SCRS/1984/076 Col.Vol.Sci.Pap. ICCAT, 23 (2): 254-263 (1985)



# METHODOLOGY AND ASSUMPTIONS OF MULTIPLE COHORT ANALYSIS TECHNIQUE USED BY THE WORKING GROUP ON JUVENILE TROPICAL TUNAS

P. Kleiber
NMFS, Southwest Fisheries Center

## SUMMARY

A complex method of carrying out a multiple cohort analysis was used by the Working Group on Juvenile Tropical Tunas (WGJTT) in assessing the state of skipjack, yellowfin and bigeye stocks in the tropical Atlantic Ocean. The purpose of the complexities was to constrain the solutions of individual cohort analyses for a series of cohorts such that the fishing mortality estimates would vary approximately in proportion to the fishing effort. In this paper, the technique is explained in detail. The assumptions and the logical coherence of the technique are explored, and modifications are suggested to remove some of the inconsistencies. Also, alternative analytical approaches are suggested for incorporating fishing effort data into multiple cohort analysis.

## RESUME

Une méthode complexe de réalisation d'une analyse de cohortes a été employée par le Groupe de travail sur les Thonidés tropicaux juvéniles (WGJTT) pour evaluer l'état des stocks de listao, albacore et thon obèse dans l'Atlantique tropical. Le but visé par la complexité du procédé était de forcer la solution des analyses de cohortes individuelles pour une série de cohortes de façon à ce que les estimations de la mortalité par pêche varient de façon approximativement proportionnelle à l'effort de pêche. Une explication détaillée de la technique est fournie dans le présent document. Les hypothèses et la cohérence logique de la technique sont examinées, et des modifications sont suggérées pour en retirer certaines contradictions. D'autres méthodes d'analyses sont également suggérées pour l'incorporation de données d'effort dans les analyses pluri-cohortes.

## RESUMEN

El Grupo de Trabajo sobre Túnidos Tropicales Juveniles utilizó un método complejo a fin de desarrollar un análisis de cohorte múltiple para evaluar la situación de las poblaciones de listado, rabil y patudo en el Atlántico tropical. El propósito de este método es limitar las soluciones de los análisis individuales de cohortes - para una serie de cohortes - de forma que las estimaciones por pesca varíen aproximadamente en proporción al esfuerzo de pesca. En este documento se explica detalladamente la técnica seguida, y se estudian los supuestos y la coherencia lógica, sugiriéndose algunas modificaciones para eliminar varias contradicciones. Asimismo, se proponen diversos enfoques alternativos para incorporar datos de esfuerzo de pesca en los análisis de cohortes múltiples.

## INTRODUCTION

A series of multiple cohort analyses were recently performed on catch data for the three principal species in the tropical Atlantic tuna fisheries: skipjack (Cayre' and Diouf, 1984), yellowfin (Fonteneau, 1984), and bigeye (Pereira, 1984). I call this work multiple cohort analysis because a series of cohorts (year classes) were analyzed rather than a single cohort. The analyses were a part of the recent effort of the Working Group on Juvenile Tropical Tunas (WGJTT). Results from these analyses were used in assessing the state of Atlantic stocks of the three species and in predicting the likely consequences of alternative management schemes (report of July, 1984 meeting of WGJTT).

Cohort analyses based only on catch data must be constrained in some way, otherwise there are more quantities to be estimated from the analysis than there are data used in the analysis. This is true whether a single cohort or a series of cohorts is being analyzed. Such constraint is typically realized by assuming a value for natural mortality and either a recruitment value for a young age class or a fishing mortality value for the oldest age class. However, with a series of cohorts, multiple cohort analysis can be constrained in a variety of other ways. In particular, when fishing effort data as well as catch data are available, we may wish to constrain a multiple cohort analysis in such a way that estimates of fishing mortality vary with time in the same way that effort varied. Scientists involved in the work of the WGJTT developed a procedure for imposing this sort of constraint. Since the results are central to conclusions of the WGJTT, it is appropriate that a full technical description of the method be available. That is the purpose of this report.

My information on this multiple cohort analysis technique comes from somewhat cursory descriptions in WGJTT papers (Cayre' and Diouf, 1984; Fonteneau, 1984; Pereira, 1984; Report of July, 1984 meeting of WGJTT) and from discussion with scientists who carried out the analyses. I will at times allude to definitions and assumptions that have not been explicitly mentioned in previous descriptions of the technique. These definitions and assumptions are all implicitly involved in this cohort analysis method, and making them explicit makes the method easier to understand.

For the purposes of describing the technique, we shall assume that the following data are already available: a time-age catch matrix, a time series of data on catch per unit effort, and a time series of data on standardized fishing effort. Data processing methods used to assemble these data for the WGJTT analyses have been adequately described in previous WGJTT and SCRS reports.

## TERMINOLOGY

Notation for matrices will always be in boldface type (eg. **F**), and notation for scalar values, including elements of matrices and vectors, will always be in normal type (eg. $F_{y,t}$ is the element in row y and column t of matrix **F**).

Analytical methods that deal with age-apecific population parameters are often confusing because parameter values can vary in three important dimensions: pure time, pure age, and both time and age. To help clarify our discussion we will adopt the following conventions.

Consider a time-age matrix, ie. a matrix of values of a population parameter in which each row represents a year and each column represents an age class (Figure 1). Cohorts advance in both time and age; therefore values pertaining to cohorts progress diagonally from upper left to lower right in the matrix. Each cohort is thus represented by a diagonal vector which I will call a "schedule". A vector of values along a row of the matrix represents variation with age, within a particular time period, and across cohorts. I will call each of these row vectors a "distribution". Column vectors in the matrix each represent variation with time, within a particular age class, and across cohorts. I will call each of these column vectors a "time series".

Furthermore, there are parameters that apply to either all age classes or to a subset of age classes in a particular time period (eg. effort, or total biomass). A vector of such values represents variation with time, within a range of age classes, and across cohorts, and such a vector will also be called a time series.

In the following discussion I will often refer to time or age periods as years, but the discussion is valid for other time intervals as well. The catch data in the WGJTT analyses were actually placed into quarterly age categories.

DETAILS OF THE WGJTT MULTIPLE COHORT ANALYSIS TECHNIQUE

Once a catch matrix, **C**, and a time series of effort, **f**, and of catch per unit effort, **CPUE**, are assembled, the method proceeds as follows:

1) Estimate linear trend in the effort time series. Define this linear trend as a vector, $\mathbf{f'}$, whose elements are given by

$$f'_y = a+by \qquad (1)$$

where y is the year and where a and b are linear function parameters. It was not clear whether this straight line was fit to the time series, **f**, by linear regression or by eye.

2) Define $^{*}\mathbf{F}$ for a particular year to be the average of the first three years of the fishing mortality schedule (diagonal vector) for the cohort recruited in that year. That is,

$$^{*}F_y = [F_{y,1} + F_{y+1,2} + F_{y+2,3}]/3 \qquad (2)$$

where the F's are elements in a fishing mortality matrix. (In the analyses performed for WGJTT, $^{*}\mathbf{F}$ was really an average of 12 quarterly F's). It is assumed that $^{*}F_y$ for any year class is proportional to $f'_y$ for that year. Thus

$$^{*}F_y = q^{*}f'_y \qquad (3)$$

where $q^*$ is a proportionality constant. This constant would be a type of catchability parameter representing the first three years in the fishery of any cohort (assumed to be constant over cohorts).

3) Assume a natural mortality schedule, **M**, which is constant over all cohorts. For skipjack the natural mortality schedule was assumed to be a constant for all ages in the fishery, and for yellowfin and bigeye the natural mortality was assumed to drop after a certain time in the fishery.

4) Define a recruitment function

$$R_y = R({}^*F_y, \mathbf{M}, \mathbf{C}^*_y) \qquad (4)$$

where $R_y$ is a recruitment value for year class y which is put into an individual cohort analysis with a natural mortality schedule, **M**, and a catch schedule, $\mathbf{C}^*_y$, for year class y. The resulting fishing mortality schedule, averaged over the first three years gives the value ${}^*F_y$. In other words, for a given **M** and $\mathbf{C}^*_y$, any chosen value of ${}^*F_y$ will uniquely determine a value of $R_y$ by virtue of simple cohort analysis. In practice, values of the function, R, were obtained by constructing a table using simple cohort analysis to calculate ${}^*F_y$ values from a range of $R_y$ values for each of a series of year classes (a different $\mathbf{C}^*$ for each year class). R values were then picked from this table for any given year class, y, and ${}^*F_y$ value. (A computer subprogram that would directly implement this function would be easy to construct).

5) Given the definitions and assumptions in step 2 above, a time series of $\mathbf{{}^*F}$ values, $\mathbf{{}^*F'}$, can be determined from $\mathbf{f'}$ if a value for the proportionality factor, $q^*$, is chosen. In practice, values of $q^*$ were not explicitly used. Instead, a value of ${}^*F'_{last}$ was chosen for the final year of the analysis. The value was chosen to be reasonably close to the value of $\mathbf{{}^*F}$ thought to prevail in later years. $\mathbf{{}^*F'}$ values for earlier years were then calculated in proportion to $\mathbf{f'}$ (implicit use of $q^*$). Thus from equations 1 and 3 the elements of $\mathbf{{}^*F'}$ are given by

$${}^*F'_y = q^*(a+by) \qquad (5)$$

where

$$q^* = {}^*F'_{last}/f'_{last}.$$

6) A time series of recruitments, **Rt**, was obtained from the time series, $\mathbf{{}^*F'}$, using the function defined in step 4. Thus from equations 4 and 5, elements in **Rt** are given by

$$Rt_y = R(q^*(a+by), \mathbf{M}, \mathbf{C}^*_y) \qquad (6)$$

for each year, y.

7) By-products of the individual cohort analyses associated with the recruitment series in step 6 are a fishing mortality schedule for each cohort and a schedule of the average number of fish at large in each cohort. Such schedules for all the cohorts can be combined into a fishing mortality-at-age matrix, F, and a population-at-age matrix, **N**. These

matrices and the recruitment time series, **Rt**, could be taken as the solution of the multiple cohort analysis. As a check on the results, a total population time series, **Nt**, is calculated from the row sums of **N**, and **Nt** is then compared to the catch per effort time series, **CPUE**, under the assumption that **CPUE** is an index of population size. It was not mentioned, but presumably an arbitrary scaling of either the **CPUE** or the **Nt** values is allowed in the comparisons between these two series.

In order to improve the comparison between **Nt** and **CPUE**, the individual cohort analyses are redone a number of times using scaled versions of **Rt**, and **Nt** is recalculated and compared to **CPUE** each time until the best fit is found. In other words the fudge factor, x, in

$$\mathbf{Nt} = \mathbf{Nt}(x\mathbf{Rt},\mathbf{M},\mathbf{C}) \qquad (7)$$

is adjusted to get the optimum fit between **Nt** and **CPUE**. The criterion for goodness of fit (least squares, judgement by eye, or other) was not mentioned in previous discussions of this method.

8) The best fitting time series of scaled recruitment values, **Rt** from step 7 and associated by-products of individual cohort analysis, **F** and **N** are taken as the final solution to the multiple cohort analysis.

## CRITIQUE OF THE MULTIPLE COHORT TECHNIQUE

The fundamental reason for assuming a linear trend in effort and thereby in fishing mortality is that in this procedure the cohorts are dealt with individually. That is, individual fishing mortality schedules (diagonals in the fishing mortality matrix) are estimated one at a time, and the results are reconciled afterwards. As a result (see Figure 2), the constraint of each effort datum does not affect the corresponding fishing mortality distribution (row in the fishing mortality matrix) as it ideally should. Instead, each effort datum affects the first three elements of a fishing mortality schedule (diagonal in the matrix), and only one element in this schedule represents the same time period represented by the effort datum. In other words, the fishing effort in any given year, y, affects $^*F_y$, the $^*F$ value for that year, but $^*F$ is an average of fishing mortalities in year y and in two subsequent years, y+1 and y+2, during which effort levels are likely to be different than in year y. If effort varied consistently from year to year, it would be possible to express the effort levels in years y+1 and y+2 as functions of effort in year y. From equation 2

$$^*F = [q_1 f_y + q_2 f_{y+1} + q_3 f_{y+2}]/3$$

$$= [q_1 f_y + q_2 h_1(f_y) + q_3 h_1(f_y)]/3 \qquad (8)$$

where $q_t$ is the catchability in age class, t, and where $h_1$ and $h_2$ are functions that transform effort in year y to effort in the two subsequent years. That is,

$$h_1(f_y) = f_{y+1}$$

$$h_2(f_y) = f_{y+2}$$

Equation 8 asserts that effort in year y is an appropriate constraint on ${}^*F$ if and only if functions $h_1$ and $h_2$ exist. These functions can exist only if a consistent pattern of change in effort occurred. A linear trend is one such consistent pattern of change and is the pattern assumed in this method.

A possible problem with the assumption of a linear trend in effort is that such a trend may not actually have happened. For the method to be useful, the effort data should presumably show at least an approximate linear trend with time. How the effort did in fact vary in time is of course determined by the activities of fishermen. In the case of the eastern tropical Atlantic surface tuna fisheries during the years included in the analysis there does not appear to be a marked curvilinear trend (Figure 3), but there is considerable scatter about the linear trend. Deviations of true effort values from the assumed linear trend imply errors in the corresponding estimates of recruitment, which means that the time series of recruitment estimates from this method may not be an accurate picture of the true variation in recruitment.

A second problem with assuming a linear trend in effort is that this assumption is not consistent with the assumption that elements of ${}^*F$ vary in proportion to effort as in equation 3. We can demonstrate this inconsistency by postulating that fishing mortality in year y and age class t is proportional to the effort in that year, ie. that

$$F_{y,t} = q_t f_y \qquad (9)$$

If effort varies linearly with year, the ratio of effort in year y to effort n years later would be

$$f'_{y+n}/f' = 1 + n\Delta f/f'_y \qquad (10)$$

where $\Delta f$ is the yearly change in effort. However, by definition of ${}^*F$ (equation 2)

$$\begin{aligned} {}^*F_y &= [q_1 f'_y + q_2 f'_y + q_3 f'_{y+2}]/3 \\ &= [q_1 f'_y + q_2(f'_y+\Delta f) + q_3(f'_y+2\Delta f)]/3 \qquad (11) \end{aligned}$$

Therefore

$${}^*F_{y+n}/{}^*F_y = 1 + n\Delta f/(f'_y + K) \qquad (12)$$

where

$$K = (q_2+2q_3)\Delta f/(q_1+q_2+q_3) > 0. \qquad (13)$$

By inspecting equations 10 and 12 we can see that the ratio of ${}^*F$ values will always be smaller than the corresponding ratio of $f'$ values. In other words, if the elements of $f'$, increase by some multiplicative factor in a number of years, the elements of ${}^*F$ should increase by a lesser factor, but equation 3 asserts that the elements of ${}^*F$ and $f'$ should change by the same factor. Thus we cannot have a linear trend in effort and also have ${}^*F$ values varying in proportion to that effort without violating either the definition of ${}^*F$ (equation 2) or the postulate in equation 9. This incon tency could be resolved by assuming that the effort changes geometrically in time instead of linearly, but again, such a pattern of change in effort may not actually have happened. Judging from Figure 3, there is little evidence that effort in the fishery increased geometrically.

The use of the fudge factor, x, in equation 6 introduces another inconsistency in the method. This factor is adjusted to maximize the comparability between the **Nt** and **CPUE** time series. Step 5 postulates that $^*F$ is proportional to linearized effort, $f'$, but that postulate is violated when the resulting recruitment time series is rescaled by adjusting x in equation 7. This is because the $^*F$ values corresponding to the rescaled recruitments no longer vary in proportion to effort. Figure 4 shows an example where the original $^*F$ is proportional to $f'$ but the $^*F$ vectors calculated from the rescaled recruitments are not. In that particular example, the WGJTT skipjack analysis, the $^*F$ values from the rescaled recruitments happen to increase more slowly than the $f'$ values. This is what we would expect given the proportionality problem (conflict between equations 1 and 3 discussed in the previous paragraph). In this case the proportionality problem and the fudge factor problem cancel each other to some degree, but this will not necessarily always be the case. I have not worked out if the two problems balance each other quantitatively, but there is no guarantee that such would occur in all cases.

The fudge factor problem, which destroys the originally postulated relationship between $^*F$ and $f'$, could be solved by rescaling the $^*F'$ series in equation 6 instead of the **Rt** series in equation 7. That is, $q^*$ could be adjusted in equation 6 in order to maximize the comparability of **Nt** and **CPUE**.

## ALTERNATE TECHNIQUES

It is possible to devise constrained cohort analysis techniques in which the cohorts are analyzed simultaneously rather than individually. This approach allows a given constraint to act across cohorts (down columns or across rows in the time-age matrix) rather than within cohorts (diagonals in the time-age matrix), and the problems alluded to above can be avoided. I will not give a detailed account of such a technique, but will refer to two examples in the literature. Similar techniques could certainly be devised to apply to the Atlantic tropical tuna data.

Doubleday (1976) demonstrates a cohort analysis method in which fishing mortality distributions (rows) are constrained to have a similar shape, though different magnitudes, for all years in the analysis. However, effort data are not used in this method to influence the magnitudes of fishing mortalities. Therefore it is not necessary to make assumptions about how effort varies in time and how effort is related to fishing mortality.

Fournier and Archibald (1982) present an example in which effort data can be used to constrain the fishing mortality estimates for all age classes in a given time period. With this method it is necessary to assume a relationship between effort and fishing mortality, but it is the relationship between effort in a given year and fishing mortality in that same year for all age classes. Assumptions about how effort varies in time are not required.

## CONCLUSIONS

I have described the multiple cohort technique used by WGJTT in some detail in hopes that it will be easily understood by all scientists reviewing the WGJTT results. I have commented rather critically on the mathematical logic of the method, pointing out problems with assuming a linear trend in effort with time and the problem with rescaling the recruitment vector. However, I have not questioned the results so far generated by the method and the conclusions therefrom reported by the WGJTT. The problems I have discussed may not significantly affect the multiple cohort analyses performed for WGJTT and the stock assessments based on those results. Nevertheless, it is important to determine how robust the stock assessments are to those problems. Before directly embarking on such a project, though it would be useful to consider other methods of multiple cohort analysis. The most efficient way to check the validity of conclusions based on the WGJTT multiple cohort analyses may be to reanalyse the data using a different method that does not contain the inconsistencies present in the current WGJTT technique. Such methods are available, and I provide references to two of them.

## REFERENCES


CAYRE, P. and T. DIOUF. 1984. Etat des stocks de listao (katsuwonus pelamis) de l'Atlantique est par analyse des cohortes et modele de production par recrue. Collective Volume of Scientific papers, International Commission for the Conservation of Atlantic Tuna, Madrid, XXI(2):38-63.

DOUBLEDAY, W.G. 1976. A least squares approach to analyzing catch at age data. Res. Bull. Int. Comm. NW Atlant. Fish. 12:68-81.

FONTENEAU, A. 1984. Analyse de l'etat des stocks d'albacore au 30 mai. Collective Volume of Scientific papers, International Commission for the Conservation of Atlantic Tuna, Madrid, XXI(2):80-101.

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

PEREIRA, J. 1984. Etat des stocks du patudo de l'Atlantique par analyse des cohortes, analyse de production par recrue et par le modele global. Collective Volume of Scientific papers, International Commission for the Conservation of Atlantic Tuna, Madrid, XXI(2):155-183.

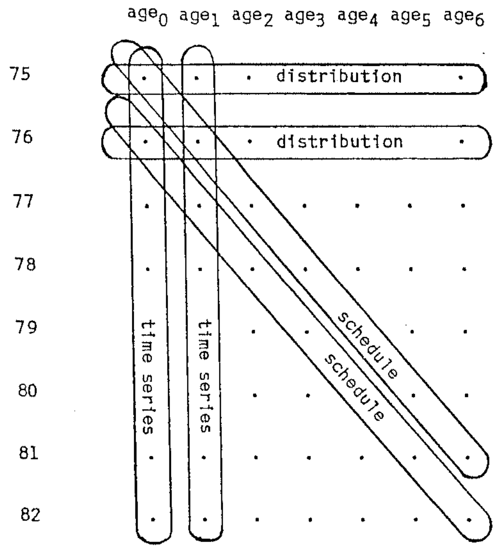


Figure 1. Time-age matrix of any population parameter indicating two examples each of a distribution, a schedule, and a time series.

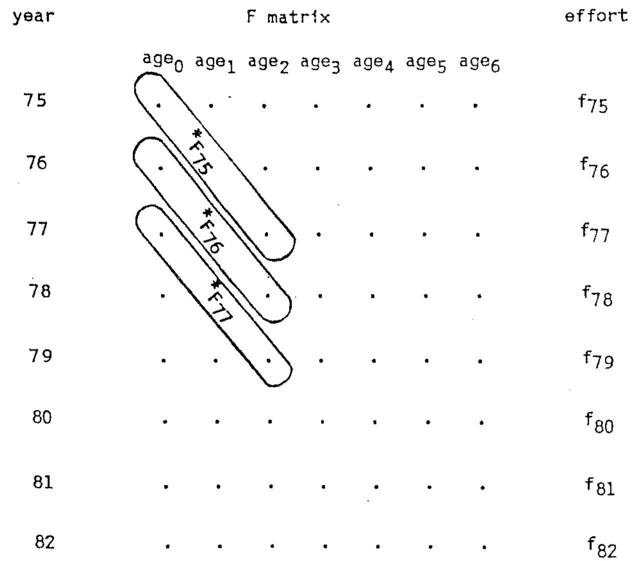


Figure 2. Connection of $^*F$ time series to effort time series. Each element of $^*F$ is an average of three diagonal elements in the matrix. Only the first of these three matches the actual year in which the corresponding effort occurred.

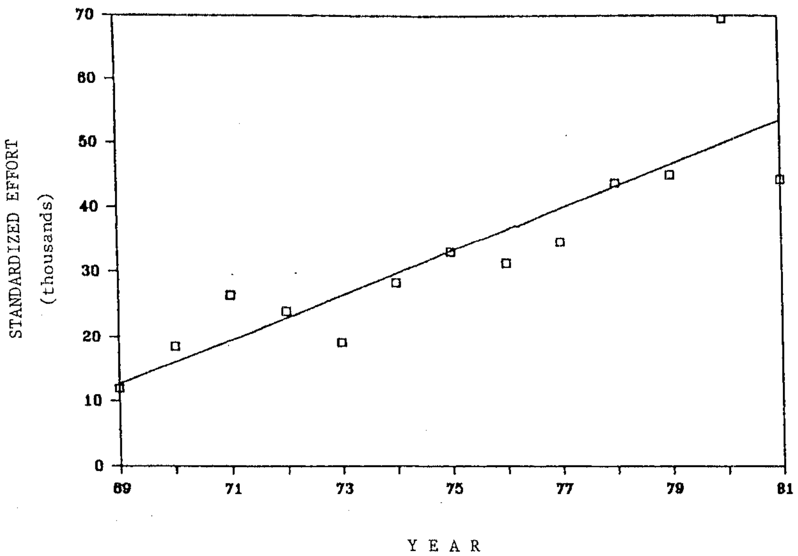


Figure 3. Standardized effort values, f, from skipjack data (Cayre' and Diouf, 1984) with linear regression line.

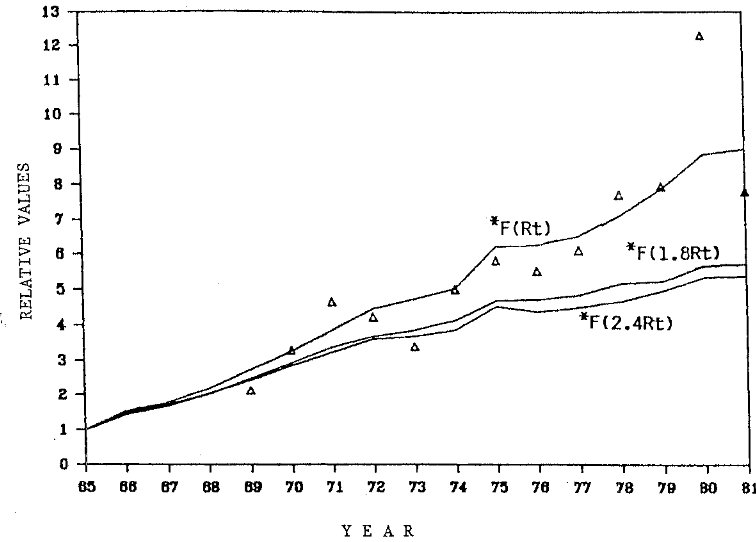


Figure 4. Time series of *F (curves) and standardized effort (triangles) from skipjack analysis (Cayre' and Diouf, 1984). The three *F time series correspond to three recruitment time series, the original series, **Rt**, and two rescaled series, **1.8Rt** and **2.4Rt**. Values are standardized to unity in 1965. The *F lines are not straight as they theoretically should be because the values were picked without interpolation from the recruitment and *F table given by Cayre' and Diouf. The points are standardized effort values for 1969 through 1981. These values are scaled so that their mean value equals the mean of the ***F(Rt)** time series over the same time period. ***F(Rt)** corresponds well with the effort trend. However, **1.8Rt** was the final choice for a recruitment time series in the Cayre and Diouf skipjack analysis, and ***F(1.8Rt)** clearly does not follow the trend in effort.