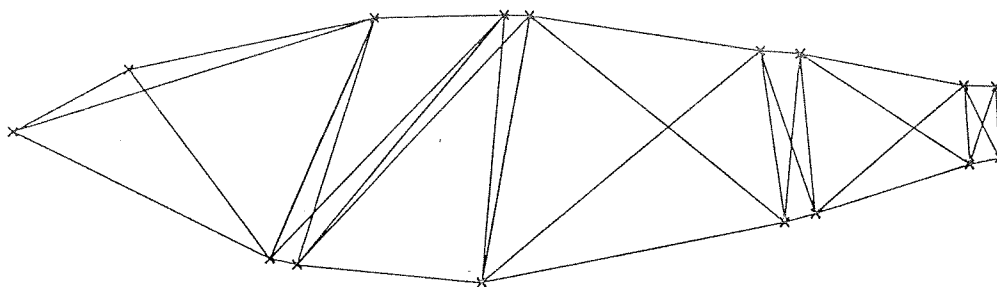
Figure 11. Plot of the first and second principal components (PC1 and PC2) from an analysis of pollock; 1 = Gulf of Maine; 2 = Scotian Shelf.

## POLLOCKALLSETS

a

TOTAL FISH LENGTH IS 25.0CM

DISTORTION IS 0.068

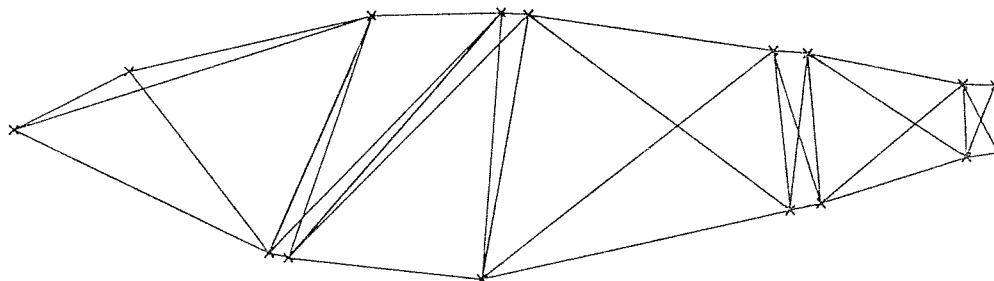


b

## POLLOCKALLSETS

TOTAL FISH LENGTH IS 50.0CM

DISTORTION IS 0.037



c

## POLLOCKALLSETS

TOTAL FISH LENGTH IS 75.0CM

DISTORTION IS 0.027

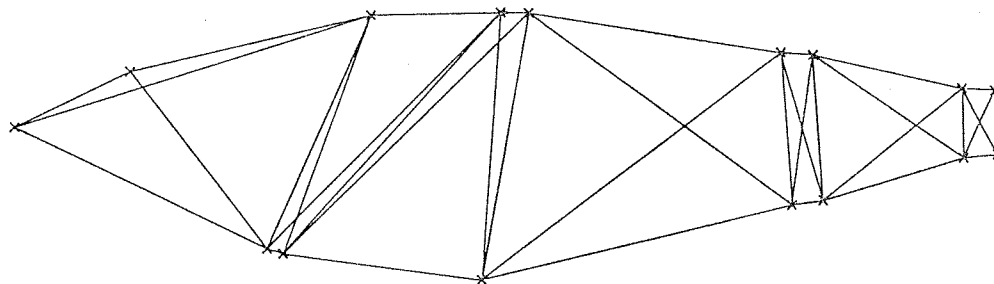


Figure 12. Composite forms for pollock of fork lengths a) 25 cm; b) 50 cm; and c) 75 cm using all sets combined.

Table 2. Regression coefficients for pollock for the first principal component loadings of each truss measure and the log-transformed measures versus total length, for all sets combined used to reconstruct a composite truss form.

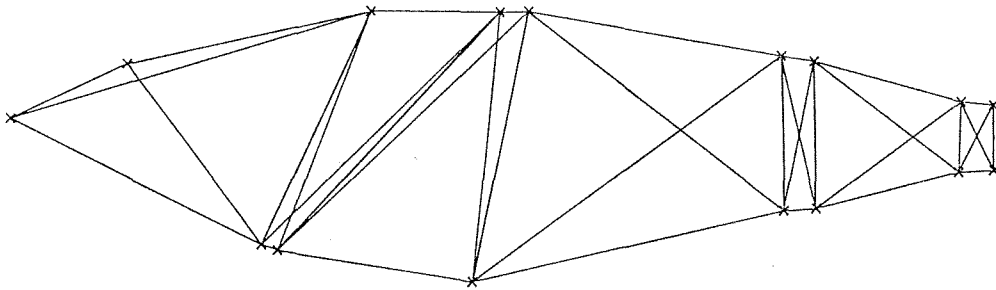
Intercept	Slope	r	X Variable	Y Variable
-7.19444	4.24707	.90	Total Fork Length	1st PC scores
1.09960	.18352	.99	1st PC scores	Truss (1,1)
.75694	.18565	.98	1st PC scores	Truss (2,1)
1.03883	.18900	.99	1st PC scores	Truss (3,1)
1.05631	.18522	1.00	1st PC scores	Truss (4,1)
1.00253	.18234	.99	1st PC scores	Truss (5,1)
1.21779	.18934	.99	1st PC scores	Truss (6,1)
-.04523	.08531	.37	1st PC scores	Truss (1,2)
1.05631	.18522	1.00	1st PC scores	Truss (2,2)
.75805	.18761	.98	1st PC scores	Truss (3,2)
1.15565	.19050	1.00	1st PC scores	Truss (4,2)
1.05257	.18765	.99	1st PC scores	Truss (5,2)
1.16835	.18673	1.00	1st PC scores	Truss (6,2)
.93428	.20661	.97	1st PC scores	Truss (1,3)
1.15565	.19050	1.00	1st PC scores	Truss (2,3)
.08225	.21378	.77	1st PC scores	Truss (3,3)
1.08282	.18861	.99	1st PC scores	Truss (4,3)
1.07562	.18863	.99	1st PC scores	Truss (5,3)
1.18056	.19194	.99	1st PC scores	Truss (6,3)
1.14781	.19487	.99	1st PC scores	Truss (1,4)
1.08282	.18861	.99	1st PC scores	Truss (2,4)
1.04183	.20728	.98	1st PC scores	Truss (3,4)
.85108	.16474	.99	1st PC scores	Truss (4,4)
1.15744	.19206	.99	1st PC scores	Truss (5,4)
1.21054	.20103	.91	1st PC scores	Truss (6,4)
.14437	.18440	.79	1st PC scores	Truss (1,5)
.85108	.16474	.99	1st PC scores	Truss (2,5)
.18378	.14433	.70	1st PC scores	Truss (3,5)
.82763	.16964	.99	1st PC scores	Truss (4,5)
.85329	.16743	.99	1st PC scores	Truss (5,5)
.84496	.16661	.99	1st PC scores	Truss (6,5)
.82821	.17016	.98	1st PC scores	Truss (1,6)
.82763	.16964	.99	1st PC scores	Truss (2,6)
.84594	.17200	.99	1st PC scores	Truss (3,6)
.51014	.16379	.97	1st PC scores	Truss (4,6)
.92343	.16812	.93	1st PC scores	Truss (5,6)
.91105	.17103	.99	1st PC scores	Truss (6,6)
.15698	.21953	.90	1st PC scores	Truss (1,7)
.51014	.16379	.97	1st PC scores	Truss (2,7)
.15315	.20347	.92	1st PC scores	Truss (3,7)
.47520	.17005	.93	1st PC scores	Truss (4,7)
.53589	.17442	.93	1st PC scores	Truss (5,7)
.53483	.16935	.96	1st PC scores	Truss (6,7)

a

POLLOCKSETS3AND12

TOTAL FISH LENGTH IS 68.0CM

DISTORTION IS 0.063



b

POLLOCKSET18

TOTAL FISH LENGTH IS 37.0CM

DISTORTION IS 0.080

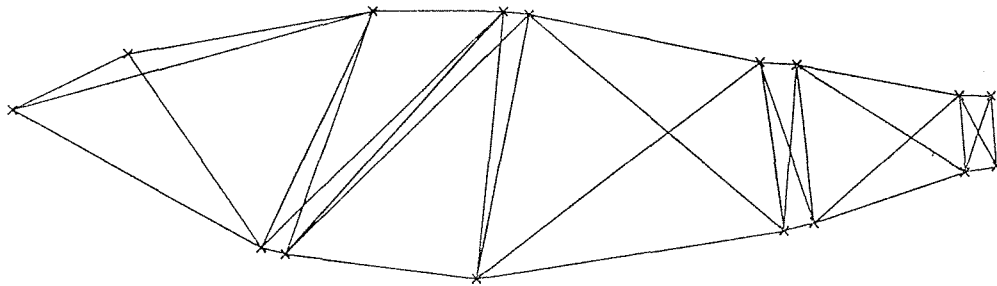


Figure 13. Composite forms for pollock from a) the Scotian Shelf and b) the Gulf of Maine corresponding to a mean fork length of 68 cm and 37 cm, respectively.

A principal components analysis was again undertaken on the correlation matrix, for all the sets from all the cruises combined. The first principal component accounted for 89.1% of the variance (Table 3). This component had high positive loadings for all the variables including B1. As was true of the pollock data then, most of the variance in the haddock data could be attributed to variation in size. The second principal component had large positive loadings from C1, E1, E3, G1, and G3, a large negative loading from B1, and accounted for an additional 2.4% of the variance (Table 3). The third principal component accounted for only an additional 1.4% of the variance so will not be discussed.

BMDP6D can be used to label the points on the plots of the principal component scores in different ways depending on the hypotheses about stock delineation. We initially used a different symbol for each of the sets but the resulting plots showed no evidence of clustering by set, so we have not included them. We then decided to aggregate the sets into four geographic groupings: X containing sets 3, 4, 6, 9, 10, 11, 31, 37, 41, 42, 43, and 88 (Browns Bank - LaHave Basin), F containing sets 28, 29, and 89 (mouth of the Bay of Fundy), W containing set 70 (Emerald Bank), and Y containing set 79 (Gulf of Maine) (Fig. 9). We drew an envelope around the extreme points for each group to facilitate interpretation of the plots.

On the plot of the first principal component scores versus the second principal component scores (Fig. 14), groups W (Emerald) and Y (Gulf of Maine) are separated from groups X (Browns - LaHave) and F (Fundy) along the axis of the first principal component. Fishes from groups W and Y are on average larger than fishes from groups X and F, but it will not be possible to tell if this difference is real or a sampling artifact until more than one set is available from each of areas W (Emerald) and Y (Gulf of Maine). There was no separation of these four groups along the axis of the second principal component although group Y has a more restricted distribution than the other three groups.

To define stocks on the basis of morphometric characters alone, the robustness of the classification should be tested by classifying fishes of unknown origin to one of the previously identified stocks. Before this can be done a classification function must be constructed from a linear combination of the morphometric variables for each of the stocks. We have done this for the four geographical groupings of haddock using a discriminant analysis program (BMDP7M, Appendix C) which uses a stepwise method of entering and removing the variables into the classification functions in order to select the most discriminating set of variables. The use of these four groupings is meant as an example only -- we are not making implications about the stock structure of haddock on the Scotian Shelf. With the F value to enter set at 2.00 and the F value to remove set at 1.99, the classification functions contained coefficients for 18 of the 37 variables on step 18, the final step (Table 4). The jackknifed classification matrix shows that

these classification functions correctly classified an average of 70.7% of the fish into the correct geographical grouping (Table 5).

Group X (Browns - LaHave) had the highest number of its fish correctly classified (83.1%) while group F (Fundy) had the lowest (34.4%). Fish from the Browns - LaHave group were most often misclassified into the Fundy group and vice-versa. Fish from Emerald and from the Gulf of Maine were most often misclassified into the Browns - LaHave group. Most of the dispersion between the centroids of the groups (86.5%) was accounted for by the first two canonical variables (Table 6). The plot of these variables shows while the centroids of group W (Emerald) and group Y (Gulf of Maine) are well separated from those of the other groups, the envelopes drawn around the extreme points show considerable overlap (Fig. 15). We suspect that at least some of this separation of the centroids is attributable to differences in mean size among the groups, especially in light of the results obtained from the principal component analysis.

We constructed composite forms for haddock at three sizes: 25 cm, 50 cm, and 75 cm using the data combined for all sets from all cruises. The 25 cm form is proportionately much longer in cell A and proportionately shorter in cell C and D than the 50 cm form (Figs. 16a, b, c). As a haddock grows from 25 cm to 75 cm its head gets proportionately smaller and its body region below the 2nd and 3rd dorsal fin gets proportionately longer. We have constructed composite forms at 50 cm for each of the four geographical groupings. All four forms are similar (Figs. 17a, b, c, d).

#### POLLOCK VERSUS HADDOCK

The pollock data, combined for all sets, were compared with the haddock data, combined for all sets. A principal components analysis was performed on the correlation matrix of the combined data from both species.

From the results (Table 7) we can see that the first principal component has large positive loadings from all the variables except B1. The coefficient for B1 is small relative to those of the other variables indicating it is relatively independent of general size. The first principal component explains 88.3% of the variance; even though we are now dealing with data from two species most of the variation in the data can still be attributed to variation in size. The second principal component explains an additional 3.7% of the variance. This component has a very high positive loading from B1 and high negative loadings from E1 and E3. Fish with a large score on this component will have a large value of B1 and a small value of E1 and E3. The third principal component accounts for an additional 1.4% of the variance and has large positive loadings from B1, C3, E1, E3, G1, and G3, and small negative loadings from most of the other variables (Table 7). Haddock and pollock show good separation on the axis of the 2nd principal component but not on the axis of the first (Fig. 18a). It is interesting that the best line separating the two clusters of points would be a diagonal one; this suggests the scores on the second principal component are somewhat size

Table 3. Principal component scores for each truss measure for haddock, the variance explained and the cumulative percentage for the first three unrotated components (PC1, PC2, PC3) (using correlation matrix).

	PC1	PC2	PC3
A1	.940	-.007	.016
A2	.971	.024	-.057
A3	.972	-.014	.019
A4	.992	-.025	-.012
A5	.978	-.019	.005
A6	.991	.001	-.011
B1	.731	-.221	-.041
B3	.954	-.085	-.014
B4	.990	-.027	-.008
B5	.990	-.024	-.011
B6	.991	-.050	-.014
C1	.953	-.006	-.022
C3	.834	.182	-.009
C4	.987	-.033	-.033
C5	.985	-.032	-.025
C6	.992	-.014	-.019
D1	.975	-.073	.024
D3	.975	-.084	.044
D4	.985	-.058	-.032
D5	.979	-.062	.013
D6	.976	-.060	.001
E1	.701	.449	-.396
E3	.704	.470	-.223
E4	.988	-.069	-.023
E5	.989	-.037	-.035
E6	.988	-.057	-.030
F1	.975	-.093	.015
F3	.977	-.067	-.009
F4	.967	-.061	-.017
F5	.972	-.081	.004
F6	.976	-.090	.024
G1	.751	.419	.406
G3	.827	.301	.331
G4	.964	-.060	-.027
G5	.973	.030	.074
G6	.962	.007	.053
Variance Explained	32.095	.854	.506
Cumulative %	89.1%	91.5%	92.9%

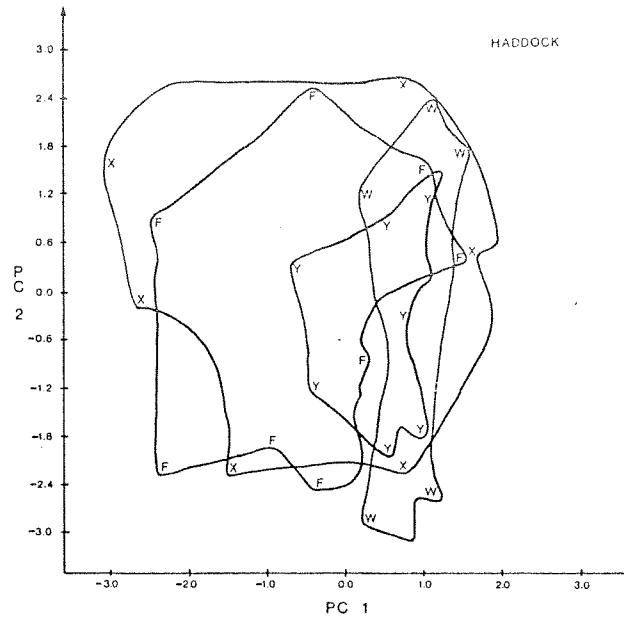


Figure 14. Plot of the first and second principal components (PC1 and PC2) from an analysis of haddock; X = Browns-LaHave Banks; F = mouth of the Bay of Fundy; W = Emerald Bank; and Y = Gulf of Maine.

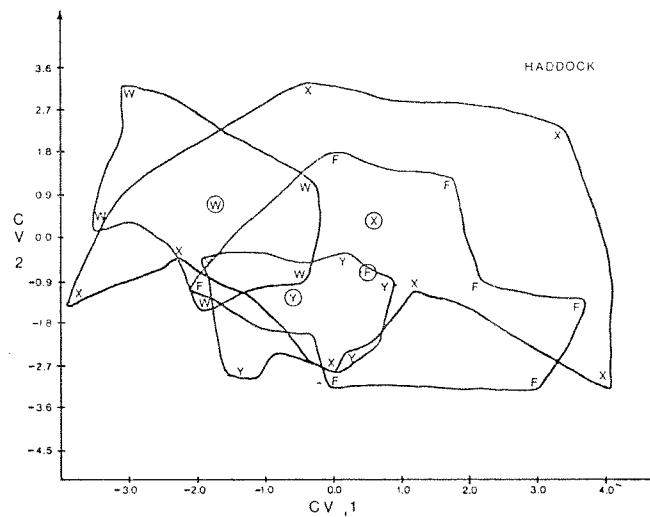


Figure 15. Plot of the first and second canonical variates (CV1 and CV2) from a discriminant function analysis of haddock; the centroids for each group are circled -- X = Browns-LaHave Banks; F = mouth of Bay of Fundy; W = Emerald Bank; and Y = Gulf of Maine.

Table 4. Results from the discriminant function analysis of haddock:  
classification functions for the four groups.

Variable	Browns-LaHave Banks (X)	Bay of Fundy (F)	Emerald Banks (W)	Gulf of Maine (Y)
A1	54.96812	46.88943	61.67882	55.48672
A2	-210.93507	-219.25172	-182.07651	-199.99408
A4	-230.67607	-216.12962	-214.52070	-157.65714
A5	111.34489	113.79213	135.39158	80.29010
A6	1252.60022	1223.34600	1228.68963	1185.74304
B1	-20.22112	-28.50871	-21.75976	-23.49483
B3	11.26280	2.24227	.60841	1.80206
C1	-207.84874	-203.06384	-197.26514	-184.05957
C3	-87.61634	-86.74213	-91.02581	-88.53851
C4	154.62978	158.01670	125.64992	124.75467
D1	-4.06531	1.03590	-5.55832	17.84244
E5	84.28142	78.13561	71.13342	56.48189
E6	-235.61437	-205.05537	-244.52142	-190.75460
F3	-127.75501	-134.83338	-121.06742	-141.51438
F4	-210.14415	-209.03612	-195.96999	-204.57436
G3	-79.56089	-85.86851	-85.07580	-83.33117
G5	-56.11385	-48.09416	-42.02010	-51.11458
G6	-69.74275	-64.21309	-75.84045	-71.38562

Table 5. Jackknifed classification matrix for the discriminant function analysis of the four groups of haddock.

Group	Percent Correct	Number of Cases Classified into Group			
		Browns-LaHave	Bay of Fundy	Emerald Bank	Gulf of Maine
Browns-LaHave	83.1	275	16	19	21
Bay of Fundy	34.4	47	33	2	14
Emerald Bank	72.0	17	6	72	5
Gulf of Maine	63.3	26	6	4	62
TOTAL	70.7	365	61	97	102

Table 6. Results from the discriminant function analysis of haddock.

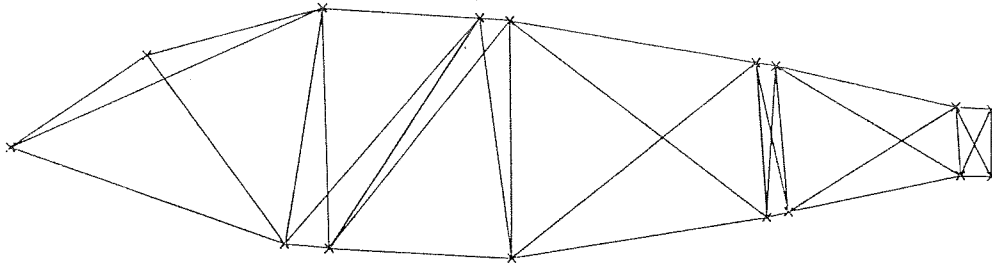
<u>Cumulative Proportion of Total Dispersion</u>			
	.53509	.86519	1.00000
<u>Canonical Correlations</u>			
	.64485	.55240	.38995
Variable	Coefficients for Canonical Variables		
A1	-3.43925	3.24647	4.91256
A2	-13.30696	3.87387	5.45256
A4	-14.21527	-31.63220	19.26072
A5	-4.95893	18.55716	-20.73224
A6	14.93822	30.63949	-3.98424
B1	.19512	3.04430	4.97345
B3	4.42015	3.81279	4.50425
C1	-6.66723	-9.08509	5.18741
C3	1.51346	-.59383	-.39484
C4	15.31848	5.54516	-10.39193
D1	-1.76055	-11.13310	6.48660
E5	8.03565	10.56778	-5.49537
E6	.62617	-28.93087	-.51023
F3	-1.48241	9.24386	-2.33259
F4	-6.04808	.58291	-1.09264
G3	1.98011	1.87047	4.06284
G5	-5.26035	-.82892	-6.47440
G6	3.09786	-2.00742	-3.65146
Group	Canonical Variables Evaluated at Group Means		
Browns-LaHave Banks	.55319	.36521	.16341
Bay of Fundy	.45503	-.67510	-.86231
Emerald Bank	-1.69539	.63118	-.22055
Gulf of Maine	-.58418	-1.21624	.51782

a

## HADDOCKALLSETS

TOTAL FISH LENGTH IS 25.0CM

DISTORTION IS 0.109

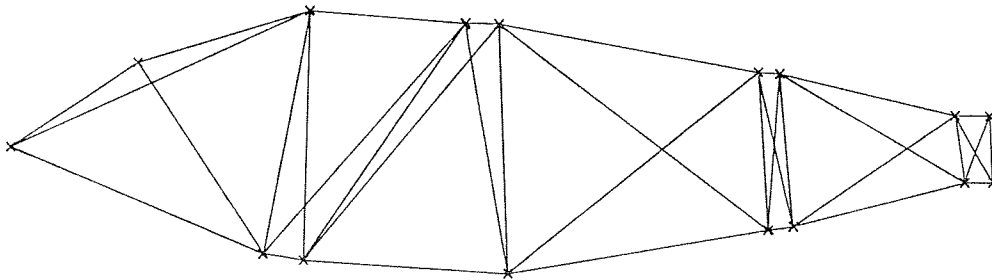


b

## HADDOCKALLSETS

TOTAL FISH LENGTH IS 50.0CM

DISTORTION IS 0.161



c

## HADDOCKALLSETS

TOTAL FISH LENGTH IS 75.0CM

DISTORTION IS 0.596

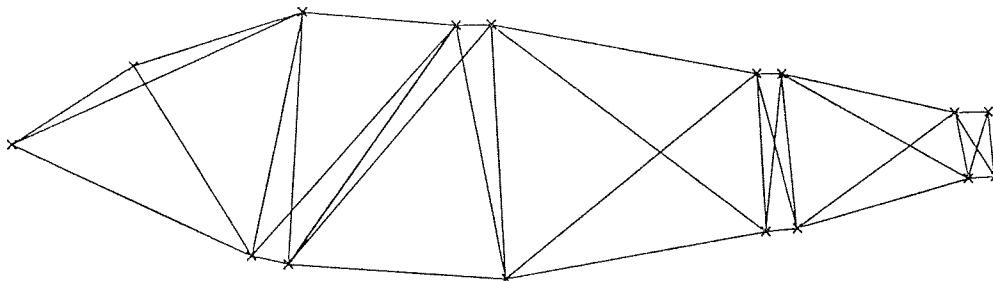


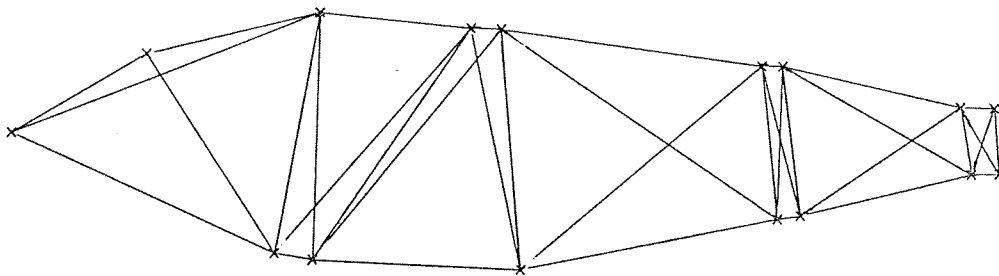
Figure 16. Composite forms for haddock of fork lengths a) 25 cm; b) 50 cm; and c) 75 cm using all sets combined.

a

HADDOCKGROUPX

TOTAL FISH LENGTH IS 50.0CM

DISTORTION IS 0.060



b

HADDOCKGROUPF

TOTAL FISH LENGTH IS 50.0CM

DISTORTION IS 0.450

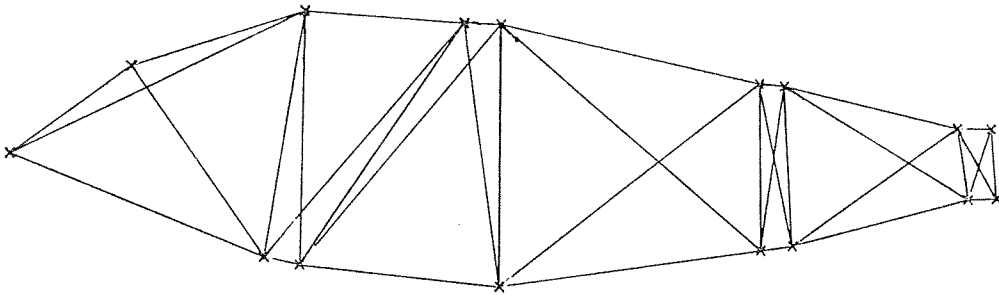


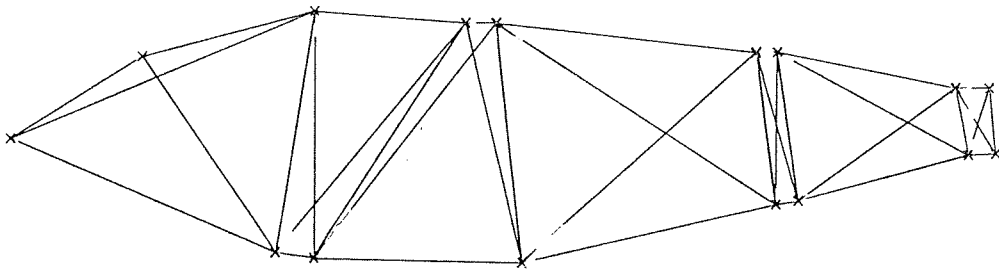
Figure 17. Composite forms for haddock from a) Browns-LaHave Banks (X); b) mouth of the Bay of Fundy (F); c) Emerald Bank (W); and d) Gulf of Maine (Y) all scaled.

c

HADDOCKGROUPW

TOTAL FISH LENGTH IS 50.0CM

DISTORTION IS 0.060



d

HADDOCKGROUPY

TOTAL FISH LENGTH IS 50.0CM

DISTORTION IS 0.015

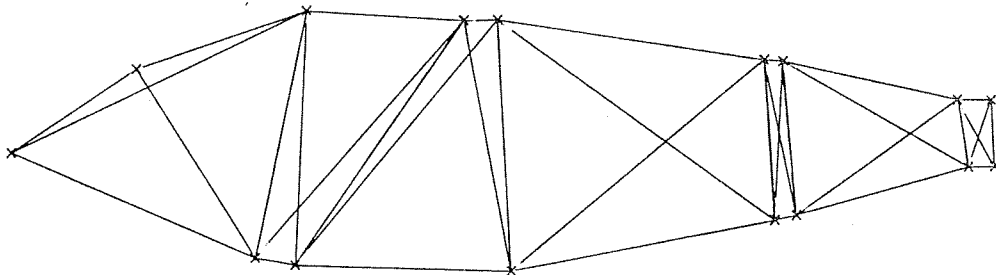


Figure 17. (Continued).

Table 7. Principal component scores for each truss measure for pollock and haddock, the variance explained and cumulative percentage for the first three unrotated components (PC1, PC2, PC3) (using correlation matrix).

	PC1	PC2	PC3
A1	.964	-.024	-.038
A2	.947	.171	.032
A3	.951	-.226	-.054
A4	.994	-.019	-.036
A5	.984	.011	-.041
A6	.986	-.094	-.025
B1	.364	.820	.240
B3	.938	.210	-.051
B4	.989	-.082	-.055
B5	.992	-.001	-.037
B6	.992	-.039	-.043
C1	.960	.025	-.046
C3	.793	.130	.276
C4	.987	-.056	-.046
C5	.988	-.028	-.040
C6	.991	-.076	-.044
D1	.973	-.098	-.059
D3	.975	.093	-.048
D4	.987	.015	-.056
D5	.981	.072	-.047
D6	.954	-.075	-.067
E1	.771	-.348	.202
E3	.736	-.476	.175
E4	.989	.043	-.062
E5	.990	.006	-.046
E6	.989	.019	-.056
F1	.960	.181	-.064
F3	.967	.163	-.041
F4	.971	-.033	-.067
F5	.948	.163	-.066
F6	.978	.097	-.064
G1	.833	-.175	.360
G3	.863	-.006	.306
G4	.959	.023	-.023
G5	.964	.010	.070
G6	.966	-.059	.033
Variance Explained	31.797	1.352	.504
Cumulative %	88.3%	92.0%	93.4%

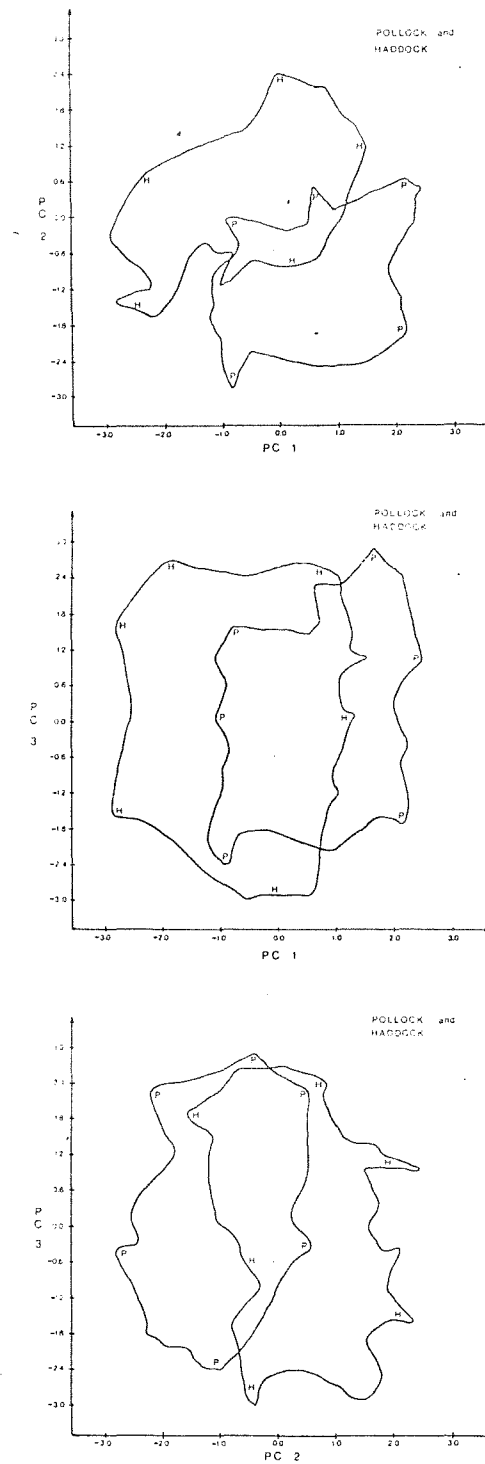


Figure 18. Plots of the first, second, and third principal components (PC1, PC2, and PC3) from an analysis of pollock (P) and haddock (H); a) PC1 versus PC2; b) PC1 versus PC3, and c) PC2 versus PC3.

dependent. This plot also shows the same separation of the pollock into two clusters, based on size, that we saw in the analysis of pollock only. The plot of the third principal component scores versus the first principal component scores shows fair separation of the two species along the axis of the first principal component but none along the axis of the third (Fig. 18b). The plot of the second principal component against the third confirms that there is no separation of the two species along the axis of the third principal component (Fig. 18c).

The patterns of growth in pollock can be usefully compared with those in haddock using the composite forms we discussed previously (Figs. 12 and 16). The 25 cm pollock form is proportionately narrower in cell A, and proportionately shorter in cell F and sides B1, C3, and proportionately longer in sides D1, E1, and E3 than the 25 cm haddock. The same trends are noticeable when the 50 cm pollock form is compared to the 50 cm haddock and the 75 cm pollock form is compared to the 75 cm haddock. The most noticeable differences in the growth of the pollock and the haddock, then, are in the body streamlining and in the distance between the pectoral and the pelvic fins. The shape of a pollock is such that the body does not reach its widest point until the end of the first dorsal fin whereas the haddock body reaches its widest point at the beginning of the first dorsal fin. As a pollock grows from 25 cm to 75 cm the distance between the pectoral and pelvic fins becomes proportionately shorter whereas in the haddock this distance stays in proportion. It would therefore be interesting to understand how these differences in morphology relate to ontogeny and the different environments that these fish live in -- the pollock is semi-pelagic and the haddock largely benthic -- although these ideas are beyond the scope of this report. In short, however, the techniques outlined above can differentiate the growth patterns observed within and between species, and can be used to give a quantitative estimate of the differences.

### CONCLUSIONS

Biologists can as a rule be divided into those who strive to capture diversity, and those who seek underlying unities in the more than one million species of living organisms. D'Arcy Wentworth Thompson was amongst the latter, as he searched for the basic immutable patterns of a bauhaus design in the organisms he studied. But as many biologists realize, these approaches are simply aesthetic styles that affect the modus operandi of science rather than different theories of biology. Indeed anyone who studies diversity would admit to common generating patterns, and any analysts of unity recognise the importance of particular expressions of a pattern. Thus the differences lie in the application and final intent. Population biologists tend generally to think in terms of point by point, or character by character differences, whilst systematicists generally look for continued similarities in overall characteristics. The point of departure is thus the event horizon against which the results are posited; in fisheries management the horizon is generally five to ten years, in ichthyology it can be marked in thousands of years.

From the analyses given above, it is clear that truss data sets provide a clear and reliable view of shape differences in gadoids. Moreover, comparisons with conventional morphological data sets suggest that truss data capture more information about local body proportions (pers. comm., J. McGlade). In the examples used in this study, point by point interspecific differences in the ontogeny of shape were identified, and certain areas isolated with respect to causal links between morphology and life-history. Thus the data from trusses provide a view for both the population biologist and the systematicist interested in long-term evolutionary patterns. The fact that the trusses can be used to identify those areas of body development which increase most rapidly with age, and describe the actual form of such growth is of real importance in a fishery where mesh controls are used as the conservation measure. Indeed, it would be possible to predict the shape of a selectivity ogive given the cross-dimensions of the mesh and a series of trusses for fish of different sizes. The impact of a mesh-regulation on a multispecies fishery could thus be modelled.

The long-term evolutionary view is given through comparisons of growth patterns for different species; as was shown in this study, growth rates of the area in front of the caudal peduncle and of the head region were different in haddock and pollock. Such differences may in fact represent an overall response to a semi-pelagic versus a benthic way of life -- hypotheses that could be open to analysis through experiments on swimming and observation of prey preference. More important, however, is that the truss analysis can be used to transcend the somewhat reductionist approach that generally ensues from taking each character and presuming in some naive way that every variation represents an optimal design. And this is clearly the most important conclusion that studies of morphology must realise, for organisms are directed and limited by their past, and hence imperfect and unpredictable in their form and function.

### ACKNOWLEDGEMENTS

The authors wish to acknowledge Toby Hay for the statistical help and criticism, V. Myra for typing the manuscript, Diane Beanlands for drawing the figures, and Paul Fanning for reviewing parts of the manuscript..

### REFERENCES

- Atchley, W.R., C.T. Gaskins, and D. Anderson. 1976. Statistical properties of ratios I empirical results. *Syst. Zool.* 25: 137-148.
- Beverton, R.J.H., and S.J. Holt. 1957. On the dynamics of exploited fish populations. *Fish. Invest., Ser. II, Vol. XIX.* London, 533 p.
- BMDP Statistical Software 1983 Ed. BioMedical Computer Programs Series, Univ. Calif., Berkeley, 734 p. BMDP-83 for NOS documentation, 1985 release Vogelback

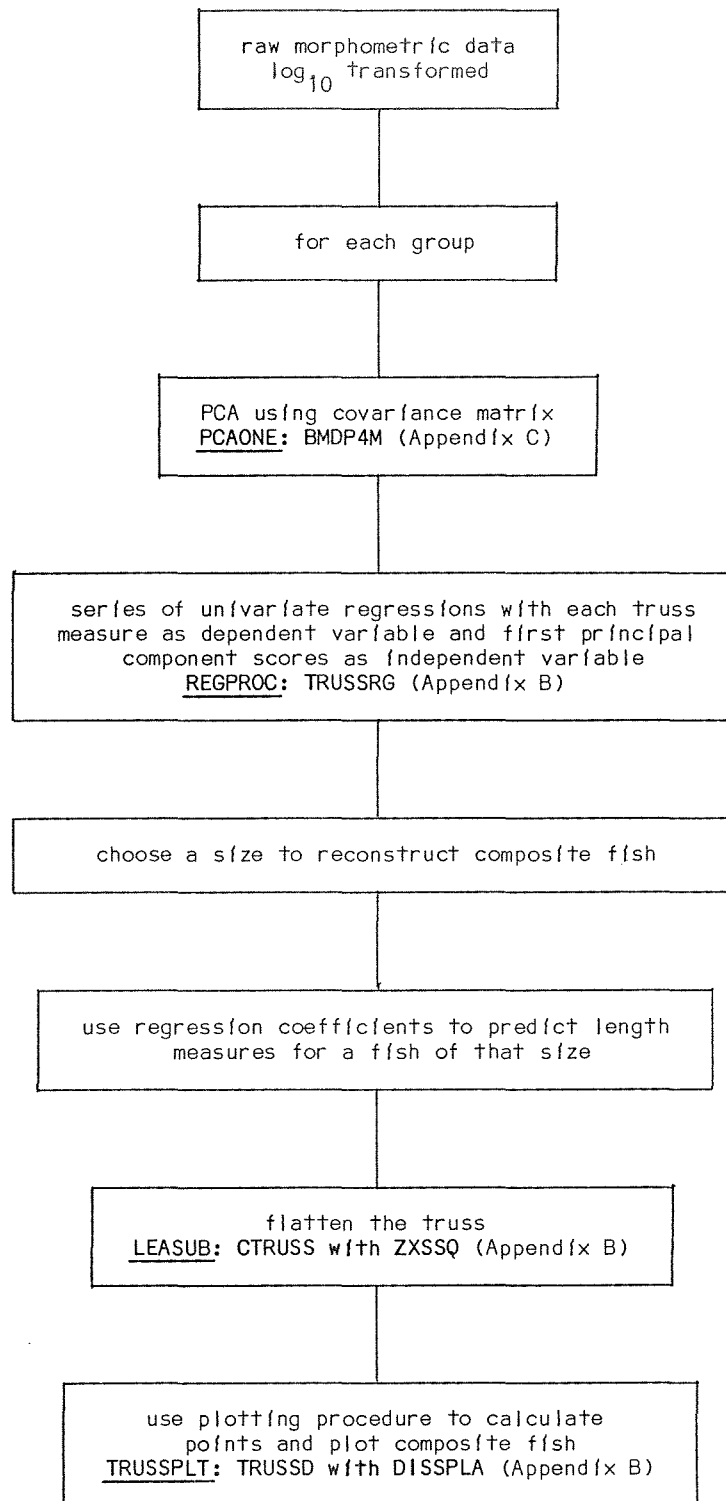
- Computing Centre, Northwestern University, Illinois.
- BMDP Users Digest 1981 Ed. BioMedical Computer Programs Series, Univ. Calif., Berkeley.
- Booke, H.E. 1981. The conundrum of the stock concept -- are nature and nurture definable in fishery science. *Can. J. Fish. Aquat. Sci.* 38: 1479-1480.
- Bookstein, F.L. 1978. The measurement of biological shape and shape change. Lecture Notes in Biomathematics No. 24, Springer-Verlag, N.Y. 191 p.
- Bookstein, F.L. 1982. Foundations of morphometrics. *Ann. Rev. Ecol. Syst.* 13: 451-470.
- Bookstein, F.L., B. Chernoff, R.L. Elder, J.M. Humphries, G.R. Smith, and R.E. Strauss. 1985. Morphometrics in evolutionary biology: The geometry of size and shape with examples from fishes. Special Publication 15 Academy of Natural Science Philadelphia: 277 pp.
- Castle, W.E. 1914. The nature of size factors as indicated by a study of correlation. *Pub. Carnegie Inst.* 196: 51-55.
- Cock, A.G. 1966. Genetic aspects of metrical growth and form in animals. *Q. Rev. Biol.* 41: 131-190.
- Dickie, L.M. 1979. Perspectives on fisheries biology and implications for management. *J. Fish. Res. Bd. Canada* 36: 838-844.
- D.I.S.S.P.L.A. Display Integrated Software System and Plotting Language Users Manual Edition 8.2. 1970. Integrated Software Systems, San Diego.
- Gould, S.J. 1966. Allometry and size in ontogeny and phylogeny. *Biol. Rev.* 41: 587-640.
- Habbema, J.D.F., and J. Hermans. 1977. Selection of variables in discriminant analysis by F -- statistic and error rate. *Technometrics* 19: 487-493.
- Heuts, M.J. 1949. Racial divergence in fin ray variation patterns in Gasterosteus aculeatus. *J. Genet.* 49: 183-191.
- Humphries, J.M., and F.L. Bookstein, B. Chernoff, G.R. Smith, R.L. Elder, and S.G. Poss. 1981. Multivariate discrimination by shape in relation to size. *Syst. Zool.* 30: 291-308.
- Huxley, J.S. 1924. Constant differential growth ratios and their significance. *Nature* 11: 895-896.
- Huxley, J.S. 1932. Problems of relative growth. The Dial Press, N.Y.. 276 p.
- Ihssen, P.E., H.E. Booke, J.M. Casselman, J.M. McGlade, N.R. Payne, and F.M. Utter. 1981. Stock identification: Materials and methods. *Can. J. Fish. Aquat. Sci.* 38: 1838-1855.
- I.M.S.L. Library 1 Reference Manual 1984. Edition 9.2, 2 Vols. International Mathematical and Statistical Library, Houston. Z-1,3; ZXSSQ-1,7.
- Jennrich, R., and P. Sampson. 1983. P7M stepwise discriminant analysis. p. 519-537. In: W.J. Dixon [ed.] BMDP Statistical Software 1983 Ed., Univ. Calif., Berkeley.
- Jolliffe, P. 1963. The multivariate generalization of the allometry equation. *Biometrics* 19: 497-499.
- Kendall, M.G., and A. Stuart. 1966. The advanced theory of statistics. Vol. 3, Griffith and Co., Lond. 552 p.
- Kuhry, B., and L.F. Marcus. 1977. Bivariate linear models in biometry. *Syst. Zool.* 26: 201-209.
- Lachenbruch, P.A., and M.R. Mickey. 1975. Computations for variable selection in discriminant analysis. *Technometrics* 17: 103-109.
- Lachenbruch, P.A., and M. Goldstein. 1979. Discriminant analysis. *Biometrics* 35: 69-85.
- Larkin, P.A. 1972. The stock concept and management of Pacific salmon. p. 11-15. In: R.C. Simon and P.A. Larkin [eds.] The stock Concept in Pacific salmon, H.R. MacMillan Lectures in Fisheries. Univ. British Columbia, Vancouver.
- Lee, P.J. 1971. Multivariate analysis for the fisheries biology. *J. Fish. Res. Bd. Canada* Tech. Rep. No. 244.
- Mahalanobis, P.C. 1936. On the generalised distance in statistics. *Proc. Nat. Inst. Sci., India* 2: 49-55.
- McGlade, J. 1981. Genotypic and phenotypic variation in the brook trout, Salvelinus fontinalis (Mitchill). Unpubl., Ph.D. dissertation, Univ. of Guelph.
- Morrison, D.F. 1967. Multivariate statistical methods. McGraw-Hill, New York. 338 p.
- Morrison, D.F. 1976. Multivariate statistical methods. 2nd Ed., McGraw-Hill, Toronto. 415 p.
- Mosiman, J.E., and F.C. James. 1979. New statistical methods for allometry with application to the Florida red-winged blackbirds. *Evolution* 33: 444-459.
- Pimental, R.A. 1979. Morphometrics: the multivariate analysis of biological data. Kendall-Hunt Dubuque. 276 p.
- Rao, C.R. 1952. Advanced statistical methods in biometric research. John Wiley and Sons, N.Y. 390 p.
- Reyment, R.A., R.E. Blackith, and N.A. Campbell. 1984. Multivariate morphometrics. 2nd Ed., Academic Press, Lond. 233 p.

- Royce, W.F. 1972. Introduction to the fishery sciences. Academic Press, New York. 351 p.
- Salmon, G. 1914. A treatise on the analytic geometry of three dimensions, 2 vols. (R.A.P. Rogers, Rev. Ed.) Longmans and Green, Lond. 470 + 334 p.
- Schmidt, J. 1921. Racial investigation VII - VIII. C.R. Trav. Lab. Carlsberg 14. Cited in: Tåning, A.V. 1946. Nature 157: 594-595.
- Simpson, G.G. 1961. Principles of animal taxonomy. Columbia Univ., N.Y. 247 p.
- Sneath, P.H., and R.R. Sokal. 1973. Numerical taxonomy. W.H. Freeman, San Francisco. 573 p.
- Spangler, G.R., A.H. Beist, and J.F. Koonce. 1981. Perspectives and policy recommendations on the relevance of the stock concept to fishery management. Can. J. Fish. Aquat. Sci. 38: 1908-1914.
- Sprent, P. 1972. The mathematics of size and shape. Biometrics 28: 23-27.
- Strauss, R.E., and F.L. Bookstein. 1982. The truss: body form reconstructions in morphometrics. Syst. Zool. 31: 113-135.
- Tåning, A.V. 1946. Stage of determination of vertebrae in teleostean fishes. Nature 157: 594-595.
- Tåning, A.V. 1952. Experimental studies of meristic characters in fishes. Biol. Rev. Cambridge Philos. Soc. 27: 169-193.
- Tessier, G. 1936. Les lois quantitatives de la croissance. Ann. Physiol. Physicochim. Biol. 12: 527-579.
- Thompson, D.W. 1917, 1942. On growth and form. Cambridge Univ., Cambridge. 1116 p.
- Thorpe, J.E., R.I.G. Morgan, E.M. Ottaway, and M.S. Miles. 1980. Time and divergence of growth groups between potential 1+ and 2+ smolts among sibling Atlantic salmon. J. Fish. Biol. 17: 13-21.
- Thorpe, R.S. 1976. Biometric analysis of geographic variation and racial affinities. Biol. Rev. 51: 407-452.
- Thorpe, R.S., and L. Leamy. 1983. Morphometric studies in inbred and hybrid house mice (Mus sp.): multivariate analysis of size and shape. J. Zool. Lond. 199: 421-432.
- Wilks, S.S. 1932. Certain generalizations in the analysis of variance. Biometrika 24: 471-494.
- Winans, G.A. 1984. Multivariate morphometric variability in the Pacific salmon: technical demonstration. Can. J. Fish. Aquat. Sci. 41: 1150-1159.
- Wright, S. 1932. General, group, and special size factors. Genetics 17: 605-619.

## APPENDIX A:

## RECONSTRUCTION OF COMPOSITE TRUSS

**\*\*(CYBER procedure given in bold type and underlined)**



## APPENDIX B:

## FORTRAN V PROGRAMS USED IN MORPHOMETRIC ANALYSIS.

THE PROGRAMS IN THIS APPENDIX ARE WRITTEN IN FORTRAN V AS INSTALLED ON THE CYBER MAINFRAME COMPUTER AT THE BEDFORD INSTITUTE OF OCEANOGRAPHY. THE OPERATING SYSTEM OF OUR CYBER IS NOS 2.4 WHICH IS WRITTEN BY THE CONTROL DATA CORPORATION.

THE ENTIRE NOS PROCEDURE IS LISTED HERE. FOR PEOPLE WITH A COMPUTER SYSTEM EXACTLY LIKE OURS AT B.I.O., YOU RUN THE NOS PROCEDURES BY TYPING THE PROCEDURE'S NAME THEN PUSHING THE CARRIAGE RETURN. THE CYBER WILL PROMPT YOU FOR THE INFORMATION IT NEEDS TO RUN THE PROGRAM. NOTE THAT AUTOSAV IS A PROCEDURE WE HAVE IN OUR LIBRARY AT M.F.D. THAT DECIDES IF A FILE IS DIRECT OR INDIRECT THEN USES EITHER RETURN OR SAVE AS APPROPRIATE. YOU CAN REPLACE THESE TWO LINES WITH "SAVE,FILENAME." UNLESS YOU HAVE THOUSANDS OF FISH IN WHICH CASE "RETURN,FILENAME." IS NECESSARY. SOME COMPUTER CENTRES WITH A CYBER (SUCH AS DALHOUSIE UNIVERSITY) KEEP BMDPID IN A FILE CALLED SOMETHING OTHER THAN BMDPID - YOU'LL HAVE TO CHECK. ALSO WATCH OUT FOR DIFFERENCES ON OTHER CYBERS IN THE TWO LINES THAT FOLLOW THE "/JOB" STATEMENT; THESE RELATE TO ACCOUNTING OF COMPUTER FUNDS, ALLOCATION OF MEMORY AND TIME FOR THE JOB RUN BY THE PROCEDURE.

IF YOUR COMPUTER HAS A DIFFERENT OPERATING SYSTEM THESE NOS PROCEDURES WILL NOT RUN. YOU SHOULD BE ABLE TO MODIFY THE FORTRAN V PROGRAMS SO THAT THEY WILL RUN ON YOUR COMPUTER. IF YOU ARE INTERESTED IN DOING THIS PLEASE NOTE THE FOLLOWING:

- 1) THE FORTRAN PROGRAMS THEMSELVES ARE AT THE VERY END OF THE PROCEDURE.
- 2) THE FIRST FIVE OR SIX LINES OF THE NOS PROCEDURE BEGIN WITH A NOS VARIABLE NAME THAT MAY BE USED IN THE FORTRAN V PROGRAM LISTINGS. THESE ARE USED TO PROMPT THE USER FOR ARRAY SIZES, NUMBER OF FISH, DATA FILE NAMES, AND OUTPUT FILE NAMES. IF YOU ARE MODIFYING THE FORTRAN V PROGRAMS FOR ANOTHER SYSTEM YOU WILL HAVE TO REMOVE THE NOS VARIABLES FROM THE FORTRAN PROGRAMS AND REPLACE THEM WITH EXACT NUMBERS AND NAMES. OF COURSE EVERY OPERATING SYSTEM HAS MINOR DIFFERENCES IN THEIR INSTALLATION OF FORTRAN V THAT YOU WILL HAVE TO LOOK OUT FOR. PARTICULARLY WATCH OUT FOR THE WAY YOUR OPERATING SYSTEM INTERFACES YOUR FORTRAN V PROGRAM WITH THE DATA FILES AND OUTPUT FILES IT USES.
- 3) THE GRAPHICS PROGRAM "DISSPLA" USED BY THE FORTRAN V PROGRAM TRUSSD IN TRUSPLT AND THE IMSL SUBROUTINE "ZXSSQ" USED BY THE FORTRAN V PROGRAM CTRUSS IN LEASUB MAY NOT BE AVAILABLE ON ALL SYSTEMS OR EVEN ON ALL NOS SYSTEMS. BUT AT LEAST IF YOU HAVE A NOS SYSTEM YOU HAVE THE OPTION OF BUYING DISSPLA AND IMSL FROM THE COMPANIES THAT PRODUCE THEM.
- 4) IN THE PROCEDURE REGPROC THERE IS A FORTRAN V SUBROUTINE "TRANSF" THAT INTERACTS WITH BMDPID IN ORDER TO WRITE OUT SOME VARIABLES IN FROM THE BMDP SAVE FILE CREATED IN PCAONE (YOU NEED THE PRINCIPAL COMPONENT SCORES TO DO THESE REGRESSIONS). FOR A DIFFERENT OPERATING SYSTEM YOU WILL WANT TO CONSULT THE BMDP MANUAL TO FIGURE OUT THE BEST WAY OF DOING THIS.

```

.PROC,REGPROC*I,ARRAY"FOUR DIGITS GREATER THAN NUMBER OF FISH"=(#S4(1234567890)),
NUMCELLS"NUMBER OF CELLS IN YOUR TRUSS"=(#S1(4567)),
BMDFILE"CONTAINING DATA AND SCORES FROM PCA"=(#F),
RESIDULS"DESIRED NAME OF OUTPUT FILE RESIDUALS"=(#F),
COEFFS"DESIRED NAME OF OUTPUT FILE REGR COEF"=(#F),
MEANS"DESIRED NAME OF OUTPUT FILE BMDPID"=(#F).
* THIS DOES 37 REGRESSIONS AND SAVES RESIDUALS
* AND REGRESSION COEFFICIENTS.
* FOR MORE INFORMATION SEE COMMENTS IN
* FORTRAN PROGRAM BELOW.
SUBMIT,SUBJOB.
.DATA,SUBJOB.
/JOB
REG,CM177000,T100.
/READ,MFDU1
GET,BMDFILE/NA.
IFE,.NOT.FILE(BMDFILE,AS),NOGET.
    ATTACH,BMDFILE.
ENDIF,NOGET.
ATTACH,BMDPID/UN=LIBRARY.
FTNS.
BMDPID,L=MEANS,B,W=15000.
PACK,BMDPOUT.
REWIND,BMDPOUT.
REWIND,LGO.
FTNS.
LGO.
GET,AUTOSAV/UN=LIBRARY.
REPLACE,MEANS.
ROUTE,MEANS,DC=LP.
REPLACE,RESID=RESIDULS.
REPLACE,COEF=COEFFS.
DAYFILE,OUTPUT.
REPLACE,OUTPUT=REGOK.
ENQUIRE,F.
EXIT.
DAYFILE,OUTPUT.
REPLACE,OUTPUT=REGBOMB.
REPLACE,MEANS.
ENQUIRE,F.
/EOB

```

```

      SUBROUTINE TRANSF(X,KASE,NPROB,USE,NVAR,XMIS)
C   THIS FORTRAN V SUBROUTINE INTERFACES WITH A BMDPID PROGRAM TO WRITE OUT
C   VARIABLES FROM THE BMDP SAVE FILE CONTAINING
C   THE PRINCIPAL COMPONENT SCORES TO AN ASCII FILE THAT CAN BE
C   READ BY THE SUBSEQUENT FORTRAN PROGRAM "TRUSSG".
C   FROM THE FIRST PRINCIPAL COMPONENT ANALYSIS OF THE TRUSS MEASURES.
C   THERE MUST NOT BE ANY MISSING DATA FOR ANY OF THE CASES (FISH).
      DIMENSION X(NVAR)
      INTEGER OUTFIT(35),OUTBUF(512),NCELL,NTRUSH,J,ILA1
      CHARACTER GROUP(5)*110
      PARAMETER(NCELL=NUMCELLS,NTRUSH=5*NCELL+1)
C   DETERMINE POSITION OF FIRST VARIABLE IN FILE
      ILA1=NVAR-5-NTRUSH
C   CALCULATE NUMBER OF LINES OF DATA PER CASE
      L=1+INT(((NVAR-ILA1+1)/11)+1)
C** DEFINE AND OPEN FILE ONLY FOR THE FIRST CALL
      IF (KASE .EQ. 1) THEN
          CALL FILESD (OUTFIT,'LFN','BMDPOUT','RT','Z',
& 'BT','C','FL',110,'FWB',OUTBUF(1),
& 'BFS',512,'DFC',3)
          CALL OPENM(OUTFIT,'OUTPUT')
          CALL CLOSEM(OUTFIT,'N')
          ENDIF
C** FOLLOWING IS EXECUTED FOR EVERY CASE
          CALL OPENM(OUTFIT,'OUTPUT','N')
          WRITE(GROUP(1),200)X(1),X(2)
          WRITE(GROUP(2),201)(X(J),J=ILA1,NVAR )
          DO 20 I=1,L
              CALL PUT(OUTFIT,GROUP(I), 110 )
20      CONTINUE
          CALL CLOSEM(OUTFIT,'N')
200  FORMAT(2X,A4,1X,F3.0)
201  FORMAT (11F10.4/11F10.4/11F10.4/9F10.4 )
          RETURN
          END
/EOB
/PROBLEM TITLE='WRITING OUT MEANS'.
/INPUT
      FILE=BMDFILE.
      CODE=SECRET.
/TRAN
/END
/EOB

```

## PROGRAM TRUSG

```

C*****
C
C THIS PROGRAM DOES 37 REGRESSIONS USING THE LOG-TRANSFORMED TRUSS
C MEASURES AND THE FACTOR SCORES FROM THE FIRST PRINCIPAL COMPONENT
C IT SAVES THE COEFFICIENTS FROM THE REGRESSIONS FOR LATER USE
C IN TRUSS RECONSTRUCTION AT A STANDARD SIZE.
C THE RESIDUALS FROM THE REGRESSIONS ARE ALSO SAVED IN A SEPARATE FILE.
C
C WRITTEN BY ELIZABETH BOULDING
C MARINE FISH DIVISION
C BEDFORD INSTITUTE OF OCEANOGRAPHY.
C
C VERSION MAY 4,1984.
C
C
C REAL XVAR(ARRAY),YVAR(ARRAY),TRTRUS(36,ARRAY),TLENG(ARRAY)
C REAL FACTR1(ARRAY),FACTR2(ARRAY)
C INTEGER NTRUSH,NCALL,NCASE,NCASES,NTAG(ARRAY),NSET(ARRAY),NCELL
C PARAMETER(NCELL=NUMCELLS,NTRUSH=NCELL*5+1)
C
C OPEN (8,FILE='BMDPOUT')
C** THIS FILE CONTAINS THE ORIGINAL TRUSS MEASURES AFTER THEY HAVE BEEN
C LOG-TRANSFORMED AND THE FACTOR SCORES FROM THE PRINCIPAL COMPONENT
C ANALYSIS. IT IS THE INPUT DATA FOR THIS PROGRAM.
C
C OPEN (10,FILE='COEF')
C** THIS FILE CONTAINS THE REGRESSION COEFFICIENTS THAT WILL BE
C USED IN THE TRUSS
C RECONSTRUCTION FOR A FISH OF A STANDARD SIZE.
C THE FIRST VALUE ON EACH LINE IS THE INTERCEPT AND THE SECOND
C IS THE SLOPE.
C
C OPEN (12,FILE='RESID')
C** THIS FILE CONTAINS THE RESIDUALS FROM THE REGRESSIONS OF THE
C LOG-TRANSFORMED TRUSS MEASURES AGAINST THE FIRST PRINCIPAL COMPONENT
C SCORES.
C
C FORMAT STATEMENTS
C 100 IN THIS PROGRAM MUST MATCH 200 AND 201 IN THE PRECEDING SUBROUTINE
C TRANSF THAT WROTE THE DATA INTO THE ASCII FILE READ BY THIS PROGRAM.
C 104 IS THE FORMAT IN WHICH THE RESIDUALS FROM THIS PROGRAM ARE WRITTEN
C TO AN ASCII FILE.
C 105 IS THE FORMAT IN WHICH THE REGRESSION COEFFICIENTS ARE WRITTEN TO AN
C ASCII FILE.
C
C** INITIALIZE VARIABLES
C NCALL=0
C NCASE=1
C 99 READ (8,100,ERR=16,END=18)NTAG(NCASE),NSET(NCASE),
C &(TRTRUS(I,NCASE),I=1,NTRUSH),
C &TLENG(NCASE),FACTR1(NCASE),FACTR2(NCASE)
C** ASSUME THAT ALL CASES HAVE NO MISSING DATA
C NCASE=NCASE+1
C GO TO 99

```

```

100 FORMAT (2X,I4,1X,I2,/,11F10.4,/,11F10.4,/,11F10.4,/,6F10.4)
103 FORMAT (' ERROR WHILE READING, CHECK YOUR DATA')
16 WRITE(10,103)
18 NCASES=NCASE-1
C** DO FIRST REGRESSION WHICH WILL BE USED TO OBTAIN THE
C FACTOR1 SCORE OF A FISH OF TOTAL LENGTH X .
DO 10 J=1,NCASES
C** XVAR IS THE INDEPENDENT VARIABLE
XVAR(J)=TLENG(J)
C** YVAR IS THE DEPENDENT VARIABLE
YVAR(J)=FACTR1(J)
10 CONTINUE
CALL REGRES(TRTRUS,XVAR,YVAR,NCASES,NCALL)
C
C** DO INDIVIDUAL REGRESSIONS FOR EACH TRUSS MEASURE WITH FACTOR 1
C SCORES AS THE INDEPENDENT VARIABLE.
C
C** THE INDEPENDENT VARIABLE IS THE SAME FOR ALL SUBSEQUENT REGRESSIONS
DO 15 L=1,NCASES
XVAR(L)=FACTR1(L)
15 CONTINUE
C** DO THE FOLLOWING FOR EACH TRUSS MEASURE
DO 20 K=1,NTRUSM
C** HOWEVER THE DEPENDENT VARIABLE DOES CHANGE FOR EACH REGRESSION
DO 30 M=1,NCASES
YVAR(M)=TRTRUS(K,M)
30 CONTINUE
CALL REGRES(TRTRUS,XVAR,YVAR,NCASES,NCALL)
20 CONTINUE
C** WRITE THE RESIDUALS INTO A FILE
DO 40 IN=1,NCASES
WRITE(12,104)NTAG(IN),NSET(IN),TLENG(IN),FACTR1(IN),FACTR2(IN),
&(TRTRUS(N,IN),N=1,NTRUSM)
40 CONTINUE
WRITE(10,105)NCASES
104 FORMAT(2X,I4,1X,I2,3F10.4,/,15F8.4,/,15F8.4,/,6F8.4)
105 FORMAT('THE NUMBER OF CASES (FISH) IS ',I4)
WRITE(10,*)'REGRESSIONS WERE SUCCESSFUL!'
STOP
END
SUBROUTINE REGRES(TRTRUS,XVAR,YVAR,NCASES,NCALL)
REAL XVAR(ARRAY),YVAR(ARRAY),XDEV(ARRAY),YDEV,INTCPT,SLOPE,R
REAL RESY,SUMX,SUMY,SUMXX,SUMYY,SUMXY,XMEAN,YMEAN,TRTRUS(36,ARRAY)
INTEGER NCALL,NCASES,NVAR
C** INITIALIZE AND SUM VARIABLES
NCALL=NCALL+1
SUMY=0.0
SUMYY=0.0
SUMXY=0.0
C IF THIS IS LESS THAN THE THIRD CALL TO THIS SUBROUTINE
C INITIALIZE ALL VARIABLES ASSOCIATED WITH INDEPENDENT VARIABLE
IF (NCALL.LT.3) THEN
SUMX=0.0
SUMXX=0.0
ENDIF
ENDIF

```

```

      DO 10 I=1,NCASES
        IF (NCALL.LT.3) THEN
          SUMX=SUMX+XVAR(I)
        ENDIF
C** CALCULATE THE SUM OF Y FOR ALL CALLS TO REGRESS
        SUMY=SUMY+YVAR(I)
10      CONTINUE
        IF (NCALL.LT.3) THEN
          XMEAN=SUMX/NCASES
        ENDIF
        YMEAN=SUMY/NCASES
        DO 20 J=1,NCASES
          IF (NCALL.LT.3) THEN
            XDEV(J)=XVAR(J)-XMEAN
            SUMXX=SUMXX+XDEV(J)*XDEV(J)
          ENDIF
          YDEV=YVAR(J)-YMEAN
          SUMYY=SUMYY+YDEV*YDEV
          SUMXY=SUMXY+XDEV(J)*YDEV
20      CONTINUE
        SLOPE=SUMXY/SUMXX
        INTCPY=YMEAN-(SLOPE*XMEAN)
C** CALCULATE R THE CORRELATION COEFFICIENT
        R=SUMXY/((SUMXX*SUMYY)**0.5)
C** NCALL REFERS TO NO. OF CALLS TO REGRESS
        CALL COEFSV(SLOPE,INTCPY,R,NCALL)
        IF (NCALL.GT.1) THEN
          DO 30 K=1,NCASES
C** REPLACE INDIVIDUAL TRUSS MEASURES WITH RESIDUALS FROM REGRESSIONS
            RESY=(YVAR(K)-(SLOPE*XVAR(K)+INTCPY))
            NVAR=NCALL-1
            TRTRUS(NVAR,K)=RESY
30      CONTINUE
          ENDIF
        RETURN
      END

```

```

      SUBROUTINE COEFSV(SLOPE,INTCPT,R,NCALL)
      REAL OLDCEF(3),SLOPE,INTCPT,R
      INTEGER SIDE,CELL,NCALL
C** INITIALIZE SIDE AND CELL
      IF (NCALL.EQ.1) THEN
        SIDE=0
        CELL=0
      ENDIF
C** WRITE THE COEFFICIENTS FROM THE REGRESSIONS INTO THE FILE
C** DO THE FOLLOWING ONLY FOR FIRST REGRESSION
      IF ((SIDE.EQ.0).AND.(CELL.EQ.0)) THEN
        WRITE(10,104) INTCPT,SLOPE,R
104    FORMAT(2F10.5,2X,'CORRELATION COEFFICIENT IS ',F4.2)
      ELSE
C** DO THE FOLLOWING FOR ALL SUBSEQUENT REGRESSIONS
      IF (SIDE.EQ.4) THEN
C** MUST STORE COEFFICIENTS FOR SHARED SIDE
        OLDCEF(1)=INTCPT
        OLDCEF(2)=SLOPE
        OLDCEF(3)=R
      ENDIF
      WRITE (10,105) INTCPT,SLOPE,R,SIDE,CELL
105    FORMAT(2F10.5,2X,'CORRELATION COEFFICIENT IS ',F4.2,2X,
      &'TRUSS(',I1,',',I1,',',I1,')')
C** MUST AUTOMATICALLY FILL IN COEFFICIENTS FOR SHARED SIDE
      IF ((SIDE.EQ.1).AND.(CELL.GT.1)) THEN
        SIDE=SIDE+1
        WRITE(10,105)(OLDCEF(J),J=1,3),SIDE,CELL
      ENDIF
      ENDIF
C** DO THE FOLLOWING FOR ALL REGRESSIONS
      IF ((SIDE.EQ.0).OR.(SIDE.EQ.6)) THEN
        SIDE=1
        CELL=CELL+1
      ELSE
        SIDE=SIDE+1
      ENDIF
      RETURN
      END

```

```

.PROC, LEASUB*I, TLENGTH"DESIRED SIZE OF COMPOSITE FISH INTEGER"=(%S3(1234567890)),
NUMCELLS"NUMBER OF CELLS IN YOUR TRUSS"=(%S1(4567)),
COEFFFILE"INPUT FILE CONTAINING REGRESSION COEF"=(%F),
ADJUSTED"TRUSS DESIRED NAME OF OUTPUT FILE **"=(%F),
RUNMESS"DESIRED NAME OF FILE WITH INSL MESSAGES"=(%F).
.* THIS PROCEDURE RUN THE FORTRAN PROGRAM THAT FLATTENS THE
.* TRUSS OF AVERAGE TRUSS MEASURES FROM A FISH POPULATION.
.* IT USES THE INTL. MATH. STAT. LIBRARY (SUBROUTINE ZXSSQ).
.* FOR MORE DETAILED COMMENTS REFER TO FORTRAN PROGRAM BELOW.
SUBMIT, SUBJOB.
.DATA, SUBJOB.
/JOB
LEAST, CM177000, T32.
/READ, MFDU1
ATTACH, INSLIB/UN=LIBRARY.
LIBRARY, INSLIB.
GET, COEF=COEFFFILE.
FTNS.
LDSET(LIB=INSLIB, PRESET=ZERO)
LGO.
GET, AUTOSAV/UN=LIBRARY.
REPLACE, OUT=ADJUSTED.
BEGIN,, AUTOSAV, LDBUG, RUNMESS.
DAYFILE, OUTPUT.
REPLACE, OUTPUT=LEAOK.
ENQUIRE, F.
EXIT.
DAYFILE, OUTPUT.
REPLACE, OUTPUT=LEABOMB.
SAVE, LDBUG.
ENQUIRE, F.
/EOR

```

## PROGRAM CTRUSS

```

*****
C  CTRUSS CALCULATES THE TRUSS FOR AN AVERAGE FISH AT A STANDARD SIZE
C  USING THE REGRESSION COEFFICIENTS FROM THE REGRESSIONS OF THE TRUSS MEASURES
C  OF THE INDIVIDUAL FISH AGAINST THE FIRST WITHIN-GROUP PRINCIPAL COMPONENT.
C  IT THEN "FLATTENS EACH CELL OF THE TRUSS BY ADDING OR SUBTRACTING A
C  SMALL CORRECTION TO EACH OF THE SIX SIDES.
C  THESE CORRECTIONS ARE CALCULATED BY MINIMIZING THE VALUE OF A
C  DETERMINANT WHICH WILL BE EQUAL TO ZERO WHEN THE TRUSS IS
C  PERFECTLY PLANER. THE ADJUSTED TRUSS MEASURES
C  ARE THEN IN A FORM WHERE THEY CAN BE RECONSTRUCTED IN X-Y SPACE AND
C  TRANSFORMED INTO A PLOT FILE BY OUR COMPANION PROGRAM TRUSSD.
C
C  DEVELOPED 1984 BY:
C      JACQUELINE MCGLADE (M.F.D., B.I.O.)
C  AND      ELIZABETH G. BOULDING (M.F.D., B.I.O.)
C  WITH ASSISTANCE FROM:
C      TOBY KEITH HAY (BIOLOGY, DALHOUSIE UNIVERSITY)
C  GENERAL IDEA SIMILAR TO THAT IN STRAUSS & BOOKSTEIN (1982, SYSTEMATIC
C  ZOOLOGY 31:113-135).
C
C** THIS PROGRAM WORKS FOR TRUSSES OF FOUR TO SEVEN CELLS EACH CONTAINING
C** SIX MEASUREMENTS.
C    THE SIX MEASUREMENTS ARE LISTED IN THE ARRAY "TRUSS"
C    IN THE FOLLOWING ORDER:
C
C
C      I      I      I      I      I      I
C      1      2      3      4      5      6
C      ----      \      /
C      I      I      I      I      I      I
C
C
C  NOTE THAT SIDE4 IS THE SAME AS SIDE 2 OF THE NEXT CELL.
C  THIS PROCEDURE CALLS THE SUBROUTINE ZYSSQ FROM THE
C  INTL. MATH. STAT. LIBRARY WHICH CAN ONLY BE CONNECTED
C  BY CLEARING ONE OF OUR EXISTING LIBRARIES. THIS IS
C  DONE NEAR THE BEGINNING OF THIS PROCEDURE.
C
C*****

```

```

C
C  INTERPRETATION OF CONVERGENCE CRITERION AND ERROR MESSAGES FROM
C  IMSL'S ZSSQ SUBROUTINE ARE IN THE FILE RUNMESS.
C
C  INFER=0  CONVERGENCE FAILED
C  SEE VARIABLE IER FOR EXPLANATION.
C
C
C  CONVERGENCE SATISFIED
C  INFER=1  CRITERION THAT THE PARAMETER ESTIMATES AGREE TO NSIG DIGITS
C           ON SUCCESSIVE ITERATIONS WAS SATISFIED.
C
C  INFER=2  CRITERION THAT ON TWO SUCCESSIVE ITERATIONS THE RESIDUAL
C           SS ESTIMATES DIFFER BY NO MORE THAN EPS WAS SATISFIED.
C
C  INFER=4  CRITERION THAT EUCLIDIAN NORM OF THE APPROXIMATE GRADIENT
C           IS LESS THAN OR EQUAL TO DELTA WAS SATISFIED.
C
C** IF MORE THAN ONE CRITERION WAS SATISFIED THEN INFER WILL
C    BE EQUAL TO THE SUM OF THOSE SATISFIED.
C
C** ERROR PARAMETERS TO REFER TO WHEN INFER=0
C
C  FATAL ERRORS
C
C  IER=129  SINGULARITY WAS DETECTED IN THE JACOBIAN AND
C           RECOVERY FAILED.
C
C  IER=130  AT LEAST ONE OF M,N,IOPT,PARM(1), OR PARM(2)
C           WAS SPECIFIED INCORRECTLY.
C
C  IER=131  MARQUAND PARAMETER EXCEEDED PARM(3).
C
C  IER=132  AFTER A SUCCESSFUL RECOVERY FROM A SINGULAR
C           JACOBIAN, THE VECTOR X HAS CYCLED BACK TO THE FIRST
C           SINGULARITY.
C
C  IER=133  IMPLIES THAT MAXFN WAS EXCEEDED. WE HAVE SET
C           MAXFN(THE MAXIMUM NUMBER OF CALLS TO THE SUBROUTINE)
C           TO BE 100.
C
C** WARNING ERROR
C  IER=0 IMPLIES THAT THE JACOBIAN IS ZERO. THE
C  SOLUTION X IS A STATIONARY POINT.
C
C** FOR A MORE DETAILED DESCRIPTION SEE THE IMSL
C  MANUAL, THE RELEVANT PORTIONS OF WHICH ARE
C  APPENDED ONTO THE "TRUSS" MANUAL
C
C*****

```

```

C
C INPUT - OUTPUT INFORMATION
C
C FORTRAN UNITS
C 5= INPUT FILE
C 6= DEBUG AND RUN TIME MESSAGES
C 7= ADJUSTED TRUSS MEASURES AND CALCULATION OF DISTORTION
C     READY FOR INPUT INTO THE COMPANION PROGRAM TRUSSD.
C
C FORMAT STATEMENTS
C 102 & 103 READ THE SLOPES AND INTERCEPTS FOR THE REGRESSIONS
C     OF THE TRUSS MEASURES AGAINST THE FIRST WITH-IN GROUP
C     PRINCIPAL COMPONENT SCORES (SIZE) AND OF THESE SCORES
C     AGAINST EXPLICITLY MEASURED SIZE, TLENG. THESE WERE
C     CALCULATED BY REGPROC.
C 104 THIS WRITES OUT THE TRUSS MEASURES SO THEY CAN BE USED
C     BY TRUSPLT TO CALCULATE THE CO-ORDINATES OF THE
C     COMPOSITE FORM.
C 105 THIS WRITES OUT THE DISTORTION OR DEPARTURE FROM PLANARITY
C     OF THE COMPOSITE FORM AND THE SIZE THE FISH HAS BEEN
C     CONSTRUCTED AT.
C
C DICTIONARY OF SELECTED VARIABLES
C NOTE: CHECK TYPE DECLARATION, FIRST LETTER TYPE DECLARATION IS
C     NOT USED.
C
C B(6) - REAL ARRAY CONTAINING THE INTERCEPTS
C FSIZE - TOTAL LENGTH (CM) DESIRED FOR COMPOSITE FISH
C LFACT1 - REAL VARIABLE CONTAINING FIRST P.C. SCORE OF FISH
C LFSIZE - REAL VARIABLE LOG10-TRANSFORMED OF TOTAL LENGTH
C     OF THAT TOTAL LENGTH
C SIDE(6) - REAL ARRAY CONTAINING THE DISTANCE MEASURES OF THE
C     CELL BEING PROCESSED
C SL(6) - REAL ARRAY CONTAINING SLOPES
C TRUSS(6,7) - REAL ARRAY CONTAINING DISTANCE MEASURES OF COMPOSITE FISH
C     WITH 6 SIDES PER CELL AND UP TO 7 CELLS IN THE TRUSS
C X(6) - REAL ARRAY CONTAINING CORRECTIONS TO DISTANCE MEASURES
C XCOR(6,7) - REAL ARRAY CONTAINING CORRECTION FOR ALL CELLS
C

```

```

      REAL TRUSS(6,7),XCOR(6,7),X(6),SIDE(6),LFSIZE,LFACT1
      REAL B(6),SL(6)
      INTEGER INFER,WARN,NCELLS
      PARAMETER (NCELLS=NUMCELLS)
      DATA X/0.0,0.0,0.0,0.0,0.0,0.0,0.0/
      OPEN (5,FILE='COEF')
      OPEN (6,FILE='LDBUG')
      OPEN (7,FILE='OUT')
C** INITIALIZE VARIABLES
      DO 3 KK=1,NCELLS
        DO 4 JJ=1,6
          XCOR(JJ,KK)=0.0
        4      CONTINUE
      3      CONTINUE
C** THIS IS THE STANDARD SIZE AT WHICH THE "FISH" WILL BE RECONSTRUCTED
      FSIZE=FLOAT(TLENGTH)
      READ (5,102)BLFACT,SLFACT
      WRITE(6,*)'FOR A FISH OF STANDARD SIZE ',FSIZE
      WRITE(6,*)BLFACT,SLFACT
C  CALCULATE FIRST PRINC. COMP. SCORE OF A FISH OF DESIRED STANDARD SIZE
      LFSIZE=LOG10(FSIZE)
      LFACT1=SLFACT*LFSIZE+BLFACT
      ICELL=1
      WRITE(6,*)FSIZE,LFSIZE,LFACT1
      99  WRITE(6,*)'THE CELL NUMBER IS ',ICELL
          DO 5 I=1,6
            READ(5,103,END=990)B(I),SL(I)
          5  CONTINUE
C** CALCULATE THE AVERAGE TRUSS MEASURES FROM THE REGRESSION COEFFICIENTS.
          DO 10 J=1,6
            SIDE(J)=SL(J)*LFACT1+B(J)
            SIDE(J)=10**SIDE(J)
            WRITE(6,*)'TRUSS(',J,',',ICELL,')',SIDE(J)
          10  CONTINUE
C  NOW ADJUST THE SIDES OF THE CELL UNTIL IT IS PLANER
C  THIS IS DONE BY ADDING AND SUBTRACTING SMALL AMOUNTS
C  ONTO EACH OF THE SIX SIDES THAT COMPRISE EACH CELL.
C
C
C  THIS SUBROUTINE IS CALLED ONE FOR EACH OF THE SEVEN CELLS.
      CALL FLAT(SIDE,X,INFER)
      DO 30 K=1,6
        30  CONTINUE
          DO 40 L=1,6
            TRUSS(L,ICELL)=SIDE(L)
            IF(INFER.NE.0)XCOR(L,ICELL)=X(L)
          40  CONTINUE
            ICELL=ICELL+1
            IF (ICELL.LE.NCELLS) GO TO 99

```

```

C** IF ALL SEVEN CELLS HAVE BEEN FLATTENED CONTINUE
C CALCULATE THE DISTORTION
    DISTORT=0.0
    DO 45 M=1,NCELLS
        DO 47 N=1,6
            DISTORT=DISTORT+(XCOR(N,M)/TRUSS(N,M))**2
47    CONTINUE
45    CONTINUE
    DISTORT=(DISTORT)**0.5
    WRITE(7,105)DISTORT,F5IZE
C CALCULATE THE FLATTENED TRUSS
C ADD THE NECESSARY CORRECTIONS TO THE ORIGINAL DATA MATRIX.
    WRITE(6,*)'THE ADJUSTED TRUSS IS'
    DO 50 II=1,NCELLS
        DO 60 JJ=1,6
            TRUSS(JJ,II)=TRUSS(JJ,II)+XCOR(JJ,II)
60    CONTINUE
    WRITE(7,104)(TRUSS(KK,II),KK=1,6)
    WRITE(6,104)(TRUSS(KK,II),KK=1,6)
50    CONTINUE
102  FORMAT(2F10.5)
103  FORMAT(2F10.5)
104  FORMAT(6F6.2)
105  FORMAT(F6.4,F6.1)
990  STOP
    END
C
C

```

```

      SUBROUTINE FLAT(SIDEP,XX,INFER)
C  DRIVER FOR ZXSSQ LEAST SQUARES ESTIMATION
C  TO FORCE CELL TO BE PLANER
C  INSL VERSION FEBRUARY 2,1984.
C  *****
      EXTERNAL FUNC
      INTEGER M,N,IXJAC,NSIG,MAXFN,IOPT,I,INFER,IER,L,ITER,LL
      REAL PARM(4),X(6),F(6),XJAC(6,6),XJTJ(21),WORK(63),EPS,DELTA,
&SSQ,SIDE(6),XX(6),SIDEP(6)
      COMMON/LSQ/SIDE,ITER
      DO 3 MM=1,6
        SIDE(MM)=SIDEP(MM)
3      CONTINUE
C** INITIALIZE VARIABLES
      INFER=99
      IER=99
      M=6
      N=6
      ITER=0
      IXJAC=6
      NSIG=3
      EPS=0.1
      DELTA=0.0
      MAXFN=100
      IOPT=1
      DO 5 J=1,6
        X(J)=0.0
5      CONTINUE
      CALL ZXSSQ(FUNC,M,N,NSIG,EPS,DELTA,MAXFN,IOPT,PARM,X,SSQ,F,
&      XJAC,IXJAC,XJTJ,WORK,INFER,IER)
C
      WRITE(6,*)' '
      WRITE(6,*)'THE RESULTS OF ',ITER,' ITERATIONS ARE'
      WRITE(6,*)'IER=',IER
      WRITE(6,*)'INFER=',INFER
      WRITE(6,*)'SSQ=',SSQ
      DO 9 L=1,6
        WRITE(6,*)'F(',L,')=',F(L)
9      CONTINUE
      DO 11 LL=1,6
        WRITE(6,*)'X(',LL,')=',X(LL)
        XX(LL)=X(LL)
11     CONTINUE
      RETURN
      END

```

```

      SUBROUTINE FUNC(X,M,N,F)
C  SUBROUTINE TO CALCULATE FUNCTION REQUIRED BY ZXSSQ
      INTEGER M,N,I,J,K,ITER
      REAL X(N),F(M),Y(5),D(6),SIDE(6),XSUM,DETER
      COMMON /ZSQ/SIDE,ITER
C
C
C
      ITER=ITER+1
      WRITE(6,*)'MADE IT TO FUNC, NO. OF ITERATIONS=',ITER
      XSUM=0
      DO 10 I=1,6
          D(I)=SIDE(I)+X(I)
          D(I)=D(I)**2
          WRITE(6,*)'DSQD('',I,'')= ',D(I)
10  CONTINUE
C** WHEN THIS DETERMINANT IS ZERO THE CELL (OF THE TRUSS)
C** IS PERFECTLY PLANER.
      DETER=-2.*D(3)*D(1)**2-2.*D(1)*D(3)**2-2.*D(4)*D(2)**2
      &-2.*D(2)*D(4)**2-2.*D(6)*D(5)**2-2.*D(5)*D(6)**2
      &+2.*D(1)*D(2)*D(3)+2.*D(1)*D(2)*D(4)-2.*D(1)*D(2)*D(5)
      &+2.*D(1)*D(3)*D(4)+2.*D(1)*D(3)*D(5)+2.*D(1)*D(3)*D(6)
      &-2.*D(1)*D(4)*D(6)+2.*D(1)*D(5)*D(6)+2.*D(2)*D(3)*D(4)
      &-2.*D(2)*D(3)*D(6)+2.*D(2)*D(4)*D(5)+2.*D(2)*D(4)*D(6)
      &+2.*D(2)*D(5)*D(6)-2.*D(3)*D(4)*D(5)
      &+2.*D(3)*D(5)*D(6)+2.*D(4)*D(5)*D(6)
      WRITE(6,*)'DETER=',DETER
      DO 20 J=1,6
          XSUM=XSUM+X(J)
20  CONTINUE
      DO 30 K=1,6
          F(K)=ABS(X(K))+(100000*DETER)**2
30  CONTINUE
      RETURN
      END

```

```

.PROC,TRUSPLT#I,PLTLABEL"19 DIGITS TO LABEL TRUSS DRAWING",
NUMCELLS"NUMBER OF CELLS IN YOUR TRUSS"=(#S1(4567)),
FISHTYPE"1 FOR INDIVIDUAL FISH OR 2 FOR COMPOSITE"=(#S1(12)),
DATAFILE"INDIVIDUAL TRUSSES OR ADJUSTED COMPOSITE"=(#F),
PLOTFILE"NAME YOU WANT PLOT FILE SAVED AS"=(#F),
RUNMESS"RUN TIME MESSAGES  DETAILS OF POINT CALC"=(#F).
* THIS VERSION IS FOR TRUSSES WITH FOUR TO SEVEN CELLS.
* THIS PROCEDURE CONVERTS THE TRUSS
* MEASURES TO X-Y COORDINATES.
* IT THEN USES DISSPLA TO GENERATE A
* PLOTFILE.
SUBMIT,SUBJOB.
.DATA,SUBJOB.
/JOB
PLOT,CM177000,T32.
/READ,MFDU1
GET,TDATA=DATAFILE.
ATTACH,DISSPLA/UN=LIBRARY.
FTN5.
LDSET(LIB=DISSPLA).
LGO.
GET,AUTOSAV/UN=LIBRARY.
REPLACE,PLFILE=PLOTFILE.
AUTOSAV,PDEBUG,RUNMESS.
DAYFILE,OUTPUT.
REPLACE,OUTPUT=PLOK.
ENQUIRE,F.
EXIT.
DAYFILE,OUTPUT.
REPLACE,OUTPUT=PLBOMB.
SAVE,PDEBUG=RUNMESS.
ENQUIRE,F.
/EOR

```

## PROGRAM TRUSSD

```

C *****
C PROGRAM TRUSSD RECONSTRUCTS THE TRUSS IN X-Y SPACE IN A FORM WHICH
C CAN BE PLOTTED ON A GRAPHICS DEVICE. THE INPUT DATA CAN EITHER
C BE TRUSS MEASURES FROM INDIVIDUAL FISH OR AVERAGE "FLATTENED"
C TRUSS MEASURES FROM A RECONSTRUCTED AVERAGE FISH AT A
C STANDARD SIZE.
C
C
C DEVELOPED 1984 BY:
C             JACQUELINE MCGLADE (M.F.D., B.I.O.)
C AND        ELIZABETH G. BOULDING (M.F.D., B.I.O.)
C WITH ASSISTANCE FROM:
C             TOBY KEITH HAY (BIOLOGY, DALHOUSIE UNIVERSITY)
C
C             ORIGINAL: FEBRUARY 1984
C             REVISED: JULY 1984 & MAY 1985
C GENERAL IDEA SIMILAR TO THAT IN STRAUSS & BOOKSTEIN (1982, SYSTEMATIC
C ZOOLOGY 31:113-135).
C
C** THIS PROGRAM IS FOR TRUSSES WITH FOUR TO SEVEN CELLS
C** EACH CONTAINING SIX MEASUREMENTS.
C     THE MEASUREMENTS ARE LISTED IN THE ARRAY "TRUSS"
C     WHERE TRUSS IS A TWO DIMENSIONAL ARRAY WITH SIDE NUMBER
C     BEING THE FIRST DIMENSION AND CELL NUMBER BEING THE SECOND
C     (IE., TRUSS(SIDE,CELL).
C     CELL 1 IS NEAREST THE HEAD AND CELL 7 IS NEAREST THE TAIL.
C     THE SIDES ARE NUMBERED IN THE FOLLOWING ORDER:
C
C           I       -----       I       \       /
C           1       I 2       3       4 I       \ 5       6 /
C           -----       I       I       \       /
C
C NOTE THAT SIDE 4 IS THE SAME AS SIDE 2 OF THE NEXT CELL.
C
C INPUT - OUTPUT INFORMATION
C
C FORTRAN UNITS
C 4= INPUT FILE
C 7= DEBUG AND RUN TIME MESSAGES
C THE PLOT OF THE TRUSS IS AUTOMATICALLY GENERATED BY
C DISPLA INTO THE FILE PLFILE
C
C FORMAT STATEMENTS
C 100 MUST MATCH DATA FILE WITH RAW TRUSS MEASURES FOR
C     INDIVIDUAL FISH.
C 200 & 201 MUST MATCH OUTPUT DATA FILE FROM THE FORTRAN
C     CTRUSS CONTAINING THE "FLATTENED" TRUSS MEASURES
C     FROM THE RECONSTRUCTED COMPOSITE FISH.
C

```

C  
C DICTIONARY OF SELECTED VARIABLES  
C  
C NOTE: CHECK TYPE DECLARATION, FIRST LETTER TYPE DECLARATION  
C IS NOT USED.  
C  
C CELL - INTEGER VARIABLE, CURRENT CELL NUMBER INDEX.  
C DISTOR - REAL VARIABLE, DISTORTION OF COMPOSITE TRUSS READ  
C FROM CTRUSS PROGRAM.  
C IFISH - INTEGER VARIABLE, EQUALS 1 FOR INDIVIDUAL TRUSSES AND  
C EQUALS 2 FOR COMPOSITE.  
C M - INTEGER PARAMETER, NUMBER OF SIDES (ALWAYS 6).  
C N - INTEGER PARAMETER, NUMBER OF CELLS IN THE TRUSS.  
C NPNTS - INTEGER PARAMETER, NUMBER OF POINTS NEEDED TO PLOT TRUSS.  
C SET - INTEGER VARIABLE, READ FOR INDIVIDUAL FISH.  
C SIDE - INTEGER VARIABLE, CURRENT SIDE NUMBER INDEX.  
C PLABEL - CHARACTER ARRAY, LABEL FOR TRUSS DRAWING.  
C TRLONG - REAL VARIABLE, ACTUAL LENGTH OF TRUSS  
C TLENG - REAL VARIABLE, LENGTH OF FISH TO BE DRAWN IN CM, READ IN.  
C TRUSS - REAL ARRAY, CONTAINS TRUSS MEASURES.  
C XPOINT - REAL ARRAY, CONTAINS X CO-ORDINATES OF TRUSS.  
C YPOINT - REAL ARRAY, CONTAINS Y CO-ORDINATES OF TRUSS.  
C  
C  
C

```

C
C** MAIN PROGRAM *****
C
      REAL TRUSS(6,7),TLENG,XPOINT(16),YPOINT(16),DISTOR,TLENG
      CHARACTER PLABEL*19
      INTEGER CELL,SIDE,M,N,SET,L,LL,NFISH,NPNTS
      COMMON TRUSS,TLENG,SET,XPOINT,YPOINT,NFISH,DISTOR,TLENG
      PARAMETER (M=6,N=NUMCELLS,NPNTS=2*N+2)
C  INITIALIZE VARIABLES
      SET=0
      DISTOR=0.0
      DATA PLABEL/'              '/
      DATA PLABEL/'PLTLABEL'/
C
C** INPUT FILE DEPENDS ON WHETHER INDIVIDUALS OR COMPOSITE.
      IFISH=FISHTYPE
      IF (IFISH.EQ.1) THEN
C  USED ONLY FOR INDIVIDUAL FISH
        OPEN (4,FILE='TDATA',RECL=150)
        ELSE
C  USED ONLY FOR COMPOSITE FISH
        OPEN (4,FILE='TDATA')
        ENDIF
        OPEN (7,FILE='PDEBUG')
        CALL COMPRS
        NFISH=1
C
      IF (IFISH .EQ. 1) THEN
C  FOLLOWING USED FOR INDIVIDUAL FISH
        99 READ (4,100,ERR=16,END=18)SET,((TRUSS(SIDE,CELL),SIDE=1,M),
          &CELL=1,N),TLENG
        100 FORMAT (5X,12,/,7X,7F4.1,T20,F4.1,T36,5F4.1,T40,F4.1,T56,5F4.1,
          &T60,F4.1,T76,5F4.1,T80,F4.1,T96,5F4.1,T100,F4.1,T116,4F4.1,
          &F3.1,T121,F3.1,T135,4F3.1,F4.1)
        WRITE(7,104)SET,((TRUSS(SIDE,CELL),SIDE=1,M),CELL=1,N),TLENG
        CALL FINDP
        WRITE(7,105)(XPOINT(L),L=1,NPNTS),(YPOINT(LL),LL=1,NPNTS)
        CALL DRAWIT(PLABEL)
        NFISH=NFISH+1
        IF(NFISH.GT.100)GO TO 17
        GO TO 99

```

```

C
  ELSE
C  USE FOR COMPOSITE FISH
    READ(4,201)DISTR, TLENG
199  DO 10 CELL=1,N
      READ (4,200,ERR=16,END=18)(TRUSS(SIDE,CELL),SIDE=1,M)
10   CONTINUE
200  FORMAT(6F6.2)
201  FORMAT(F6.4,F6.1)
      WRITE(7,104)SET,((TRUSS(SIDE,CELL),SIDE=1,M),CELL=1,N),TLENG
      CALL FINDP
      WRITE(7,105)(XPOINT(L),L=1,NPNTS),(YPOINT(LL),LL=1,NPNTS)
      CALL DRAWIT(PLABEL)
      NFISH=NFISH+1
      GO TO 199
    ENDIF
C
16  WRITE(7,103)
    STOP
17  WRITE(7,*)'DO YOU REALLY WANT TO DRAW MORE THAN 100 TRUSSES?'
18  CALL DONEPL
103  FORMAT ('ERROR WHILE READING, CHECK YOUR DATA')
104  FORMAT (12,2X,24F5.1,/,12F5.1,F4.1,5F4.1,F5.1,/)
105  FORMAT ('X COORDINATES: ',16F6.1/'Y COORDINATES: ',16F6.1)
    STOP
  END
C
C

```

```

      SUBROUTINE FINDP
C
C THIS SUBROUTINE CHANGES THE TRUSS DISTANCES INTO X,Y COORDINATES
C SUITABLE FOR PLOTTING.
      REAL TRUSS(6,7),TLENG,XPOINT(16),YPOINT(16),RADNXT,RADLST,TRLENG
      REAL YNEXTL,YLAST,K1,K2,A,B,C,THETA,XSUB,XADD,YSUB,YADD,IM,D,DY,DX
      REAL XMID,YMID,XLAST,XNEXTL,THETB,TEMP,TEMP2,PI,DISTOR
      INTEGER II,MISS,NTEMP,I,J,IJ,MM,NN,KK,NTEMP3
      INTEGER LPOS(7),CELL,NEW,SET,JK,NFISH
      COMMON TRUSS,TLENG,SET,XPOINT,YPOINT,NFISH,DISTOR,TRLENG
      PARAMETER (N=NUMCELLS,NPNTS=2*N+2,PI=3.141592653)
      DATA LPOS(1),LPOS(2),LPOS(3),LPOS(4),LPOS(5),LPOS(6),LPOS(7)
      &/-1,2,1,1,2,-1,-1/
C ONE SIDE OUT OF SIX IS REDUNDANT AND IS NOT NEEDED TO CALCULATE
C THE X,Y CO-ORDINATES OF EACH CELL. THE SHORT SIDES SHOULD BE
C USED LEAST THEY BE SQUEEZED OUT OF EXISTENCE.
C WHEN LPOS IS ONE, THE TOP SIDE OF THE CELL IS NOT USED,
C WHEN IT IS -1 THE BOTTOM SIDE IS NOT USED.
C AND WHEN IT IS TWO THE END IS NOT USED.
C
      DO 10 II=1,NPNTS
        XPOINT(II)=0.0
        YPOINT(II)=0.0
10    CONTINUE
        YPOINT(2)=TRUSS(2,1)
C THE VALUES OF THE FIRST TWO COORDINATES ARE NOW DEFINED
      DO 20 J=1,N
        CELL=J
        NEW=2*CELL
        DO 30 I=1,2
          MISS=LPOS(CELL)
          IF (I.EQ.1) THEN
C DO THE FOLLOWING ONLY FOR THE FIRST TRIANGLE
            NEXTL=2*CELL-1
            XNEXTL=XPOINT(NEXTL)
            YNEXTL=YPOINT(NEXTL)
            LAST=2*CELL
            XLAST=XPOINT(LAST)
            YLAST=YPOINT(LAST)
            IF ((MISS.EQ.1).OR.(MISS.EQ.2)) THEN
C DO THE FOLLOWING IF THE TOP SIDE IS NOT USED
              RADNXT=TRUSS(1,CELL)
              RADLST=TRUSS(5,CELL)
              NEW=NEW+1
            ELSE
C DO THE FOLLOWING IF THE BOTTOM SIDE IS NOT USED
              RADNXT=TRUSS(6,CELL)
              RADLST=TRUSS(3,CELL)
              NEW=NEW+2
            ENDIF
          ELSE
            NEW=NEW+1
          ENDIF
        END DO
      END DO
      ELSE

```

```

C DO THE FOLLOWING ONLY FOR THE SECOND TRIANGLE
  IF (MISS.EQ.1) THEN
C DO THE FOLLOWING IF THE TOP SIDE IS NOT USED
  XLAST=XPOINT(NEW)
  YLAST=YPOINT(NEW)
C XNEXTL & YNEXTL RETAIN THE SAME VALUE AS FOR THE FIRST TRIANGLE
  RADNXT=TRUSS(6,CELL)
  RADLST=TRUSS(4,CELL)
  NEW=NEW+1
  ELSE IF (MISS.EQ.2) THEN
C DO THE FOLLOWING IF THE END IS NOT USED
C NB. NOTE THAT XLAST,YLAST,XNEXTL, & YNEXTL STAY THE SAME.
  RADNXT=TRUSS(6,CELL)
  RADLST=TRUSS(3,CELL)
  NEW=NEW+1
  ELSE
C DO THE FOLLOWING IF THE BOTTOM SIDE IS NOT USED
  XNEXTL=XPOINT(NEW)
  YNEXTL=YPOINT(NEW)
C XLAST & YLAST RETAIN THE SAME VALUE AS FOR THE FIRST TRIANGLE
  RADNXT=TRUSS(4,CELL)
  RADLST=TRUSS(5,CELL)
  NEW=NEW-1
  ENDIF
  ENDIF
  K1=(XNEXTL**2+YNEXTL**2+RADLST**2-RADNXT**2-XLAST**2
&-YLAST**2)/(2*(YNEXTL-YLAST))
  K2=(XLAST-XNEXTL)/(YNEXTL-YLAST)
  A=1+K2**2
  B=2*K1*K2-2*K2*YNEXTL-2*XNEXTL
  C=XNEXTL**2+K1**2+YNEXTL**2-2*K1*YNEXTL-RADNXT**2
  IM=B**2-4*A*C
  IF (IM.LT.0.0) GO TO 999
  IM=(SQRT(IM))/(2*A)
  XADD=(-1.0*B/(2*A))+IM
  XSUB=(-1.0*B/(2*A))-IM
  IF (I.EQ.1) THEN
C IF THIS IS THE FIRST TRIANGLE, TAKE THE LARGEST X VALUE
  XPOINT(NEW)=XADD
  YPOINT(NEW)=K1+K2*XADD
  ELSE
C IF THIS IS THE SECOND TRIANGLE
  YADD=K1+K2*XADD
  YSUB=K1+K2*XSUB
  IF (MISS.EQ.2) THEN
    XPOINT(NEW)=XADD
    YPOINT(NEW)=YADD
  ELSE IF ((MISS.EQ.1.AND.YADD.GT.YSUB).OR.
& (MISS.EQ.-1.AND.YADD.LT.YSUB)) THEN
C IF THE TOP SIDE IS NOT USED WE WANT THE LARGEST VALUE OF Y

```

```

C WHEREAS IF THE BOTTOM SIDE IS NOT USED WE WANT THE SMALLEST.
      XPOINT(NEW)=XADD
      YPOINT(NEW)=YADD
    ELSE
      XPOINT(NEW)=XSUB
      YPOINT(NEW)=YSUB
    END IF
    IF (CELL.LT.N) THEN
C IF NOT THE LAST CELL DO THE FOLLOWING
C MUST ROTATE THE COMPLETE CELL NOW SO THAT IT IS ON X AXIS
C TELLS US WHICH WAY IS UP.
      M=-1+2*CELL
      MN=1+2*CELL
      THETB=ATAN((YPOINT(MN)-YPOINT(M))/(XPOINT(MN)-XPOINT(M)))
      DO 40 IJ=1,NPNTS
        TEMP=XPOINT(IJ)
        XPOINT(IJ)=XPOINT(IJ)*COS(THETB)+YPOINT(IJ)*SIN(THETB)
        YPOINT(IJ)=YPOINT(IJ)*COS(THETB)-TEMP*SIN(THETB)
40      CONTINUE
C NOW ADJUST THE SHARED SIDE TO ITS VALUE IN THE NEXT TRUSS
      NTEMP=CELL+1
      D=(TRUSS(2,NTEMP)-TRUSS(4,CELL))
C VALUE OF LOWER POINT
      MM=1+2*CELL
C VALUE OF UPPER POINT
      NN=2+2*CELL
      THETA=ATAN((YPOINT(NN)-YPOINT(MM))/
&      (XPOINT(NN)-XPOINT(MM)))
      IF (THETA.LT.0.0) THEN
        THETA=PI+THETA
      ENDIF
      DY=0.5*D*SIN(THETA)
      DX=0.5*D*COS(THETA)
      XPOINT(MM)=XPOINT(MM)-DX
      XPOINT(NN)=XPOINT(NN)+DX
      YPOINT(MM)=YPOINT(MM)-DY
      YPOINT(NN)=YPOINT(NN)+DY
    ENDIF
  ENDIF
30  CONTINUE
20  CONTINUE
C ROTATE COMPLETED TRUSS BACK TO STANDARD ORIENTATION
C BISECT "FISH" AT PSEUDOLATERAL LINE
C FIND POINT BISECTING SIDE 4 IN LAST CELL.
      NTEMP3=NPNTS-1
      XMID=XPOINT(NTEMP3)+0.5*(XPOINT(NPNTS)-XPOINT(NTEMP3))
      YMID=YPOINT(NTEMP3)+0.5*(YPOINT(NPNTS)-YPOINT(NTEMP3))
      THETA=ATAN(YMID/XMID)
      DO 50 JK=1,NPNTS
        TEMP2=XPOINT(JK)
        XPOINT(JK)=XPOINT(JK)*COS(THETA)+YPOINT(JK)*SIN(THETA)
        YPOINT(JK)=YPOINT(JK)*COS(THETA)-TEMP2*SIN(THETA)
50  CONTINUE
      TRLENG=((XMID-XPOINT(1))**2+(YMID-YPOINT(1))**2)**0.5
C TRANSLATE POINTS TO ABOVE X AXIS

```

```
      DO 60 KK=1,NPNTS
        YPOINT(KK)=YPOINT(KK)+0.25*TLENG
60    CONTINUE
      RETURN
999  WRITE(7,*)'THE ROOTS OF THE QUADRATIC ARE IMAGINARY, '
      WRITE(7,*)'CHECK YOUR DATA!!!!!!'
      RETURN
      END
C
C
```

## SUBROUTINE DRAWIT(PLABEL)

```

C
C**THE SUBROUTINE DRAW AND DRAWIT PLOT THE POINTS WHOSE CO-ORDINATES
C ARE IN XPOINT AND YPOINT USING THE DISSPLA GRAPHICS PACKAGE BY
C ISSCO. YOU WILL HAVE TO MODIFY THEM FOR OTHER GRAPHICS PACKAGES.
C
C
      REAL CURVX1(30),CURVY1(30),CURVX2(14),CURVY2(14)
      REAL XPOINT(16),YPOINT(16),TRUSS(6,7),TLENG,DISTOR,TRLENG
      INTEGER PENPS1(30),PENPS2(14),I,J,SET,ISET,N,NPNTS,NCURV1,NCURV2
      CHARACTER PLABEL*19
      COMMON TRUSS,TLENG,SET,XPOINT,YPOINT,NFISH,DISTOR,TRLENG
      PARAMETER (N=NUMCELLS,NPNTS=2*N+2,NCURV1=2*NPNTS-2,NCURV2=NPNTS-2)
      DATA PENPS1/1,4,3,2,1,3,6,4,5,6,7,5,8,7,10,8,9,10,
&11,9,12,11,14,12,13,14,15,13,16,15/
      DATA PENPS2/2,4,3,5,6,8,7,9,10,12,11,13,14,16/
C
      ISET=SET
      DO 10 I=1,NCURV1
        CURVX1(I)=XPOINT(PENPS1(I))
        CURVY1(I)=YPOINT(PENPS1(I))
10  CONTINUE
      DO 20 J=1,NCURV2
        CURVX2(J)=XPOINT(PENPS2(J))
        CURVY2(J)=YPOINT(PENPS2(J))
20  CONTINUE
      CALL DRAW (CURVX1,CURVY1,CURVX2,CURVY2,ISET,NFISH,DISTOR,
&PLABEL,TLENG,TRLENG)
      RETURN
      END
C
C

```

```

SUBROUTINE DRAW (CURVX1,CURVY1,CURVX2,CURVY2,ISET,NFISH,DISTOR,
&PLABEL,TLENG,TRLENG)
C
CHARACTER PLABEL*19
REAL CURVX1(30),CURVY1(30),CURVX2(14),CURVY2(14),DISTOR,TLENG
REAL XXSTP,XXMAX,YYSTP,YYMAX,TRLENG
INTEGER N,NPNTS,NCURV1,NCURV2,ISET
PARAMETER (N=NUMCELLS,NPNTS=2*N+2,NCURV1=2*NPNTS-2,NCURV2=NPNTS-2)
C
CALL BGNPL(NFISH)
CALL TITLE (PLABEL,-19,0,0,0,10.0,6.0)
C SCALE THE PLOT FROM 10.0 X 6.0 INCHES TO UNITS OF CO-ORDINATES.
XXSTP=(TRLENG+(TRLENG*0.1))/10.0
XXMAX=XXSTP*10.0
YYSTP=XXSTP
YYMAX=YYSTP*6.0
CALL GRAF(0.,XXSTP,XXMAX,0.,YYSTP,YYMAX)
CALL MARKER(4)
CALL CURVE (CURVX1,CURVY1,NCURV1,1)
CALL CURVE (CURVX2,CURVY2,NCURV2,0)
CALL MESSAG ('TOTAL FISH LENGTH IS      CM',28,6.0,5.0)
CALL REALNO (TLENG,103,8.5,5.0)
CALL MESSAG ('DISTORTION IS ',13,6.0,4.5)
CALL REALNO (DISTOR,104,7.7,4.5)
CALL ENDPL(0)
RETURN
END

```

## APPENDIX C:

## BMDP PROGRAM CONTROL LANGUAGE LISTINGS FOR MORPHOMETRIC ANALYSIS.

## NOTES ON USING BMDP STATISTICAL SOFTWARE.

THE CONTROL LANGUAGE LISTINGS IN THIS APPENDIX ARE THE ONES WE USED TO RUN PRINCIPAL COMPONENTS ANALYSIS, OBTAIN PLOTS OF THE PRINCIPAL COMPONENTS SCORES THAT WERE LABELLED BY GROUP, AND TO DO DISCRIMINANT ANALYSIS. WE INCLUDE THESE LISTINGS ONLY AS A GUIDE, YOU WILL HAVE TO MAKE MANY CHANGES IN THEM TO USE THEM FOR YOUR DATA. FOR EXAMPLE IN THE LISTING FOR PRINCIPAL COMPONENT ANALYSIS YOU WILL HAVE TO PROBABLY CHANGE THE FORMAT AND IF YOU HAVE OTHER THAN 7 CELLS IN YOUR TRUSS YOU WILL HAVE TO CHANGE THE NUMBER OF VARIABLES, THE ADD STATEMENT, THE USE STATEMENT, THE VARIABLE NAMES, AND THE NUMBER OF TRANSFORMATION STATEMENTS.

FOR THE LISTING TO PRODUCE THE LABELLED PLOTS OF PRINCIPAL COMPONENT SCORES YOU WILL HAVE TO CHANGE THE ABOVE AS WELL AS THE VARIABLE NUMBER IN BRACKETS INCLUDED IN THE MAX,MIN,CODE AND NAMES STATEMENT. IN THE TWO EXAMPLES OF LISTINGS FOR PCAPlot WE GIVE THERE ARE TWO DIFFERENT METHODS OF DEFINING GROUPS TO LABEL WITH A DISTINCT SYMBOL: ONE USING THE TRAN PARAGRAPH TO SUBDIVIDE BY CASE NUMBER AND ONE USING THE VALUE OF A VARIABLE; YOU WILL WANT TO CHOOSE THE APPROPRIATE ONE FOR YOUR DATA.

FOR THE DISCRIMINANT ANALYSIS LISTING YOU WILL HAVE TO CHANGE ALL OF THE ABOVE PLUS THE CODE AND NAME STATEMENT TO CORRESPOND TO THE NUMBER OF GROUPS YOU HYPOTHEZIZE FOR YOUR DATA. YOU WILL ALSO HAVE TO CHANGE THE PRIOR STATEMENT (NOTE THAT FOR A GIVEN GROUP THAT THE PRIOR PROBABILITY IS EQUAL TO THE NUMBER OF FISH IN THAT GROUP DIVIDED BY THE TOTAL NUMBER OF FISH IN THE ANALYSIS). DEPENDING ON THE RESULTS YOU MAY ALSO WANT TO ADJUST THE F TO ENTER AND REMOVE.

TO MAKE THE ABOVE CHANGES YOU WILL WANT TO REFER TO THE 1983 REVISED PRINTING OF THE BMDP STATISTICAL SOFTWARE MANUAL. THERE IS ALSO A HANDY QUICK REFERENCE MANUAL THE BMDP USER'S DIGEST. FOR THE CYBER INSTALLATION OF BMDP YOU WILL WANT TO REFER TO THE ONLINE DOCUMENTATION (AVAILABLE AT YOUR COMPUTER CENTRE IF THEY HAVE BMDP) "BMDP-83 (CDC VERSION) FOR NOS OPERATING SYSTEMS" WHICH WAS RELEASED BY NORTHWESTERN UNIVERSITY IN APRIL 1985.

TO RUN EITHER THE PCAONE PROCEDURE OR THE DISCONE PROCEDURE ON A CYBER WITH BMDP83 AND NOS 2.4 YOU JUST NEED TO MAKE THE CHANGES DESCRIBED ABOVE TO CUSTOMIZE IT FOR YOUR DATA, TYPE THE PROCEDURE NAME AND ANSWER THE INTERACTIVE QUESTIONS ABOUT THE NAMES OF YOUR INPUT AND OUTPUT FILES. NOTE THAT AUTOSAV IS A PROCEDURE WE HAVE IN OUR LIBRARY AT M.F.D. THAT DECIDES IF A FILE IS DIRECT OR INDIRECT THEN USES EITHER RETURN OR SAVE AS APPROPRIATE. YOU CAN REPLACE THESE TWO LINES WITH "SAVE,FILENAME." UNLESS YOU HAVE THOUSANDS OF FISH IN WHICH CASE "RETURN, FILENAME." IS NECESSARY. THE TWO LINES OF THE PROCEDURE AFTER /JOB RELATE TO ACCOUNTING OF COMPUTER FUNDS AND ALLOCATION OF MEMORY AND TIME AND MAY DIFFER ON OTHER CYBERS. ASK YOUR CONSULTANTS BUT INFORM THEM THAT MULTIVARIATE BMDP PROGRAMS REQUIRE LOTS OF TIME AND MEMORY.

NOTE THAT SOME COMPUTER CENTRES WITH CYBERS, SUCH AS DALHOUSIE UNIVERSITY, PUT ALL THE BMDP PROGRAMS IN ONE GIANT FILE WHICH MAY HAVE A DIFFERENT NAME THEN THE ONE ASSUMED HERE; CONSULT YOUR COMPUTER CENTRE THEN CHANGE THE ATTACH COMMAND. NOTE THAT PEOPLE ON COMPUTERS THAT USE OTHER OPERATING SYSTEMS WILL NOT USE THESE PROCEDURES. HOWEVER THE BMDP CONTROL LANGUAGE (EVERYTHING AFTER THE "/EOR") IS ALMOST IDENTICAL ON MOST TYPES OF COMPUTERS.

THE LABELLED PLOTS OF THE PRINCIPAL COMPONENT SCORES (PCAPLOT) ARE BEST MADE BY RUNNING BMDP6D INTERACTIVELY. TO DO THIS ON OUR CYBER:

GET,SAVEFILE. (IT COULD BE DIRECT ACCESS IF IT IS REALLY LARGE).

ATTACH,BMDP6D/UN=LIBRARY.

BMDP6D,L=OUTPUT,W=15000.

AFTER THE JOB HAS RUN YOU CAN SEND OUTPUT TO THE PRINTER.

```

.PROC,PCAONE#I,DATAFILE"INPUT FILE WITH RAW DATA"=(#F),
SAVEFILE"DESIRED NAME OF BMDP SAVE FILE #"=(#F),
RESULTS"DESIRED NAME OF FILE WITH BMDP OUTPUT"=(#F),
OUTTITLE"LABEL FOR BMDP OUTPUT LESS THAN 20 CHAR".
* THIS DOES PCA ON LOG-TRANSFORMED DATA.
SUBMIT,SUBJOB.
.DATA,SUBJOB.
/JOB
PCA1,CM177000,T290.
/READ,MFDU1
GET,DATAFILE/NA.
IFE,.NOT.FILE(DATAFILE,AS),NOGET.
    ATTACH,DATAFILE.
ENDIF,NOGET.
ATTACH,BMDP4M/UN=LIBRARY.
BMDP4M,L=OUTPUT,W=20000.
BOMB.
EXIT.
ENQUIRE,R.
DAYFILE,DAYPCA1.
REPLACE,DAYPCA1.
GET,AUTOSAV/UN=LIBRARY.
AUTOSAV,OUTPUT,RESULTS.
AUTOSAV,SAVEFILE.
REWIND,OUTPUT.
/EOR
/PROBLEM TITLE='PCA LOG-TRANS OUTTITLE'.
/INPUT VARIABLE=39.
    FILE=DATAFILE.
    FORMAT='(28X,3X,/,1X,A4,F2.0,31F4.1,5F3.1,F4.1)'.
/VARIABLE
    NAMES=TAG,SET,A1,A2,A3,A4,A5,A6,B1,B3,B4,B5,B6,C1,C3,C4,C5,C6,D1,
        D3,D4,D5,D6,E1,E3,E4,E5,E6,F1,F3,F4,F5,F6,G1,G3,G4,G5,G6,TLENG,
    LA1,LA2,LA3,LA4,LA5,LA6,LB1,LB3,LB4,LB5,LB6,LC1,LC3,LC4,LC5,LC6,
    LD1,LD3,LD4,LD5,LD6,LE1,LE3,LE4,LE5,LE6,LF1,LF3,LF4,LF5,LF6,
    LG1,LG3,LG4,LG5,LG6,LTLENG.
    ADD=37.
USE=LA1 TO LG6.
    LABEL=TAG.
/TRAN

```

```

IF(SET EQ 0) THEN USE=-1.
    LA1=LOG(A1).
    LA2=LOG(A2).
    LA3=LOG(A3).
    LA4=LOG(A4).
    LA5=LOG(A5).
    LA6=LOG(A6).
    LB1=LOG(B1).
    LB3=LOG(B3).
    LB4=LOG(B4).
    LB5=LOG(B5).
    LB6=LOG(B6).
    LC1=LOG(C1).
    LC3=LOG(C3).
    LC4=LOG(C4).
    LC5=LOG(C5).
    LC6=LOG(C6).
    LD1=LOG(D1).
    LD3=LOG(D3).
    LD4=LOG(D4).
    LD5=LOG(D5).
    LD6=LOG(D6).
    LE1=LOG(E1).
    LE3=LOG(E3).
    LE4=LOG(E4).
    LE5=LOG(E5).
    LE6=LOG(E6).
    LF1=LOG(F1).
    LF3=LOG(F3).
    LF4=LOG(F4).
    LF5=LOG(F5).
    LF6=LOG(F6).
    LG1=LOG(G1).
    LG3=LOG(G3).
    LG4=LOG(G4).
    LG5=LOG(G5).
    LG6=LOG(G6).
    LTENG=LOG(TLENG).
/FACTOR FORM=COVA.
    NUMBER=5.
    CONST=0.0125.
/ROTATE METHOD=NONE.
/SAVE CODE IS SECRET.
    NEW.
    FILE=SAVEFILE.
    FORM=BMDP.
    CONT=DATA.
/END

```

```

/PROBLEM TITLE IS 'HADDOCK: PLOT OF PC SCORES, RAW DATA (MORPHOMETRICS)'.
/INPUT FILE=BHADALL.
  CODE IS SECRET.
/VAR
  MAXIMUM IS (77)3, 3, 3.
  MINIMUM IS (77)-3, -3, -3.
  GROUPING IS SET.
/GROUP
  CODES(2)=3,4,6,9,10,11,28,29,31,
           37,41,42,43,70,79,88,89.
  NAMES(2)=X,X,X,X,X,X,F,F,X,
           X,X,X,X,W,Y,X,F.
/PLOT
  YVAR ARE FACTOR2,FACTOR3,FACTOR3.
  XVAR ARE FACTOR1,FACTOR1,FACTOR2.
  GROUP IS W,X,Y,F.
  GROUP IS X.
  GROUP IS Y.
  GROUP IS W.
  GROUP IS F.
  SIZE=70,56.
/END
/PROBLEM TITLE IS 'POLLOCK: PLOT OF PC SCORES, RESIDUALS CORRELATION MATRIX'.
/INPUT FILE=BTOT.
  CODE IS SECRET.
/VAR
  ADD=1.
  MAXIMUM IS (39)3, 3, 3.
  MINIMUM IS (39)-3, -3, -3.
  NAMES=(44)SPECIES.
  GROUPING IS SPECIES.
/TRAN
  IF(KASE LT 233) THEN SPECIES=2.
  IF(KASE GE 233) THEN SPECIES=1.
/GROUP
  CODES(44)=1,2.
  NAMES(44)=H,P.
/PLOT
  YVAR ARE FACTOR2,FACTOR3,FACTOR3.
  XVAR ARE FACTOR1,FACTOR1,FACTOR2.
  GROUP IS H,P.
  GROUP IS H.
  GROUP IS P.
  SIZE=70,56.
/END

```

```

.PROC,DISCONE*I,DATAFILE"NAME OF STANDARD INPUT FILE"=(#F),
OUTFIL"FILE THAT WILL CONTAIN BMDP OUTPUT"=(#F).
* THIS PROGRAM DOES DISCRIMINANT ANALYSIS ON THE MORPHOMETRIC DATA
* YOU WILL HAVE TO CHANGE THE GROUP PARAGRAPH DEPENDING ON WHICH
* DATASET YOU USE.
* IT IS CURRENTLY SET UP FOR THE HADDOCK DATA.
* YOU MUST CHANGE THE CODES,NAMES, AND PRIOR STATEMENTS
SUBMIT,SUBJOB.
,DATA,SUBJOB.
/JOB
DISCR,T100,CM177000.
/READ,MFDUI
GET,DATAFILE/NA.
IFE,.NOT.FILE(DATAFILE,AS),NOGET.
    ATTACH,DATAFILE.
ENDIF,NOGET.
ATTACH,BMDP7M/UN=LIBRARY.
BMDP7M,L=OUTPUT,W=35000.
BOMB.
EXIT.
ENQUIRE,R.
DAYFILE,DAYDISC.
GET,AUTOSAV/UN=LIBRARY.
AUTOSAV,OUTPUT,OUTFIL.
REWIND,OUTPUT.
ROUTE,OUTPUT,DC=LP.
AUTOSAV,DAYDISC.
/EOR
/PROBLEM TITLE IS 'DISCRIMINANT ANALYSIS OF HADDOCK'.
/INPUT VARIABLE=39.
    FILE=DATAFILE.
    FORMAT='(2BX,3X,/,1X,A4,F2.0,31F4.1,5F3.1,F4.1)'.
/VARIABLE
    NAMES=TAG,SET,A1,A2,A3,A4,A5,A6,B1,B3,B4,B5,B6,C1,C3,C4,C5,C6,D1,
        D3,D4,D5,D6,E1,E3,E4,E5,E6,F1,F3,F4,F5,F6,G1,G3,G4,G5,G6,TLENG,
    LA1,LA2,LA3,LA4,LA5,LA6,LB1,LB3,LB4,LB5,LB6,LC1,LC3,LC4,LC5,LC6,
    LD1,LD3,LD4,LD5,LD6,LE1,LE3,LE4,LE5,LE6,LF1,LF3,LF4,LF5,LF6,
    LG1,LG3,LG4,LG5,LG6,LTLENG.
    ADD=37.
    USE=LA1 TO LTLENG.
    LABEL=TAG.
    GROUP=SET.
/TRAN

```

```

LA1=LOG(A1).
LA2=LOG(A2).
LA3=LOG(A3).
LA4=LOG(A4).
LA5=LOG(A5).
LA6=LOG(A6).
LB1=LOG(B1).
LB3=LOG(B3).
LB4=LOG(B4).
LB5=LOG(B5).
LB6=LOG(B6).
LC1=LOG(C1).
LC3=LOG(C3).
LC4=LOG(C4).
LC5=LOG(C5).
LC6=LOG(C6).
LD1=LOG(D1).
LD3=LOG(D3).
LD4=LOG(D4).
LD5=LOG(D5).
LD6=LOG(D6).
LE1=LOG(E1).
LE3=LOG(E3).
LE4=LOG(E4).
LE5=LOG(E5).
LE6=LOG(E6).
LF1=LOG(F1).
LF3=LOG(F3).
LF4=LOG(F4).
LF5=LOG(F5).
LF6=LOG(F6).
LG1=LOG(G1).
LG3=LOG(G3).
LG4=LOG(G4).
LG5=LOG(G5).
LG6=LOG(G6).
LTLENG=LOG(TLENG).
/GROUP
CODES(2) ARE 3,4,6,9,10,11,28,29,31,37,41,42,
              43,70,79,88,89.
NAMES(2) ARE X,X,X,X,X,X,F,F,X,X,X,X,X,W,Y,X,F.
PRIOR=.5296,.1536,.16,.1568.
/DISC ENTER=2.0,2.0.
      REMOVE=1.99,1.99.
      JACKKNIFE.
/PRINT POST.
      POINT.
/END

```