Scientia Marina 72(2)
June 2008, 221-229, Barcelona (Spain)
ISSN: 0214-8358

Developing a size indicator for fish populations

YONG CHEN [2,1], XINJUN CHEN [1] and LIUXIONG XU [1]

[1] College of Marine Sciences and Technology, Shanghai Fisheries University, 334 Jungong Road, Shanghai, China.
[2] School of Marine Sciences, University of Maine, Orono, Maine 04469, USA. E-mail: ychen@maine.edu

SUMMARY: Monitoring temporal and/or spatial variations in fish size-at-age data can often provide fisheries managers with important information about the status of fish stocks and therefore help them identify necessary changes in management policies. However, due to the multivariate nature of size-at-age data, commonly used single-age-based approaches ignore covariance between sizes of different age groups. Different results may therefore be derived when evaluating temporal variations using different age groups for the comparison. The possibility of atypical errors in size-at-age data due to ageing and measurement errors further complicates the comparison. We propose a two-step approach for developing an indicator for monitoring temporal and/or spatial variation in size-at-age data. A robust approach, minimum volume ellipsoid analysis, is used to identify possible outliers in size-at-age data. Then a weighted principal component analysis is applied to the data with the identified outliers down-weighted. An indicator is defined from the resultant principal components for monitoring temporal/spatial variations in size-at-age data. We illustrate the proposed approach with size-at-age data for cod (*Gadus morhua*) in the northwest Atlantic, NAFO subdivision 3Ps. The overall size-at-age indicator identified shows that the pre-1980 year classes tend to have a much higher size-at-age than the post-1980 year classes.

Keywords: size-at-age, robust, principal component analysis, minimum volume ellipsoid analysis, size indicator.

RESUMEN: Desarrollo de un indicador de talla para poblaciones de peces. – El seguimiento de las variaciones temporales y/o espaciales de datos de talla por edad en peces puede, a menudo, aportar información a los gestores de pesquerías sobre el estado de explotación de los *stocks* de peces y ayudarles a identificar los cambios necesarios en políticas de gestión. Sin embargo, debido a la naturaleza multivariante de los datos de talla por edad, las aproximaciones tradicionalmente empleadas, basadas en el análisis de una sola clase de edad, ignoran la covarianza entre tallas de distintos grupos de edad, lo que puede generar distintos resultados cuando se analizan variaciones temporales mediante la comparación de distintos grupos de edad. La posible existencia de errores atípicos en datos de talla por edad, debidos a errores de atribución de edad o errores de medida, puede complicar más la comparación. Proponemos una aproximación en dos etapas para el desarrollo de un indicador para el seguimiento de variaciones temporales o espaciales en datos de talla por edad. Una aproximación robusta, conocida como análisis de elipsoide de volumen mínimo, nos permite identificar los posibles valores aberrantes en datos de talla por edad, y a continuación aplicamos el análisis ponderado de componentes principales a los datos con los valores aberrantes debidamente ponderados. Los componentes principales resultantes permiten definir el indicador para el seguimiento de las variaciones espacio-temporales en datos de talla por edad. Ilustramos la aproximación propuesta con datos por edad de bacalao (*Gadus morhua*) in el Atlántico noroccidental, correspondiente a la subdivisión 3Ps de la NAFO. El indicador general de talla por edad obtenido muestra que las clases de edad anteriores a 1980 tienden a tener una talla por edad mucho mayor que las clases de edad posteriores a 1980.

Palabras clave: talla-por-edad, robusto, análisis de componentes principales, análisis de elipsoide de volumen mínimo, indicador de talla.

INTRODUCTION

Size-at-age data are important in understanding the dynamics of fish populations. They are essential in estimating fish stock biomass and productivity. The growth in size between two ages can be estimated by evaluating the differences in the sizes of these two age groups. This can provide fisheries managers with important information such as the fish growth rate and the age at which fish attain their highest growth rate (Nikolskii, 1965; Paloheimo and Dickie, 1965; Myers *et al.*, 1997). Such information is essential for formulating management policies (Hilborn and Walters, 1992). Many fish stock assessment models, such as yield-per-recruit models and delay-difference models, require size-at-age as input data (Ricker, 1975; Hilborn and Walters, 1992).

Many biotic and abiotic environmental variables can affect growth in size, and subsequently fish size-at-age values (Nikolskii, 1965; Paloheimo and Dickie, 1965; Moreau, 1987). A direct consequence of such a process is fluctuation in fish size-at-age values between different year classes (Beacham, 1983; Chen and Harvey, 1995). Closely monitoring temporal/spatial changes in size-at-age values reveals some important information about the status of fish stocks and can help fishery managers identify the necessary changes in management policies (Beverton and Holt, 1957; Ricker, 1975; Charnov, 1993). For example, a substantial decrease in fish stock biomass or overexploitation may lead to a decrease in the age of fish when they attain maturity, which may in turn result in a decrease in size-at-age (Nikolskii, 1965; Roff, 1984; Chen and Harvey, 1994; Jensen, 1996).

It can be difficult to evaluate temporal/spatial changes in fish size-at-age data due to its multivariate nature. In practice, such an evaluation is often conducted by examining the size of fish at each age separately or by evaluating the size of fish in an age group arbitrarily selected by researchers (Beacham, 1983; Lilly, 1996). This single-age-based approach ignores the covariance between sizes of different ages (e.g. size at age 1 affects size at age 2). Different results may arise from using different age groups for evaluating temporal/spatial changes in sizes.

Multivariate fisheries data are often analyzed using principal component analysis (PCA; Manly, 1991; Jackson, 1993; Chen and Harvey, 1995). This method is one of the most commonly used data-exploratory multivariate ordination techniques and allows data relationships and reductions in dimensionality to be studied (Rao, 1964; Jackson, 1993). This multivariate approach can reduce the size of data (i.e. the number of variables), while retaining the essential information inherent in the original data.

It is very likely that there will be atypical errors in size-at-age data as a result of errors in ageing, small sample sizes of an age group, and measurement errors (Chen and Mello, 1999). This leads to erroneous results when evaluating temporal/spatial patterns of fish size-at-age data (Chen and Harvey, 1994; Chen *et al.*, 1994). Thus, it is important to evaluate the possible existence of atypical data when analyzing size-at-age data.

In this study, we propose using PCA to summarize size-at-age data. As fisheries data tend to be subject to atypical errors (Chen and Harvey, 1994; Chen *et al.*, 1994), a two-step procedure is proposed: a robust multivariate approach is applied to size-at-age data to identify outliers in the data, and then a weighted PCA is applied with the defined outliers down-weighted. The resultant principal components (PCs) are interpreted with respect to the original variables (Manly, 1991; Jackson, 1993). An indicator is then identified from the resultant PCs for monitoring temporal/spatial variations in size-at-age data. The proposed approach is applied to size-at-age data from cod (*Gadus morhua*) in the northwest Atlantic, NAFO subdivision 3Ps. An overall size-at-age indicator is developed for evaluating changes in size-at-age values of cod between different year classes.

METHODS AND MATERIALS

Identifying outliers for multivariate data

Fisheries data are commonly subject to errors of various sources (Hilborn and Walters, 1992; Chen and Paloheimo, 1998; Jackson and Chen, 2003). This may result in outliers when modelling fisheries data (Chen *et al.*, 1994; Jackson and Chen, 2003). Commonly used statistical methods such as PCA can be severely biased by the existence of outliers in the data (Rousseeuw and Leroy, 1987). Outliers are much more difficult to identify in a multivariate analysis than in a univariate analysis (Rousseeuw and Leroy, 1987). The squared Mahalanobis distance (Krzanowski, 1988) is the commonly used method for identifying outliers in multivariate analyses. For a data matrix with K variables and each variable has n observations,

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{i1} & \cdots & x_{K1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{1j} & \cdots & x_{ij} & \cdots & x_{Kj} \\ x_{1n} & \cdots & x_{in} & \cdots & x_{Kn} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \cdot \\ \mathbf{x}_j \\ \cdot \\ \cdot \\ \mathbf{x}_n \end{pmatrix}$$

the squared Mahalanobis distance is calculated as

$$MD^2 (\mathbf{x}_j, \mathbf{X}) = (\mathbf{x}_j - \mathrm{T}(\mathbf{X}))\, \mathrm{C}^{-1}(\mathbf{X})(\, \mathbf{x}_j - \mathrm{T}(\mathbf{X}))^t$$

for each point $\mathbf{x}_j$, where bold letters are vectors or matrices, T(**X**) is the arithmetic mean of the data set **X** and **C**(**X**) is the classical covariance estimate. These are calculated as

$$T(\mathbf{X}) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$$

$$\mathrm{C}(\mathbf{X}) = \frac{1}{n-1}\sum_{i=1}^{n} \left(x_j - \mathrm{T}(\mathbf{X})\right)^t \left(x_j - \mathrm{T}(\mathbf{X})\right)$$

Points with large MD^2 ($\mathbf{x}_j$, **X**) values are identified as outliers and are subsequently deleted from the sample used for further analyses. This approach works well if there is only one single outlier (Rousseeuw and van Zomeren, 1990), but it may not work properly when there are more than one outlier because one distant outlier can cause all the other outliers to have small MD^2 ($\mathbf{x}_j$, X). Some refinements of this approach have been proposed, such as iterative deletion, iterative trimming, and depth trimming (Campbell, 1980; Devlin *et al.*, 1981; Rousseeuw, 1985). However, there are still problems associated with the MD^2 approach in these methods (Rousseeuw and van Zomeren, 1990; Jackson and Chen, 2003).

Rousseeuw (1984, 1985), proposed a robust method, the minimum volume ellipsoid (MVE), for identifying outliers when estimating means and covariance for multivariate data. Although it is not uncommon that data from fisheries or ecological studies are contaminated by outliers, the effects of outliers on multivariate analyses (e.g. PCA, canonical correspondence analysis, and multiscaling methods) have received little attention. The MVE has recently been used in other research fields, such as engineering and economics (Rousseeuw and Leroy, 1987), but its application in fisheries or ecological studies is limited (Jackson and Chen, 2003).

An algorithm that involves extensive computer subsampling has been suggested for the MVE analysis (Rousseeuw and Leroy, 1987). This algorithm can be summarized as follows:

(1) For a multivariate data matrix **X** with K variables and n observations (as described above), draw a subsample of K+1 different observations, indexed by J = (j_1, ..., j_{K+1}), and calculate the arithmetic mean and the corresponding covariance matrix as

$$T_J = \frac{1}{K+1}\sum_{j\in J} \mathbf{x}_j$$

and $\mathbf{C}_J = \frac{1}{K}\sum_{j\in J}\left(\mathbf{x}_j\text{-}\mathrm{T}_J\right)^t\left(\mathbf{x}_j\text{-}\mathrm{T}_J\right)$ where $\mathbf{C}_J$ is nonsingular;

(2) Calculate $m^2_J = [(\mathbf{x}_j - \mathrm{T}_J)\, \mathbf{C}^{-1}_J(\mathbf{x}_j - \mathrm{T}_J)^t]_{h:n}$ where h = (n + K +1)/2; in the above computation the ellipsoid should be inflated or deflated to contain exactly h points (out of n points);

(3) Calculate $P_J = (\det(m^2_J\, \mathbf{C}_J))^{1/2}$;

(4) Repeat the above procedure for a large number of subsample J, and retain the one with the lowest P_J;

(5) For this retained subsample J, compute $T(\mathbf{X}) = T_J$ and $\mathbf{C}(\mathbf{X}) = c^2$ (n, K)$(\chi^2_{K,0.50})^{-1} m^2_J \mathbf{C}_J$, where c^2 (n, K) is a small-sample correction term calculated as $[1+15/(n\text{-}K)]^2$ and $\chi^2_{K,0.50}$ is the median of the χ^2 distribution with K degrees of freedom.

The T(**X**) and **C**(**X**) calculated in step (5) are the MVE-estimated mean and covariance matrices.

Intensive sampling and computation are necessary in order to find the solution in the MVE analysis. The total number of subsampling required depends on the values of *K* and *n* (Rousseeuw and Leroy, 1987). It increases quickly with an increase in *K* and/or *n*. Based on the MVE-estimated mean T(**X**) and covariance **C**(**X**), the following statistic, which is similar to MD^2, can be calculated,

$$W_j^2 = \left(\mathbf{x}_j - \mathrm{T}(\mathbf{X})\right)\mathbf{C}^{-1}(\mathbf{X})\left(\mathbf{x}_j - \mathrm{T}(\mathbf{X})\right)^t.$$

For a data point $\mathbf{x}_j$, if $W_j^2 > \chi^2_{K,0.975}$, it is defined as an outlier, otherwise it is defined as a "normal" observation.

Principal component analysis

Principal component analysis is a multivariate technique for examining the relationship between several quantitative variables. Giving a data set with *K* numerical variables $\mathbf{X}_1$, $\mathbf{X}_2$, …, and $\mathbf{X}_K$, each of which has *n* individuals, PCA linearly transforms the variables $\mathbf{X}_1$, $\mathbf{X}_2$, …, and $\mathbf{X}_K$, to new variables $\mathbf{Y}_1$, $\mathbf{Y}_2$, …, and $\mathbf{Y}_K$,. These new variables are the principal components (PC). The original data observation $\mathbf{X}^{(i)}$, which is the observation vector for the i^th^ individual denoted as $\mathbf{X}^{(i)}=(X_{i1}, X_{i2}, \ldots, X_{iK})^t$, is transformed to the corresponding PC scores $\mathbf{Y}^{(i)}=(\, Y_{i1}, Y_{i2}, \ldots, Y_{iK})^t$. Each PC is a linear combination of the original variables, with coefficients equal to the eigenvectors of the correlation or covariance matrix of the *K* variables. The principal components are sorted by descending order of the eigenvalues, which are equal to the variances of the components (Rao, 1964).

Principal component analysis is often used to summarize multivariate data and reduce the number of variables (Rao, 1964; Cooley and Lohnes, 1971). Dimensionality reduction is effective when q ($q<K$) of the components $\mathbf{Y}$ convey most of the sample information inherent in $\mathbf{X}$. In this case the original observations $\mathbf{X}$(i) can be replaced by the first q elements of the corresponding PC scores. The number of variables measured in a fisheries or ecological study is often large. A PCA can be used to replace a large number of the original variables with a few PCs. These derived PCs are then used for further regression analyses with other variables (e.g. principal component regression analysis: Hill *et al.*, 1977; Mason and Gunst, 1985; Vogt and Kolsett, 1987).

A two-step procedure is proposed for developing an indicator for monitoring temporal/spatial variations in size-at-age data. The robust MVE procedure is applied first to size-at-age data to identify possible outliers in the data. In the next step a weighted PCA (SAS, 1987) is applied to the size-at-age data. For the weighted PCA, data identified as outliers in the MVE analysis are given a weight of 0 (thus effectively removing the impact of these data in the PCA) and the other "normal" data are given a weight of 1. For size-at-age data with K age groups observed for n years (or n year classes), the number of PCs derived from the PCA is K. The resultant K PCs are interpreted with respect to the original size-at-age data using eigenvector values calculated from the PCA. This can establish the relationship between the PCs and original size-at-age data. If the correlation between the sizes of age groups is high, the first PC, which always explains the largest proportion of variance inherent in the original data between all PCs, will be a good indicator of the sizes of fish of all the age groups included in the analysis. This PC can then be interpreted as an overall indicator of fish size-at-age. Temporal changes in size-at-age can be evaluated using the scores of the first PC.

Application

Previous studies have shown that cod in many areas of northwest Atlantic Canada have experienced pronounced changes in growth over the last 20 years (Beacham, 1983; Hutchings and Myers, 1994; Lilly, 1996; Shelton *et al.*, 1996; Myers *et al.*, 1997). Declining size-at-age was observed in some stocks during this period. However, previous studies that evaluated temporal changes in size-at-age data did not consider the multivariate nature of the data. Temporal variations in cod size-at-age data were usually evaluated separately for each age group (Beacham, 1983; Lilly, 1996). Such an approach disregards covariance in sizes between different age groups. As temporal patterns may differ for different age groups, inconsistency may arise when different age groups are used for evaluating temporal variation in size-at-age values.

In this study, the proposed two-step approach was applied to cod size-at-age data collected from a fishery-independent bottom trawl survey in the northwest Atlantic, NAFO subdivision 3Ps. Size-at-age data were available for 20 year-classes from 1971 to 1990. The between-cohort variations in size-at-age data of cod were examined using the PCs derived from the proposed method. Since cod is a long-lived fish species, to avoid the problem of nonlinearity, only the first 6 age groups of size data were included in the analysis. We also log-transformed the data because, like many other fishery variables, size-at-age data tend to follow a log-normal distribution and normality is assumed in PCA.

In order to make the size scales in different age groups comparable, log-transformed size-at-age data were standardized using the following formula:

$$y_{ij} = \frac{x_{ij} - \bar{x}_i}{S_i} ,$$

where x_{ij} is log-size at age i for year class j, $\bar{x}_i$ and S_i are the mean and standard deviations of logarithm sizes at age i across all year classes, and y_{ij} is standardized log-size at age i for year class j. This standardization did not change the temporal variation patterns in size-at-age, but it did ensure that size-at-age data had the same scale for different age groups.

RESULTS

Variations in size were observed between year classes included in this study for each age group (Fig. 1). The size-at-age values of the recent year classes tended to be smaller when compared with those of the 1970s year classes. However, because data were log-transformed and there were large differences in sizes between different age groups, the differences in temporal variations between age groups were difficult to evaluate from Figure 1 as commonly done when showing temporal variations graphically.

The standardized log-transformed size-at-age data showed differences in temporal variations in sizes in different age groups (Fig. 2). Temporal variations

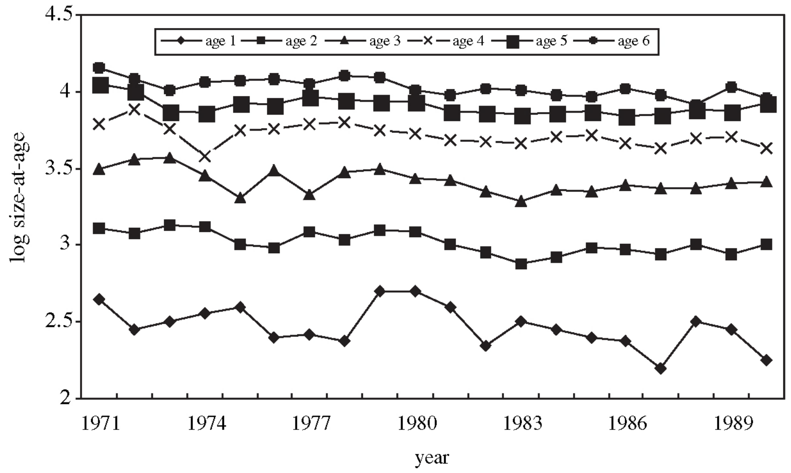

FIG. 1. – Variations in size (cm TL) of six age groups for the 1971 to 1990 year classes.

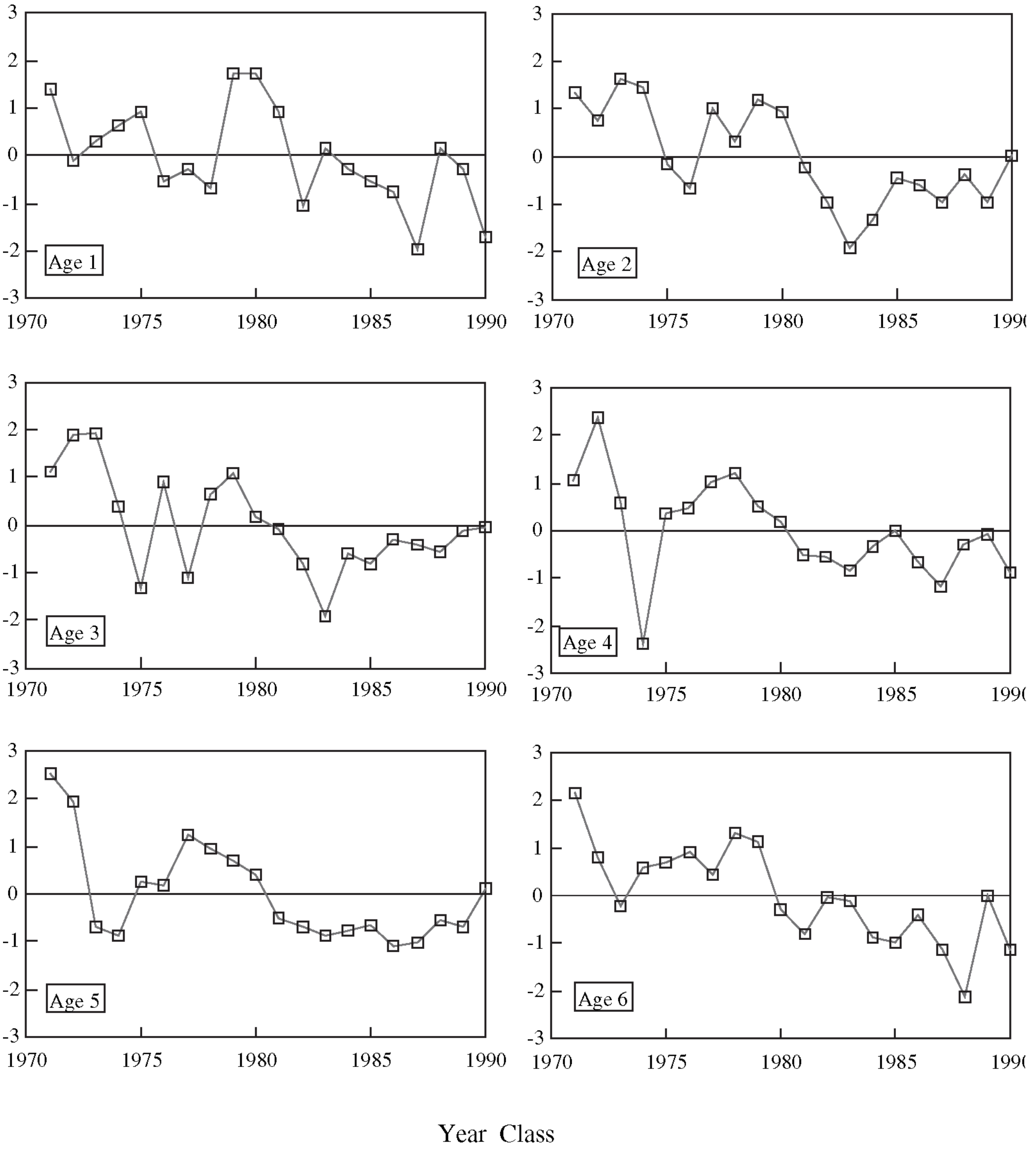

FIG. 2. – Plot of standardized size-at-age data for six age groups of 20 year-classes from 1971 to 1990.

TABLE 1. – Eigenvectors for the first three components in the principal component analysis of logarithmic sizes (cm TL) at ages 1 to 6 for cod in 3Ps.

	Principal component		
Log size at age	PC I	PC II	PC III
1	0.30	0.03	0.90
2	0.45	0.41	0.07
3	0.32	0.70	-0.31
4	0.50	-0.16	-0.11
5	0.45	-0.34	-0.11
6	0.40	-0.46	-0.25

were large for age groups 1, 2, 3 and 6, while the variations were relatively small for ages 4 and 5 (Fig. 2). In general, we conclude that the size-at-age tends to decrease for recent year classes. However, different interpretations could be derived with respect to detailed temporal variations when different age groups were used.

Three year-classes, 1971, 1972, and 1974, were identified as outliers in the MVE analysis of the size-at-age data. They were subsequently given a weight of 0 in the PCA. Eighty-six percent of the variance in the size-at-age data was explained by the first three principal components in the PCA. The first PC explained 56% of the variance, and the second and third PCs explained 17% and 13% respectively. PC4, PC5 and PC6 together only explained 14% of the variance inherent in the original size-at-age data, and were thus not important to this study.

The correlation coefficients between size-at-age variables and the first PC in the eigenvector ranged from 0.50 for age 4 to 0.30 for age 1 (Table 1). Such a small range in the correlation coefficients suggests that the first component was an overall indicator of sizes for all six age groups. The correlation coefficients between the size-at-age variables and the first PC were positive for all six age groups (Table 1). This implies that a year-class with a larger score for the first PC tends to have a larger size.

The scores for the first PC varied greatly among year classes (Fig. 3). The 1971 year-class, which was identified as an outlier in the MVE analysis, had the largest value for the scores of the first PC. This indicates that the 1971 year class had the largest size prior to age 7 among the cohorts included in this study. The score values (thus sizes) decreased for cod from year class 1971 to 1974, and then increased from year class 1974 to 1979. The sizes decreased again from the 1979 year class. From year class 1981 to the most recent year class included in this study, the sizes of cod prior to age 7 were much smaller than those for year classes in the 1970s (Fig. 3).

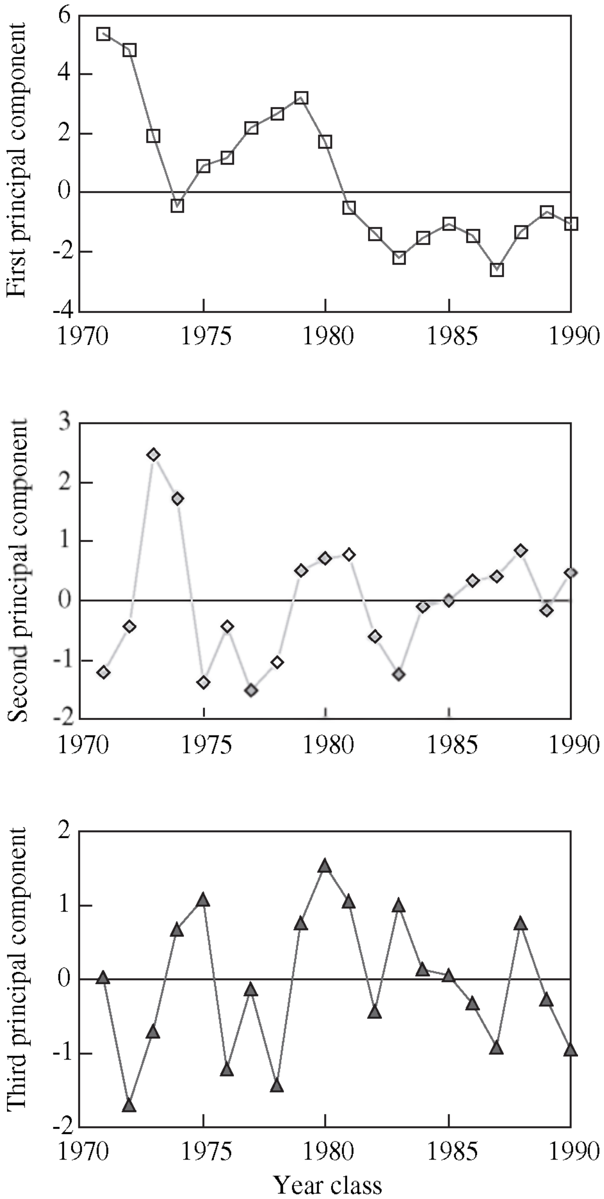

FIG. 3. – Plot of the first three principal components (PC) derived in the proposed principal component analysis of size data of age groups 1 to 6 for the 1971 to 1990 year classes.

The correlation between the second PC and size at age 2 was positive and much higher than the correlation between the second PC and other age groups (Table 1). Thus, the second PC was positively related to size at age 3. A year class with a large score for the second PC tended to have a large size at age 3. The correlation between the third PC and size at age 1 was positive and much higher than the correlation between the third PC and other age groups (Table 1). Thus, the third PC was positively related to size at age 1. A year class with a large score for the third PC tended to have a large size at age 1. The scores for both the second and third PCs fluctuated for year classes from 1971 to 1990 with no clear-cut patterns (Fig. 3). An evaluation of the correlation coefficient between the size-at-age variables versus the second and/or third PC suggests that neither the second nor the third PC are a good indicator for sizes of all age groups, and thus cannot be used for evaluating overall temporal variations in size-at-age.

DISCUSSION

Principal component analysis is commonly used to reduce the size of a large set of data without losing information inherent in the data (Jackson and Chen, 2003). Of all the principal components derived in PCA, PC1 always describes the largest proportion of variations inherent in the original data, followed by PC2 and PC3. Due to this characteristic of PC1, it is more likely to be a good overall indicator of size-at-age data than the other PCs. This study suggests that PC1 is a good overall size indicator, which is consistent with the above argument.

The size of fish in one age group is often chosen in fisheries studies as an indicator of fish size without considering the size of fish in other age groups. This may result in different interpretations of spatial or temporal change in size-at-age data if different age groups are chosen. The approach we proposed in this study considers variations in all the age groups and derives an overall size indicator which reflects overall variations in sizes in all age groups better.

A simulation study is often carried out in fisheries with data that has been simulated with a prior knowledge of the statistical attributes in order to evaluate the performance of a proposed modelling approach (e.g. Chen and Paloheimo, 1998; Jackson and Chen, 2003). This kind of simulation study is a necessary step for developing and evaluating a new stock assessment model with unknown statistical properties. However, PCA is a standard multivariate statistical method and its statistical properties are well known. The performance of PCA in association with MVE has been evaluated in an extensive simulation study conducted by Jackson and Chen (2003). Thus, it is not necessary to run a simulation study to evaluate the performance of PCA when analyzing size-at-age data in this study.

An assumption implied in a PCA is that the relationship between variables included in the PCA is linear (Rao, 1964; Manly, 1991). This assumption may be violated when analyzing fish size-at-age data because the growth rate for size tends to decrease with age and the relationship between sizes of different ages may be nonlinear, especially for long-lived fish (Ricker, 1975). Two approaches can be used to avoid this problem: transforming size-at-age data (e.g. logarithm) and grouping age classes so as to ensure that the relationship between sizes of age classes within each group is linear. These two approaches can be used together. If grouping of age classes is used, PCA should be conducted separately for each age group. In this study, we only include the first six age groups. Since cod age groups tend to grow quickly (Chen and Mello 1999), their relationship is more likely to be linear.

The MVE analysis showed that three year-classes (i.e. 1971, 1972 and 1974) were outliers. This might have resulted from exceptional values in some age groups for these year classes. The 1970 year class had exceptionally high values for sizes at ages 5 and 6. The 1971 year class had high values for sizes at ages 4 and 5, while the 1974 year classes had exceptionally low sizes at age 4. However, without knowing the level of measurement errors associated with these size data, it is difficult to tell whether these year classes were defined as outliers as a result of exceptionally large measurement errors or for other reasons. If it is reasonable to assume that measurement errors are more or less the same for all data included in the study, the exceptional values may result from exceptional growth or the inclusion of a large number of samples from a different stock with different growth patterns for these three year-classes (Rollet *et al.*, 1995; Lilly, 1996).

Different hypotheses have been developed to explain the decrease in size-at-age observed in cod populations in recent years. These hypotheses include large scale temporal variations in water temperature (Beacham, 1983; Hutchings and Myers, 1994; Gomes *et al.*, 1995), changes in stock biomass (Hanson and Chouinard, 1992; Swain, 1993), stock overfishing (Trippel, 1998) and variation in prey species biomass (Krohn *et al.*, 1997). Regardless of the factors causing the decrease in size-at-age, most authors suggest that this phenomenon is indicative of population stress (Kovtsova, 1995; Trippel, 1995; 1998). If observations/data are available for these environmental variables, the overall indicator for size-at-age identified in the proposed PCA can be used as a variable representing fish size in regression analysis with these environmental variables. A principal component regression analysis can identify whether the temporal variations in fish size are related to the temporal variations of environmental variables (Hill *et al.*, 1977; Vogt and

Kolsett, 1987). This approach can reduce the number of variables without losing information inherent in the original variables. When size-at-age data are used directly in an analysis with environmental variables, an arbitrary decision has to be made when determining which age group should be included in the analysis. Information for other age groups that are not included in the analysis is thus lost.

The size-at-age-1 data were obtained from small sample sizes (Lilly, 1996), and were thus less reliable. However, although this may undermine any reliable interpretation when using this kind of data series in a conventional univariate analysis, it is less disruptive to the method proposed in this study because it assesses the cumulative effect of size-at-age over the total period of time the cohorts are considered in the study. The ability to minimize disruptions from data quality or availability issues is an important strength of the methodology proposed in this study.

ACKNOWLEDGEMENTS

Financial support for this study was partially provided by the Program for New Century Excellent Talents in University and Shanghai Leading Academic Discipline Project (Project # T1101). The manuscript has been modified from the authors' work included in the Canadian Stock Assessment Research Documents. We thank two anonymous reviewers and the Editor, F. Maynou, for constructive, helpful comments that improved the manuscript.

REFERENCES

Beacham, T.D. – 1983. Growth and maturity of Atlantic cod (*Gadus morhua*) in the southern Gulf of St. Lawrence. *Can. Tech. Rep. Fish. Aquat. Sci.*, 1142.

Beverton, R.J.H. and S.J. Holt. – 1957. On the dynamics of exploited fish populations. *Fish. Invest. Ser. 2 Mar. Fish. G.B. Minist. Agric. Fish. Food* ., 19.

Campbell, N.A. – 1980. Robust procedures in multivariate analysis I: robust covariance estimation. *Applied Stats.*, 29: 231-237.

Charnov, E. – 1993. *Life History Invariants*. Oxford University Press, New York.

Chen, Y. and H.H. Harvey. – 1994. Maturation of white sucker, *Catostomus commersoni*, populations in Ontario. *Can. J. Fish. Aquat. Sci.*, 51: 2066-2076.

Chen, Y. and H.H. Harvey. – 1995. Growth, abundance, and food supply of white sucker. *Trans Am. Fish. Soc.*, 124: 262-271.

Chen, Y. and G.S. Mello. – 1999. Growth and maturation of cod (*Gadus morhua*) of different year classes in NAFO Subdivision 3Ps in Northwest Atlantic. *Fish. Res.*, 42: 87-101.

Chen, Y. and J.E. Paloheimo. – 1998. Can a more realistic model error structure improve the parameter estimation in modelling the dynamics of fish populations? *Fish. Res.*, 38: 9-17.

Chen, Y., D.A. Jackson and J.E. Paloheimo. – 1994. Robust regression approach to analyzing fisheries data. *Can. J. Fish. Aquat. Sci.*, 51: 1420-1429.

Cooley, W.W. and P.R. Lohnes. – 1971. *Multivariate Data Analysis*. John Wiley and Sons, New York.

Devlin, S.J., R. Gnanadesikan and J.R. Kettenring. – 1981. Robust estimation of dispersion matrices and principal components. *J. Am. Stat. Ass.*, 76: 354-362.

Gomes, M.C., R.L. Haedrich and M.G. Villagarcia. – 1995. Spatial and temporal changes in the groundfish assemblages on the northeast Newfoundland/Labrador Shelf, Northwest Atlantic, 1978-1991. *Fish. Oceanogr.*, 4: 85-101.

Hanson, J.M. and G.A. Chouinard. – 1992. Evidence that size-selective mortality affects growth of Atlantic cod (*Gadus morhua* L.) in the southern Gulf of St. Lawrence. *J. Fish. Biol.*, 41: 31-41.

Hilborn, R. and C.J. Walters. – 1992. *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Chapman and Hall, New York.

Hill, R.C., T.H. Fomby and H.H. Johnson. – 1977. Component selection norms for principal component regression. *Commun. Stats. Theor. Methods*, 6: 309-334.

Hutchings, J.A. and R.A. Myers. – 1994. Timing of cod reproduction: interannual variability and the influence of temperature. *Mar. Ecol. Prog. Ser.*, 108: 21-31.

Jackson, D.A. – 1993. Stopping rules in principal component analysis: a comparison of heuristical and statistical approaches. *Ecol.*, 74: 2204-2214.

Jackson, D.A. and Y. Chen. – 2003. Robust principal component analysis of ecological data. *Environmetrics*, 14: 1-11.

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

Kovtsova, M.V. – 1995. Changes in growth and maturation of the Barents Sea plaice (*Pleuronectes platessa* L.) in 70s-90s. *ICES Council Meeting Papers* 12. ICES, Copenhagen (Denmark).

Krohn, M., S. Reidy and S. Kerr. – 1997 Bioenergetic analysis of the effects of temperature and prey availability on growth and condition of northern cod (*Gadus morhua*). *Can. J. Fish. Aquat. Sci.*, 54 (Suppl. 1): 113-121.

Krzanowski, W.J. – 1988. *Principal of Multivariate Analysis*. Clarendon Press, Oxford, UK

Lilly, G.R. – 1996. Growth and condition of cod in Subdivision 3Ps as determined from trawl surveys (1972-1996) and sentinel surveys (1995). *DFO Atlantic Fish. Res. Doc.*, 96/69.

Manly, B.F.J. – 1991. *Randomization and Monte Carlo Methods in Biology*. Chapman and Hall, New York.

Mason, R.L. and R.F. Gunst. – 1985. Selecting principal components in regression. *Stat. Prob. Lett.*, 3: 299-301.

Moreau, J. – 1987. Age and growth of fish. In: R.C. Summerfelt and G.E. Hall [eds.], *Fish Growth*, pp. 101-143. Iowa State University Press, Ames, Iowa.

Myers, R.A., G. Mertz and P.S. Fowlow. – 1997. Maximum population growth rates and recovery times for Atlantic cod, *Gadus morhua. Fish. Bull.*, 95: 762-772.

Nikolskii, G.V. – 1965. *Theory of Fish Population Dynamics*. Oliver and Boyd, Edinburgh, U.K.

Paloheimo, J.E. and L.M. Dickie. – 1965. Food and growth of fishes. I. A growth curve derived from experimental data. *J. Fish. Res. Bd. Can.*, 22: 521-542.

Rao, C.R. – 1964. The use and interpretation of principal component analysis in applied research. *Sankhya A*, 26: 329-358.

Ricker, W.E. – 1975. Computation and interpretation of biological statistics of fish populations. *Fish. Res. Board Can.*, 191.

Roff, D.A. – 1984. The evolution of life history parameters in teleosts. *Can. J. Fish. Aquat. Sci.*, 41: 989-1000.

Rollet, C., J-C. Brêthes and A. Fréchet. – 1995. Spatial and temporal variation in cod length frequencies and length at 50% maturity in Divisions 3Pn, 4RS and 3Ps. *DFO Atl. Fish. Res. Doc.*, 95.

Rousseeuw, P.J. – 1984. Least median of squares regression. *J. Am. Stat. Ass.*, 79: 871-880.

Rousseeuw, P.J. – 1985. Multivariate estimation with high breakdown point. In: W. Grossmann, G. Pflug, I. Vincze and W. Wertz (eds.), *Mathematical Statistics and Applications* (Vol. B), pp. 283-297. Reidel Publishing, Dordrecht, The Netherlands.

Rousseeuw, P.J. and A.M. Leroy. – 1987. *Robust Regression and Outlier Detection*. John Wiley and Sons Inc. New York.

Rousseeuw, P.J. and B.C. van Zomeren. – 1990. Unmasking multivariate outliers and leverage points. *J. Am. Stat. Ass.*, 85: 633-644.

SAS. – 1987. SAS *Guide for personal computers.* SAS Institute, Cary, North Carolina.

Shelton, P.A., D.E. Stansbury, E.F. Murphy, J. Brattey and G.R. Lilly. – 1996. An assessment of the cod stock in NAFO subdivision 3Ps. *DFO Atl. Fish. Res. Doc.*, 96/91.

Swain, D.P. – 1993. Age- and density-dependent bathymetric pattern of Atlantic cod (*Gadus morhua*) in the southern Gulf of St. Lawrence. *Can. J. Fish. Aquat. Sci.*, 50: 1255-1264.

Trippel, E.A. – 1995. Age at maturity as stress indicator in fisheries. BioSci., 45: 759-771.

Trippel, E.A. – 1998. Egg size and variability and seasonal offspring production of young Atlantic cod. *Trans. Am. Fish. Soc.*, 127: 339-359.

Vogt, N.B. and K. Kolsett. – 1987. Composition activity relationships _CARE. Part III. Polynomial principal component regression and response surface analysis of mutagenicity in air samples, 1981. Report 87: 747, Chemometrics and Indelligent Lab Systems.

Scient. ed.: F. Maynou.
Received June 29, 2007. Accepted November 16, 2007.
Published online April 9, 2008.