THE VALUE OF CATCH STATISTICS AND RECORDS OF GUANO HARVESTS FOR MANAGING CERTAIN SOUTH AFRICAN FISHERIES.

by

Michael Olaf Bergh

Submitted for the degree of Ph. D.
Faculty of Science
University of Cape Town

March 1986

TABLE OF CONTENTS

PAGE

ABSTRACT

Research into species inhabiting the southern part of the Benguela current region has been biassed towards those of commercial importance and those which are either more visible or accessible than others. A crude evaluation of material flows and standing stocks, including the lesser studied species, was made. Estimates suggest that the productivity of juvenile hake (3 million metric tonnes per year or more), squid (> 0.5 million metric tonnes per year), and midwater shoaling fish (together > 0.5 million metric tonnes per year) is substantial compared to the total annual landing of marine fish off South Africa which is in the region of 0.5 million metric tonnes. The consumption of anchovy and other epi-pelagic shoaling fish is dominated by piscivorous fish, particularly snoek. Snoek may consume as much as 0.4 million metric tonnes of anchovy per year, out of a total anchovy production of in the region of 2 million metric tonnes wet mass per year. Anchovy are the main fish consumers of plankton, with an estimated consumption of more than 20 million metric tonnes wet mass per year. Total annual plankton production is in excess of 800 million metric tonnes wet mass. About 32% of this enters higher trophic levels in the epi-pelagic region, mainly through zooplankton grazing. It is suspected that the bulk of production in the zooplankton and hake communities is recycled by omnivorous feeding in the case of zooplankton and cannibalism in the case of hake. The amount of cannibalised hake is about 10 times the amount taken by trawlers. The biomass, production and consumption rates of apex predators such as seals, whales, dolphins and seabirds (sizes of the order of metres to 10's of metres) is quantitatively unimportant in comparison with species consisting of smaller individuals (sizes of the order of 10's of centimetres). Due to the complexity of the food web, a conceptual model based on size may be preferable to one involving trophic levels. Overall, however, the complexity of the system, and the sketchiness of knowledge about it mitigates against the use of multispecies models for making useful predictions.

A maximum likelihood estimation approach to Virtual Population Analysis (VPA) is adopted, developed and implemented. It is shown from results obtained using artificial fisheries data that the database conventionally considered adequate for VPA work, viz. the catch mass and age structure, does not lead to reliable estimates of stock abundances for recent years. Use of further data related to stock adundance, such as effort data in situations where catch per unit effort is proportional to abundance, is investigated. Results using simulated effort data show dramatic improvements in abundance estimator precision. It is therefore argued that VPA should not be attempted in the absence of data complementary to the catch masses and age structures. With this in mind, VPA is applied to three stocks of commercial importance which occur off the South African coastline: anchovy, pilchard and hake.

Results are presented which show that anchovy biomasses are determined with reasonable precision if catch rates can be assumed to be proportional to stock abundance, and if the selectivity functional form is known. However it is argued that it is unlikely that catch rates are proportional to anchovy abundances. Furthermore, recent surveys of the anchovy stock size and habitat range suggest that the relationship between age and selectivity is different to that used in the analysis, and is

essentially unknown. Accordingly further analyses are performed, excluding the effort data, and replacing the initial one parameter selectivity function with unique selectivity parameters for each age class seen in the catch. Very imprecise biomass estimates are obtained when the likelihood function is formulated in this way, due to the now enhanced instability of the position of the maximum of the VPA likelihood function. Separate simulations illustrate that the bias in abundance estimates introduced by using the same age length key in all years for converting catch length distributions to catch age distributions may be as much as 50%. This, the aforementioned imprecision of VPA biomass estimates, unassessed but probable errors in the single age length key which is used, and finally very recent ageing re-evaluation studies suggesting that the catch consists almost entirely of 0-year-olds, effectively rule out the possibility of obtaining management-orientated information from the anchovy VPA.

Complementary data for the pilchard VPA is obtained by analysis of records of guano harvests which are made at a number of islands off the South African coastline. Evidence is presented suggesting that the harvest at Bird Island, Lambert's Bay is approximately linearly related to pilchard recruitment and these guano harvests are thus included in the likelihood function for the pilchard VPA. The resultant biomass estimates are plausible in terms of the likely system carrying capacity for pelagic fish. The biomass trend does not reflect the major peak in biomass in 1960 which earlier work consistently suggested. An appropriate sum of squares minimisation procedure is adapted from the Ludwig-Walters' formulation for fitting the pilchard stock recruit function to the recruitment and spawning biomass data from the VPA. The surplus production curve is estimated by combining the estimated recruitment function parameters with a yield per recruit analysis. The results indicate a mean virgin spawning biomass of 2.4 million metric tonnes, and an MSY spawning biomass level of 1.0 million metric tonnes. The maximum sustainable annual catch, for a 20 year scenario starting with the population at half the unexploited biomass, and ensuring a 10% maximum risk of obtaining a stock size of less than 0.2 of the pristine size, is 187 thousand metric tonnes.

The surplus production curve for the hake stock off the west coast of South Africa (ICSEAF Div. 1.6) is estimated using the same methods as applied to pilchard. No substantial difference between the VPA results and those produced using dynamic catch-effort models could be demonstrated - both VPA and the Schaefer model predict a maximum sustainable yield of about 140 thousand metric tonnes. Closer analysis shows however that the additional age structure information taken into account by the VPA has not led to better estimates than are provided by the catch mass and effort data alone. The suspected past practice of discarding small hake at sea potentially invalidates all sustainable yield calculations, even those based on the dynamic catch-effort models, by introducing spurious effects into the catch-based statistical records. The future in hake stock assessment may therefore lie with (as is the case for the pelagic species) direct survey techniques.

1. INTRODUCTION

1.1 OVERALL PERSPECTIVES

Past management of the fish stocks inhabiting the pelagic region of the southern Benguela ecosystem is notable for significant differences between scientific recommendations and quotas awarded (Alant et al, 1983). The conclusion that the final quotas have simply been a political compromise between various interest groups therefore seems unavoidable. Underlying this, there are no convincing purely ecological rationales for management, except possibly that species extinction may be undesirable.

Gordon (1954) attempted to explain why fish stocks were persistently reduced far below "reasonable" levels, and why the fishing industry was economically inefficient. The simple explanation offered by Gordon (op. cit.) hinges on the observation that fish stocks are not privately owned; instead they are regarded as the common property of those who choose to engage in fishing. The economic conseqence of this is that the fishery reaches bio-economic equilibrium at the average social rate of profit. Hardin (1968) has described this dissipation of potential economic rent as the "tragedy of the commons". Clark (1973) presents a simple bio-economic model in this regard, which predicts that, for certain values of the economic discount and per capita stock production rates, monopolistic overfishing appears to be a meaningful management option. Thus, for either open or closed access fisheries (i.e. competitive or monopolistic), severe stock depletion may result from the pursuit of economic goals.

The "tragedy", in Hardin's terms, for the open access fishery, is twofold. First, the bio-economic equilibrium state often occurs at severely depleted population sizes, which may be "ecologically undesirable". Secondly, there is good reason to expect that by reducing fishing effort below the bio-economic equilibrium effort level, larger catches and profits may be made (Clark, 1985). This, and the extensive worldwide empirical evidence of economically inefficient overfishing (Clark, op. cit.), suggests that appropriate statutory control may be of considerable benefit to both the industry and the state.

Thus fishery management is frequently undertaken at a national level, as is the case in South Africa. The local pelagic industry developed after the end of World War II, but quotas set between 1950 and 1974 had little scientific basis. Limitation by quota was in fact dropped between 1960 and 1971, although effort restrictions, various closed seasons and minimum mesh size regulations (Butterworth, 1983) were still enforced during that time.

In the 1970's quantitative scientific assessments were made which suggested global sustainable pelagic yields of 360 (Newman et al, 1974) and 375 (Crawford, 1979) thousand metric tonnes (see also Butterworth, 1983) and these analyses played a partial role in quota decisions in the late 1970's. Since that time, scientists have argued for caution in exploiting stocks (scientific quota recommendations are summarised by Alant et al, 1983).

However, recent research has shown that policies resulting from such an attitude can limit understanding of stock dynamics (Walters, 1984), leading perhaps to underexploitation of a resource. Thus it is not necessarily the case that a "conservationist" viewpoint need always require a cautious approach (the term conservation is used here in a broad sense including the possibility of rational exploitation of the resource). In this context, a risk averse tendency is essentially a prejudice, which is not necessarily backed by sound scientific arguments.

This risk averse tendency on the side of scientists has resulted in an unhealthy and polarised relationship with the fishing industry. At the root of this conflict is the lack of any reliable insight concerning and quantification of the population dynamics of the stocks of commercial interest (anchovy and pilchard population size trends are poorly determined, see Chapters 5 and 7; overall predictive capabilities are relatively underdeveloped). This is due partly to the complexity of the ecosystem, which has fuelled a debate, seldom properly articulated within local scientific circles, about the appropriate scientific response to this complexity.

Two viewpoints can be identified in this regard. The first (implicitly if not explicitly) supports the development of complex models which take into account the numerous system variables and interactions , both biotic and abiotic, at either the whole system or subsystem level, and this appears to have been a justification for the existing multidimensional framework of the Benguela Ecology Programme (Siegfried and Field, 1982). A brief examination of important species groups and material flows in the southern Benguela ecosystem is conducted in Chapter 2, the aim being

(a) to overcome previous research biases with respect either to commercially important or visible and accessible species by analysing existing data on diet, biomass, consumption and production, and

(b) to illustrate the complexity of the biotic system and suggest appropriate ways of modelling it.

The study method involved a large number of interviews with a broad cross-section of biologists, and combining this information with additional published and unpublished material. Results show that there are a large number of groups of species which should be considered as interacting in the southern Benguela ecosystem. If one considers the minimum number of parameters required to define the set of differential equations describing the dynamics of this ecosystem, one is still faced with estimating of the order of 30 to 40 parameter values. The sampling requirements for obtaining initial estimates of all these parameters to a reasonable degree of precision are anticipated to be excessive, and it is thus pointed out (Chapter 2) that the success of multispecies models is thwarted by severe statistical limitations.

The alternative view of ecosystem modelling is that because of measurement constraints, the only models that are feasible and usable (i.e. with some predictive capability, which is essential for management purposes), are very simple models embodying a few fundamental assumptions. The definitive example of such an approach is the single species population model.

1.2 THE SINGLE SPECIES MODEL

The term "single species model" describes a set of models which treat populations in isolation from other stocks which may impinge on their dynamics. Links with the ecosystem are made via a few very basic assumptions.

As a first attempt at defining the population dynamics of a resource, it is assumed that the population can be adequately described by a single variable indicative of its size, usually chosen to be its biomass, $B(t)$. The dynamics is then described by the derivative, $dB(t)/dt$, which is taken to be a function of $B(t)$ only. It is argued that unexploited populations fluctuate around some average equilibrium population size, K, since there must be some environmental limitation to population size. These fluctuations are due to biological interactions and environmental variations which are assumed to have a stochastic nature - this will be discussed in due course. For the moment, considering the deterministic limit, there is an equilibrium at $B(t)=K$, where $dB(t)/dt = 0$. Furthermore, it must be true that;

$$dB(t)/dt = 0 \text{ for } B(t) = 0. \qquad \ldots 1.1$$

Therefore, for $B(t)$ between 0 and K, $dB(t)/dt$ must be positive over at least part of that range; for $B(t)$ greater than K, $dB(t)/dt$ will be negative, so that $B(t)=K$ is a stable equilibrium situation in the absence of exploitation.

The simplest nonlinear function which captures these essential features is a quadratic, which leads to what is commonly known as the (Verhulst) logistic growth equation for describing resource dynamics in the absence of exploitation;

$$dB(t)/dt = rB(t)(1-B(t)/K) \qquad \ldots 1.2$$

where the parameter r is the specific (i.e. per capita) stock growth rate at stock sizes approaching zero. Other more complex forms have been suggested for the surplus production function (equation 1.2) by for example Pella and Tomlinson (1969) and Shepherd (1982). These other forms can usually still be expressed in terms of parameters r and K, where

$$r = \lim_{B \to 0} [dB/dt\ (1/B)].$$

Equation 1.2 can be modified to include the possibility of fishing;

$$dB(t)/dt = rB(t)(1-B(t)/K) - C(t) \qquad \ldots\ 1.3$$

where C(t) is the instantaneous catch rate. Depending on the time dependent behaviour of C(t), equation 1.3 may be solved for B(t) in terms of t only, and a number of parameters. The parameters would include r, K and an arbitrary boundary value for B(t) at t=T, B(T). The seemingly modest task of estimating r, K and B(T) (or at least combinations of these reflected in sustainable yields or TAC's) is a crucial concern of most fishery management models.

The bulk of this thesis is concerned with the estimation of the last quantity B(T), the stock size (Chapters 3, 4, 5, 7 and 8), and examines the usefulness of the existing fisheries data base for doing this. The estimation of the surplus production function is treated in somewhat less detail (Chapters 7 and 8; using yield per recruit techniques with a fitted stock/recruit relation rather than an empirically fitted functional form such as that of equation 1.2). The problem of estimating fish population sizes is known in fisheries parlance as stock assessment (to a lesser extent this also implies estimation of sustainable yields).

Techniques of stock assessment can be broadly categorised as either direct or indirect. Direct stock assessment techniques involve some attempt at estimating population sizes from direct observations of fish density and habitat range, while indirect techniques use alternative data (which in some sense index the population biomass), typically pertaining to the landed catches (such as the catch mass, catch effort and catch age structure) to infer population sizes. Although all direct estimation techniques are partly indirect, it is usual to refer to hydroacoustic and egg surveys as direct assessment methods and to mathematical techniques such as VPA (Virtual Population Analysis) and dynamic production models based on CPUE (Catch Per Unit Effort) data, as indirect methods. At the time when the research reported here was initiated (1982), the SFRI (Sea Fisheries Research Institute - the government research body) was heavily committed to using a particular indirect method of stock assessment, known as Virtual Population Analysis (VPA) and sometimes termed Cohort Analysis (subsequently effort has shifted towards using direct stock assessment techniques). Virtual Population Analysis techniques of stock assessment are examined in detail in this thesis, and subsequently applied to three stocks of commercial importance.

1.3 VIRTUAL POPULATION ANALYSIS

The research into VPA which is reported here was initiated at a critical time for local anchovy stock assessment procedures. (Anchovy is the dominant pelagic shoaling fish species in the ecosystem as well as in the pelagic catch.) Prior to this time, VPA offered the only technique of stock assessment; CPUE had essentially been disregarded because of the difficulties encountered in defining effort in a pelagic fishery and allocating it between different fish species, and also because the catch rate for shoaling fish may be poorly related to abundance.

The VPA method was mathematically appealing, and it produced a single biomass estimate for each year since the inception of fishing, which gave the (misleading) impression of much greater reliability compared to direct survey estimates not yet attempted at that time, but for which it was known that variance would be large. Nevertheless, there was some debate about the appropriateness of some of the basic assumptions and application procedures for VPA. For example, the assumption of constant natural mortality appeared to be questionable. Likewise the absence in some analyses of an objective criterion for setting terminal fishing mortality values was cause for considerable concern. The need for the research reported here thus sprang from the requirement to place VPA abundance estimates in a more meaningful perspective.

At the same time, again because of the uncertainty of VPA abundance estimates, the SFRI started work on a programme of combined hydroacoustic - midwater trawl surveys for anchovy. Apart from providing preliminary anchovy abundance estimates, this survey work has very recently provided some revolutionary insights into the anchovy biomass and habitat range. This, coupled with recent otolith daily growth ring studies, has suggested that past anchovy catches consisted almost entirely (more than 90% by number) of recruits (0-year-olds) rather than about 60% as had been previously thought. Thus the fishery has effectively sampled recruits only, and VPA must therefore be disregarded as being of any use for anchovy stock assessment.

To a large extent these findings render the estimates of VPA abundance precision reported here outdated. Nevertheless, the techniques used in assessing the VPA abundance estimate precision are instructive both in respect of other stocks for which indirect techniques such as VPA are still used (such as hake), and also for comparison with other assessment techniques. It would thus be prudent to subject the direct survey results (hydroacoustic and egg-production methods of biomass assessment are now being used for anchovy) to similarly rigorous statistical analyses. Furthermore, it is important to dispel a number of myths concerning historic anchovy and pilchard biomass trends, which have entered the scientific literature.

In fishery mathematics, the term cohort refers to all the fish in a population which have the same birthdate, normally to within one year. The basic model of cohort dynamics is a simple account of the decay rate of cohort numbers, N(t) (eg. Gulland, 1965), which assumes that depletion rates are proportional to number present;

$$dN(t)/dt = -FN(t) - MN(t) \qquad \ldots\ 1.4$$

where FN(t) is the death rate due to fishing, and MN(t) that due to natural predation. F is known as the fishing mortality and M as the natural mortality. Both F and M have units of t^{-1}. Equation 1.3 can be solved for N(t);

$$N(t) = N(0)\ \exp(-(F+M)t) \qquad \ldots\ 1.5$$

and a complete description of cohort decay is possible if the parameters N(0), F and M are given. Knowledge of the status and dynamics of all the cohorts in a population can be used to define the population size trends over the period that data exist. Cohort analysis techniques will not be dealt with in depth at this stage (see Chapter 3 instead). It is to be noted however that a fixed value is normally assumed for the parameter M, and F can be obtained from catch information. N(t) is given in terms of N(T), where T is some other point in time, by the relationship;

$$N(t) = N(T)\ \exp(-(F+M)(t-T)) \qquad ..\ 1.6$$

so that the problem of estimating N(0) is equivalent to estimating N(t) for any convenient t. A fishery which has exploited a stock with a maximum age of m years, for n years, has exploited a total of n+m-1 cohorts (the number of diagonals in an nxm matrix). The estimation of these n+m-1 cohort sizes (for any convenient stage t in each of the cohorts lifespan) is of crucial concern to Virtual Population Analysis techniques. This problem may also be viewed as the problem of estimating the set of n+m-1 terminal fishing mortalities (Chapter 3).

The development of the original VPA technique (Gulland, 1965) for application to populations consisting of a number of cohorts has been attempted by Doubleday (1976), Gray (1977), Pope (1977), Fournier and Archibald (1982) and Pope and Shepherd (1982), to name a few. Their papers are mainly concerned with estimating the n+m-1 crucial parameters (cohort sizes at some age, or terminal fishing mortalities) by use of objective statistical procedures (for which the crucial model assumption is that the fishing mortality trends with age in any year is the same as those in other years).

Two of these techniques (Pope and Shepherd, 1982; Fournier and Archibald, 1982) are reviewed in Chapter 3. It is argued that the method of Fournier and Archibald (1982) (insofar as the assumptions on which it is based are appropriate to the data set under consideration) provides a statistical treatment which is superior to Pope and Shepherd's (1982) procedure. This is then investigated further using simulated fisheries data sets. Techniques are set up to evaluate abundance estimate variance for typical anchovy populations, the aim being to determine the adequacy of the classical VPA data, the catch-at-age estimates. In further developments a resampling procedure is suggested for evaluating abundance estimate variance in practical situations where only one data set exists (compared to simulations where numerous data sets are generated for obtaining variance estimates). Applications of VPA, using the Fournier and Archibald (1982) approach, to three South African fish stocks, the anchovy (Chapter 5), pilchard (Chapter 7) and hake (Chapter 8) are reported.

The application of VPA techniques to the South African pilchard and anchovy stocks was first attempted by Centurier-Harris (1977). Results for both anchovy and pilchard are quoted in many of the papers recorded in ICSEAF Reports, and most recently by Butterworth (1983), Crawford et al (1983) and Shelton and Armstrong (1983). VPA results for hake are similarly presented in ICSEAF documents, most recently by Butterworth and Andrew (1984).

In the context of this thesis, the results quoted in most of these papers are seen as poorly founded. This is because either the n+m-1 critical VPA parameters (the terminal fishing mortalities) are not determined in an objective fashion, or because the objective technique used is inadequate (for reasons discussed in Chapters 3). Furthermore, in most cases (excepting Butterworth, 1983; Butterworth and Andrew, 1984) the input data consist only of the estimates of numbers of fish caught at age. Results presented in Chapter 4 and by Pope and Shepherd (1982) suggest that these data are insufficient for obtaining reliable estimates of biomass trajectories (see also Cooke, 1983). For example, using the VPA with only catch mass and age structure information, it is sometimes impossible to tell whether a stock is increasing or decreasing at the present time (see Chapter 5). Other results (Chapter 4) show that supplementary data must normally be used before reliable biomass estimates are obtained.

For the stocks under consideration here - pilchard, anchovy and hake - the additional data over and above catch mass and age structure information which are immediately available are measures of effort for the demersal and pelagic fisheries, and records of commercial guano yields. (Effort is assumed proportional to fishing mortality - although this assumption is open to question for the pelagic fishery) . The relevance of the guano yields to pilchard population dynamics follows from an argument (to be presented in Chapter 6) that they reflect recruitment trends. Records of commercial guano harvests dating back to 1896 are analysed (Chapter 6) in an attempt to extract an abundance indicator for pilchard, the preferred prey for certain breeding seabirds.

It is therefore only because a limited amount of relative abundance and recruitment information about a few South African fish stocks exist that the VPA work presented here is of potential value. The main advances achieved in indirect stock assessment are thus the recognition of limitations in the catch at age data, the application of a technique (developed by Fournier and Archibald, 1982) which incorporates additional data in an objective fashion, and finally the application of a resampling procedure for determining VPA estimator variance.

For anchovy and hake, estimates of the age structure of the catch are biassed (for a variety of reasons to be discussed, Chapters 5 and 8). Results are presented to illustrate the impact of these biases on final biomass estimates; these results highlight the need for foresight in planning data sampling strategies for eventual use in stock assessment procedures.

The spawning biomass and recruitment data produced by the VPA can be used in separate calculations for determining maximum sustainable yields (MSY's) (via a yield per recruit approach combined with an estimated stock-recruitment relation) and the MSY and unexploited biomass levels, B_{MSY} and K respectively. (MSY was chosen as a measure providing some indication of the sustained production capabilities of the resource - some of the complexities of such a concept given the existence of extensive recruitment variability are treated for the case of the pilchard stock in Chapter 7.) Estimates of MSY, B_{MSY} and K for pilchard and anchovy using this approach were previously reported by Butterworth (1983) and results for hake from a dynamic catch-effort model by Butterworth and Andrew (1984).

Appropriate procedures for fitting a stock recruit function are developed here (Chapter 7), based on a method reported by Ludwig and Walters (1981). Their method is modified to account for differences between the errors in measuring recruitment and spawning biomass using VPA. This case is not treated in detail by Ludwig and Walters. Furthermore the simultaneous occurrence of model prediction and observational errors in recruitment estimates, treated in the final model, was not dealt with by Butterworth (1983).

Although Butterworth and Andrew (1984) report VPA estimates of hake abundance, they do not use these data to obtain estimates of MSY, but instead they employ a dynamic Catch Per Unit Effort fitting procedure for estimating MSY, B_{MSY} and K. Stock - recruitment curve related yield per recruit estimates of MSY produced in Chapter 8 are also compared to their results. Estimates of MSY, B_{MSY} and K for pilchard and hake are presented in Chapters 7 and 8 respectively. Likely correction factors which should be applied to obtain unbiassed estimates of MSY, B_{MSY} and K are estimated from simulations (Chapter 7).

A recurrent theme which will be encountered in this research is the question of model complexity in the context of estimation models. The two estimation models used for hake, the VPA and the dynamic production model using only CPUE data employed by Butterworth and Andrew (1984), represent two examples in this regard: of a rather complex model on the one hand (VPA), and a very simple one (dynamic production) on the other. From an input data standpoint, the VPA model using effort can be viewed as a CPUE model incorporating additional catch age structure information. VPA model complexity is apparent in the large number of associated model parameters to be estimated (between 20 and 80 for typical situations), while dynamic production models of the sort used by Butterworth and Andrew may require only a small number of parameters (typically about 3). The crucial issue, when reasonable catch mass, age structure and effort data are available, is whether to use the age structure information rather than just to follow a dynamic CPUE approach? (i.e. does the extra information provided by the age data more than compensate for the additional parameters that require estimation?) Some comments about this are made with regard to the hake stock on the west coast of South Africa.

Overall discussion and conclusions drawn from the thesis are summarised in Chapter 9.

In the chapters which follow, a large number of mathematical symbols are used. Their different meanings are described in Appendix 1.1, which includes a section on English symbols, Greek symbols, mathematical conventions and acronyms. An attempt has been made to be consistent, and to avoid the duplication of symbols, as far as possible. Nevertheless a certain amount of duplication has been unavoidable, but in such cases the distinction will be clear from the context in which a symbol is used.

2. AN OVERVIEW OF MATERIAL FLOWS IN THE SOUTHERN BENGUELA PELAGIC ECOSYSTEM

2.1 INTRODUCTION

Like most other ecosystems the southern Benguela ecosystem is extremely complex and this has produced a wide range of responses from scientists concerned about renewable resources in this region. Two extreme viewpoints are presented here, since the results of the material balance set out in this Chapter have a bearing on this debate.

Proponents of the first extreme argue that because of the multitude of possible interactions between species and the environment, these should first be understood individually at a microscopic level before the whole system can be understood. Thus, the argument goes, the multitude of species interactions requires that each and every interaction should be fully understood before a complete understanding of the dynamics of one species of particular importance can be reached. Strictly speaking this is absolutely correct; however, the logical extension of this type of argument ignores the reality of contraints imposed by statistical sampling theory, since it requires that every source of variation be explained. Alternatively, and on the basis of statistical limitations, other researchers argue that the system is so complex that it will never be understood at the level of species interaction, and instead they promote the use of simpler models with very few parameters.

These two approaches are mirrored by multispecies fish management models on the one hand and single species models on the other. Although seldom articulated in quite this way, the aim of an eventually useful multispecies model appears to have shaped and justified much of the research carried out on the Benguela ecosystem. However, it is clear from statistical considerations that at some level or other , understanding of reality is, and always will be, empirical. This has been formally realised in other scientific fields where far simpler models (eg. Boyle's Gas Laws) provide adequate working descriptions of physical systems for which behaviour at the microscopic (or molecular) level is extremely complex.

It is felt therefore that ecosystem models of the multispecies type may ignore fundamental statistical limitations, and that absolute limits to knowledge about ecosystems probably exist because of practical measurement constraints. In local scientific research directed towards fishery management a major dispute is centred on the level at which the determination and understanding of fundamental underlying processes is necessary.

There is growing support for very simple management models, although opponents of this approach argue that they ignore possibly important processes. This problem reflects a more general problem encountered in statistics, in which the model complexity is described by n, the number of parameters which define the model. These parameters are estimated from k observed data values. In general it is expected that the closer n/k approaches unity, the larger the variance of the parameter estimates but the smaller their bias. In contrast, as n/k approaches 0, the variance of the parameter estimates may decrease drastically but bias is substantially increased. The optimum model or regression equation is one which minimises the sum of bias and variance, and this sum is therefore logically defined as the efficiency of the parameter estimator (eg. Wonnacott and Wonnacott, 1977). For fishery management this implies some compromise between models which are either too simple or too complex.

It is hoped that the analysis presented here will help decisions about what level of resolution of ecodynamics is practical for the Benguela ecosystem. Thus it helps to tentatively identify key aspects of the system, which may for example facilitate the construction of more appropriate limited parameter stock assessment models. Thus in carrying out the data search for the material balance, additional information which could be used with Virtual Population Analysis (VPA) techniques may be identified (this is taken up in more detail in Chapter 6).

Finally, improved limited parameter models may be suggested by this approach. It is possible for example that general relations (analogous for example to Boyle's Gas Laws) such as Sheldon et al's (1972) particle size model exist in marine ecosystems and a whole system approach may help to identify these.

It is then in this general context that the following very crude ecosystem material balance is presented.

2.2 AREA OF INTEREST

Upwelling of nutrient-rich water off the west coast of southern Africa creates a suitable habitat for a variety of productive pelagic shoal fish, which form the basis of important fisheries in the Benguela current region. The Benguela current system is bounded in the north near Cape Frio (18°S) by the southward moving Angolan Current, and in the south near Cape Agulhas (35°S) by the Agulhas current retroflection (Nelson and Hutchings, 1983). However, from the ecological point of view the system needs to be defined in terms of the populations of interest, in this instance the pelagic shoal fish.

Traditionally, the South African and Namibian fisheries have been managed separately, and Stander and le Roux (1968) concluded that Cape and Namibian pilchard (Sardinops ocellata) stocks are reproductively isolated; further, even though tag returns demonstrated limited interchange of individuals, Newman (1970) concluded that they were practically separate stocks. There is also the suggestion from egg and larval surveys that anchovy (Engraulis capensis) stocks off Namibia (King, 1977; le Clus, 1983) and the Cape (Crawford, 1981a,b; Shelton and Hutchings, 1982) are separate. Recent starch gel electrophoretic studies of population genetics have been less conclusive about separating Namibian and Cape anchovy stocks (Armstrong et al., 1983), but comparatively little interchange is needed for this approach to lose its power of discrimination.

Boyd and Cruickshank (1983) have shown that there is an extensive zone of semi-permanent upwelling between about 25°S and 31°S, which probably forms a low-temperature barrier separating the Benguela system into two sub-systems or "environmental basins" (MacCall, 1983) of preferred habitat for both species. These stocks may on occasion tend to merge by migration of adults and/or transport of larvae when the stock sizes increase or the thermal barrier disappears through reduced upwelling (Shannon and Field, 1985). This thesis is concerned mainly with the southern stocks and the area south of the suggested thermal barrier (Fig. 2.1). The exact area involved is only of relevance for plankton for which gross standing stock and production estimates must be

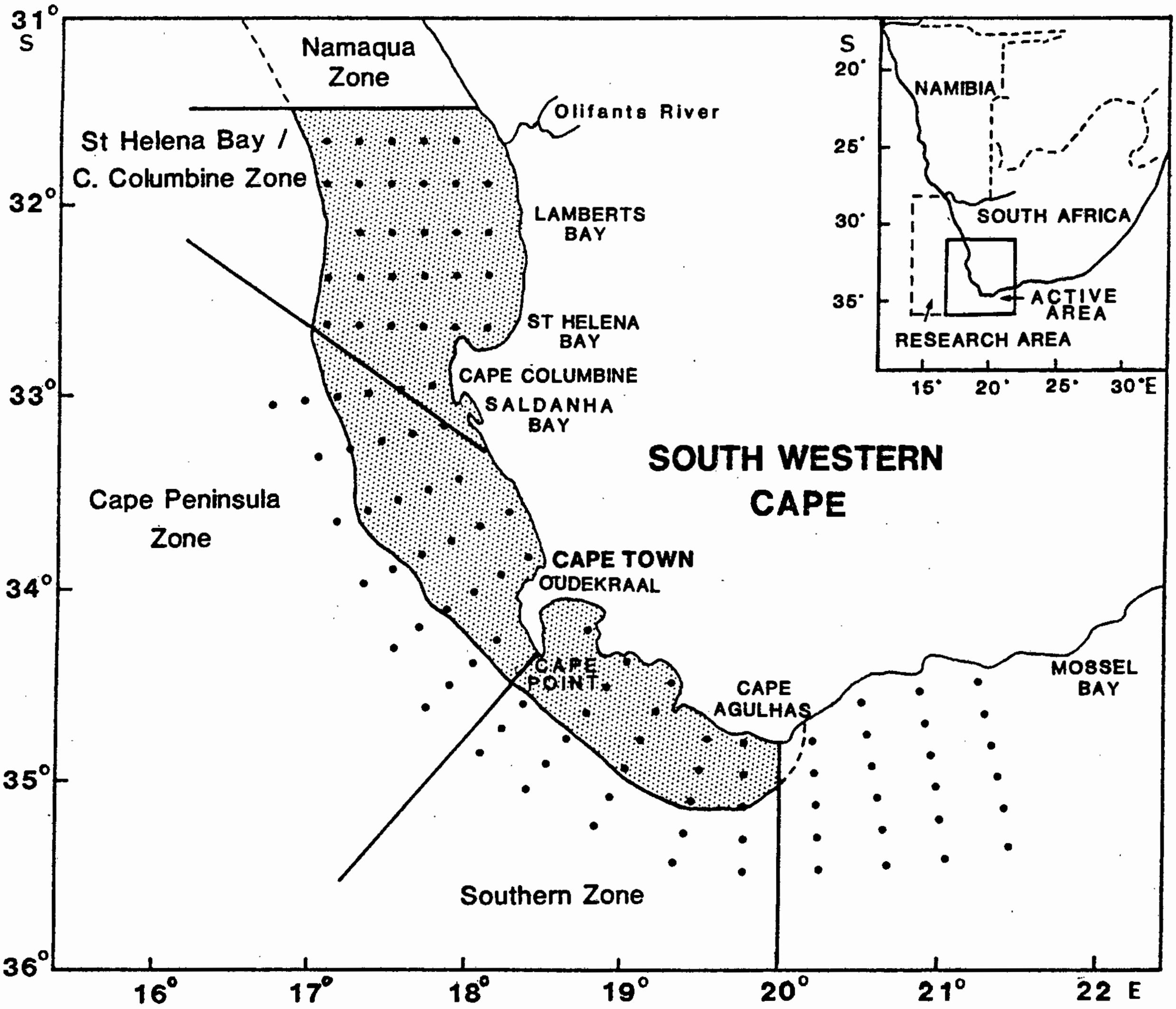


Fig. 2.1. The southern Benguela region showing positions of stations occupied monthly for 12 months for chlorophyll measurements (Shannon et al, 1983), and the four upwelling zones. The "productively active area" (see text) is shaded. The total research area and the productively active area are shown in the insert.

obtained by multiplying estimates in terms of unit area by the total area of interest. In the case of larger species such as fish, the exact area is less significant, since gross population biomass and production for these species is given independently of the precise area of their habitat and in any case this area is different for each species considered here.

For plankton, the area of interest is given by the productively active zone (defined in Section 2.5.1.1), shown shaded in Fig. 2.1 (the basis for definition of this zone is given later in this section). For the purposes of crude production estimates this area can be subdivided into four oceanographically distinct zones (Shannon et al., 1983; Shannon and Field, 1985). With reference to Fig. 2.1 these are :

1. The southern zone over the Agulhus Bank. During the 6 month summer season a fairly concentrated sub-surface layer of phytoplankton at the thermocline is sustained by nutrients mixing across the thermocline. This is about 25 % of the total area.

2. The Cape Peninsula zone from the Cape Peninsula to just south of Saldanha Bay. This forms about 15% of the total area. Plumes of water upwell nearshore in pulses related to the south-easterly winds, and move offshore. Blooms of phytoplankton develop in the upwelled water as it warms and moves offshore. The process is disrupted by downwelling, caused by cessation of southerly winds or by north-westerly winds (Andrews and Hutchings, 1980).

3. The Cape Columbine/St Helena Bay zone from north of Saldanha Bay to the Olifants River. Plumes develop as in region 2 but at a steady, slower rate, without pulsing, allowing the development of large zooplankton standing stocks. This forms about 25% of the area and is the region where recruiting pelagic fish are caught.

4. The Namaqua zone from the Olifants River to the Orange River. This is a zone of perenniel upwelling and strong winds, which may form a cool thermal barrier to pilchard and anchovy. Few pelagic fish are caught in this zone, which forms some 35% of the total area.

Shannon and Field (1985) have shown that most of the pelagic shoal fish are caught within the "productively active area" of the southern Benguela system, an area that includes the highest chlorophyll concentrations and fastest rates of primary production stretching from 31°S to 20°E; this subsystem narrows towards the south and remains close inshore (Fig. 2.1). The boundary of this active area was taken as the point at which the NPRL/CZCS-deduced nearsurface chlorophyll a concentration was higher than 0.1 mg m^{-3} or in the case of shipboard measurements where the median concentration of chlorophyll a during 60% or more of the surveys was greater than 50 mg m^{-2} (Shannon and Field, 1985). The active area is some 40 000 km^2 in extent (see Table 2.2: Shipboard estimate is 38 800 km^2; Satellite estimate is 32 400 km^2, however the shipboard estimate is more reliable).

The pelagic shoal fish cannot be considered in isolation, since they are also preyed upon by populations of larger fish; therefore when these interactions are considered this paper will be concerned with a larger system extending further offshore. Accordingly midwater populations such as lanternfish (Myctophidae) and the demersal populations such as hakes (Merluccius spp.) which are common over the continental slope to depths of some 1000 m are also considered. This larger total southern Benguela area is some 140 000 km^2 in extent; however the exact area is not particularly relevant since it does not come into the calculations in any way (cf. phytoplankton, where total active area is important).

2.3 THE HISTORY OF THE PELAGIC PURSE-SEINE FISHERY

Fig. 2.2 illustrates the history of the pelagic fishery in the southern Benguela region. Initially, the fishery was mainly for pilchard, but catches declined dramatically after 1963, and in 1964 small (12.7 mm) mesh nets were introduced in substantial numbers to enable anchovy to be exploited. Thereafter anchovy dominated the catches, with annual landings averaging over 250 000 tonnes between 1972 and 1983. The figure includes the combined biomass of small pelagic shoal fish (pilchard, anchovy, horse mackerel Trachurus trachurus, chub mackerel Scomber japonicus, and round herring Etrumeus whiteheadii) as estimated by Virtual Population Analysis (V.P.A.) (from data in Crawford et al., 1983) and hydro-acoustic survey estimates of anchovy spawning biomass (Armstrong and Butterworth, 1986).

It should be noted that the VPA calculations shown here are based purely on catch mass and age structure information, and this data base is considered (Pope and Shepherd, 1982, and Chapter 4 this thesis) to be insufficient for estimating recent stock abundances at least. It is possible that for longer lived species historical trends are revealed by VPA calculations. However the species just mentioned are not in this category, and in fact hydroacoustic and egg-production (Armstrong and Butterworth, 1986) estimates of the anchovy spawning biomass are very different to these VPA results.

The record of seabird guano collected commercially from islands around the southern Benguela coast (Crawford and Shelton, 1978)

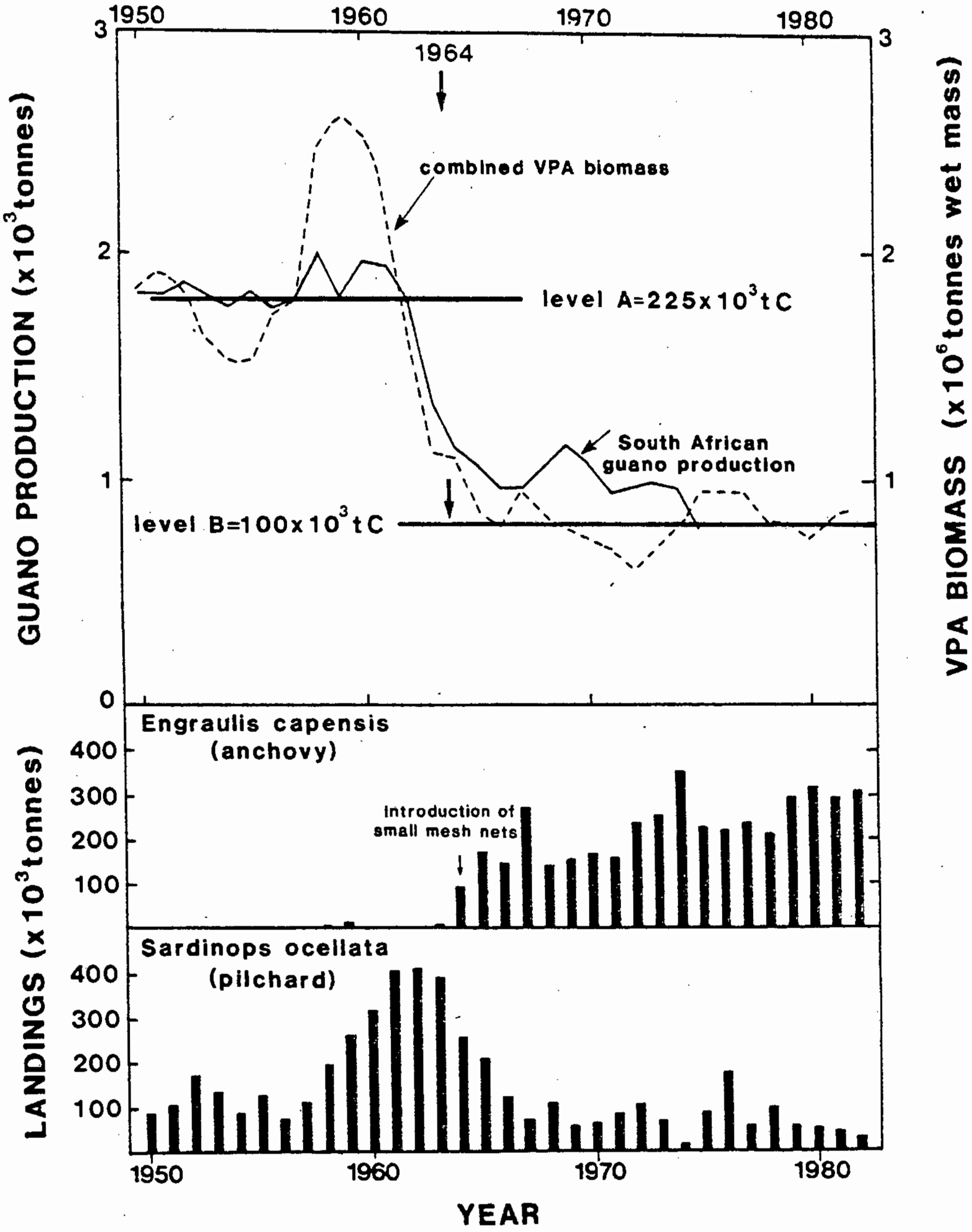


Fig. 2.2. Landings of pilchard (Sardinops ocellata) and anchovy (Engraulis capensis) in South Africa, with virtual population analysis estimates of the combined biomass of pilchard, anchovy and other exploited small pelagic shoal fish since the fishery commenced in 1950 (dashed line). Guano production is also shown. Levels A and B respectively indicate the best estimates of average total pelagic fish biomass before and since 1964, when small-mesh nets were introduced. Carbon equivalents of the biomass are also given. (From Shannon and Field, 1985; data from Crawford et al., 1983).

is also shown, agreeing remarkably well with the estimated biomass trends, apart from the biomass peak in 1959. This figure suggests that the total pelagic fish community supported before 1964 had a relatively constant biomass of approximately 1.8 million tonnes (equivalent to 225 thousand tonnes of carbon) which was dominated by pilchard, and that the system then changed rapidly to a new state dominated by anchovy, with a fairly stable biomass of approximately half the previous level (Shannon and Field, 1985) which has persisted to the present.. Again, the validity of the VPA calculations upon which these deductions are based is highly questionable, as will be shown in more detail in later chapters.

The present chapter aims to study the pelagic fishery in an ecological manner, by considering the trophic links between the main components in the pelagic ecological system. Since so little is known about the trophic interactions in the pelagic system, many estimates that have to be made are the first of their type, but it is argued that initial crude estimates may provide a useful framework within which to view the fishery.

2.4 STUDY METHODS

In order to obtain the best available information about important flows through the food web, a questionnaire was circulated to key biological researchers concerned with different aspects of the Benguela pelagic region. In this way it was hoped that additional qualitative and semi-quantitative information based on both unpublished material and years of experience would be obtained.

A crude summary matrix of some important interactions between predator and prey components in the pelagic ecosystem is given in

Table 2.1. To quantify this matrix, some knowledge of species biomass, production, food consumption rates and diet is required. For most species, more is known about their life-histories and taxonomy than about their quantitative behaviour as populations. A series of assumptions, described below, was based on the literature and on discussions with biologists who have a good feel for probable annual mean diets. Certain other assumptions are better documented and are based on experimental and observational results.

2.5 FLOWS IN THE PRODUCTIVELY ACTIVE AREA

This section is concerned with the epi-pelagic region of the area where most of the pelagic fish catches are made, i.e. zones 1-3 above (see Fig. 2.1).

2.5.1 Estimation of biomass, production and consumption

2.5.1.1 Phytoplankton

The phytoplankton standing stock in the active area was estimated by a combination of shipboard and satellite techniques detailed in Shannon and Field (1985). Chlorophyll a in the upper 50 m of the water column was measured spectrophotometrically at eight depths at 120 stations around the south western Cape coast (see Fig. 2.1) every month for 12 months (Shannon et al, 1983). The productively active area was demarcated by considering those stations where the median depth-integrated concentration of chlorophyll a during 60% or more of the surveys was greater than 50 mg m^{-2}. Near-surface chlorophyll a concentrations were also estimated from Nimbus-7 Coastal Zone Colour Scanner (CZCS) satellite ocean colour images to give synoptic maps of phytoplankton standing stock (Shannon et al., 1983). Depth-

Table 2.1. A matrix of important predator (consumer) - prey (producer) interactions in the southern Benguela pelagic ecosystem. Asterisks indicate that a flow of carbon or energy is believed to occur from the producer (row) to the consumer (column).

PREDATORS/CONSUMERS

PRODUCERS PREY

	Phytoplankton 1	Bacteria 2	Zooplankton 3	Pilchard 4	Anchovy 5	Mackerel 6	Maasbanker 7	R. Herring 8	Myctophids 9	Garfish 10	Squid 11	Hake 12	Snoek etc. 13	Seals 14	Dolphins 15	B. Whale 16	Gannets 17	Cormorants 18	Penguins 19	Man 20
1. Phytoplankton		*	*	*	*															
2. Bacteria		*	*																	
3. Zooplankton		*	*	*	*	*	*	*	*	*	*	*	*							
4. Pilchard		*				*					*		*	*	*	*	*	*	*	*
5. Anchovy		*				*					*		*	*	*	*	*	*	*	*
6. Mackerel		*				*					*	*	*	*	*	*	*	*	*	*
7. Maasbanker		*				*					*	*	*	*	*	*	*	*	*	*
8. Round Herring		*				*					*	*	*	*	*	*	*	*	*	*
9. Myctophids		*				*					*	*	*	*	*	*	*	*	*	*
10. Garfish		*				*					*	*	*	*	*	*	*	*	*	*
11. Squid		*				*					*	*	*	*	*	*	*	*	*	*
12. Hake		*				*					*	*	*	*	*	*				*
13. Snoek etc.		*											*	*	*	*	*	*	*	*
14. Seals		*											*							*
15. Dolphins		*																		
16. Bryde's Whale		*																		
17. Gannets		*																		
18. Cormorants		*																		
19. Penguins		*																		
20. Man		*																		

Note: The category labelled "snoek etc." includes other predatory fish and also sharks. This is the reason for including the asterisk representing predation of seals by "snoek etc.", since sharks prey on seals.

integrated chlorophyll a values correlated well with surface measurements (r = 0.70-0.84, n = 126 - 208) for three oceanographic zones in summer and winter situations (Shannon and Field, 1985). The "productively active area" defined by both conventional shipboard and satellite image methods agreed reasonably in each zone (see Table 2.2), totalling approximately 40 000 km^2 for shipboard estimates for the whole area shaded in Fig. 2.1.

Chlorophyll a values were converted to carbon equivalents by assuming a carbon : chlorophyll a ratio of 60 for the period September through February (the main upwelling season in the southern Benguela region) and 100 for the period March through August when phytoplankton blooms are less prevalent (Andrews and Hutchings, 1980; Carter, 1983; Shannon and Field, 1985). Table 2.2 also shows the productively active area of each oceanographic zone and the phytoplankton standing stock of each zone expressed in carbon units.

Some recent field estimates of primary production in the southern Benguela region are summarised in Table 2.3. The best estimate for the active area (i.e. the average excluding the low values on the Agulhas Bank) is 2.8 g C.m^{-2}.day^{-1}, which combined with the average estimate of phytoplankton standing stock in Table 2.2 of 11 g C.m^{-2}, gives a production to biomass ratio of 0.25 day^{-1}. This is equivalent to a turnover time of four days. These are annual averages over the whole active area, and vary considerably with locality and season (Shannon and Field, 1985).

2.5.1.2 Zooplankton

Cushing (1969) estimated the zooplankton biomass in the southern Benguela region to be about 3.6 g C. m^{-2} (or some 32% of the

Table 2.2. Comparison of satellite and shipboard estimates of the "productively active area" (see text) and mean annual phytoplankton standing stock (g C m^{-2}) in three oceanographic zones distinguished in the southern Benguela system (see Fig. 2.1). (from Shannon and Field, 1985).

Zone	Cape Columbine/ St Helena Bay	Cape Peninsula	Southern
	Active area (km^2)		
Shipboard	17 200	11 800	9 800
Satellite	16 500	9 500	6 400
	Standing stock (g C m^{-2})		
Shipboard	14.1	8.6	9.7
Satellite	11.7	7.3	-

Table 2.3. Some recent estimates of phytoplankton primary production in the southern Benguela area based on oxygen production and carbon-14 uptake experiments. The number of observations (n), mean and standard deviation (s) of daily carbon production measurements in g C.M^{-2}.day^{-1} is given (from Shannon and Field, 1985).

Source	Area	Technique	n	Mean	s
Carter (1982)	Inshore	O_2	29	2.6	1.76
Brown (1983a)	Cape Peninsula				
	(Spring)	^{14}C	12	4.0	2.4
	(Late Summer)	^{14}C	25	2.9	1.4
	Peninsula-Columbine	^{14}C	12	3.5	2.8
Brown (1983b)	Agulhas Bank	^{14}C	5	1.1	0.9

phytoplankton standing stock estimate). For the Cape Peninsula zone which experiences pulsed summer upwelling of an erratic nature, Andrews and Hutchings (1980) observed mean summer and winter values of 0.6 and 1.5 g C.m^{-2} respectively (they estimated zooplankton biomass as some 12% of phytoplankton biomass).

The estimate of zooplankton biomass used here has been taken to be just 16% of phytoplankton biomass (as assumed by Shannon and Field, 1985), and this is in the range of the values estimated by Cushing (1969) and Andrews and Hutchings (1980). Production has been estimated from mean turnover times (B/P ratios) of 10 days based on energy budget measurements of Shushkina et al (1978) and Sorokin and Mikheev (1979) off Peru, and the review of Conover (1978). Zooplankton consumption rates have been estimated to be about 50% of body mass per day (Shushkina et al, 1978; Conover, 1978) which amounts to roughly 32% of primary production (Shannon and Field, 1985).

2.5.1.3 Bacteria

For the purposes of gauging the possible significance of bacteria in the system a ratio of bacterial biomass to phytoplankton biomass of 1:5 was inferred from Newell and Linley (1984), as an annual average in the euphotic zone. Assuming a mean B/P ratio of 2 days for bacteria (both active and inactive), bacterial production may be of the order of 1.6 x 10^7 mt carbon per year over the total area.

2.5.1.4 Planktivorous pelagic shoal fish

Table 2.4 shows estimates of the live biomass of the main pelagic shoal fish. Shannon and Field (1985) have argued that under present conditions, the shoal fish have been fished down to a

Table 2.4. Biomass estimates of the principal stocks of small pelagic shoal fish in the productively active area of the southern Benguela region.

Population	Biomass Estimate (metric tonnes)	Source
Anchovy Engraulis capensis	1 300 000	Armstrong & Butterworth, 1986
Pilchard Sardinops ocellata	100 000	Butterworth, 1983 Crawford et al, 1983
Chub mackerel Scomber japonicus	30 000	Crawford et al, 1983
Horse mackerel Trachurus trachurus	10 000	Crawford et al, 1983
Round Herring Etrumeus whiteheadii	70 000	Crawford et al, 1983

Table 2.5. Estimated biomass and mean annual production and consumption of southern Benguela pelagic shoal fish. Level A refers to the pre-1964 situation when pilchard was the dominant stock, and level B to the present situation. Production was based on estimated turnover rates, B/P, and consumption from the assumed gross growth efficiency (Shannon and Field, 1985)

	Biomass (tonnes)	Production (tonnes per year)	Consumption (tonnes per year)
Level A (pilchard up)	225 000	225 000	1125 000
Level B (anchovy up)	100 000	200 000	1000 000

state in which the system is dominated by short-lived anchovy and juvenile pilchard, whereas the larger, longer lived fish of higher quality for human consumption (mackerel, horse mackerel and pilchard) have been reduced in abundance. Thus the present stocks have on average a faster turnover rate than in the pre-1964 pilchard-dominated system. Table 2.5 summarises their estimates of fish production and food consumption requirements at two states of the system, suggesting that the same pelagic fish production is being maintained with similar food requirements. Circumstantial evidence of changes in the length at first maturity of both pilchard and anchovy with changes in their population sizes (Crawford et al., 1983) lends support to the hypothesis that food supply limits pelagic fish productivity (Shannon and Field, 1985).

Fig. 2.3 summarises the carbon flows in the productively active region of the southern Benguela system. The estimated consumption requirements of zooplankton and pelagic fish account for some 13 million tonnes of carbon per year, leaving 27 million (or 68%) of phytoplankton primary production to form detritus which is either decomposed in the water column or is exported from the active area to be deposited on the sea floor. The existence of extensive mud belts in the St Helena Bay area (north of Cape Columbine) and especially on the continental slope under the front between the upwelled water inshore and the warmer oceanic water offshore (Bailey, 1985) lends support to this hypothesis. The paradox that fish production may be limited by food supply, yet 68% of the phytoplankton becomes detritus, may be due to the "match-mismatch" mechanism, whereby fish shoals (and zooplankton) are unable to respond to patchy phytoplankton blooms fast enough to utilise them fully.

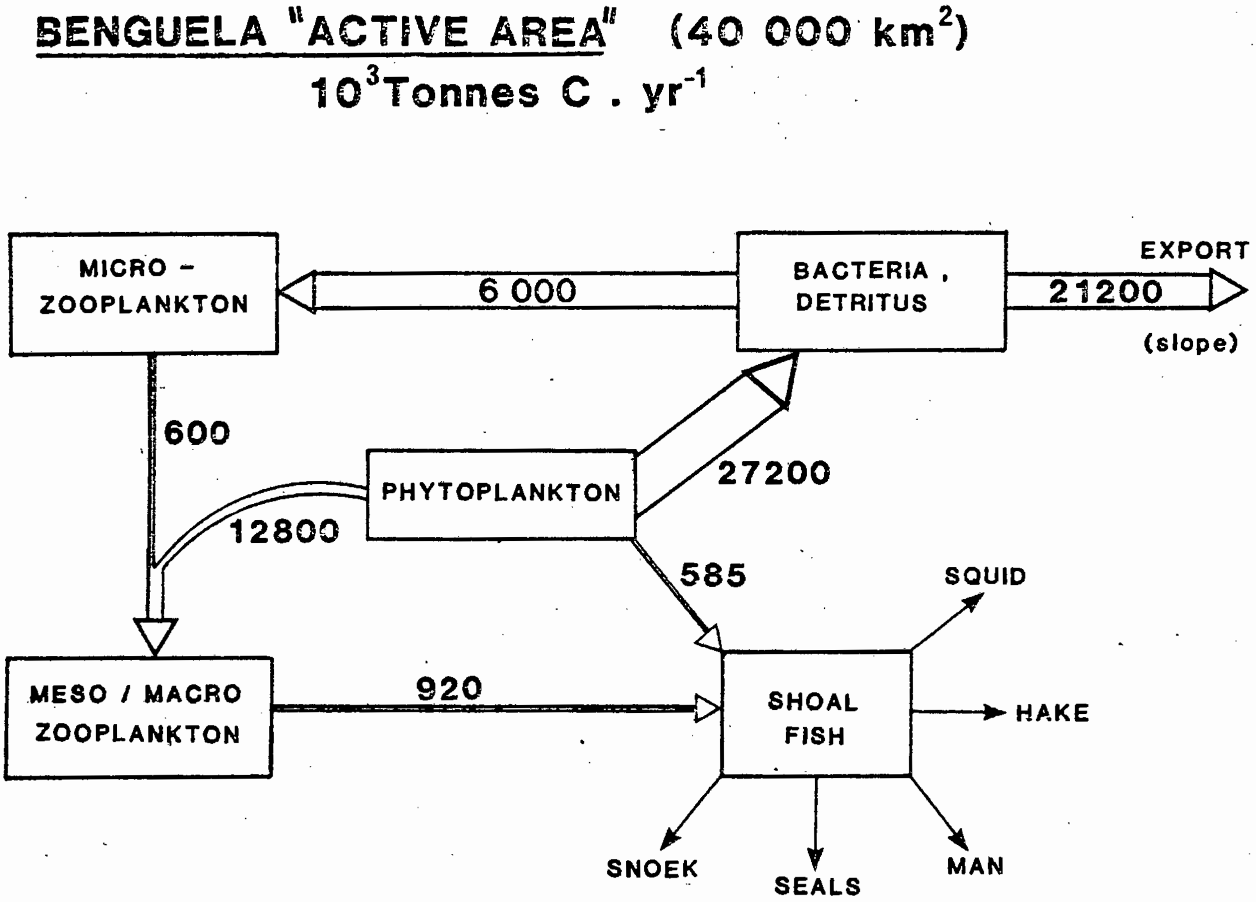


Figure. 2.3. Estimated flows of carbon in the productively active area (40 000 km²) of the southern Benguela system (thousands of metric tonnes). The estimated food requirements of meso- and macro-zooplankton and small shoal fish account for only 14% of phytoplankton production, the remaining 34 million tonnes is assumed to become detritus which is exported from the nearshore surface waters (see text). Metabolic losses and population growth are not indicated on the figure, hence no balance across compartments is intended.

2.6 FLOWS IN THE WHOLE SYSTEM

In this section the system is extended to include the whole southern Benguela area from the Orange River in the north to Cape Agulhas in the south (see Fig. 2.1), offshore to the edge of the continental slope at some 1000 m depth, and including also the midwater and demersal fish, birds and mammals.

For the species freshly encompassed by this larger area, reliable estimates of biomass only really exist for hakes, for which there are Virtual Population Analysis and Catch Per Unit Effort model estimates of biomass (Crawford et al., 1983, Butterworth and Andrew, 1984), whereas the population size of other predators can only be roughly estimated from annual catches. Most of the estimates of this section are therefore very crude, and rely on the enquiry approach (see above) combined with a crude input-output analysis of flows to predators and from prey populations (Platt et al., 1981). A similar study has been made off the California Current Ecosystem (Greene, 1978).

2.6.1 Estimation of biomass, production and consumption

For simplicity, wet mass values are used in this section and converted to carbon equivalents using a wet mass : carbon ratio of 20 for plankton (Cushing, 1969; Platt and Irwin, 1973), and 10 for fish (Mills and Fournier, 1979) in the final input - output table.

Scomber japonicus (Chub mackerel) have been studied by Baird (1978a, b) who found lantern fish (Lampanyctodes hectoris) to be their main fish prey off Cape Columbine, while Nepgen (1982), working in False Bay, found anchovy to be important in their diet. Both workers found zooplankton to comprise the rest of their food intake. Assuming that mackerel consume 12% of their body mass in food per day [best estimate from mackerel figures of Hatanaka and Takahashi (1960) and Hatanaka et al. (1957)], and allocating anchovy and lantern fish each 10% of the diet with the remaining 80% zooplankton, provides estimates using the VPA generated biomass estimate of 30 000 mt wet mass(Table 2.4):

total annual consumption by mackerel: 1 250 000 mt wet food per year (85 000 mt carbon equivalent)
consumption of anchovy: 125 000 mt wet (12 500 mt carbon equivalent)
consumption of lantern fish: 125 000 mt wet (12 500 mt carbon)
consumption of zooplankton: 1 000 000 mt wet zooplankton (60 000 mt carbon)

Trachurus trachurus (Horse mackerel) feed mainly on zooplankton (Sea Fisheries Research Institute, unpublished data), probably at a rate of some 7% of body mass per day (general estimate for pelagic fish - see Conover, 1978). Using the biomass estimate of Table 2.4, the estimated consumption is therefore some 300 000 mt wet zooplankton per year, or 15 000 mt carbon equivalent.

Etrumeus whiteheadii (round herring) are estimated to have a biomass of some 70 000 mt (Table 2.4), consume zooplankton (Sea Fisheries Research Institute, unpublished data) at a rate of 7% of body mass per day (Conover, 1978; Caddy, 1985) and therefore probably consume about 2 000 000 mt wet zooplankton (100 000 mt carbon) per year.

Merluccius capensis and M. paradoxus (hake) are the most significant demersal fish in terms of biomass and the fishery.

Furthermore, they are particularly important to this analysis, because they are top predators and their diet offers some valuable clues about the ecosystem.

Botha (1980) has described their biology and diet in detail, while the most recent biomass estimates for the ICSEAF subdivisions are given by Butterworth and Andrew (1984). For the larger research area being considered in this work (ICSEAF Division 1.6 plus part of 1.5), the biomass of hake is estimated to be about 350 000 tonnes.

The diet of hake changes distinctly beyond the age of about 4 years (Botha, 1980; Chłapowski, 1976; Assorov and Kalinina, 1979), and they become far more piscivorous than before. A large fraction of the fish diet of these older hake is smaller cannibalised hake. The fraction of hake older than 4 years old was estimated from VPA results which are discussed in Chapter 7. The mass proportion of older piscivorous hake is approximately 0.25 for recent years. Using this and assuming a 2:1 ratio of *M. paradoxus* : *M. capensis*, diet composition data from Botha (1980), and a daily ration of 3% body mass per day (Lane et al, 1979), the estimates of annual consumption of various prey by large and small hake are shown in Table 2.6. Note the large amount of cannibalism, and the importance of squid, lanternfish and round-herring in hake diets.

Thyrsites atun (snoek) are major predators of pelagic fish in the Benguela region (Crawford and de Villiers, 1984), but no sound biomass assessments have been made yet. From the annual catches of 5 000 mt by handline inshore and 10 000 mt demersally by trawlers in the study area, as a first estimate the stock may be about 75 000 mt, or 5 times the catch.

Table 2.6. Estimated annual consumption of various prey by hake, *M. paradoxus* and *M. capensis*, in thousand metric tonnes of wet food in the southern Benguela system.

PREDATORS

		M cap. juv.	*M cap.* adults	*M par.* juv.	*M par.* adults	Total
	Total Consumption	968	330	1930	638	3866
	M paradoxus juveniles	247	169	385	507	1308
	M capensis juveniles	25	35	-	-	60
P	*Merluccius sp*	131	-	-	-	131
R	Chub Mackerel	-	23	50	30	103
E	Ribbonfish	14	23	-	-	37
Y	Decapods	21	-	121	-	142
	Lanterfish	68	-	486	24	578
	Lightfish	58	-	169	-	227
	Red eye	294	18	-	13	325
	Rattail	-	6	-	12	18
	Squid	-	5	586	8	599
	Euphausiids	39	-	80	-	119
	Amphipods	32	-	46	-	78
	Stomatopods	13	-	-	-	13
	Dragonets	15	-	-	-	15
	Other	11	51	7	44	113

The snoek migrate southwards inshore and northwards offshore, spending only part (say half) of the year in the study area (Crawford and de Villiers, 1984), which one may assume is split about equally between feeding on the inshore pelagic fish and the offshore demersal and midwater fish. Assuming that the snoek in pelagic surface waters consume about 7% of body mass per day (from typical data cited in Conover, 1978), the epi-pelagic consumption is estimated to be a total of 480 000 mt fish. Using the data of Nepgen (1979) from False Bay, the diet is assumed to be 59% anchovy and 16% pilchard by mass, resulting in an estimate of 280 000 tonnes of anchovy and 80 000 mt wet mass of pilchard consumed by snoek per year.

Similarly, the demersal prey consumed by snoek can be estimated by assuming that at the cooler deeper temperatures snoek need to consume only 5% of their body mass per day (Caddy, 1985). If the deeper living prey are split in four equal proportions between hake juveniles, light-fish (Maurolicus muelleri), lantern-fish and crustacea (euphausiids, stomatopods, and amphipods), the average annual consumption would be some 85 000 mt wet mass of each. Table 2.7 shows the assumptions involved in estimating the consumption of other predatory fish. The linefish include Argyrosomus hololepidotus (kob), Atractoscion aequidens (Cape salmon), and Pomatomus saltatrix (elf).

Cape gannets (Sula capensis), Cape cormorants (Phalacrocorax capensis) and Jackass penguins (Spheniscus demersus) are the main resident seabirds which prey on pelagic fish, and together with non-resident oceanic birds they are estimated to take an exceedingly small fraction of the total annual anchovy production: about 50 000 mt per year wet mass out of the total anchovy production of 2 000 000 mt wet mass (only 2.5%).

Table 2.7. Assumptions and preliminary estimates of biomass and consumption of predatory fish in the southern Benguela system, derived from catch data (Anon, 1977-1980) and the enquiry system. Biomass has been estimated to be some multiple of the mean annual catch (between 2 and 10 times): consumption is evaluated from the biomass, % of the year in the study area, % body mass per day needed in food, and the proportions of each prey species in the diet, mainly from Nepgen (1979, 1982). Linefish include *Argyrosomus* (kob), *Atractoscion* (Cape salmon) and *Pomatomus* (elf).

Predator	Catch (units:	Biomass tons wet)	% Year	% B.W./day	Diet	Consumption (tons wet)
Snoek *Thyrsites atun*						
Pelagic	15 000	75 000	25%	7%	anchovy	280 000
					pilchard	80 000
Demersal			25%	5%	small hake	85 000
					lanternfish	85 000
					lightfish	85 000
					crustacea	85 000
Yellowtail *Seriola lalandii*						
	1 500	15 000	25%	7%	anchovy	28 000
					pilchard	10 000
					squid	28 000
Tuna *Thunnus spp.*						
	2 000	20 000	10%	15%[1]	anchovy	?2 000
					total	100 000
Linefish						
	15 000	30 000	20%	5%	anchovy	15 000
					horse mackerel	4 000
					chub mackerel	4 000
					pilchard	4 000
					round herring	4 000
					squid	15 000
					total	100 000

[1] Magnuson (1969)

Similarly, from crude biomass estimates, the consumption of anchovy by Bryde's whales (Balanoptera edeni), seals and dolphins is estimated to be 30 000 mt (geometric mean of upper and lower estimate cited in Best et al, 1984), 75 000 mt, and 6 000 mt, respectively. Preliminary diet studies of Cape fur seals (Arctocephalus capensis) indicate large amounts of hake (probably scavenged from trawlers) and considerable amounts of squid in their diet (J.H.M. David, pers. commn), from which the trophic flows used below were estimated.

2.7 THE INPUT-OUTPUT TABLE

Table 2.8 presents the input-output matrix of estimated annual carbon flows between some of the main components in the southern Benguela system. The matrix reflects some of the current concepts of the pelagic ecosystem. For example, it suggests that about 32% of phytoplankton production is consumed by zooplankton, which is compatible with zooplankton grazing experiments in the Cape Peninsula zone performed by Olivieri and Hutchings (1983). Note that the annual flows are calculated assuming that diet compositions are constant in time and space; however this is only true as an average over a long period of time, since daily feeding is probably highly variable.

The matrix presentation potentially provides a check for consistency, but in doing this the assumptions underlying various estimates should be considered. Each row total gives an estimate of the total production (P) of that component, and this can be checked against the biomass estimate (B) to see if the P/B ratio is realistic compared to published values. Similarly, the columns add up to the total consumption (C) by that species, and one can check this for realism by comparing the P/C ratio with K_1

Table 2.8. Input-output matrix of estimated annual carbon flows from producers or prey (rows) to consumers or predators (columns) in the southern Benguela system. Units are thousands of metric tonnes of carbon per year, assuming that for plankton 20 g wet mass is equivalent to 1 g carbon, and that for other components the ratio is 10:1 (see text). The first column gives the biomass estimate (B) in thousands of tonnes of carbon. Rows sum to give total production (P), with production/biomass (P/B) in the last column (see continuation overleaf). Columns sum to give total consumption (C), with Gross Growth Efficiency (K_1 = P/C, see text) given in the bottom row. Caution: note that two total production values, viz. for phytoplankton and bacteria, and one total consumption value, viz for squid, are not given by totalling the respective rows and columns. These are labelled 1, 2 and 3 in the table. The value for squid (total amount of material consumed by squid = 250 thousand tonnes of carbon) is just taken to be three times the 83 thousand tonnes of carbon consumed by squid. The basis for values 1 and 2 is discussed in the text.

PREDATORS/CONSUMERS

		BIOMASS (B)	1	2	3	4	5	6	7	8	9	10
	1. Phytoplankton	440		27200	12800	120	465					
	2. Bacteria	88										
	3. Zooplankton	70				80	840	60	15	100	?	?
	4. Pilchard	10						?				
	5. Anchovy	130						13				
	6. Chub Mackerel	3						?				
P	7. Horse Mackerel	1						?				
R	8. Round Herring	7						?				
O	9. Lanternfish	?						?				
D	10. Saury	?						?				
U	11. Squid	?						?				
C	12. Hake	35						?				
E	13. Snoek etc.	13										
R	14. Seals	1.7										
S	15. Dolphins	0.1										
	16. Bryde's Whale	0.15										
	17. Gannets	<0.02										
	18. Cormorants											
	19. Penguins											
	20. Man											
	Total = Consumption (C)			27200	12800	200	1305	>73	>15	100	?	?
	P/C			0.6	0.09	0.08	0.08	0.15	0.18	0.36	?	?

Table 2.8 cont. ...

PREDATORS/CONSUMERS

		11	12	13	14	15	16	17	18	19	20	Prodn (P)	P/B
	1. Phytoplankton											40 000^{1}	91
	2. Bacteria											16 000^{2}	182
	3. Zooplankton		10	15								1 120	16
	4. Pilchard	?		14.3	?	?	?	?	?	?	1.6	16·3	1.6
	5. Anchovy	?		52.6	7.5	0.60	3	0.8	2.20	1.88	30	110	0.85
P	6. Chub Mackerel	?	10	1.5	?	?	?	?	?	?	0.1	>11	>3.7
R	7. Horse Mackerel	?	?	1.5	?	?	1	0.006	?	?	0.2	> 2.7	>2.7
O	8. Round Herring	?	32.4	1.5	?	?	?	0.04	?	?	2	>36	>5.1
D	9. Lanternfish	?	58	18		?						76	?
U	10. Saury	?	?	?	?	?	?	0.24	?	?		?	?
C	11. Squid	?	60	14.3	9	?	?	?	?	?	0.15	>83	?
E	12. Hake	?	130	9	15	?	?	?	?		13	167	4.8
R	13. Snoek etc.	?	?	?	?	?	?	0.01	?	?	2	> 2	>0.15
S	14. Seals			?							0.1	> 0.1	>0.06
	15. Dolphins												
	16. Bryde's Whale												
	17. Gannets												
	18. Cormorants												
	19. Penguins												
	20. Man												
	Total = C	250^{3}	300	>130	>33	>0.6	>4	>1.1	>2.2	>2	≈ 48		
	P/C:	0.33	0.56	0.02	0.003	?	?	?	?	?	-		

(gross growth efficiency) values in the literature (e.g. Conover, 1978; Mann, 1982). However, this line of reasoning povides no real check if P and C are both calculated from B according to the literature. For example, if C is obtained using plausible percentage body mass per day values and P from plausible turnover times for the category in question, then it is likely that P/C will agree with typical values quoted in the literature. Rather, it is useful to compare P/C with literature values of K_1 when P or C or both are estimated by addition along a row or column in the input - matrix (Table 2.8) and similarly for P/B.

For biomass estimates which are obtained using VPA (pilchard for example), there is an inbuilt assumption of stock productivity, since M, the natural mortality, is assumed. Thus plausible B/P ratios for these species are a direct result of assuming plausible values of M, the natural mortality, if P is calculated from the VPA as well (rather than from predator consumption rates obtained from Table 2.8).

The bacterial P/C ratio is calculated to be 0.6, suggesting that bacteria are 60% efficient in converting phytoplankton carbon into bacterial production. This figure may be too high (Newell and Linley, 1984). However bacteria are also likely to feed on faeces and corpses of zooplankton, fish and other organisms, carbon flows which are not reflected in this matrix. Therefore bacterial consumption would be underestimated and may in fact be twice the value quoted here (for consumption of phytoplankton alone).

On the other hand, the ratios for anchovy and pilchard agree with the ranges in the literature, indicating that the assumptions are at least consistent with one another. The P/B ratio for hake of

4.8 appears to be high for a fish that lives for 9+ years (Botha, 1980), but it may be real, and the fact that much of this production circulates within the hake community (Table 2.6) would make it undetectable in catch and effort statistics and models based thereon.

Predation on zooplankton adds up along the row to a sum production of 1 120 000 metric tonnes carbon, giving a P/B ratio for zooplankton of 16 turnovers per annum. The P/C ratio for zooplankton (0.09) is small, but nevertheless lies within the range of literature values (Conover, 1978 indicates values in the neighbourhood of 0.05 - 0.5). One might expect that P/C for zooplankton would be estimated too low by the method used here due to the known complexity of the zooplankton community.

It should be noted that the zooplankton category spans a wide size range from nauplius larvae of copepods to large chaetognaths, euphausiids and larval fish. Thus much of the carbon flow in the zooplankton community is likely to occur in the smaller size classes <u>within</u> the zooplankton (Azam et al., 1983), and will not be reflected by transfers of production to fish, the prime interest of this study. This is analagous to cannibalism in hake referred to above, i.e. rapid turnover of carbon within the category, which would not have been detected without the studies of hake diet referred to above. Only 13 000 tonnes of hake carbon is harvested annually by trawlers, compared to 130 000 tonnes that is estimated to be consumed cannibalistically by adult hake. Thus total zooplankton production has probably been estimated too low because it fails to take into account the complexity of the zooplankton microcosm, as indicated by the low P/C ratio.

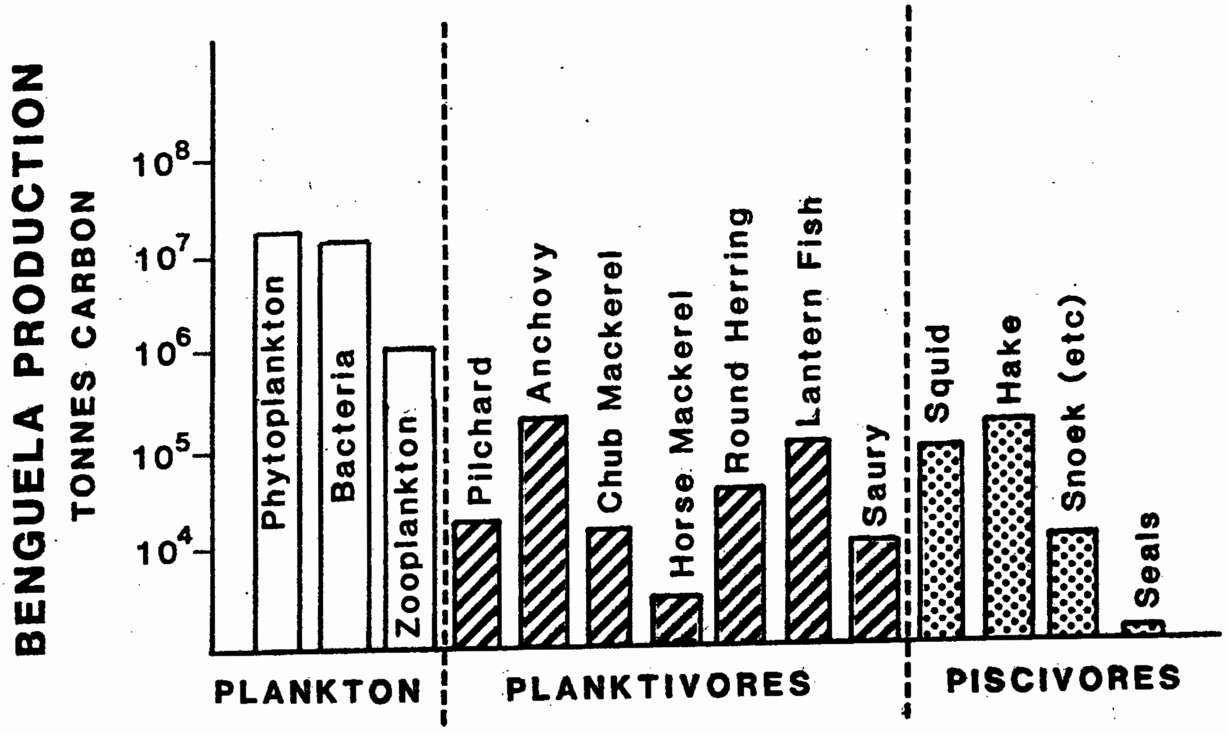


Fig. 2.4. Estimated production of some major components of the southern Benguela ecosystem plotted on a logarithmic scale, summarised from the input-output matrix. The category "snoek etc." includes other predatory linefish (see text) and sharks.

2.8 A SUMMARY OF KEY FEATURES IDENTIFIED

2.8.1 Key components

Fig. 2.4 summarises the production of the main components in the input-output matrix on a logarithmic scale. The largest standing stocks and carbon flows in the system involve the phytoplankton, bacteria and zooplankton. For bigger "particles" such as anchovy and larger species, there are major gaps in current knowledge; in particular the biomass and feeding of midwater lanternfish, lightfish, sauries, and squid need quantification.

Although little is known about Bryde's whales and particularly dolphins, the estimates presented here suggest that they are relatively unimportant quantitatively, and the same applies to seabirds. However, Crawford and Shelton (1978) have shown that seabirds are invaluable as independant indicators of pelagic fish population dynamics. The same may apply to marine mammals which, although long lived so that their total population may respond to prey population changes only slowly, may nevertheless provide more sensitive indicators of such changes through indices of their reproductive success. Cetaceans (without the restriction of land based breeding colonies) may also provide clues to prey distribution changes. Furthermore, these top predators may play an important stabilising role in the ecosystem by feeding on both the smaller predators and small fish and switching among them according to their relative abundance.

The data of Botha (1980) on hake diet and calculations made here indicate that horse mackerel, round herring, lanternfish, lightfish and particularly squid are important components of the ecosystem. Hake consumption of squid is considerable, and suggests that squid production is some 100 000 tonnes carbon annually (1 million tonnes wet mass). In other marine ecosystems, squid have been recognised as important predators of small shoal fish. For example, Greene (1978) estimated that squid accounted for up to 30% of the total mortality of small school fishes in the California Current. Likewise, Idyll (1973) mentions squid as a predator on anchoveta off Peru. In the Southern Ocean, Everson (1977) estimated that squid production is larger than fish production. It is possible therefore that squid are also important in the southern Benguela system, where they are relatively unexploited (see yield to man in Table 2.8).

Squid may however be overrepresented if their beaks are used to estimate the frequency of occurrence in diet, since the beaks are particularly difficult to digest. Hence the findings with regard to the importance of squid in the ecosystem need to be regarded with some caution.

2.8.2 Key processes

Fig. 2.5 illustrates the consumption estimates derived from the input-ouput matrix. Again the large role of lanternfish is prominent. The major roles of squid (which is questionable) and hake are also evident, and both are able to meet a portion of their dietary requirements cannibalistically. Fig. 2.6 illustrates the relationships between hake, squid and small school fish and their larger predators in a flow diagram, indicating the central role of squid, both as prey of hake and predators of small fish. It seems therefore that changes in squid and hake exploitation may have a direct bearing on the dynamics of the small shoaling fishes.

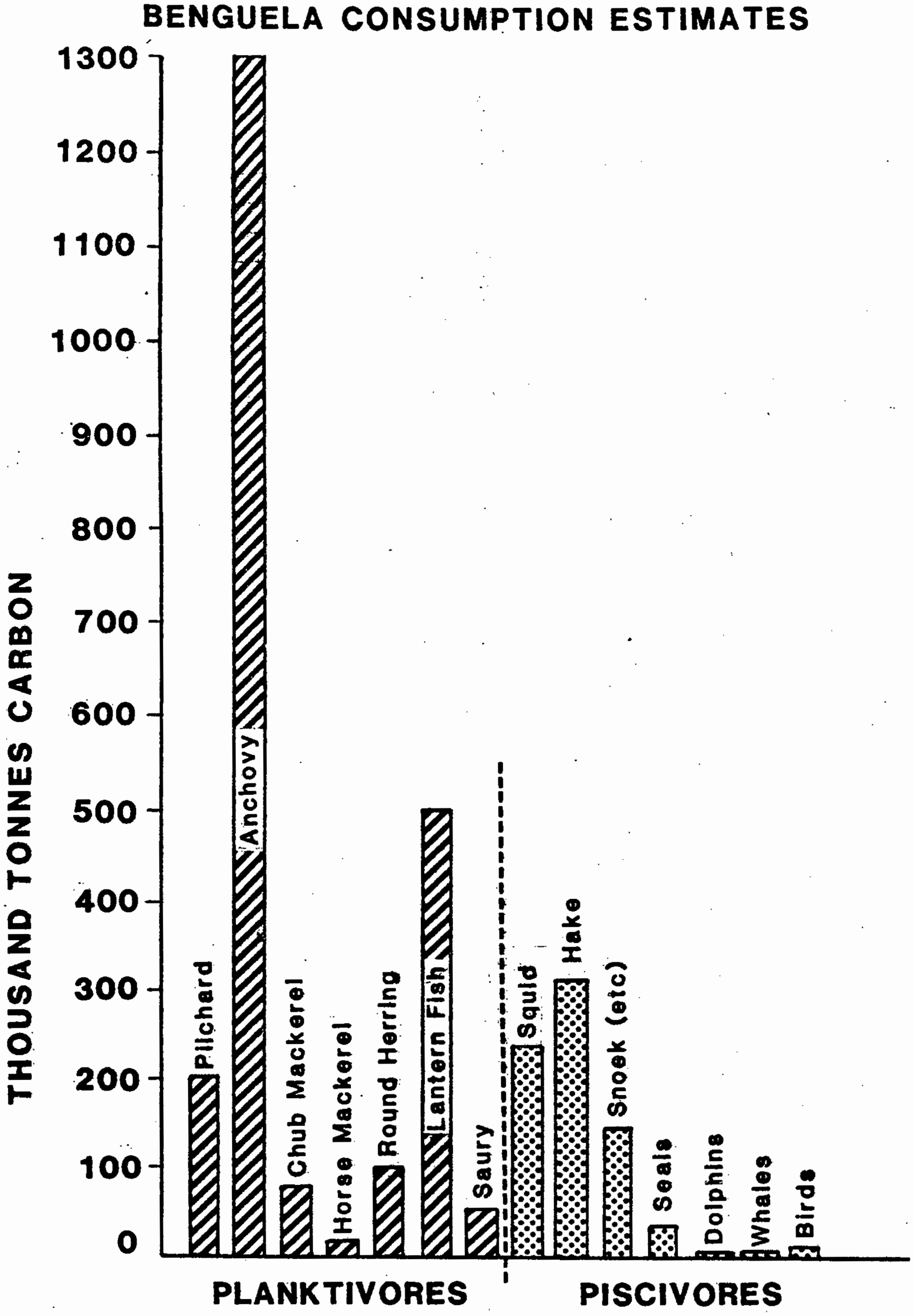
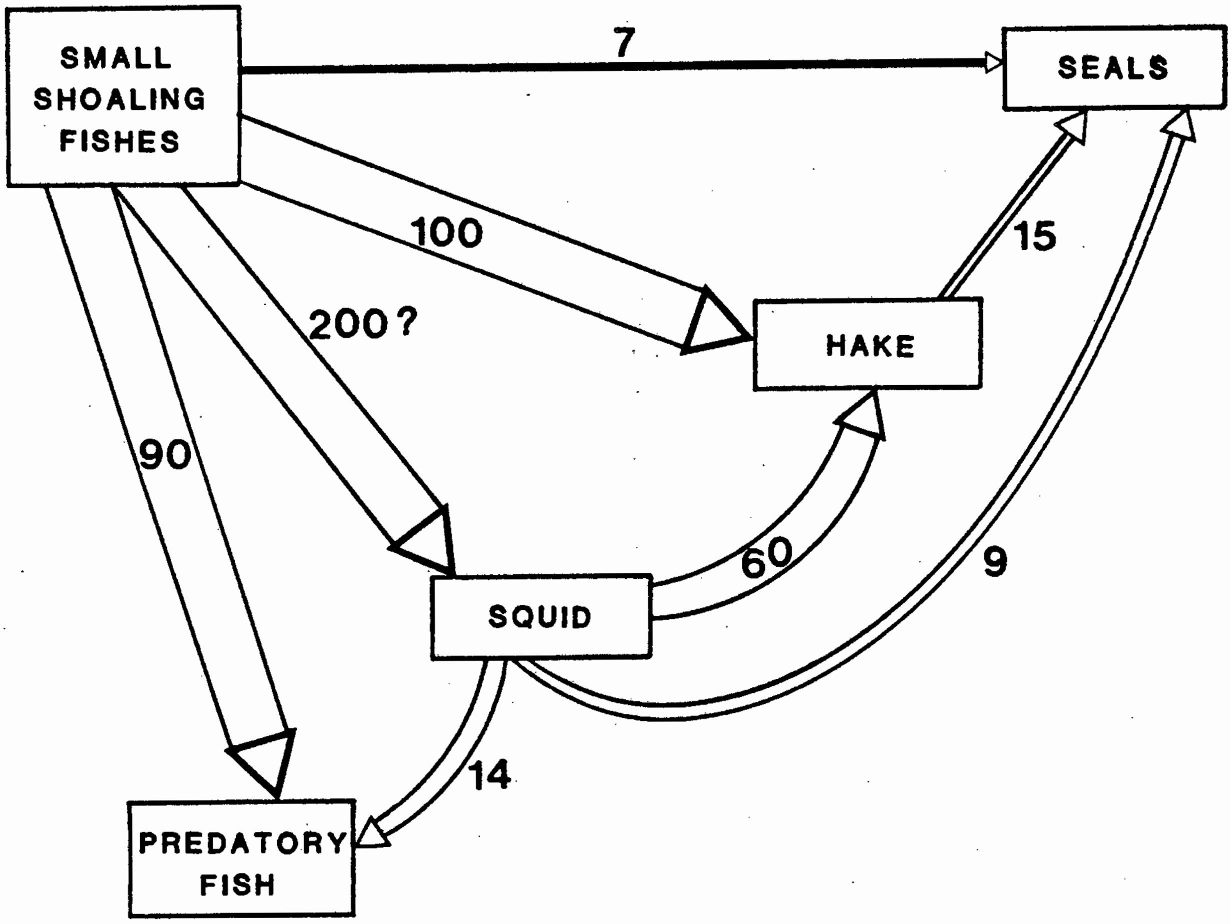


Fig. 2.5. Estimated annual consumption requirements of some major components in the southern Benguela system, summarised from the input-output matrix in thousands of metric tonnes of carbon. The category "snoek etc." includes other predatory linefish and sharks (see text). Multiplying by 10 converts the given quantities to wet mass units.



SOUTHERN BENGUELA SYSTEM

(Tonnes Carbon x 10^3. yr^{-1})

Fig. 2.6. Estimated annual flows of carbon (thousand tonnes carbon per year) between some major components in the southern Benguela system (area 140 000 km²). Note the central position of squid, which are prey of hake, snoek and other fish, and seals on the one hand, and predators on small shoaling fish on the other. (See text for discussion).

Piscivorous fish, of which snoek are the major component (see Table 2.7), are shown to be major predators particularly of anchovy.

2.8.3 Comments about appropriate limited parameter models

The study of flows between conventional compartments, the species stocks, cannot be usefully classified in terms of transfers between trophic levels. This is due to the complexity of the food web, especially when it comes to dealing with the phenomenon of cannibalism (eg. squid and hake) which appears to be common in low diversity upwelling systems (Caddy, 1985). A complementary approach which may avoid some of these difficulties might be to model the system in terms of the "particle-size spectrum" (Sheldon et al., 1972; Platt et al., 1981; Azam et al., 1983).

For hake, cannibalism has important consequences for indirect stock assessment using VPA techniques, since these conventionally assume natural mortality to be constant, independent of year and age. Predation of juveniles by adults makes this assumption highly questionable, and a realistic VPA requires additional comprehensive diet information which is not available for all years (see Lleonart et al, 1984).

Overall, this analysis has had some success in achieving the goals originally set out. Areas of research neglect which have been revealed do seem to have a bearing on the development of rational management strategies for the renewable resources. For hake, the implications of extensive cannibalism needs to be assessed in terms of current harvesting strategies. The relationship between snoek and anchovy is important and generally the piscivorous fish seem to be dominant predators. Untapped midwater resources (lanternfish/lightfish) are shown from predator diets to be highly productive, and squid may be a significant component of the ecosystem.

The complexity of the food web has been illustrated, and this does not encourage the use of multispecies models of fish dynamics. This is related to comments about model complexity made in the introduction. It would seem that a working description of the ecosystem outlined in Table 2.1 would require the estimation of quite an overwhelming number of parameters. This should be seen in the context of work concerned with stock assessment using VPA methods reported in the remainder of this thesis, which illustrates the problems associated with obtaining reliable estimates of just some of these parameters.

CHAPTER 3. AN EVALUATION OF TWO METHODS OF VIRTUAL POPULATION ANALYSIS

3.1 INTRODUCTION

In this chapter two different approaches to the VPA problem are described. Only the second technique (Fournier and Archibald, 1982) is developed and applied in subsequent chapters and the reasons for this are set out here. However the first method described (Pope and Shepherd, 1982) provides a useful example of a specific solution to the problem, and produces similar basic results to those obtained in this thesis. The practical application of VPA techniques is dealt with in subsequent chapters.

3.1.1 Background

The Sea Fisheries Research Institute (SFRI) of South Africa has sampled catches of pelagic fish since about 1950. The sampling scheme was not specifically designed for the subsequent application of stock assessment techniques such as the method of Virtual Population Analysis (VPA). Age structure estimates have been obtained from a large sample estimate (about 20 000 fish per year) of the length frequency distribution of the catch, and smaller subsample estimates of the age distribution in different length classes, known as the age-length key. The total number of fish caught each year, C_i^{obs}, is estimated from catch mass and length distribution estimates

$$C_i^{obs} = {}^{m}C_i^{obs} \Big/ \sum_{l=1}^{lmax} w_l f_{i,l}^{obs} \qquad \ldots\ 3.1$$

where ${}^{m}C_i^{obs}$ = the observed catch mass in year i

$f_{i,l}^{obs}$ = the fraction of the catch observed to fall in length class l in year i;

lmax = the total number of length classes, where

$$\sum_{l=1}^{lmax} f_{i,l}^{obs} = 1$$

w_l = the mass of a fish with a length in the middle of length class l, where the length classes are generally of width 0.5 cm.

The observed breakdown of the catch by numbers-at-length in year i, $C_{i,l}^{obs}$, is given by

$$C_{i,l}^{obs} = f_{i,l}^{obs} C_i^{obs} \qquad \ldots\ 3.2$$

The age-length key for year i, $a_{i,l,j}^{obs}$ is the fraction of fish in length class l observed to lie in age class j (i.e. are aged between j and j+1 years old). Thus

$$\sum_{j=1}^{m} a_{i,l,j}^{obs} = 1 \qquad \ldots\ 3.3$$

for all i,l where m is the oldest fish recorded over the history of the fishery. The proportions of fish caught in year i and estimated to be aged j, the quantities $P_{i,j}^{obs}$, are therefore given by the relation

$$P_{i,j}^{obs} = \sum_{l=1}^{lmax} a_{i,l,j}^{obs} f_{i,l}^{obs} \qquad \ldots\ 3.4$$

These sampling and calculation procedures have an important bearing on the formulation of VPA techniques which are applied to South African fish stocks in subsequent chapters. To avoid confusion later on it is important to note that the sum

$$\sum_{j=1}^{m} P_{ij}^{obs} C_i^{obs} w_j \qquad \ldots\ 3.5$$

where w_j is the midyear mass of an individual fish in age class j, and the sum

$$\sum_{l=1}^{lmax} f_{i,l}^{obs} C_i^{obs} w_l \qquad \ldots\ 3.6$$

where lmax is the largest length class observed, and w_l is the mid-length class fish mass, are not necessarily equal. Note that for the two sums to be equal it is required that

$$\sum_{j=1}^{m} \left[\sum_{l=1}^{lmax} a_{i,l,j} f_{i,l} \right] w_j = \sum_{l=1}^{lmax} f_{i,l} w_l \qquad \ldots\ 3.7$$

so that

$$\sum_{l=1}^{lmax} \left[\sum_{j=1}^{m} a_{i,l,j} f_{i,l} w_j \right] = \sum_{l=1}^{lmax} f_{i,l} w_l \qquad \ldots\ 3.8$$

For this always to be satisfied it is necessary (for arbitrary cases) that

$$\left[\sum_{j=1}^{m} a_{i,l,j}\ w_j \right] = w_l$$

which requires that w_l is year dependent contrary to conventional calculation procedures. Thus the equality 3.7 is not in general satisfied. Note further that the sum in 3.5 is an approximation for the integral;

$$\int_{t=0}^{t=\infty} w(t)\ C_i(t)\ dt$$

where $C_i(t)$ is the instantaneous catch rate in year i, and t refers to age. The sum in 3.6 is an approximation for the integral;

$$\int_{l=0}^{l=\infty} w(l)\ C_i(l)\ dl$$

The two integrals are equal, so that if fish ages and length classes were considered over infinitesimal time and length intervals, then the equality in 3.7 would hold. However, ages over the range [j, j+1] yrs are represented by a single mass and so are lengths over the range [l, l+0.5] cm. The time discretization is more severe than that for length, since for a species like anchovy the first year step spans a large number of length classes.

This discrepancy should not normally be a problem, because the VPA calculation considers fish numbers only. The final abundance estimates are then calculated for a fixed time of the year (usually the beginning) for which the appropriate estimates of w_j can be given. However, catch mass is in many cases a basic observation, so that a mass quantity is sometimes incorporated in the VPA likelihood function (see Chapter 5). Thus it becomes necessary to define catch mass by summing over ages (equation 3.5). This is not exactly equal to the true catch mass because of the discretization error noted above, but it is considered that this quantity is adequate for use in the the objective function as used in Chapter 4 because the difference is usually quite small.

In the VPA, historic estimates of cohort numbers are made by a process of backcalculation by adding up the estimated numbers caught at each age and allowing for natural predation. The number of live fish aged j at the beginning of year i, $N_{i,j}$, those caught in year i, $C_{i,j}$, the fishing mortalities, $F_{i,j}$, and the natural mortalities, $M_{i,j}$, are related by the catch equation:

$$C_{i,j}=N_{i,j}F_{i,j}[1-\exp(-M_{i,j}-F_{i,j})]/[F_{i,j}+M_{i,j}] \quad \ldots\ 3.9$$

This equation is the result of assuming that removals via predation and fishing follows a classical exponential decay model. The differential equation for this is;

$$dN(t)_{i,j}/dt=-M_{i,j}N(t)_{i,j}-F_{i,j}N(t)_{i,j} \quad \ldots\ 3.10$$

Pope (1972) has shown that (unless fishing has made a very insubstantial impact on the stock) the error caused by a poor estimate of the terminal cohort size (or associated terminal fishing mortality) decreases as one calculates further back in time. Thus one expects that, in the absence of systematic departures from model assumptions, recruitment is fairly accurately determined compared to numbers of older fish.

Essentially, a VPA calculation can be performed if estimates of $F_{i,j}$ for the oldest recorded age of a cohort are given. Using equation 3.9, the fact that

$$N_{i+1,j+1}=N_{i,j}\exp(-F_{i,j}-M_{i,j}) \quad \ldots\ 3.11$$

and assuming some value for $M_{i,j}$ (usually taken to be one constant value), all the $N_{i,j}$ and $F_{i,j}$ can be attained by the process of backcalculation already described. However, the current year biomass is given only in terms of the catches and the terminal fishing mortalities for that year, which must somehow be estimated. Thus the precision of the current year biomass is critically dependent on the precision of the terminal fishing mortalities, as are the estimates of the oldest cohort numbers.

For the local pelagic fish species, these terminal fishing mortalities have previously been extremely crudely determined (Centurier-Harris, 1977; Crawford et al, 1983; Shelton and Armstrong, 1983; Butterworth, 1983), without taking cognisance of the errors in the basic data. For example, in some cases the terminal fishing mortalities for the last column in the catch-at-age matrix (i.e. for fish born m years prior to the last year of fishing) have all been given the same fixed value (Centurier-Harris, 1977; Crawford et al, 1983; Shelton and Armstrong, 1983), and this is tantamount to assuming the value of the current year biomass (see equation 3.7).

It is thus not surprising that some previously estimated anchovy biomass trajectories have matched catch trends (Crawford et al, 1983; Shelton and Armstrong, 1983), since this is a consequence of year invariant catch age structures and using a single value for all the terminal fishing mortalities.

Alternatively, the terminal fishing mortalities could be determined by iteratively forcing

$$F_{l,k}=F_{l-1,k}$$

$$F_{l,k}=F_{l,k-1} \qquad \ldots\ 3.12$$

for the terminal year and age fishing mortalities respectively (l=terminal year, k= terminal age), which allows some measure of objectivity. Butterworth (1983) actually averaged the last 3 ages and years for the two equations shown in 3.12 above;

$$F_{l,k}=(F_{l-1,k}+F_{l-2,k}+F_{l-3,k})/3$$

$$F_{l,k}=(F_{l,k-1}+F_{l,k-2}+F_{l,k-3})/3 \qquad \ldots\ 3.13$$

However, this gives considerable statistical weight to age structure information for old fish, which is known to be poorly determined (see Chapter 4).

More recently, methods have been developed which allow a more objective interpretation of the available information. In particular, these estimate the terminal fishing mortalities within the VPA calculation. Two examples of this are the separable VPA devised by Pope and Shepherd (1982), and the more general approach of Fournier and Archibald (1982), based on maximum likelihood theory. These methods are discussed in some detail in Section 3.3 and 3.4.

Pope and Shepherd (1982) conclude that VPA is only really of use as an abundance estimation technique if other information about the fishery, for example relative abundance information such as CPUE, is available. It is only in this context that further development of the application of VPA techniques for the South African fish stocks considered here is justified. Furthermore, there is a great need to demonstrate the inaccuracy and imprecision of South African stock size estimates, even when the most appropriate statistical techniques are used, to break through the impasse that has been reached by the indiscriminate application of deterministic techniques and subjective assumptions such as mentioned above with regard to the estimation of critical parameters. Thus the originality of this work rests on the specific and practical application of existing techniques rather than the development of fundamentally new approaches.

3.2 A GENERAL PARAMETERISATION OF THE VPA PROBLEM

Typically, the fisheries mathematician is presented with an m x n matrix of values of $P_{i,j}^{obs}$, of the proportion by number of fish aged j in the catch of year i (where m is the maximum age detected over all the n years of the fishery) and the n observations of total catch mass, ${}^{m}C_i^{obs}$, or total catch number, C_i^{obs}, for each year. The purpose of the VPA is to use this information to obtain estimates of $N_{i,j}$, the number of fish aged j at the beginning of year i. With additional information about fish masses at different ages, w_j, this can be used to obtain stock abundance estimates, B_i:

$$B_i = \sum_{j=1}^{m} w_j N_{i,j} \qquad \ldots\ 3.14$$

If the only model assumptions are those of exponential mortality (equation 3.10), then the VPA can be regarded as a problem in estimating 3nm+n parameters from nm+n pieces of information. The model parameters to be estimated are the nm fishing mortalities, $F_{i,j}$, the nm natural mortalities, $M_{i,j}$, the nm proportions of fish of M ages for n years in the catch, $P_{i,j}$, and the n catch mass quantities, ${}^{m}C_i$, or catch number quantities, C_i, $(P_{i,j}=C_{i,j}/C_i)$.

However, it will be shown that because the terms, $C_{i,j}$ (or $P_{i,j}C_i$), for the same year class are all related by catch equations (equation 3.16), and because for each year the $P_{i,j}$ sum to 1, the nm $P_{i,j}$ parameters and n C_i parameters are in fact reducible to n+m-1 unique parameters. Thus the total number of parameters that must be estimated is 2nm+n+m-1. Nevertheless, it is clear that the problem as stated is underdetermined, because the number of parameters (2nm+n+m-1) exceeds the amount of data (nm+n).

It is mathematically necessary to formulate a problem that is at least perfectly determined, and statistically desirable that the problem is well over-determined. By making a number of assumptions, the number of parameters can be radically reduced to achieve this successfully. Given an over-determined formulation of the problem, it is further necessary to devise an objective function such as the sum of squares of the residuals between the fitted and observed catch mass and age structure data, whose minimum defines the optimum choice of parameters.

To start with, it is usually assumed that natural mortality is constant over all ages and for all years. The nm natural mortality parameters are thus replaced by one parameter, M. This is probably the most sweeping assumption one can make because it reduces the whole gamut of mortality-causing interactions with other species (even with itself if cannibalism occurs) and with the environment into a single value. It can only be hoped that the extreme complexity of these other effects results in an approximately random distribution of the actual nm natural mortalities around M with no marked trends with age or stock biomass.

The total number of parameters is now nm+n+m, if M is included, so that the problem is still underdetermined (data = nm+n). At this stage, however, the parameterisation procedures adopted by different workers diverge, and two of these will be discussed separately. The first is due to Pope and Shepherd (1982).

3.3 THE METHOD OF POPE AND SHEPHERD

Apart from the assumption of constant natural mortality over all years and ages, Pope and Shepherd (1982) assume that the fishing mortality for a cohort in any one year may be represented by the product of two effects, one dependent on age and the other on the year. Thus they propose that the nm fishing mortalities, $F_{i,j}$, can be separated into the product of n year effects (usually related to the amount of fishing effort in that year), F_i, and m age effects (selectivities), S_j:

$$F_{i,j} = F_i S_j \qquad \ldots\ 3.15$$

This equation requires that the average age specific selectivity pattern of the fishery, which is just the relative susceptibility of fish of different ages to being caught, is unchanged from year to year. This fixed selectivity assumption is often difficult to justify, because fish stock habitats are frequently age stratified, and in the South African pelagic fishery for example, fish catches of significantly different age compositions are off-loaded at different ports. Thus the selectivity pattern could vary systematically from year to year if the quota allocations to companies based at different ports change differentially over time.

Pope and Shepherd assume further, although this is not explicitly stated in their paper, that the total number of fish recorded as caught each year, C_i^{obs}, is known exactly. This eliminates the n parameters, C_i, the total numbers caught which would best fit the model, so that C_i now equals C_i^{obs}.

The nm parameters defining the proportion by number in the catch ($P_{i,j}$'s) consist of a smaller number of unique parameters (i.e. the number of age classes minus 1). This is because for each cohort, all the $P_{i,j}$'s can be expressed in terms of any one $P_{i,j}$, given the n F_i's, m S_j's and n C_i's, by a relationship involving the catch equation :

$$\ln P_{i,j} + \ln C_i - \ln P_{i+1,j+1} - \ln C_{i+1} = \ln D_{i,j} \qquad \ldots\ 3.16$$

where $D_{i,j}$ is the ratio

$D_{i,j} = P_{i,j}C_i/[P_{i+1,j+1}C_{i+1}]$ ... 3.17

and the $D_{i,j}$ quantities are expressed in terms of the model parameters F_i and S_j from the catch equation

$P_{i,j}C_i=N_{i,j}F_iS_j(1-\exp(-F_iS_j-M))/(F_iS_j+M)$ ... 3.18

$P_{i+1,j+1}C_{i+1}=N_{i+1,j+1}F_{i+1}S_{j+1}(1-\exp(-F_{i+1}S_{j+1}-M))/(F_{i+1}S_{j+1}+M)$ ... 3.19

There are n+m-1 diagonals in an mxn matrix (Fig. 3.1), yielding n+m-1 $P_{i,j}$ parameters. In addition, for each year the $P_{i,j}$'s sum to one;

$$\sum_{j=1}^{m} P_{i,j} = 1 \qquad \text{... 3.20}$$

so that a further n $P_{i,j}$ parameters are eliminated. This leaves m-1 unique $P_{i,j}$ parameters. Pope and Shepherd do not make use of this contraint probably because this complicates the formulation and minimisation of their objective function. They retain the n+m-1 unique $P_{i,j}$ parameter formulation, and these will be termed "fundamental $P_{i,j}$'s" hereafter.

The VPA problem has thus been redefined in terms of 2(n+m) parameters and nm+n observations, and these are summarised in Table 3.1. In this table, the $P_{i,j}^{obs}$ represent the observed proportion by number of fish of age j in the catch in year i. Prospects of a solution exist if the available amount of data exceeds the number of parameters i.e. if nm+n > 2(n+m). For

m=1 no solution is possible
m=2 a solution exists for n>4
m=3 a solution exists for n>3
m>3 a solution exists for n>1

so that for most fishery data banks solutions are in principle potentially attainable.

Table 3.1. A breakdown of the parameters which must be estimated and the input data for the Pope and Shepherd (1982) Virtual Population Analysis model, using only catch mass and age structure data.

	PARAMETERS	QUANTITY	INPUT DATA	QUANTITY
	F_i	n	C_i^{obs}	n
	$P_{q,k}$	n+m-1	$P_{i,j}^{obs}$	nm
	S_j	m		
	M	1		
TOTAL		2(n+m)		nm+n

3.3.1 Method of parameter estimation

Pope and Shepherd (1982) propose that the F_i and S_j parameters and the fundamental $P_{i,j}$ parameters be estimated by two separate least squares minimisations. It is assumed that some reasonable estimate of M, the natural mortality constant, can be made independently.

For the fishing mortality parameters, F_i and S_j, the least squares objective function suggested by Pope and Shepherd (1982) is

$$\sum_{i=1}^{n-1} \sum_{j=1}^{m-1} (1/D_{i,j}^{obs}-1/D_{i,j})^2 \qquad \ldots\ 3.21$$

$$D_{i,j}^{obs}=P_{i,j}^{obs}C_i^{obs}/[P_{i+1,j+1}^{obs}C_{i+1}^{obs}] \qquad \ldots\ 3.22$$

and $D_{i,j}$ is calculated from F_i and S_j according to the catch equations. Note that when $D_{i,j}$ is expressed using the catch equations

$$D_{i,j}=\exp(F_{i+1}S_{j+1}+M)F_iS_j(F_{i+1}S_{j+1}+M)(1-\exp(F_iS_j-M))/[\ \ ]$$

$$[\ \ ]=F_{i+1}S_{j+1}(F_iS_j+M)(1-\exp(-F_{i+1}S_{j+1}-M)) \qquad \ldots\ 3.23$$

it has only F_i, S_j, F_{i+1}, S_{j+1} as parameters, and as has been mentioned, M is regarded as a constant value for which a reasonable estimate is input.

The main justification for this particular objective function is its mathematical convenience. It is independent of the fundamental $P_{i,j}$ parameters, so that the F_i and S_j parameters can be obtained without considering the parameters $P_{i,j}$ (these are obtained in a separate calculation). The objective function in question leads to a convenient iterative algorithm which produces positive definite estimates of F_i and S_j (Pope and Shepherd, 1982). Their computational algorithm requires, however, that the last year value of F_i and some S_j value (F_n and S_a respectively) be specified, provided that the S_j's are normalised. By trial and error F_n and S_a are chosen to minimise the least squares sum of residuals (the objective function, expression 3.21, which is shown to be very insensitive to F_n and S_a, as is discussed later on).

The fundamental $P_{i,j}$'s, the n+m-1 $P_{p,k}$'s, are estimated by minimising the sum

$$\sum_{p=b}^{b+m} (P_{p,p-b+1}^{obs}\ {}^{n}C_{p}^{obs} - P_{p,p-b+1}\ {}^{n}C_{p}^{obs})^2 \qquad \ldots\ 3.24$$

for each cohort in turn (i.e. along each diagonal in the catch - age matrix), where b is the cohort birth date. An exact analytical solution to this can be found by differentiating the sum for each cohort with respect to its fundamental $P_{i,j}$. Pope and Shepherd (1982) choose these to be the first $P_{i,j}$ of the diagonal or cohort, the $P_{i,1}$'s and $P_{1,j}$'s, but any of the $P_{i,j}$'s down a diagonal can be used.

From the estimates of F_i, S_j and $P_{i,j}$, a matrix $N_{i,j}$, the number of fish of age j at the beginning of each interval i, can be calculated using the catch equations

$$N_{ij}=P_{i,j}^{fit}\ {}^{n}C_{i}^{obs}(F_{i,j}+M)/[F_{i,j}(1-\exp(-F_{i,j}-M))] \qquad \ldots\ 3.25$$

Together with fish masses at age, the $N_{i,j}$ provide estimates of stock biomass at the beginning of each year, B_i, by using equation 3.13, as required.

Pope and Shepherd find that even for quite small deviations from input data that fits the model exactly, the least squares objective function (3.21) becomes very flat, so that it is very difficult to fix F_n and S_a precisely. It is on this basis that they conclude that catch mass and age structure information is insufficient for estimating stock abundance trends, and that additional information about the fishery is required to determine F_n and S_a precisely.

If this additional information can be incorporated easily, then the overall VPA technique under consideration has much to recommend its use. First of all, the independence of the objective function for the F_i and S_j parameters from that of the $P_{i,j}$ parameters permits a quick estimate of the relative catchability of different age classes, the quantities S_j, to be made. The existence of a reliable convergent algorithm for estimating fishing mortalities and an explicit formula for calculating the fundamental $P_{i,j}$'s simplifies the minimisation procedure substantially, so that a microcomputer can usually complete the calculations within a reasonable period of time.

However, given the accessibility of computer packages on mainframe computers that can solve many parameter problems with relative ease, methods which consider weighting of the various input data in a more appropriate manner can be applied. In this regard it can be shown that the two objective functions used by Pope and Shepherd ignore reasonable assumptions that should be made about the error structure of the input information. For example, the homoscedasticity assumption for $1/D_{i,j}^{obs}$ (this is equivalent to $P_{i,j}^{obs}C_i^{obs}/(P_{i+1,j+1}^{obs}C_{i+1}^{obs})$) gives equal weight to age information regardless of the contribution that fish of different ages make to the catch.

Instead, one expects the coefficient of variation (C.V.) of

$P_{i,j}^{obs}$, the observed proportion by number of different aged fish in the catch, to increase as the true proportion by number of fish in the catch, $P_{i,j}$, decreases. This follows directly if one makes some simple assumptions about the error structure of $P_{i,j}^{obs}$. For example if Q fish are randomly sampled each year for aging, the coefficient of variation (C.V.) for $q_{i,j}^{obs}$, the number of fish thought to be aged j is

$$C.V.=[(1-P_{i,j})/(QP_{i,j})]^{0.5}. \qquad \ldots\ 3.26$$

This increases as $P_{i,j}$, the true proportion, decreases. Likewise, when an age length key is used, it can be shown that the coefficient of variation of $P_{i,j}^{obs}$ increases as $P_{i,j}$ decreases (see Chapter 4).

3.3.2 Weighting

The practical application of the method of Pope and Shepherd (1982) frequently requires that certain catch information, namely that for ages contributing small amounts to the catch but large residuals to the residual sum of squares, be discarded because of the destabilising effect these residuals have on the minimisation algorithm (J. Pope, pers. comm.). These are usually the very young or very old age classes. This in itself is a manifestation of the large variance associated with age classes which are infrequent in the catch.

However, instead of discarding this information it may be preferable to weight it on the basis of reasonable assumptions about its error structure. The objective functions devised by Pope and Shepherd do not provide an inbuilt objective method of weighting the catch information. This is more conveniently achieved by directly maximising the likelihood function, which has inbuilt weighting, rather than attempting an approximation to the likelihood function such as the residual sum of squares used by Pope and Shepherd. The method developed by Fournier and Archibald (1982) attempts to do this by constructing a single likelihood function for all the parameters, and can include some of the weighting constants as potentially estimable parameters.

3.4 A MAXIMUM LIKELIHOOD APPROACH TO VPA

In contrast to the previous method, Fournier and Archibald (1982) regard the true number of fish caught, C_i, as parameters that require to be estimated. The estimated numbers caught, C_i^{obs}, are thus random variables.

Fournier and Archibald (1982) have suggested that the selectivities, S_j, can be represented by a number of parameters less than m, the number of age classes. They propose that a 2 parameter function be used for this. An asymptotic relationship, different to their proposed function, has been used in calculations in subsequent chapters:

$$S_j=aj/(b+j) \qquad \ldots\ 3.27$$

but is modified in the case of hake (caught by trawlers) to include an inflexion in the catchability versus age curve and time dependence in the point of inflexion:

$$S_j = 1/[1+\exp(-b(j-j_o))] \qquad \ldots\ 3.28$$

where j_o is (linearly) related to i (see Chapter 8). It is not essential that the selectivities be described by only a few parameters in this way, and the full set of m selectivities can all be treated as parameters as done by Pope and Shepherd. The fishing mortalities are now defined by only n+1 unique parameters, since the parameter a in the selectivity function can be absorbed in the F_i parameters

$$
\begin{aligned}
F_{i,j} &= F_i aj/(b+j) \\
&= F_i^* j/(b+j) \qquad \ldots\ 3.29
\end{aligned}
$$

As before, it is assumed that natural mortality is constant for all ages and years. The parameters are thus the n F_i^* parameters and b. Together with M and the n+m-1 parameters associated with $P_{i,j}$ and C_i, the total number of parameters are thus 2n+m+1, m-1 less than Pope and Shepherd's (1982) model (the reduction arising because selectivities are now represented by a functional form). These are summarised in Table 3.2. For this model solutions exist if the available data exceeds the number of parameters i.e. if nm+n > 2n+m+1. For

m=1 no solution is possible
m=2 a solution exists for n>3
m=3 a solution exists for n>2
m>3 a solution exists for n> or = 1

and as with Pope and Shepherd's model solutions are in principle potentially attainable, for typical fishery data banks.

3.4.1 <u>The likelihood function</u>

There are two terms in the likelihood function, one for each set of observations. These are the set of $[q_{i,j}^{obs}$: i=1...n, j=1...m$]$, the number of fish which were sampled and judged to be of age j, and the set of $[C_i^{obs}$: i:1...n$]$. C_i is positive definite and this is conveniently assured by assuming that C_i^{obs} is log-normally distributed about C_i, with mean $\ln C_i$ and variance σ^2 (i.e. $\ln C_i^{obs} \approx N(\ln C_i, \sigma^2)$). The skewness in the lognormal distribution results in C_i^{obs} on average overestimating C_i ,

Table 3.2. A breakdown of the parameters which must be estimated and the input data for the Fournier and Archibald (1982) Virtual Population Analysis model (selectivities parameterised).

	PARAMETERS	QUANTITY	INPUT DATA	QUANTITY
	$P_{q,k}C_q$	n+m-1	C_i^{obs}	n
	F_i	n	$P_{i,j}^{obs}$	nm
	b	1		
	M	1		
TOTAL		2n+m+1		nm+n

since for a random variable X, with

$\ln X \approx N(\xi, \sigma^2)$

the expected value of X is

$E(X) = \exp(\xi+\sigma^2/2)$

and $\mathrm{var}(X) = \exp(2\xi+\sigma^2).(\exp(\sigma^2)-1)$

so that the variance of C_i^{obs} is proportional to C_i^2 and $E[C_i^{obs}]$ is $\exp(\sigma^2/2)$ times bigger that the true catch, C_i. Arguing the reverse, if C_i^{obs} is well represented by a symmetrical distribution, then the log transformation employed (under assumptions of lognormality) will tend to produce a negative bias in estimates of C_i. However, it is reasonable to assume here that the coefficient of variation of the catch estimate is small, say with a C.V. of about 5% (since landings in the South African pelagic fishery are individually weighed), so that σ^2 is probably in the region of 0.0025. The bias factor is therefore exp(0.0025/2), which is about 1.0013, and so the bias is likely to be minimal. The error assumption for C_i^{obs} has introduced a further parameter, σ, and this must somehow be estimated.

Fournier and Archibald (1983) assume that Q fish are sampled each year for aging purposes. The compartmentalisation of this sample into different age classes provides the values $q_{i,j}^{obs}$, the number of fish in the sample which are assessed to be of age j in year i. $P_{i,j}^{obs}$ is just the normalised vector of $q_{i,j}^{obs}$ for each year

$$P_{i,j}^{obs} = q_{i,j}^{obs}/Q \qquad \ldots\ 3.30$$

Thus the $P_{i,j}^{obs}$ are multinomially distributed about the true age proportions, the $P_{i,j}$'s. If one ignores constant terms, and for the situation sketched so far, with M and σ^2 considered as given, the likelihood function is;

$$\prod_{i}\prod_{j} P_{i,j}^{q_{i,j}^{obs}} \quad \prod_{i} \exp[-(\ln C_i^{obs}-\ln C_i)^2/(2\sigma^2)]/(2\pi\sigma)^{0.5} \qquad \ldots\ 3.31$$

and the log-likelihood function is given by

$$\sum_{i=1}^{n}\sum_{j=1}^{m} q_{i,j}^{obs}\ln P_{i,j} - \sum_{i=1}^{n} [\ln C_i^{obs}-\ln C_i]/(2\sigma^2) \qquad \ldots\ 3.32$$

for σ regarded as given, and ignoring other constant terms.

Maximisation of 3.32 has still to be achieved subject to the constraints represented by equations 3.16-3.19 and 3.20, viz. the catch equation relationships for cohorts and the normalisation requirement for $P_{i,j}$. It is convenient to make the following notational changes (Fournier and Archibald, 1982);

$\beta_{i,j}=\ln(P_{i,j}C_i/C_i^{obs})$

$\gamma_i=\ln C_i^{obs}$

$\Phi_{i,j}=\ln D_{i,j}$ expressed in terms of F_i and S_j

since, if equation 3.16 is written as

$$\ln D_{i,j} = \ln P_{i,j}C_i/C_i^{obs}+\ln C_i^{obs}-\ln P_{i+1,j+1}C_{i+1}/C_{i+1}^{obs}-\ln C_{i+1}^{obs} \qquad \ldots\ 3.33$$

then it is given in terms of β's, γ's and Φ's, by

$$\beta_{i,j}+\gamma_i-\beta_{i+1,j+1}-\gamma_{i+1}=\Phi_{i,j} \qquad \ldots\ 3.34$$

This equation is identical to the previous diagonal relationships (equations 3.32 and 3.16). However, the advantage of using the β, γ and Φ variables is that the parameters $P_{i,j}$ and C_i need no longer be considered separately, since these are substituted in the likelihood function (Fournier and Archibald, 1982) by the set of β parameters only, through use of the following relationships which follow from the constraint $\sum_{j=1}^{m} P_{i,j} = 1$:

$$C_i = C_i^{obs} \sum_{j=1}^{m} (\exp(\beta_{i,j})) \qquad \ldots\ 3.35$$

$$P_{i,j} = \exp(\beta_{i,j}) / \sum_{j=1}^{m} (\exp(\beta_{i,j})) \qquad \ldots\ 3.36$$

All the β's of a particular diagonal can be calculated given an estimate of the fundamental β for that cohort, and estimates of F_i and S_j to the necessary fishing mortalities, by using equation 3.34. It makes sense to choose as the fundamental q,k location for the β's that point for which $P_{i,j}^{obs}$ is largest along a diagonal, since this will accelerate the convergence to the maximum.

Ignoring constant terms, the log-likelihood function can now be written simply as

$$\sum_{i=1}^{n} \sum_{j=1}^{m} q_{i,j}^{obs}[\beta_{i,j} - \ln[\sum_{j=1}^{m} \exp(\beta_{i,j})]] - \sum_{i=1}^{n} \ln^2[(\sum_{j=1}^{m} \exp(\beta_{i,j}))]/(2\sigma^2)$$

$$\ldots\ 3.37$$

This expression permits the constraints 3.16 - 3.20 to be met in a straightforward manner (since only $\beta_{i,j}$ parameters are considered), rather than being bound by the awkward set of rules which would apply in attempting to fit $P_{i,j}$ and C_i as separate parameters (note that neither $P_{i,j}$ or C_i occur explicitly in 3.37).

The model has some important deficiencies. For one, it has been assumed that the annual observed catches-at-age, the products $P_{i,j}^{obs}$ $^{n}C_{i}^{obs}$, deviate from a fixed selectivity model only by way of random sampling error. This requires that the actual catches, the terms $P_{i,j}C_{i}$ conform exactly to the fixed selectivity function. It is inevitable though that some model prediction error exists, so that the actual catches, the products $P_{i,j}C_{i}$, do not necessarily conform exactly to the fixed selectivity model.

However, to account for this in the estimation process, sampling error and model prediction error need to be distinguished from each other. This leads to a further nm parameters, which specify the deviations from the fixed selectivity model, and these cannot be estimated , since there are too few data to allow a maximisation of the objective function (or even a deterministic solution).

In terms of the present formulation, one way of illustrating the effect of increasing amounts of model error may be to reduce all the $q_{i,j}^{obs}$ quantities by a factor, which will increase the variance of the $P_{i,j}^{obs}$. This adjustment would only be reasonable if the additional model errors follow a similar error structure to the sampling errors on the age structures, viz. increasing binomially as $P_{i,j}$ decreases.

Note that the parameter σ^2, the variance of $\ln C_{i}^{obs}$, has been overlooked in the final maximisation process (Table 3.2). At this stage it has been assumed that σ^2 can be reliably estimated independently. In practice this is seldom possible, although one expects that the amount of fish caught each year in the South African pelagic fishery is fairly precisely determined, say with a C.V. of less than some 5% so that $\sigma^2 < 0.0025$. Attempts at including σ^2, and similar parameters required for further developments of the likelihood function formulation of the problem, in the parameter set to be estimated by maximising the likelihood function will be made later. First however, it is worthwhile investigating the behaviour of solutions corresponding to the maximum of the objective function for some typical simulated data, for the case where σ^2 and M are taken to be estimated independently.

4. SIMULATIONS OF THE BEHAVIOUR OF SOLUTIONS MAXIMISING THE VPA LIKELIHOOD FUNCTION

4.1 INTRODUCTION

This chapter describes the results of fitting the generalised VPA model described in Chapter 3 to simulated fisheries data. The fitting procedure is tested and developed to a point at which there is confidence that the iterations have converged to the true maximum of the likelihood function. Initially, the aim is to reproduce the findings of Pope and Shepherd (1982) and others (Gray, 1977; Doubleday, 1976) regarding the inadequacy of catch-at-age information alone for stock assessment. Thereafter, the effect on estimator variance of including additional relative abundance information is shown. Imprecision in estimates of abundance is estimated using a resampling procedure closely allied to the technique known as bootstrapping (Efron, 1981) and this procedure is validated to a limited extent in a series of simulations.

The simulation study has been specifically designed to investigate VPA as applied to the South African anchovy. Throughout the simulation study, a single true parameter set (i.e. one corresponding exactly to the model parameter values chosen, with zero sampling error) has been used, from which further data sets are generated by random sampling techniques (the true spawner stock abundance trend is shown in Table 4.2). Thus the conclusions that are drawn from this work are specific to this parameter set, and hence to the South African anchovy situation.

4.2 SIMULATION STUDY METHODS

The simulation of an assessment based on typical fisheries data consists of two consecutive steps. In the first, the underlying true parameters, the number of fish caught annually, C_i, and the age structure of the catch by number, $P_{i,j}$, are calculated according to predefined recruitment and mortality model parameters (section 4.2.1 below). These are then corrupted with sampling noise to obtain the simulated data sets C_i^{obs} and $P_{i,j}^{obs}$ (section 4.2.2 below).

4.2.1 Generation of catch mass and age structure data sets

The study was aimed at an idealised anchovy stock with a natural mortality of $M = 1\ yr^{-1}$ (thought to be typical for anchovy; Butterworth, 1983). If Z is the average mortality experienced, and only the first k age classes are considered in the VPA , then the fraction (by number) which the remaining fish form of the total population number (including all age classes) is $\exp(-Zk)$ (ignoring age-dependent effects). The first 5 anchovy age classes have been considered here, and assuming a Z value of 1.3, $\exp(-Zk)$ is equivalent to 0.15%, so that the remaining age classes can be safely ignored. C_i and $P_{i,j}$ were generated exactly using the 2 parameter selectivity function described previously (equation 3.27) taking b = 3 yrs, and the F_i trend and selectivities at age shown in Table 4.1. Recruitment was modelled as being log-normally distributed about a Beverton and Holt spawner-recruit relationship:

$$r_i = [\alpha y_i/(\beta+y_i)]*\exp(u_i) \qquad \ldots\ 4.1$$

r_i = recruitment

y_i = spawning biomass, calculated with a 1 year age to maturity

Table 4.1. The F_i's (years increasing from the start) and S_j's (for ages 0 to 4) used for the model anchovy population.

F_i's

1964	0.3	1969	0.4	1974	0.6
1965	0.6	1970	0.3	1975	0.3
1966	0.5	1971	0.2	1976	0.2
1967	0.4	1972	0.5	1977	0.5
1968	0.7	1973	0.7	1978	0.4

Age range (years)	S_j
0-1	0.25
1-2	0.40
2-3	0.50
3-4	0.57
4-5	0.63

Table 4.2. The true spawner biomass trajectory for the model anchovy population over the period investigated (units '000 tonnes).

1964	2677	1969	970	1974	831
1965	2189	1970	947	1975	736
1966	1577	1971	1011	1976	804
1967	1322	1972	1128	1977	939
1968	1233	1973	1030	1978	884

$u_i \approx N(0, 0.09)$ (i.e. normally distributed with mean 0 and standard deviation 0.3)

α and β are constants with the following values;

α = 150 (recruits)

β = 1640 (in '000 tonnes)

4.2.2 Generation of C_i^{obs} and $P_{i,j}^{obs}$

C_i^{obs} and $P_{i,j}^{obs}$ were generated according to the various statistical models already described (Chapter 3). The $q_{i,j}^{obs}$ are the number of fish sampled in year i which are judged to lie in the age interval $[j, j+1]$, said to be of age j. These were modelled as being obtained from random sampling of the catch, and subsequent aging of the whole sample, and it has been assumed that fish ages are correctly assessed. The various methods used for obtaining the data sets C_i^{obs} and $P_{i,j}^{obs}$ from C_i and $P_{i,j}$ are described in Appendix 4.1.

4.2.3 Numerical maximisation of the likelihood function

For the South African fish stocks dealt with in subsequent chapters-the anchovy, pilchard and hake stocks - the maximisation of the likelihood function for the VPA (equation 3.35) requires the estimation of a large number of parameters, usually between about 20 and 80. The first derivatives of the likelihood function with respect to these parameters are all complicated non-linear functions and thus a straightforward solution is not possible. It is also difficult to write down analytic expressions for the first derivatives due to their complexity, and even the calculation of the first derivatives using these analytical expressions would be expensive in terms of computer time. Thus it was decided that the most practical and flexible approach would be to use existing derivative-free computer packages specially designed for maximising many parameter problems.

The NAG (Numerical Algorithms Group Ltd., Oxford) routine E04JAF was chosen to perform the objective function maximisations, and this was found to be adequate (Strictly this is a minimisation routine, but maximisation is readily achieved by changing the sign of the objective function, so for convenience this text will refer to maximisation throughout). Elsewhere, first derivative packages are being used (D. Fournier, pers. commn). However, if a first derivative maximisation routine is used, the addition of further terms to the log-likelihood function requires considerable software development. On the other hand, if the derivative-free approach is used, the addition of extra terms to the log-likelihood function entails only minor software adjustments. Consequently derivative-free maximisations are far more flexible than their first derivative counterparts, and this makes them compatible with the likelihood type of objective function, into which further information can be incorporated in a very flexible way. For the research reported here, software flexibility has been a high priority, and thus derivative-free calculations have been used.

Note that EOFJAF provides the option of including constraints on the parameters which are estimated. Upper and lower bounds were imposed on all the model parameters, to avoid, for example, taking the logarithm of a negative number. Final parameter estimates were checked to ensure that they did not occur at a lower or upper bound.

The use of optimisation routines like E04JAF, particularly for the number of parameters being considered here, may require associated but independent validation. For example, some criterion for the confidence of attainment of an optimum point is required, and inbuilt predefined criteria may be inadequate for the purpose intended here. These criteria normally involve some test of how close the first derivatives are to zero. It is important that the fineness of this test is appropriate to the flatness of the objective function. This problem is particularly acute here, since the VPA likelihood functions tend to be very flat in the vicinity of their maxima.

The computer offers some flexibility in this regard, since either single or double precision variables and fitting routines can be used. Since the criteria for attainment of a maximum depend on computer precision, the fineness of the test is altered by the degree of precision used. Thus preliminary model fits, using both single and double precision variables and fitting routines, can be compared. If the results for a particular situation depend quite heavily on the degree of precision used, then the double precision result is preferred. This was often found to be the case during preliminary analyses. However, given some disagreement between single and double precision results, one is still left with the problem of validating the double precision result. This can be done by testing whether parameter estimates are sensitive to the starting values used in the optimisation routine.

The NAG routine E04JAF produces a test parameter, IFAIL, for which the highest level of confidence of attainment of a local maximum occurs for IFAIL=0. The criteria for an exit with IFAIL=0 are that (B1 and B2 and B3) or B4 is satisfied, where these conditions are given as:

B1 : $\alpha^{(k)} \|\underset{\sim}{p}^{(k)}\| < (2E^{0.5}) \; x \; (1.0 + \|x^{(k)}\|)$

B2 : $|F^{(k)}-F^{(k-1)}| < (2E)*(1.0 + |F^{(k)}|)$

B3 : $\|G_z{}^{(k)}\| < (E^{1/3}+E^{1/2})*(1.0 + |F^{(k)}|)$

B4 : $\|G_z{}^{(k)}\| < 0.01*E^{1/2}$.

$\| \;\; \|$ denotes the Euclidian norm of a vector; E is the smallest number representable on the computer such that $1.0+E>1.0$. The superscript k refers to the k th iteration in the maximisation procedure, so that $x^{(k)}$ is the k th estimate of the position of the maximum, $F^{(k)}$ is the value of the function at that point, and $G_z{}^{(k)}$ is an estimate of the first derivatives of the function with respect to the z free parameters (these are the parameters which are not bounded at that stage of the calculation). $\underset{\sim}{p}^{(k)}$ is a unit vector which denotes the direction of the incremental change in $x^{(k)}$ from values for the previous iteration. The scalar $\alpha^{(k)}$ is chosen in each iteration such that $F^{(k)}(x^{(k)}+\alpha^{(k)}\underset{\sim}{p}^{(k)})$ is maximised.

Other values of IFAIL which are important regarding attainment of a maximum position are:

IFAIL=2 There have been 400xN (the number of estimable parameters) function evaluations, yet the algorithm does not seem to be converging. The calculations can be restarted from the final point held in $x^{(k)}$. This error message may also indicate that there is no maximum.

IFAIL=3 The conditions for a maximum have not all been met but a lower point could not be found and the algorithm has failed.

IFAIL=5
IFAIL=6
IFAIL=7
IFAIL=8

There is some doubt about whether the point $x^{(k)}$ found by E04JAF is a maximum. The degree of confidence in the result decreases as IFAIL increases. Thus, when IFAIL=5 it is probable that the final $x^{(k)}$ gives a good estimate of the position of the maximum, but when IFAIL=8 it is very unlikely that the routine has found a maximum.

These inbuilt measures of confidence of attainment of a (hopefully global) optimum notwithstanding, it was clearly necessary to validate the results independently. This was done by means of a sensitivity analysis which involved testing whether parameter estimates depended on the starting values given for them in the optimisation procedure.

4.2.4 Sensitivity of parameter estimates to different starting values for the F_i parameters

The sensitivity of the worst case VPA, viz. catch-at-age data only (for a typical anchovy fishery), was investigated. Preliminary results, using 15 years of fishing and 5 age classes ($M=1\ yr^{-1}$), suggested that depending on how the objective function is scaled, the IFAIL=0 result produced by the NAG routine used may be unreliable.

For fast convergence to the optimum, the routine requires that the maximum value of the log-likelihood function lies in the range [-1,1]. This is approximately the case if the log-likelihood function is either <u>divided</u> by or <u>subtracted</u> from the sum:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} q_{i,j}^{obs} \log(P_{i,j}^{obs}) \qquad \ldots\ 4.2$$

Initially, using the scaling by <u>division</u> method, rapid convergence to an IFAIL=0 result was achieved. However parameter estimates depended on the starting parameter values used (Fig 4.1). This is partly due to the large size of the first term in equation 3.37, which swamps the contribution from the term involving the total catch number estimates.

Alternatively, using the method of <u>subtraction</u>, consistent and reliable results were achieved. In Table 4.3 the true and starting parameter values for F_i which were used are set out. For two very different starting parameter sets shown in Table 4.3, estimated abundances agreed to about 6 significant figures. The IFAIL values for these results were generally not equal to 0 but lay somewhere between 5 and 8.

Table 4.4 shows the percentage changes in the value of the objective function caused by a 10 % change in each of the model parameters in turn, for scaling the likelihood function either by subtraction or division. It appears therefore that when the objective function is scaled by subtraction the log-likelihood function is of the order of 1000 times more sensitive to its constituent parameters than when scaling by division is used. In the light of this, the dependence of the estimated position of the maximum on the starting position used when scaling by division is understandable.

The different techniques of scaling the objective function are used in conjunction with a sequential parameter estimation technique, which attempts to start the optimisation routine in the vicinity of the true maximum of the likelihood function.

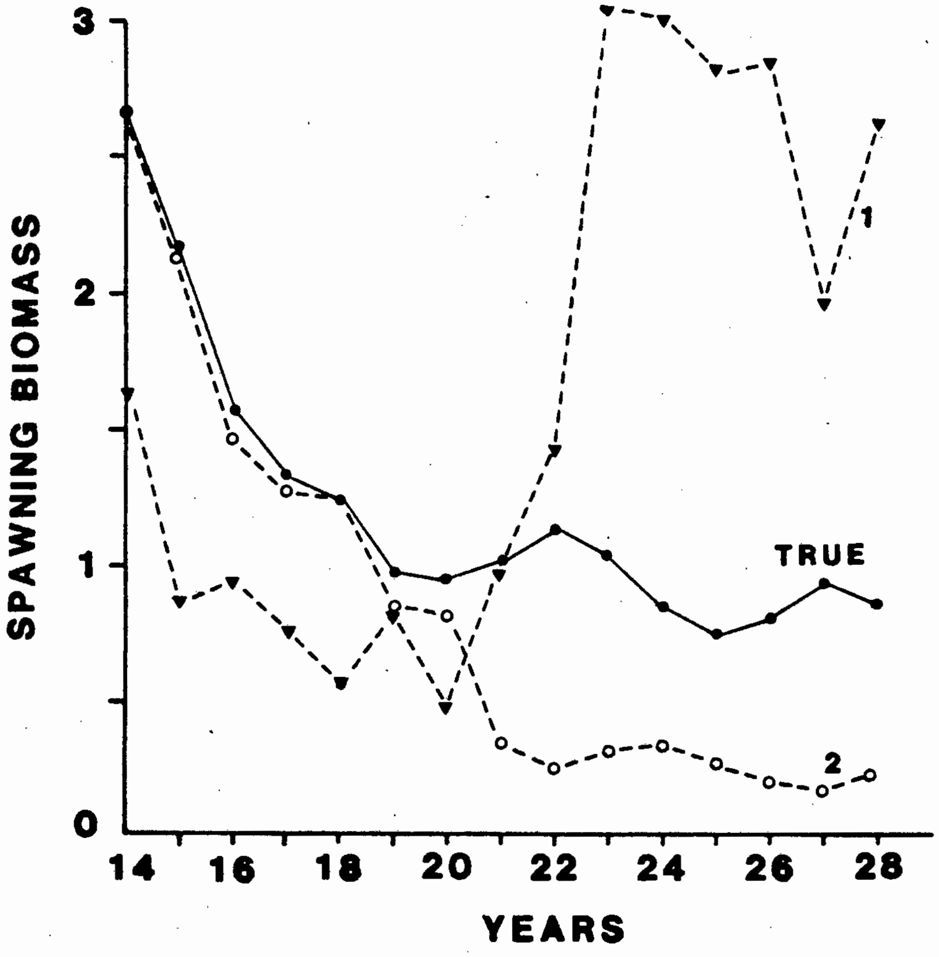


Figure 4.1. Estimated biomass trajectories which result if the objective function is scaled by the division method, and for different starting parameter value sets for F_i:

1. $F_i = 0.5$, $i<=19$; $F_i = 0.1$, $i>19$
2. $F_i = 0.5$, $i<=19$; $F_i = 2$, $i>19$

Table 4.3. True and estimated spawning biomass trajectories for a model anchovy population which is fished. The objective function was scaled by subtraction (see text), and the estimated biomass trajectory (for a single set of sampled data) is insensitive to 2 extreme starting parameter value sets (F_i', F_i'') used for F_i. The estimates are obtained by VPA using only catch at age data, which were polluted by random sampling noise.

Year	B_{true}	$B_{estimated}$	F_i'	F_i''	F_i^{true}
1964	2677	2704	0.5	0.5	0.3
1965	2189	2169	0.5	0.5	0.6
1966	1577	1579	0.5	0.5	0.5
1967	1322	1340	0.5	0.5	0.4
1968	1233	1302	0.5	0.5	0.7
1969	970	1045	0.5	0.5	0.4
1970	947	1075	0.1	4	0.3
1971	1011	1131	0.1	4	0.2
1972	1128	1237	0.1	4	0.5
1973	1030	1017	0.1	4	0.7
1974	831	745	0.1	4	0.6
1975	736	511	0.1	4	0.3
1976	804	503	0.1	4	0.2
1977	939	530	0.1	4	0.5
1978	884	272	0.1	4	0.4

Table 4.4. The percentage decrease in the value of the likelihood function for a 10% change in each of the model parameters away from the optimum point, for the model excluding effort information, and scaling either by subtraction or division.

Cohort birth date	ß parameter value at optimum (n+m-1 such parameters)	SUBTRACTION	DIVISION
		Percentage	Decrease
1960	-.66	8.3	0.011
1961	-.13	15.7	0.018
1962	-.21	18.2	0.022
1963	-.30	14.0	0.017
1964	-.39	7.4	0.0092
1965	-.56	6.6	0.0085
1966	-.55	5.8	0.0070
1967	-.56	5.8	0.0066
1968	-.50	5.0	0.0064
1969	-.52	5.8	0.0059
1970	-.52	5.0	0.0069
1971	-.56	5.8	0.0069
1972	-.49	4.1	0.0075
1973	-.50	5.0	0.0061
1974	-.46	4.1	0.0070
1975	-.51	5.0	0.0060
1976	-.53	5.0	0.0055
1977	-.52	4.1	0.0058
1978	-.52	2.5	0.0028
Year	F_i parameter (n such parameters)		
1964	.32	15.7	0.019
1965	.66	18.2	0.022
1966	.50	17.4	0.021
1967	.41	16.5	0.020
1968	.69	18.2	0.022
1969	.33	17.4	0.021
1970	.30	16.5	0.020
1971	.19	15.7	0.020
1972	.54	16.5	0.020
1973	.71	17.4	0.022
1974	.62	18.2	0.021
1975	.28	16.5	0.019
1976	.19	15.7	0.019
1977	.58	13.2	0.015
1978	.48	5.8	0.0069
	b parameter		
	2.9	26.4	0.034

4.2.5 Sequential parameter estimation

In the VPA, reasonable starting values for the ß parameters are given by:

$ß_{i,j}=\ln P_{i,j}^{obs}$

(note: this assumes that the best fit value of C_i is close to C_i^{obs})

where $ß_{i,j}$ has an upper bound of close to zero, and is not bounded below. For F_i and b, the fishing mortality parameters, no obvious starting points exist. The starting values chosen were:

$F_i = 0.5$

$b = m/2$

with both F_i and b having zero as their lower bounds. Some preliminary model fits were performed in this way, with the parameter set being the full number of 2n+m parameters. The correct values of M and σ^2 were given, and it was assumed that 1000 fish were sampled each year for ageing, which is shown in Chapter 5 to produce a similar variance on $P_{i,j}^{obs}$ as does the real anchovy age-length key, for typical $P_{i,j}$.

It was found that the routine E04JAF often failed to reach an optimum point with these as starting parameter values, when the objective function was scaled by the method of division described earlier. However no attempts at fitting the model using this approach were made with the subtraction method of scaling. The problem may therefore also have been related to the scaling procedure used.

A more reliably convergent approach was devised by recognising that apart from natural mortality and the variance on the catch estimate, there are really two types of parameters in the model. These are the n+m-1 fundamental $\beta_{i,j}$'s and the n+1 parameters associated with fishing mortalities. Two runs with E04JAF were carried out as a prelude to obtain reliable starting values for all the parameters. In these two preliminary runs, model fits were obtained first with respect to the n+1 fishing mortality parameters only, and then with respect to the n+m-1 fundamental $\beta_{i,j}$ parameters only. The parameter estimates thus obtained were then used as a starting point for the final maximisation involving all 2n+m parameters. This procedure was extremely successful in attaining fast convergence and reliable parameter estimates.

Software testing was performed for the case of no sampling error on both the catch number and the age structure input data, and the correctness of the numerical optimisation routine was confirmed when parameter estimates yielded the true biomass estimates to within about 6 significant figures.

4.2.6 Bias and variance of abundance estimates

Previous investigations into the errors in estimates of cohort size obtained from Virtual Population analyses have been primarily concerned with the propogation of errors after the terminal fishing mortalities have already been fixed on some basis. The literature in this regard is reviewed by Sampson (1984), and the following summary is based on his discussion.

Pope (1972) proposed an approximate solution to the catch equations which he used to demonstrate the pattern of errors in fishing mortalities resulting from errors in either the terminal fishing mortality estimate or catch-at-age estimates (he considered a single cohort, rather than an assemblage of different year classes). The results showed that estimates of cohort size ($N_{i,j}$) and fishing mortality ($F_{i,j}$) obtained in backcalculations were relatively insensitive to errors in the terminal fishing mortality estimate. Relative errors in $N_{i,j}$ and $F_{i,j}$ were shown to decrease with decreasing i and j, and this result has been taken as an implicit justification of the general application of VPA techniques.

Pope produced additional results concerned with the propogation of errors from inaccurate catch-at-age estimates through the calculations. For errors in the catch-at-age estimates, $C_{i,j}$, he derived a relationship between C.V.'s on the fishing mortality estimates, the $F_{i,j}$'s, and C.V.'s on the $C_{i,j}$ estimates. The results indicate that the C.V.'s on the $F_{i,j}$'s are between about 80% and 120% of the C.V.'s on the estimates of $C_{i,j}$. Further results indicate that the C.V.'s on the estimates of cohort size, the $N_{i,j}$'s, are between about 30% and 100% of the C.V.'s on the estimates of $C_{i,j}$.

Pope did not consider the case of simultaneous errors in terminal fishing mortality, catch-at-age and natural mortality estimates. Agger et al (1971) had earlier investigated the effects of simultaneous errors in the terminal fishing mortality estimate and the natural mortality estimate. For the cases they studied they concluded that estimates of $F_{i,j}$, the fishing mortalities, would be positively biassed by about 25% for M in the range 0.1 to 0.3 yrs^{-1}. Ulltang (1977) studied other sources of error, including migration, errors in natural mortality values, and the uneven temporal distribution of annual catches. His results suggest that for moderate random fluctuations in natural mortality, errors in $N_{i,j}$ and $F_{i,j}$ are slight. Problems regarding the uneven temporal distribution of annual catches have also been studied by Sims (1982). Results reported by Sims suggest that the relative error in stock size when a VPA is carried out with the supposition of catches evenly distributed throughout the year are not severe unless the natural mortality is large and/or the fishery is heavily exploited.

Sampson (1984) has followed the methodology of Pope (1972) and others (Agger et al, 1971; Ulltang, 1977) to determine the errors in $F_{i,j}$ and $N_{i,j}$ estimates due to simultaneous errors in terminal fishing mortality, catch-at-age and natural mortality values. He defines clearly the conditions under which the magnitude of the relative errors in successive cohort size estimates do not increase for decreasing cohort ages, which qualifies Pope's result. He also extends Ulltang's (1977) findings on the significance of random fluctuations in natural mortality.

In view of the stated practical aims of this thesis, there are two deficiences in the research summarised above concerning VPA estimator precision. First, the terminal fishing mortality estimates, even though at times incorrect, have been regarded as given in the analyses. However in reality some knowledge about the terminal fishing mortality variance is necessary to express VPA biomass estimate precision with any meaning. This uncertainty in the terminal fishing mortalities would normally result from their estimation from the available catch-at-age data, something which is not considered in the research reviewed in the preceeding paragraph. To achieve this, the terminal fishing mortalities have to be set up as estimable parameters. A number of techniques have been suggested (Pope and Shepherd, 1982; Gray, 1977; Doubleday, 1976; Fournier and Archibald, 1982). Amongst these, abundance variance estimates are quoted by Pope and Shepherd (1982), but any of the other methods could be extended to compute variances. What is most striking, is that the precision of the terminal fishing mortality estimate is one of the most significant components of each year's overall abundance estimate precision.

Finally, there is no evidence in the literature that any general statement on VPA precision is emerging. Attempts at producing general analytical results have not been successful (Sampson, 1984), and indiscriminate brute force techniques such as Monte Carlo experiments have all too soon to be invoked. With this background, and with a practical end result in mind, subsequent extensive use of Monte Carlo experiments and resampling procedures is made in this thesis for the purpose of estimating variances.

In order to compute variances, a large number of data sets were generated and fitted to the VPA model, the exact number depending on the computer time required for fitting data sets with different levels of sample error. From the resulting sample of abundance estimates, variances were evaluated using the standard equation (see Stage A in Fig. 4.2, particularly in relation to subsequent discussions relating to Stages B and C set out in the same figure):

$$\mathrm{var}(B_i) = \sum_{k=1}^{kt} (B_{i,k} - \sum_{k=1}^{kt} B_{i,k}/kt)^2/(kt-1) \qquad \ldots\ 4.3$$

where $B_{i,k}$ is the abundance estimate for year i from the k th data set, and kt is the total number of data sets processed. Bias was taken to be the difference between the true known abundance and the mean of the set of model fitted abundance estimates.

4.3 PRELIMINARY RESULTS

4.3.1 Without additional relative abundance information

Preliminary studies were carried out using 15 years of catch information and 5 age classes, and including a minimal amount of sampling error, viz. a 5% C.V. on the catch number estimate and randomly sampling 1000 fish each year for aging. For this situation, estimates of the coefficient of variation and the bias are shown in Table 4.5.

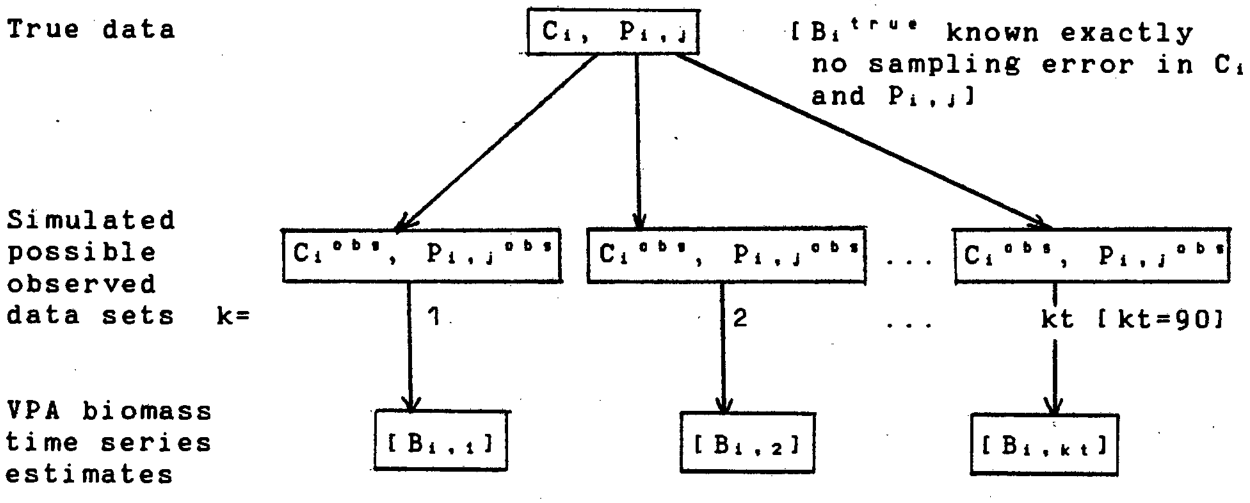
As expected from other work (Pope, 1972; Pope and Shepherd, 1982), the C.V. is highest for the most recent year, and is in fact in excess of 100 %. These results show therefore that age structure and catch mass information typical of an anchovy fishery does not provide any useful estimate of recent stock size (even given a known form for selectivity as a function of age).

Note further that there is a large positive bias in biomass estimates for more recent years, and this occurs for estimates with a large variance, notably in the more recent years. A positive bias might be expected if the biomass estimators have positively skew distributions, with the expected biomass (the mean) being larger than the modal biomass by an amount dependent on the amount of skewness, which increases with the variance. Estimates of the mode for log-normally distributed biomass estimates are given approximately by the geometric mean (imprecisely for a small sample):

$$B_{i,mode}=[\prod_{k=1}^{kt} B_{i,k}]^{1/kt} \qquad \ldots\ 4.4$$

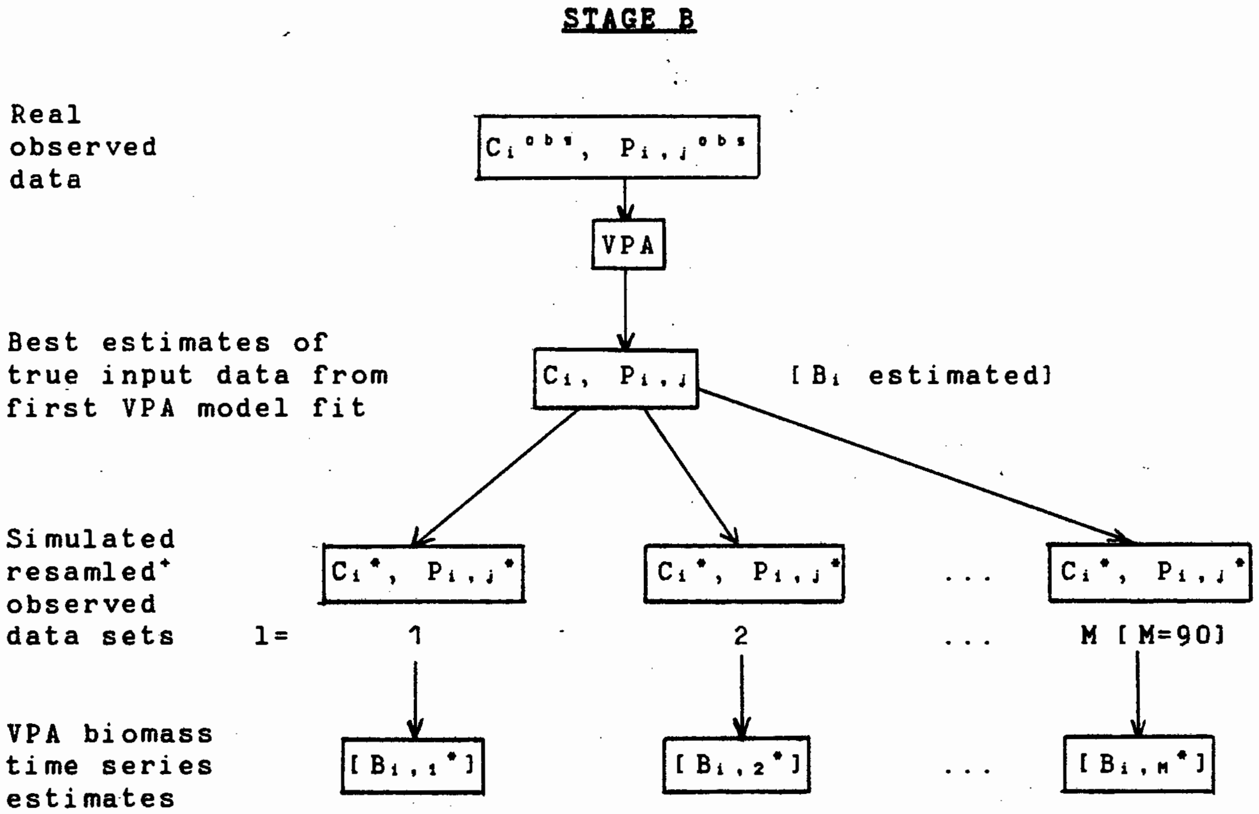
Figure 4.2 A diagrammatic representation of the procedure used to calculate the mean, variance and bias of abundance estimates resulting from the VPA (Stage A). Stage B shows how variances are calculated in practice, when the true catch at age data, and the true biomass trajectory are unknown (this procedure is utilised in Chapters 5, 7 and 8). Stage C shows how in principle the reliability (precision and accuracy) of results produced in Stage B may be assessed. This stage was found to be unrealistically expensive in terms of computer core time usage, and was not applied in full. Instead, the variances obtained in Stage B are used to infer the precision of the abundances obtained there. (Effort data are not shown here, however they are used in the same schematic way as the other data which is represented.)

STAGE A



Mean of estimates of true biomass	$\bar{B}_i = \sum_{k=1}^{kt} B_{i,k}/kt$
Variance for biomass	$\sigma_{B,i}^2 = \sum_{k=1}^{kt} (B_{i,k} - \bar{B}_i)^2/(kt-1)$
Bias of biomass estimate	$\text{Bias}=\bar{B}_i - B_i^{true}$

Figure 4.2 continued ...



Mean of resampled⁺ biomass estimates

$$\bar{B}_i^* = \sum_{l=1}^{M} B_{i,l}^*/M$$

Resampled⁺ variance estimate for biomass

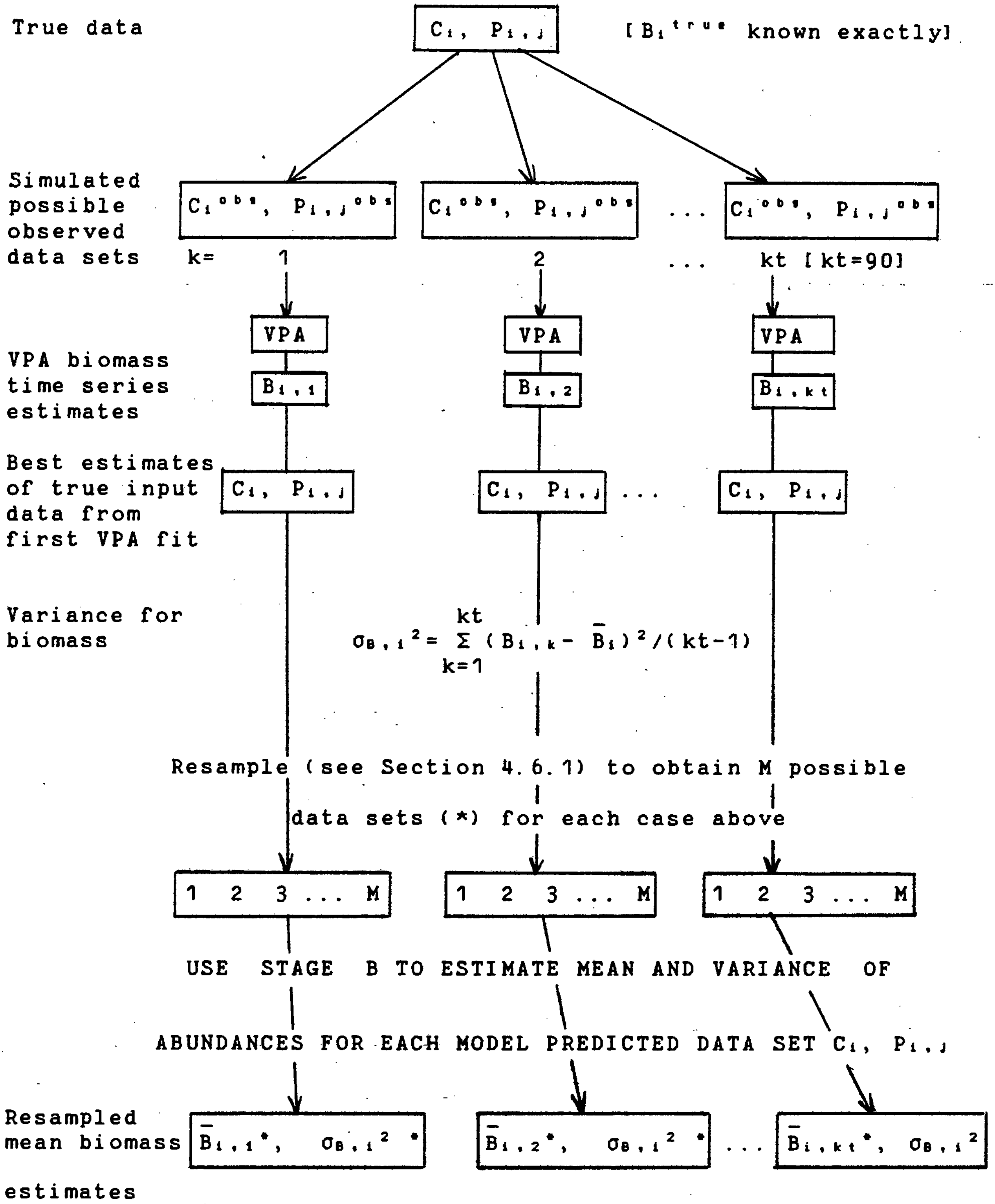
$$\sigma^2_{B,i}{}^* = \sum_{l=1}^{M} (B_{i,l}^* - \bar{B}_i^*)^2/(M-1)$$

[The true biomass will be known in a simulation only, so that in practice, with real data, the bias is unknown. However, bias could be estimated crudely from additional simulations by assuming that the best fitted catch at age data obtained from a Stage B type of approach is the true data set, and then following the calculation sequence in Stage A.]

\+ - see Section 4.6.1 for a description of the resampling procedure.

Figure 4.2 continued ...

STAGE C



Mean variance of resampled VPA biomass estimates $\quad \bar{\sigma}_{B,i}^{2\,*} = \sum_{k=1}^{kt} \sigma_B^{2\,*} / kt$

Bias of abundance variance estimates $= (\bar{\sigma}_{B,i})^{2\,*} - (\sigma_{B,i})^2$

Table 4.5. Monte Carlo estimates (kt=30) of the arithmetic mean, geometric mean and variance of VPA abundance estimates using only catch mass and age structure information. Coefficients of variation and biases (of the arithmetic mean) are estimated using the Monte Carlo mean biomass (Stage A, Fig. 4.2). The following (true) parameter values were given and fixed in the fitting procedure: $M=1\ yr^{-1}$, $\sigma=0.05$ and $Q=1000$.

YEAR	TRUE BIOMASS	STAGE A MEAN (ARITH.)	STAGE A MEAN (GEOM.)	STAGE A C.V.(%)	STAGE A BIAS (%)
1964	2677	2657	2650	8	<1
1965	2189	2178	2172	7	<1
1966	1577	1571	1567	8	<1
1967	1322	1336	1331	8	<1
1968	1233	1249	1246	7	<1
1969	970	1007	1003	8	<1
1970	947	1002	998	9	6
1971	1011	1075	1069	11	6
1972	1128	1203	1193	14	7
1973	1030	1106	1088	21	7
1974	831	963	915	38	16
1975	736	954	842	62	30
1976	804	1113	922	78	38
1977	939	1361	1068	88	45
1978	884	1498	946	117	69

Estimates of the mode obtained from this formula are also shown on Table 4.5, and these are closer to the true biomasses than are the means. Clearly the geometric mean is closer to the true biomass. However, this aspect is just of passing interest, because the large variance of the biomass estimates renders them practically useless anyway. The estimation of the mode of the probability density function for the abundance estimates should not be confused with the maximum likelihood procedure employed here for estimating the constituent model parameters. The fact is that the abundances are related to the estimated parameters in a complex non-linear fashion, and in any case the mode of the likelihood function does not necessarily coincide with the mode of the probability density function for the abundance estimates.

4.3.2 Addition of effort information

The variance of recent biomass estimates is markedly reduced by including additional information about the performance of the fishery. The preliminary results obtained using only the catch-at-age data indicate that it is essential to do this to obtain any useable estimate of true biomass trajectories, particularly as regards the most recent years. Usually, the only other relative abundance information available is obtained from a measure of annual fishing effort. It is normally assumed that the ratio of annual catch mass to fishing effort is proportional to stock biomass. A number of CPUE (catch per unit effort) models are based on this assumption. Given only the catch and effort data, they can be used to obtain estimates of r, q, K and B(t) for the logistic (Schaefer) model:

$$dB(t)/dt=rB(t)(1-B(t)/K)-qEB(t)$$

(for examples of practical applications of these methods to the

South African hake stocks see Butterworth and Andrew, 1984). These analyses permit one to specify quotas corresponding to different harvesting strategies. However it is possible that a maximum likelihood abundance estimator which utilises catch mass, age structure and effort data will be more efficient than either a CPUE model estimator or a VPA estimator alone. However, as will be seen, there are other problems associated with the additional use of catch age structure information, which may (depending on the specific case) lead one to use the CPUE and catch mass data only.

The likelihood VPA method allows one to incorporate effort data and other information about the fishery quite easily, and this is one of its main advantages. It is possible to incorporate the effort data in this manner, under the assumption of proportionality between CPUE and population biomass. This assumption is roughly equivalent to assuming that effort E_i is proportional to F_i, and this latter model leads to a convenient way of incorporating the effort information.

If one assumes that F_i is lognormally distributed about E_i (with the variance of $\ln F_i = \sigma_i^2$), then the additional term to be subtracted from the log-likelihood function is:

$$\sum_{i=1}^{n} [\ln k_e E_i - \ln F_i]^2/(2\sigma_i^2) + n\ln\sigma_i \qquad \ldots\ 4.5$$

where the weight σ_i^2 and the proportionality constant k_e are an additional parameters which must either be estimated in the maximisation or independently. The term $n\ln\sigma_i$ can be deleted if an independent estimate of σ_i^2 is available.

4.3.3 Simulation results: effort information included

Table 4.6 shows the estimated variance of the maximum likelihood abundance estimator if σ^2 and $\sigma_1{}^2$ (the variances of the catch mass estimate and the effort to fishing mortality model error) are known, and for a coefficient of variation of E_i about F_i of about 30% (i.e. $\sigma_1 = 0.3$). M is also assumed known for this case ($M=1\ yr^{-1}$). Variance estimates decrease going backwards in time from about 22% in the final year (1978) to 9% in the first year (1964). Thus it appears that the addition of catch-at-age data provides more information about the history of biomass changes than would proportionality between CPUE and abundance alone, since the abundance estimates all have variances less than the variance on the effort to F_i relationship itself. Moreover, this result further illustrates the absolute necessity of additional relative abundance information in the VPA, apart from catch mass and age structure data alone (compare Table 4.6 with 4.5, particularly the most recent years - eg. 22% vs 117% for 1978).

At this stage a serious objection to the application of the procedure presents itself. It has been assumed until now that σ, σ_1 and M are known. This is usually not the case in practice, and this problem requires some consideration.

4.4 WEIGHTING

The quantities σ and σ_1 are unknown, yet they must somehow be estimated to provide the correct weighting between the age structure, catch mass and effort information. Fournier and Archibald (1982) suggest that these cannot be efficiently estimated from the available information, and they advocate that model sensitivity be tested within a reasonable range of

Table 4.6. The Monte Carlo (i.e. Stage A results as set out in Fig. 4.2) estimates of mean and variance for the maximum likelihood abundance estimators (kt=90), with added effort information. M is taken as given, and having its correct value of 1 yr^{-1}. Similarly Q=1000, $\sigma=0.05$, and $\sigma_1=0.3$ are given. C.V.'s obtained without use of effort data (from Table 4.5) are shown for comparison.

YEAR	TRUE BIOMASS	STAGE A MEAN	STAGE A C.V. (%)	C.V. NO EFFORT
1964	2677	2730	8.9	8
1965	2189	2230	9.0	7
1966	1577	1495	9.9	8
1967	1322	1271	9.1	8
1968	1233	1223	8.9	7
1969	970	955	10	8
1970	947	973	11	9
1971	1011	994	12	11
1972	1128	1093	11	14
1973	1030	1074	9.8	21
1974	831	817	11	38
1975	736	724	14	62
1976	804	786	16	78
1977	939	918	16	88
1978	884	889	22	117

variances. This suggests that a range of possible solutions should be estimated from these results.

Kendall and Stuart (1967) have investigated the possibilities of obtaining estimates of variance parameters such as are required here. They treat a simple analogue of this aspect of the VPA problem, which is summarised below (see also Ricker, 1973).

Two mathematical variables X and Y are known to be related by a linear relationship:

$$Y=\alpha_0+\alpha_1 X \qquad \ldots\ 4.6$$

and α_0 and α_1 are to be estimated. However X and Y are unobservable, and instead the random variables ξ and η are observed, being defined as:

$$\xi_i=X_i+\varepsilon_i \qquad \ldots\ 4.7$$

$$\eta_i=Y_i+\delta_i \qquad \ldots\ 4.8$$

The following properties are assumed to apply:

$$\left.\begin{array}{l} E(\delta_i) = E(\varepsilon_i) = 0 \\ Var(\delta_i) = \sigma_\delta^2,\ Var(\varepsilon_i) = \sigma_\varepsilon^2 \end{array}\right\} \quad \text{for all } i$$

$$Cov(\delta_i, \delta_j) = Cov(\varepsilon_i, \varepsilon_j) = 0 \qquad i \neq j$$

$$Cov(\delta_i, \varepsilon_i) = 0 \text{ for all } i, j$$

Furthermore ε_i and δ_i are assumed to be normally distributed.

By equating the first derivatives of the MLF with respect to all the parameters to zero and solving this set of equations, Kendall and Stuart produce the relationship:

$$\sigma_\varepsilon^2=\alpha_1^2\sigma_\delta^2 \qquad \ldots\ 4.9$$

They conclude that (pp385, ibid):

> ".. it is not worth proceeding with the ML estimation process, for (the above relation) first deduced by Lindley (1947), shows that the ML method fails us here. We have no prior knowledge of the values of the parameters α_1, σ_δ^2, σ_ε^2 and yet (the above relation) gives a definite relation between the ML estimators, which is not true in the model as specified. In fact, (the above relation) clearly implies that we cannot be consistently estimating all three of the parameters α_1, σ_δ^2, σ_ε^2. The ML solution is therefore unacceptable here."

Thus for this case it is fundamentally impossible to estimate both σ_δ^2 and σ_ε^2 given only the set of n pairs of (ξ_i, η_i). In an additional calculation they show that efficient (i.e. minimum variance) estimators do exist if the ratio of σ_ε^2 to σ_δ^2 is specified.

For the VPA procedure developed so far, the situation for the VPA likelihood function is more complex. Specification of the ratio of σ^2 to σ_1^2 (by analogy with Kendall and Stuart's result) may be inadequate, since there is a further term in the objective function, to which weighting must be related, viz. the multinomial term defining the probability mass function for the catch at age structure estimates, $P_{i,j}^{obs}$. The effective weighting constant for this term, Q, is the number of fish which are randomly sampled for aging purposes. There are thus three weights in the likelihood function: Q, $1/\sigma^2$ and $1/\sigma_1^2$. If the analogy to the simple model investigated by Kendall and Stuart holds, then two of the three should (by analogy) be specified, but a third may perhaps be estimated from the maximum likelihood procedure. Specification of Q and σ, and maximum likelihood estimation of σ_1 was attempted first.

4.4.1 Estimating σ_i^2

The loglikelihood function, lnL, with σ_i given has the form (summations, where indicated, are from 1 to n):

$$\ln L = f(\beta_{q,k},\ F_i : i=1..n,\ b) - (1/2)\Sigma(\ln F_i - \ln k_e E_i)^2/\sigma_i^2 \qquad \ldots\ 4.10$$

Subscripts q, k denote the fundamental ß parameter sites, of which there are n+m-1. If σ_i is freed, then the loglikelihood function is lnL*, where:

$$\ln L^* = f(\beta_{q,k},\ F_i : i=1..n,\ b) - (1/2)\Sigma(\ln F_i - \ln k_e E_i)^2/\sigma_i^2 - n\ln(\sigma_i) \qquad \ldots\ 4.11$$

Note that

$$\partial \ln L^*/\partial\sigma_i = \Sigma\ (\ln F_i - \ln k_e E_i)^2/\sigma_i^3 - n/\sigma_i \qquad \ldots\ 4.12$$

The ∂ represent partial derivatives. Let the model parameters other than σ_i be represented by the vector $\underset{\sim}{p}$; the individual parameters are denoted p_l. Superficially it seems that if the position of the maximum of lnL is given by $\underset{\sim}{p}$ (components of which are F_i and k_e), then an estimate of σ_i might be from equation 4.12 set equal to zero, i.e.:

$$\sigma_i^2 = \Sigma\ (\ln F_i - \ln k_e E_i)^2/n \qquad \ldots\ 4.13$$

This would be incorrect. Note that the parameters other than σ_i - the p_l - are obtained by maximising lnL (or lnL*: the equations to be solved are independent of the additional term $n\ln\sigma_i$ included in lnL* since $\partial\ln L^*/\partial p_l = \partial\ln L/\delta p_l$):

$$\ln L = f(\underset{\sim}{p}) - (1/2)\Sigma(\ln F_i - \ln k_e E_i)^2/\sigma_i^2 \qquad \ldots\ 4.14$$

The important point is that the p_l parameters cannot be estimated independently of σ_i. A value $\sigma_i^{(g)}$ must first be given before L can be maximised. Thus $\sigma_1^{(d)}$ which is obtained using equation 4.13 is generally different to the value that would be obtained by including σ_1 as a free parameter in the maximisation procedure, using $\ln L^*$. Either 4.10 and 4.13 must be used iteratively, or 4.11 maximised directly including σ_1 in the parameter set.

Preliminary tests were performed to determine whether the log-likelihood function exhibited sufficient sensitivity to σ_1 to consider including it in the full set of parameter values. The effect of a 10% change in each of the F_i parameters (at the estimated optimum) on the log-likelihood function are compared with the effect of a 10% change in the true value of σ_1 in Table 4.7 (the best fit estimate of σ_1 was not available at this stage). Changes in the log-likelihood function due to changes in σ_1 therefore seem to be comparable to those due to changes in F_i. In addition, preliminary fits using two very different starting parameter values for σ_1 yielded identical estimates of σ_1 (see Table 4.8), which shows that the fitting procedure is insensitive to starting values given for σ_1 over a large range (0.2 - 2.0).

The variance and mean estimates for σ_1 were obtained by fitting the VPA model, including σ_1 as a parameter, to a number of equally probable data sets. Results are set out in Table 4.9. These results suggest that σ_1 may be accurately estimated. By analogy with Kendall and Stuart's (1967) analysis however, it seems unlikely that σ can also be obtained by maximising the likelihood function with σ as an additional parameter.

4.4.2 Estimating σ_1^2 and σ^2

Further investigations into the possibility of fitting both σ_1^2

Table. 4.7. The percentage change in the log-likelihood function resulting from a 10% increase in each of the F_i model parameters in turn, including σ_1. The base values for the model parameters were the estimated maximum position, and for σ_1 the true value of 0.198 was used.

Year for F_i	Percentage change in the log-likelihood function
1964	56.19
1965	63.96
1966	60.44
1967	58.66
1968	64.19
1969	60.78
1970	59.23
1971	58.51
1972	59.63
1973	65.37
1974	62.37
1975	56.93
1976	55.26
1977	45.49
1978	20.25
σ_1	17.16

Table. 4.8. Estimates of σ_1^2 in the VPA model, for different starting values of σ_1^2 (i.e. σ_1^2 is a free parameter in this case but σ^2 and M are taken as given).

σ_1^{true}:	0.198
$\sigma_1^2{}_{START}$	$\sigma_1^2{}_{ESTIMATED}$
0.2	0.104365886
2.0	0.104365881

Table 4.9. Mean and variance of σ_1 estimates obtained by maximising the VPA likelihood function using 20 different sets of equally probable data (M=1.0, σ = 0.05, Q=1000).

σ_1^{true} :	0.198	0.30
σ_1^{est} (CV):	0.176 (20%)	0.276 (18%)

and σ^2 were conducted. Table 4.10 shows the percentage change induced in the log-likelihood by a 10% change in each of the model parameters in turn. The base value for each model parameter was the best fitted value obtained by maximising the log-likelihood function without including σ_1^2 and σ^2 as estimable parameters (their true values were given: $\sigma_1^2 = 0.04$; $\sigma^2 = 0.0025$). In evaluating the results presented in Table 4.10, the term $\{-n\ln[\sigma_1\sigma]\}$ was included in the log-likelihood function.

Table 4.10 suggests firstly that the relative sensitivity of the likelihood function to σ_1 is comparable to that shown in Table 4.7. Also it suggests that σ is better determined than σ_1, since the likelihood function appears to be more sensitive to σ. Both σ and σ_1 therefore appear to be estimable, however attempts at fitting the VPA model were unsuccessful. All the attempted fits, using equally probable data sets, resulted in flag values of IFAIL=2, indicating that the criteria for convergence were not satisfied within 400N iterations. N, the number of parameters, was 38, so that the convergence criteria are unsatisfied after 15 200 iterations. Some additional mechanism, not reflected in Table 4.10, acts to render both weights inestimable. It is most likely that σ is confounded with σ_1, so that dual estimates cannot be achieved. The analogy to Kendall and Stuart's analysis is that only one of the variance parameters is potentially estimable in the best fit procedure. The numerical indications tend to suppport this for the particular likelihood function in question, viz. σ_1 (or σ) alone is potentially estimable; however σ_1 and σ cannot be fitted together.

Thus it must be assumed that a reasonable estimate of σ or σ_1 can be made independently. Reasonable estimates of Q may be obtained independently (see Chapter 5). If σ_1 is to be fitted, then the

Table 4.10. Sensitivity of the log-likelihood function to a 10% increase in $\acute{\sigma}$ and σ_1 compared to the same changes in the F_i parameters.

Year for F_i	Percentage change in the log-likelihood function
1964	7.28
1965	8.29
1966	7.84
1967	7.59
1968	8.30
1969	7.87
1970	7.68
1971	7.60
1972	7.73
1973	8.48
1974	8.09
1975	7.40
1976	7.18
1977	5.91
1978	2.60
σ_1	1.43
σ	3.22

estimate for σ^2 may just have to be an educated guess. The guess made in subsequent practical applications (see Chapters 5, 7 and 8) assumes that the quantity of fish caught each year is estimated quite precisely, with a C.V. of about 5%, which corresponds to a value of σ^2 of close to 0.0025. Note with regard to this that the mass and species composition of SA fish catches (pelagic and demersal) is monitored by a government inspectorate; imprecision and also perhaps inaccuracy is most likely to arise from species composition estimates, particularly for the early years of the fishery. There have been indications of malpractices such as underestimating the catch and species misidentification (Alant et al, 1983) but this seems more a problem for the Namibian pelagic fishery.

Alternatively, some estimate of $\sigma_1{}^2$ can be obtained from the results of dynamic CPUE model analyses, where these are available. Using this approach, a value of $\sigma_1{}^2 \approx 0.01$ is indicated for hake (Butterworth, pers. commn). Given such an independent estimate of σ_1, it would seem that σ (or alternatively Q) could then be obtained from the best fit procedure.

Another crucial parameter whose value is normally regarded as given in the VPA is M, the natural mortality constant. Attempts at estimating M as a model parameter in the VPA maximum likelihood estimation procedure are presented in the following section.

4.5 ESTIMATING NATURAL MORTALITY

Fournier (1983) attempts to estimate the natural mortality parameter, M, for pacific cod (for which typical values seem to be in the vicinity of 0.5 yrs^{-1}). After substantial manipulation of the basic VPA model, his objective function shows a weak minimum (a maximum likelihood since the sign of his objective function is opposite to that of the loglikelihood function) at a value of $M=0.65$ yrs^{-1}, which is then used in a series of other calculations.

However, from the results he presents, the slope of the objective function close to the best fitted value of 0.65 is very slight, so that a 10% change in M on either side of 0.65 yrs^{-1} produces a change in the objective function of roughly 0.02%. Fournier (1983) does not produce a variance estimate for the M value, but one suspects it to be intolerably large given the insensitivity to M revealed in the objective function. Prospects for estimating M for anchovy with M in the vicinity of 1 yr^{-1} do not therefore seem to be good, and this is borne out by results which follow.

4.5.1 Results

Preliminary results (Table 4.11) show that the likelihood function is very sensitive to M at the estimated maximum position. Table 4.12 however shows large CV's on estimates of M - in the range of between 39% and 54%. The associated CV's for biomass estimates (not shown here) are also very large, being in the region of 50% for $\sigma_1 = 0.2$ and well over 100% for $\sigma_1 = 0.3$.

These results are not particularly surprising, since relaxing the fixing of M in the model injects a potential ambiguity into the interpretation of the data set, even with additional effort data included. This concerns the relative impact made on the population by predation and by fishing operations, which the model is now required to estimate. It is difficult to imagine

Table 4.11. Sensitivity of the log-likelihood function to a 10% increase in M (1.0 → 1.1) compared to its sensitivity to the estimated F_i parameters. The percentage change in the value of the log-likelihood function (scaled by subtraction) due to a 10% increment in each of the F_i parameters and M (each amended separately) is shown.

Year of F_i	With effort $\sigma_i=0.2$ % change	With effort $\sigma_i=0.3$ % change	No effort % change
1964	15.81	10.57	20.28
1965	18.00	12.03	23.21
1966	17.01	11.37	21.93
1967	16.51	11.01	21.34
1968	18.06	12.05	23.33
1969	17.10	11.42	22.16
1970	16.70	11.13	21.60
1971	16.46	11.02	21.32
1972	16.78	11.21	21.92
1973	18.39	12.31	24.75
1974	17.55	11.73	24.42
1975	16.02	10.74	21.20
1976	15.55	10.42	20.34
1977	12.80	8.58	20.14
1978	5.70	3.77	2.14
M	259.8	173.13	346.7

Table 4.12. Estimates and CV of M obtained by maximising the VPA likelihood function for Q=1000 and $\sigma=0.05$ and $M^{true} = 1$ yr^{-1}.

σ_i^{true} :	0.2	0.3
M^{est} (%CV):	0.87 (39%)	0.88 (54%)

how the model would place realistic restrictions on the estimated M value in these circumstances (except perhaps in special cases where large data contrast has been produced by, for example, a very variable fishery), and it appears that other assumptions must be made to facilitate the estimation of M. For example independent estimates of anchovy consumption by natural predators may provide some information in this regard. However other assumptions which are sometimes introduced, such as restrictions placed on permitted per capita recruitment rates, are more meaningfuly applied to K - selected species such as marine mammals than pelagic shoaling fishes.

Note further that changes in M generally affect absolute biomass levels rather than trends. It is therefore not particularly surprising that the addition of CPUE information - a <u>relative</u> biomass measure - does not help in estimating M. This may not be as serious as it seems, as changes in M tend to affect absolute magnitude of biomass rather more than trends (which are more important from a management point of view), and the increased CV's in biomass estimates upon freeing M predominantly reflect the former of these effects. It is anticipated however that incorporating <u>absolute</u> abundance estimates from independent assessment techniques such as direct stock surveys would greatly facilitate the estimation of M.

4.6 ESTIMATES OF BIAS AND VARIANCE OF THE MAXIMUM LIKELIHOOD ABUNDANCE ESTIMATORS USING A RESAMPLING TECHNIQUE

Fournier (1983) utilises the asymptotic properties of the log-likelihood function to evaluate different VPA models. A resampling method based on the bootstrap procedure (Efron, 1981) is used here, and this offers a powerful means for evaluating biomass estimate variances obtained from different VPA procedures.

In practice, the variance of the abundance estimators, $\sigma_B{}^2$, cannot be estimated directly by techniques using the true values of the various model parameters (as done for example in section 4.3, using the procedure headed Stage A in Fig. 4.2), because their values are unknown. However, the first model fit does provide an estimate of the various model parameters and these estimates can be used in the same way that one might use true parameter values for estimating abundance estimate variance. Thus estimates of $\sigma_B{}^2$ can be obtained by a "bootstrap" procedure (Efron, 1981); this procedure, as outlined below, is Stage B in Fig. 4.2. It is necessary to subject the bootstrap procedure to a series of trials in order to validate it (this is Stage C in Fig. 4.2).

4.6.1 <u>The bootstrap procedure</u>

At a simple level, the bootstrap uses the residuals (differences between observed data and the model predicted values) as a measure of the probability density function of the true residuals (or errors). This approach has considerable merit in cases where analytical estimators for the variance of fitted parameters cannot be obtained (it can be generalised to include other statistics). The situation dealt with here is a good example of this type of regression problem.

If there are n pieces of data, $C_i{}^{obs}$, for which the logarithm of the true values, $\ln C_i$, are estimated by regression techniques, the residuals are:

$$\xi_i = \ln C_i{}^{obs} - \ln C_i \qquad \ldots\ 4.15$$

In the bootstrap approach further possible data sets, $C_i{}^*$, are generated via a random sampling procedure. A series of probable residual sets $\{\xi_i^*: i=1....n\}$ are obtained by randomly sampling k with replacement and with equal probabilities between 1 and n:

$$\xi_i^* = \xi_k \qquad \ldots\ 4.16$$

The additional (from a resample of the empirical residuals) data sets C_i^* are therefore given by:

$$C_i^* = C_i \exp(\xi_i^*) \qquad \ldots\ 4.17$$

Thus a number of data sets can be generated by considering only the initial fitted model and the single given set of observed data itself. For the VPA this allows one to obtain a series of possible biomass estimates, by fitting the VPA model to each set of resampled data in turn. Variance estimates of the biomass estimators can then be calculated using standard equations (see Stage B, Fig. 4.2). These results must however be regarded with some caution. Note that the sum $(\sum_{i=1}^{n} \xi_i^2 / n)$ is probably a biassed estimator for σ^2. For a linear model, and with the ξ_i having equal variances and all uncorrelated, the bias-corrected estimate of σ^2 is (Kendall and Stuart, 1967):

$$\sigma^2 = \sum_{i=1}^{n} \xi_i^2 / (n-p) \qquad \ldots\ 4.18$$

where n is the number of data values, and p the number of model parameters estimated from the data. In the special case of the VPA likelihood function the additional terms (involving $P_{i,j}$, F_i and a few other parameters) and mathematical relationships linking them lead to a model which is definitely not linear, so that the correction shown in equation 4.18 may be inappropriate. The extent to which bootstrap estimates of the variance of abundance estimates are biassed (a priori one might expect negative bias for the variance estimate) is therefore investigated using simulation techniques instead.

The estimates of $P_{i,j}$ from the first model fit can be taken to define the empirical distribution function of $q_{i,j}^{obs}$, since the variance of a multinomial probability function is a function only of the population means and Q, the number of samples taken. The additional sets $q^*_{i,j}$ were obtained by multinomially resampling 1000 fish with age distributions $P_{i,j}$ each year, as described before (Appendix 4.1), where $P_{i,j}$ are now obtained from the first model fit to the given data.

The accuracy (i.e. bias) of the bootstrap biomass variance estimates was tested by comparison with the variance for results obtained in the case when the true parameter values are known (see Section 4.3; see also stage C, Fig. 4.2).

4.6.2 Simulation results using the resampling procedure

The accuracy of the bootstrap estimate of variance, based on resampling the empirical residual sets (equation 4.17; Stage B, Fig. 4.2), was estimated by comparing it to the sample variance obtained in section 4.3 (Stage A, Fig. 4.2). In that process (Stage A), which has been used until now, the model parameters are all known exactly, while for the bootstrap (Stage B) the parameters are estimated from an initial fit of the model to the data (the distinction between these two calculation procedures is illustrated in Figure 4.2.

However, the number of sets of data which can be processed is limited by the required computer time, and only 90 data sets were processed. From the χ^2/d.f. distribution for s^2/σ^2, with 89 degrees of freedom, 95% of the estimates of s^2/σ^2 are expected to lie between about 0.7 and 1.3, and this sort of precision is adequate for the purposes required here. For the situation under consideration, the model was defined by:

n = 15 yrs $M = 1$ yr^{-1} b = 3 yrs
m = 5 yrs Q = 2500 fish $\sigma_1 = 0.3$
$\sigma = 0.1$

Midyear fish masses for age groups 0 to 4 years were; 20, 40, 55, 65, and 75 grams respectively, and 90 data sets were generated and processed for both Stage A and B techniques (see Fig. 4.2 for further clarity).

Recall that variances for the results in 4.3 (Stage A, Fig. 4.2) are obtained using the 90 biomass trajectories, $B_{i,k}$, obtained from fitting the VPA model to 90 unique data sets (C_i^{obs}; $P_{i,j}^{obs}$; E_i) obtained from;

$C_i^{obs} = C_i \exp(\xi_i)$
$E_i = F_i \exp(\varepsilon_i)/k_e$
$P_{i,j}^{obs}$ drawn from $P_{i,j}$ (Q=500)

(k_e is a proportionality constant)
However, for the resampling procedure, a single estimate of model parameters is provided by the first fit of the model to the raw data, producing the residuals;

$\xi_i = \ln C_i^{obs} - \ln C_i$
$\varepsilon_i = \ln k_e E_i - \ln F_i$

and the estimated age structure matrix $P_{i,j}$. A further 89 data sets were generated from;

$C_i = C_i \exp(\xi_k)$

$E_i = F_i \exp(\varepsilon_k)/k$.

with $P_{i,j}^{obs}$ now drawn from $P_{i,j}$ (Q=500)

Resampled variances were calculated from the $B_{i,k}$ biomass trajectories resulting from fitting the VPA model to these resampled data sets (see Fig. 4.2, Stage B).

Table 4.13 gives the results of the simulations as outlined. These only give an indication of likely estimator bias, as a definitive analysis would require additional determination of the precision of all the entries in Table 4.13 using for example the procedure headed Stage C in Fig. 4.2 (this was impractical because of computer time considerations).

For a linear model, one expects that the variances σ^2 and σ_1^2 estimated by $\Sigma\, \xi_i^2$ / n and $\Sigma\, \varepsilon_i^2$ / n would be negatively biassed by the factor (n-p)/n where p is the number of model parameters and n the number of given data points (for the case under consideration (n-p)/n is about 1/2). Since the residuals are therefore negatively biassed, this would cause the resampled estimates of abundance estimator variance to be negatively biassed. Results shown in Table 4.11 in general reflect this, however the effect is not large enough to be of practical concern (at least for the data/parameter ratio and configuration considered here).

Results for this section therefore suggest that the bootstrap procedure does give reliable (i.e. not markedly biassed) estimates of biomass variance.

Table 4.13. Comparison between the bootstrap (model parameters estimated - Stage B) and simulated (model parameters known exactly - Stage A) estimates of mean and variance for the maximum likelihood abundance estimators (kt=90 in both cases), with added effort information, and the true values of $M=1$ yr^{-1}, $Q=500$, $\sigma=0.1$, and $\sigma_1=0.3$ given.

YEAR	BIOMASS	STAGE A ESTIMATE	STAGE B ESTIMATE	STAGE A C.V.(%)	STAGE B C.V.(%)
1964	2677	2730	2389	8.9	9.4
1965	2189	2230	2113	9.0	7.5
1966	1577	1495	1588	9.9	8.3
1967	1322	1271	1299	9.1	8.1
1968	1233	1223	1183	8.9	8.3
1969	970	955	928	10	9.8
1970	947	973	917	11	10
1971	1011	994	900	12	9.6
1972	1128	1093	1025	11	8.0
1973	1030	1074	924	9.8	8.2
1974	831	817	761	11	8.9
1975	736	724	642	14	12
1976	804	786	707	16	13
1977	939	918	819	16	15
1978	884	889	727	22	22

4.7 INCLUDING ADDITIONAL INFORMATION IN THE VPA LIKELIHOOD FUNCTION

A considerable advantage of the maximum likelihood approach is the ease with which additional information about the fishery can be included by adding further terms to the log-likelihood function. If the additional data provides an index of recruitment or biomass (for example spawning biomass) over all or part of the history of the fishery, then this entails the estimation of 2 extra parameters. The first is the proportionality constant between the index and the cohort biomass or number. The second is the variance of the index to cohort proportionality relationship - one parameter if the time series of indices can be assumed to be homoscedastically distributed. If not, then instead n parameters - the variances for the index for each year - must be provided.

It should be recalled that an earlier result suggested that these variances had to be estimated independently of the VPA likelihood function maximisation and so the index is only really of use given these independent variance estimates. The sensitivity of the final result to these variances should be tested (Fournier and Archibald, 1982). The forms of the likelihood function for some of the more common indices are shown in the following paragraphs. (While the above has related only to the provision of one additional data set - say a relative abundance index - it is readily generalised if more than one such set is available)

4.7.1 The use of recruitment indices

It is to be expected that a recruitment index, $N_{i,1}^{ind}$, can be used to improve the quality of VPA results. However, recruitment information is probably not as useful as effort information for obtaining abundance estimates. This is because the E_i fix the F_i trend, and the F_i's define the abundance fairly rigidly from the catches, since the F_i's are closely related to the catch-to-biomass ratio. On the other hand, the recruitment index only provides an index of abundance m years after the first year of fishing considered in the VPA, since m years are necessary for recruitment to fill all age classes with an index of cohort number.

This shortcoming may be offset by the ability of the VPA to determine earlier estimates of abundance relatively precisely, on the basis of catch mass and age structure information only, as has been shown here (Table 4.5) and by Pope (1972). Furthermore, recruitment indices prior to the first year in the catch-at-age matrix can be used to impose additional constraints on fishing mortality estimates for cohorts born before year i=1, as is done for example for pilchard in Chapter 7.

It is necessary to assume some error model for the recruitment index, and to include it in the log-likelihood function according to this model. If the recruitment measurement errors in $N_{i,1}^{ind}$ are approximately log-normally distributed with constant variance, σ_r^2, then the following term must be subtracted from the log-likelihood function:

$$\sum_{i=1}^{n} [\ln k_r N_{i,1}^{ind} - \ln N_{i,1}]^2/(2\sigma_r^2) + n\ln\sigma_r \qquad \ldots\ 4.19$$

where k_r is a proportionality parameter. The $N_{i,1}$ can be expressed in terms of ß, F and S model parameters

$$N_{i,1} = \exp(\beta_{i,1}) C_i^{obs} (F_i S_1 + M) / [F_i S_1 (1 - \exp(-F_i S_1 - M))] \qquad \ldots\ 4.20$$

$$- \sum_{i=1}^{n} (\ln(B_i) - \ln(k_s B_i^{ind}))^2/(2\sigma_s^2) - n\ln\sigma_s \qquad \ldots\ 4.23$$

where k_s is the proportionality parameter between the biomass and index estimate (k_s falls away if the data are absolute abundance estimates), and a realistic independent estimate of σ_s must accompany the data.

4.8 SUMMARY

In this section the basic Fournier and Archibald (1982) VPA procedure has been considered in detail for a specific case.

Considerable care must be taken to scale the VPA objective function suitably to obtain reliable convergence to the position in parameter space of the maximum of the likelihood function. This applies to both the derivative-free and first derivative fitting procedures, although only the derivative-free approach was attempted here.

Results using the procedure (Stage A in Fig. 4.2, for which the true model parameters are all known exactly) confirm that reliable current biomass estimates cannot be obtained using catch at age data alone. When effort data are included, the increase in precision for recent biomass estimates is dramatic, and this suggests that VPA should only be used in conjunction with additional information (to catch at age data) about the fishery, such as effort. Recruitment and spawning biomass indices are further examples of data which can be used in the VPA.

The suggested resampling procedure (Stage B, Fig. 4.2, model parameters are unknown) appears to provide an adequate assessment of abundance estimator variance for the specific example studied (computer time limitations precluded an exhaustive study of this aspect).

However variance estimates should be viewed with caution when the deviation between $P_{i,j}^{obs}$ and $P_{i,j}$ is larger than one would expect on the basis of the sampling variance from multinomially distributed age data, since this is probably an indication that error in assuming selectivities fixed from year to year is substantial. This model error is not treated in Fournier and Archibald's (1982) VPA model and is the proportion of the error between $P_{i,j}^{obs}$ and $P_{i,j}$ (where $P_{i,j}$ conforms exactly to a fixed selectivity pattern) not explained by consideration of random sampling errors in obtaining $P_{i,j}^{obs}$. The most important effect of this is that abundance estimator variance derived from the resampling procedure described is negatively biassed.

It is reiterated here that the model investigated thus far has assumed a particular functional form for the age dependent selectivities. The assumption of functional form for selectivity is crucial to results obtained in real situations (see for example Butterworth et al, 1986). The effect of relaxing the fixed selectivity shape or the assumption of asymptotic flatness of the selectivity as a function of age is demonstrated in Chapter 5 in the case of the South African anchovy.

In the remainder of this thesis, a practical application of the maximum likelihood VPA procedure for the anchovy, pilchard and hake stocks which are fished in South African waters is demonstrated.

5. ABUNDANCE ESTIMATES FOR THE SOUTH AFRICAN ANCHOVY - VARIANCE AND BIAS

5.1 BACKGROUND

VPA is one of the few methods available for estimating historic biomass trends, and its usefulness for doing this for the South African anchovy resource is now investigated. The anchovy VPA calculations reported here have been carried out at a time of considerable flux in the data available and techniques used by the Sea Fisheries Research Institute (SFRI) of South Africa to assess this resource. At the outset, during 1982 and 1983, VPA provided the only basis for anchovy biomass and sustainable yield estimates. The approach used for estimating the terminal fishing mortalities was highly questionable (Butterworth, 1983), and a more rigorous statistical analysis of the data was urgently required. Essentially this work has confirmed much of what was suspected from the start, albeit in more quantitative terms, namely that VPA has little value as an assessment tool for the recent and even the historic status of the local anchovy stock.

Since late 1983, the SFRI has concentrated its efforts on hydroacoustic survey techniques, and results from such surveys have provided some revolutionary insights into anchovy biomass and habitat range, completely superceding VPA estimates of recent anchovy biomass. It should be noted, however, that previous VPA results have found their way into the scientific literature, and this work suggests that inferences about historic stock trends made from these assessments could well be false.

The likelihood VPA method which has been explored so far is based on a crude model of the error structure of the South African fisheries data bank. For example, in reality estimates of age structure for the anchovy catch are made by using an estimated age length key and a length frequency distribution, rather than by randomly sampling and ageing a fixed number of fish regardless of length. The correct likelihood function should therefore be formulated to include both the age-length key entries and the length frequency distribution as parameters. However this would involve an unmanageable number of parameters relative to that required by the current formulation of the objective function. To avoid this, it is assumed in the following that the net effect of the length histogram and age-length key estimation process can be approximated by random sampling for age. This can be achieved in quantitative terms by estimating the number of fish which need to be randomly sampled and aged to give rise to roughly the same variances of the age structure estimates, $P_{i,j}^{obs}$, as do the age-length key and length histogram.

5.2 THE PRECISION OF AGE STRUCTURE ESTIMATES

Sampling error for the estimate of the catch age structure, $P_{i,j}^{obs}$ (i=year, j=age), arises from the use of an age length key which was estimated by ageing only about 400 fish (Q) (Crawford et al, 1978), and a length frequency distribution, obtained from measuring between 10 000 and 20 000 fish per year (H). Although an annual estimate of the length frequency of the catch is made, a single age length key, which was estimated from samples taken between 1965 and 1967, has been applied to length frequencies for the whole period (1964-1983). The bias due to this, which is considerable, will be discussed later. For the present, some number Q must be estimated, so that random sampling and ageing of Q fish each year produces a variance of $P_{i,j}^{obs}$ similar to the variance resulting from the age length key and length histogram estimate. Thus the variance of $P_{i,j}^{obs}$ due to the latter process must be estimated.

Variance estimates can be obtained by assuming that the observed age length key and length histograms are exactly equal to the population means for those quantities. In addition, under random sampling, the age length key and length histograms are multinomially distributed, so that the means of the multinomials can be used to calculate the variances of $P_{i,j}^{obs}$. These variance estimates can be shown to be unbiassed (Kendall and Stuart, 1967).

The observed age length key, $a_{j,l}^{obs}$, is simply an estimate of the relative frequency distribution of fish aged j in a particular length class l, which is therefore an m (total number of age classes) by lmax (total number of length classes) matrix, such that

$$\sum_{j=1}^{m} a_{j,l}^{obs} = 1 \qquad \ldots\ 5.1$$

The following new variables are now introduced:

h_l : the number of fish sampled for ageing in length class l (i subscript suppressed)

H : the total number of fish sampled for purposes of obtaining the length frequency distribution of the catch (i subscript suppressed)

lmax : the maximum number of length classes considered

$f_{i,l}^{obs}$ has been defined previously and is the observed normalised length frequency distribution of the catch:

$$\sum_{l=1}^{lmax} f_{i,l}^{obs} = 1 \qquad \ldots\ 5.2$$

Values for numbers of fish caught at age, $P_{i,j}^{obs}C_i^{obs}$, are given by

$$P_{i,j}^{obs}C_i^{obs} = \sum_{l=1}^{lmax} (a_{j,l}^{obs}f_{i,l}^{obs}C_i^{obs}) \qquad \ldots\ 5.3$$

and therefore:

$$P_{i,j}^{obs} = \sum_{l=1}^{lmax} (a_{j,l}f_{i,l}^{obs}) \qquad \ldots\ 5.4$$

The variance of $P_{i,j}$, $Var(P_{i,j})$ is thus:

$$Var(P_{i,j}) = Var(\sum_{l=1}^{lmax} (a_{j,l}f_{i,l})) \approx \sum_{l=1}^{lmax} (Var(a_{j,l}f_{i,l})) \qquad \ldots\ 5.5$$

and assuming that the covariances are zero, this becomes:

$$Var(P_{i,j}) = \sum_{l=1}^{lmax} [Var(a_{j,l})Var(f_{i,l}) +$$

$$(E(a_{j,l}))^2 Var(f_{i,l}) + Var(a_{j,l})(E(f_{i,l}))^2] \qquad \ldots\ 5.6$$

It is assumed that the expected values (the means) of the age length key and normalised length frequency distribution are the observed quantities themselves (these are the maximum likelihood estimators of the true values):

$$E[a_{j,l}] = a_{j,l}^{obs}$$

$$E[f_{i,l}] = f_{i,l}^{obs}$$

and therefore the variances follow from binomial probability distribution considerations:

$$Var(a_{j,l})=a_{j,l}^{obs}(1-a_{j,l}^{obs})/h_l \quad \ldots\ 5.7$$

$$Var(f_{i,l})=f_{i,l}^{obs}(1-f_{i,l}^{obs})/H \quad \ldots\ 5.8$$

For the South African anchovy, the coefficients of variation of $P_{i,j}^{obs}$ increase with increasing age as $P_{i,j}^{obs}$ decreases (Table 5.1). However, the equivalent number of randomly sampled and aged fish which matches these coefficients of variation, the Q estimates ($Q=(1-P_{i,j})/(P_{i,j}(CV.)^2)$, C.V.'s from Table 5.1), are not constant for different age classes in the catch (Table 5.2). This is thus some indication of the extent to which the random sampling model approximates the age-length-key estimation procedure, and a value of between 500 and 5000 equivalent randomly sampled individual anchovy is indicated by these results. The value Q=2500 was used in fitting the anchovy model, and the effect of using 500 instead is also shown.

In general, the age length key and length frequency combination produces a smaller sampling error in $P_{i,j}$ than does randomly sampling and ageing the same number of fish as were aged to produce the age length key, particularly for the older ages which constitute a small proportion in the catch (Table 5.3). However, this is only the case if the age-length key is estimated each year.

5.3 BIAS DUE TO THE USE OF A SINGLE AGE-LENGTH KEY

The anchovy catch age structure matrix, $P_{i,j}^{obs}$, which will be used is seriously biassed by the application of a single age-length key, estimated from samples taken during 1965-1967 (Crawford et al, 1978), to observed catch length frequency distributions for all years. In general, the age-length key is specific to the age structure of the catch for a given year and so it is important to illustrate the effect of using only one key

Table 5.1. Estimates of the percentage coefficient of variation for values of the proportion of fish caught at age ($P_{i,j}^{obs}$), deduced from length frequency distributions and an age-length key as described in the text. $P_{i,j}^{obs}$ values are given in brackets.

	1965	1966	1967
AGE			
0	4.6 (0.394)	2.0 (0.780)	2.3 (0.760)
1	5.8 (0.380)	9.5 (0.168)	9.3 (0.194)
2	9.0 (0.165)	9.0 (0.039)	12.2 (0.031)
3	15.5 (0.055)	16.9 (0.012)	13.1 (0.012)
4	32.7 (0.007)	42.2 (0.001)	32.2 (0.002)

Table 5.2. A rough estimate of the number of anchovy which must be randomly sampled and aged each year to mimic the error associated with the value of $P_{i,j}^{obs}$, the observed catch age structure, by the length histogram and age length key procedure.

	1965	1966	1967
AGE			
0	742	700	573
1	483	549	477
2	617	2977	2111
3	719	2934	4660
4	1342	5228	3881

for all years.

In fact, the application of keys to years other than the specific year of age-length key estimation, is equivalent to making a number of a priori assumptions about catch age structure. If each age class is exclusively associated with a particular length range, then there will be no bias, since the age length key will be year invariant, and representable by an m x lmax matrix with lmax entries all equal to 1. However, when this is not the case - as is usual - the age-length key defines the age distribution of the catch in a particular length class, and thus the use of only one key assumes that this distribution of ages within length classes is invariant from year to year. If the overall length frequency distribution of the catch does not change much from year to year then this is tantamount to saying that overall catch age structure does not vary, which may not be the case. Alternatively, if the catch length distributions do change significantly between years, then the application of a single key over a number of years tends to produce age structures which are biased in the direction of the age structure of the catch in the year of age-length key estimation.

The 1966 (i.e. average from 1965, 1966 and 1967) age-length key for anchovy, $a_{j,l}^{66}$, is presented in Table 5.3 (Crawford et al, 1978). It can be multiplied by the catch length distribution in 1966 (the middle of the period during which it was estimated) to give a matrix, $A_{j,l}^{66}$, showing the estimated breakdown of the 1966 anchovy catch by number into age and length classes. This breakdown is shown in Table 5.4. It is commonly assumed (eg: Clark, 1981) that the normalised length distribution of the catch within each age class, $g_{j,l}^{t}$, is invariant from year to year, and may therefore be represented by $g_{j,l}$. The matrix $g_{j,l}$ is

Table 5.3. The age-length key, $a_{j,l}$, for anchovy sampled between 1965 and 1967. The h_l are the number of fish sampled and aged within each length class. Length classes are denoted by the length in centimetres of the mid-point of the length range. Each length class is 0.5 cm wide. All rows sum to 1. (Data from Crawford et al, 1978.)

	Age: 0	1	2	3	4	
Mid-length						h_l
<6.50	1*	0	0	0	0	-
6.75	1	0	0	0	0	3
7.25	1	0	0	0	0	8
7.75	0.94	0.06	0	0	0	18
8.25	0.94	0.06	0	0	0	32
8.75	0.80	0.20	0	0	0	41
9.25	0.71	0.29	0	0	0	49
9.75	0.49	0.51	0	0	0	49
10.25	0.13	0.74	0.13	0	0	38
10.75	0.15	0.63	0.17	0.05	0	41
11.25	0.02	0.40	0.53	0.05	0	55
11.75	0	0.29	0.51	0.20	0	35
12.25	0	0.03	0.52	0.39	0.06	33
12.75	0	0	0.29	0.50	0.21	24
13.25	0	0	0.11	0.67	0.22	9
13.75	0	0	0	0.67	0.33	6
14.25	0	0	0	0.50	0.50	4

* Assumed

Table 5.4. A breakdown of the 1966 anchovy catch, giving the observed $A_{j,l}^{66}$, the number of fish caught which were aged j and were in length class l (denoted by the mid-point of the length class as for Table 5.3). Units are 10^9 fish.

length l	Age				
	0	1	2	3	4
<6.50	1734	0	0	0	0
6.75	3544	0	0	0	0
7.25	3430	0	0	0	0
7.75	2971	198	0	0	0
8.25	2034	493	0	0	0
8.75	1678	671	0	0	0
9.25	851	887	0	0	0
9.75	95	533	95	0	0
10.25	91	393	106	30	0
10.75	12	252	333	34	0
11.25	0	117	211	82	0
11.75	0	7	117	90	14
12.25	0	0	12	21	9
12.75	0	0	1	5	2

The associated $g_{j,l}$ matrix is:

length l	Age				
	0	1	2	3	4
<6.50	0.106	0	0	0	0
6.75	0.216	0	0	0	0
7.25	0.209	0	0	0	0
7.75	0.181	0.056	0	0	0
8.25	0.124	0.139	0	0	0
8.75	0.102	0.189	0	0	0
9.25	0.052	0.250	0	0	0
9.75	0.006	0.150	0.109	0	0
10.25	0.006	0.111	0.121	0.115	0
10.75	0.001	0.071	0.381	0.130	0
11.25	0	0.033	0.241	0.313	0
11.75	0	0.002	0.134	0.344	0.560
12.25	0	0	0.014	0.080	0.360
12.75	0	0	0.001	0.019	0.080

therefore given in terms of $A_{j,l}^{66}$:

$$g_{j,l} = A_{j,l}^{66}/(\sum_{q=1}^{lmax} A_{j,q}^{66}) \qquad \ldots\ 5.9$$

$g_{j,l}$ and $a_{j,l}^{66}$ are related by:

$$g_{j,l} = a_{j,l}^{66}[\sum_{k=1}^{m} A_{k,l}^{66}]/(\sum_{q=1}^{lmax} A_{j,q}^{66}) \qquad \ldots\ 5.10$$

and this relationship also holds for years other than 1966, under the invariance of $g_{j,l}$ assumption:

$$g_{j,l} = a_{j,l}[\sum_{k=1}^{m} A_{k,l}]/(\sum_{q=1}^{lmax} A_{j,q}) \qquad \ldots\ 5.11$$

Thus if the true catches at age, $P_{i,j}$, are known, then under the invariant $g_{j,l}$ assumption the true catches at length in year i, $C_{i,l}$, are given (suppressing the i subscript) by the product:

$$\underset{\text{length}}{(C_1, \ldots, C_{lmax})} = \underset{\text{age}}{(C_1, \ldots, C_m)} \begin{bmatrix} g_{1,1} & \ldots & g_{1,lmax} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ g_{m,1} & \ldots & g_{m,lmax} \end{bmatrix} \qquad \ldots\ 5.12$$

If the subscripted variables are represented as matrices and vectors:

$C_{i,j} = \underset{\sim}{C}^{A}{}_{i}$

$C_{i,l} = \underset{\sim}{C}^{L}{}_{i}$

$g_{j,l} = \underline{g}$

$a_{j,l}^{66} = \underline{a}^{66}$

then equation 5.12 can be rewritten as:

$$\underset{\sim}{C}^{L}{}_{i} = \underset{\sim}{C}^{A}{}_{i}.\underline{g} \qquad \ldots\ 5.13$$

In order to solve for $\underset{\sim}{C}^{A}{}_{i}$ using equation 5.13, one must consider lmax (the number of length classes) equations in m (the number of age classes) unknowns, viz. the m entries for $\underset{\sim}{C}^{A}_{i}$. The entries for $\underline{a}$, $a_{j,l}$, can be derived from $\underset{\sim}{C}^{A}_{i}$, $\underset{\sim}{C}^{L}_{i}$ and $\underline{g}$ using the lmax x m equations resulting from inverting equation 5.11:

$$a_{j,l} = g_{j,l}[\sum_{q=1}^{lmax} A_{j,q}]/(\sum_{k=1}^{m} A_{k,l}) \qquad \ldots\ 5.14$$

However, since

$$\sum_{l=1}^{lmax} C_{i,l} = C_i \qquad \ldots\ 5.15$$

and

$$\sum_{j=1}^{m} C_{i,j} = C_i \qquad \ldots\ 5.16$$

there are in reality only lmax-1 equations in m-1 unknowns (Clark, 1981). For lmax < m, the system of equations for $\underset{\sim}{C}^{A}_{i}$ (essentially the m unknown $P_{i,j}$'s only) is indeterminate, and for lmax > m there are no unique solutions (apart from special cases). A single unique solution exists for the case lmax = m. Further the algebraic solutions do not necessarily satisfy the non-negativity requirement:

$$C_{i,j} > \text{or} = 0 \qquad \ldots\ 5.17$$

To satisfy this constraint, it is possible to estimate the m quantities, $P_{i,j}C_i$, equivalently written as $C_{i,j}$, by constrained non-linear optimisation procedures according to some best fit criterion such as maximum likelihood or least squares. A likelihood function, considering only the error structure of $f_{i,l}^{obs}$, the estimated length frequency distribution of the catch is:

$$\sum_{l=1}^{l\max} H f_{i,l}^{obs} \ln[f_{i,l}] \qquad \ldots\ 5.18$$

where H fish are randomly sampled to estimate $f_{i,l}$. The quantities $f_{i,l}$ are dependent on $P_{i,j}$ via the following relationships:

$f_{i,l} = C_{i,l}/C_i$

$C_{i,l} = \underset{\sim}{C}^{L}_{i}$ (as before), and

$C^{L}_{i} = \underset{\sim}{C}^{A}_{i}.\underline{g}$, the m entries of $\underset{\sim}{C}^{A}_{i}$ being the m $P_{i,j}C_i$.

As a first attempt, the likelihood function of equation 5.18 ignores the error involved in estimating $\underline{g}$, mainly because one otherwise becomes committed to estimating $\underline{g}$ in the fitting procedure as well, and this does not seem to be practical considering the large number of extra parameters involved.

For simulated data typical of the South African anchovy (lmax > m), numerical searches for the maximum of the log-likelihood function (5.18) were generally unsuccessful, even when only small amounts of sampling error on f_l^{obs} were included, although the correct solution was obtained in the absence of sampling error. The analysis was taken no further than this, and instead simulations have been carried out to demontrate the biases (in abundance estimates) produced by using the approximation:

$$\underset{\sim}{C}^{A}_{i} = \underset{\sim}{C}^{L}_{i}\ \underline{a}^{66} \qquad \ldots\ 5.19$$

for catch age structure.

In subsequent simulations the "true" $\underset{\sim}{C}^{A}_{i}$ $(C_{i,j})$ are generated from a hypothetical anchovy fishery model (the same one used in previous examples but with different F_i values - see Table 5.5). However, the assessment of $\underset{\sim}{C}^{A,obs}_{i}$ from the length distribution

Table 5.5. The annual year effect fishing mortality values, the F_i's, used in the simulation of the age length key effect.

Hypothetical Year	F_i
1964	0.3
1965	0.2
1966	0.35
1967	0.3
1968	0.3
1969	0.3
1970	0.7
1971	0.8
1972	0.75
1973	0.7
1974	0.6
1975	0.6
1976	0.4
1977	0.5
1978	0.4

of the catch, $\underset{\sim}{C}^{L}_{i}$, is achieved by first calculating the hypothetical $\underset{\sim}{C}^{L}_{i}$ vector:

$$\underset{\sim}{C}^{L}_{i} = \underset{\sim}{C}^{A}_{i}\ \underline{g} \qquad \ldots\ 5.20$$

assuming that $\underline{g}$ is known exactly. $\underset{\sim}{C}^{A,obs}_{i}$ is then calculated using the 1966 anchovy age length key, $\underline{a}^{66}$:

$$\underset{\sim}{C}^{A,obs}_{i} = \underset{\sim}{C}^{L}_{i}\ \underline{a}^{66} = [\underset{\sim}{C}^{A}_{i}\underline{g}]\,\underline{a}^{66} \qquad \ldots\ 5.21$$

so that in general

$$\underset{\sim}{C}^{A,obs}_{i} \not\approx \underset{\sim}{C}^{A}_{i} \qquad \ldots\ 5.22$$

The implications of this inequality for VPA abundance estimates were investigated using the VPA likelihood function, assuming the following error structure for the input data:

$Q = 2500$

$\sigma = 0.05$

5.3.1 Catch and age data only

Model fit results, using these corrupted age structure values, are shown in Fig 5.1. This one off model fit suggests that a very large positive bias is produced in the estimated biomass, due to a number of separate effects. First of all, an age-length key which was estimated at a time when the stock was in the incipient stages of exploitation underestimates the extent of subsequent heavy exploitation, because the corrupted age structure annual mean ages do not decrease as fast as in reality (note in this simulation example the input F_i values are larger from 1970 onwards compared to 1966 (Table 5.5)). Thus the fishing mortality in later years is underestimated by the VPA, which consequently overestimates stock size.

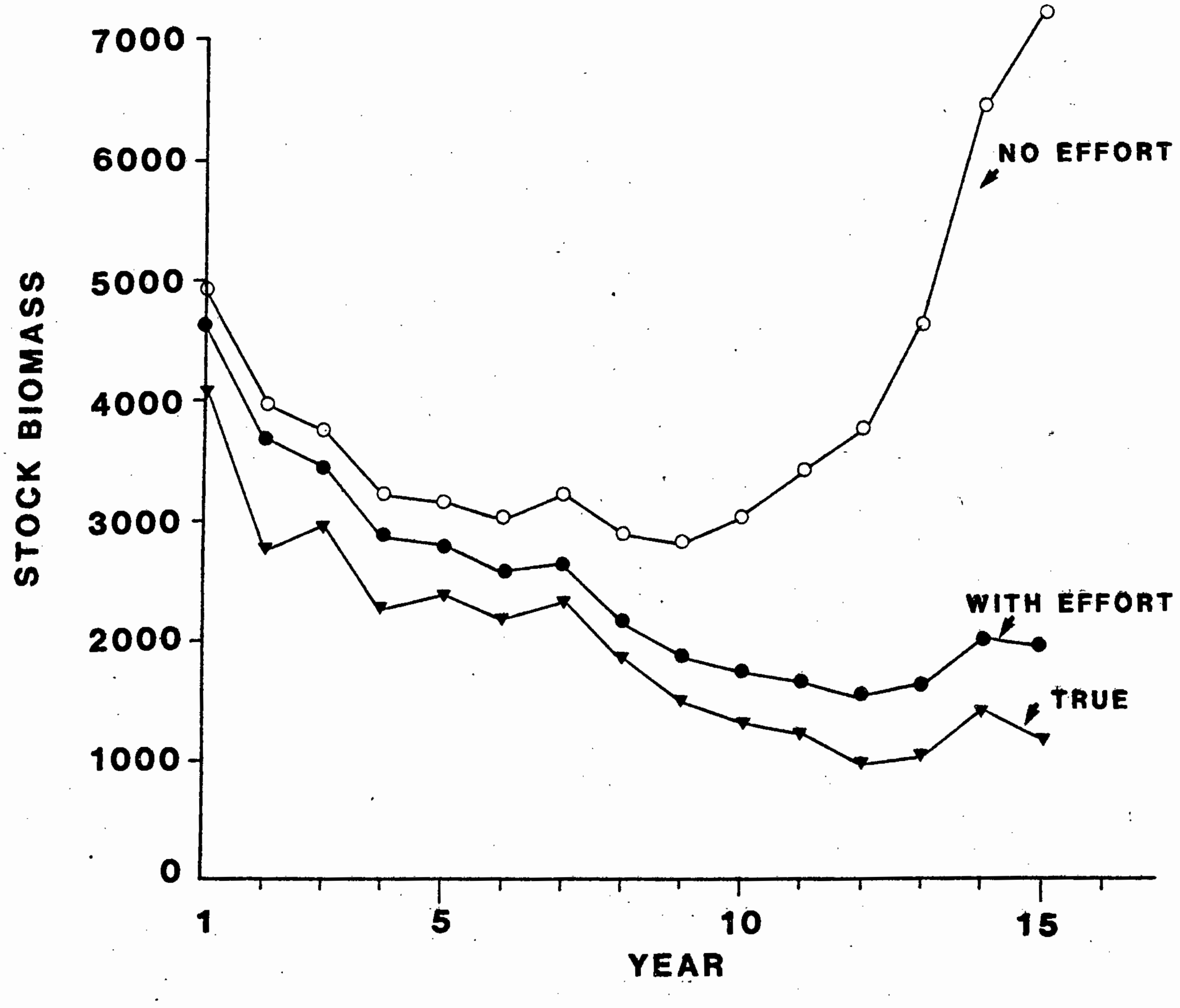


Figure 5.1. Results of single VPA model fits to simulated catch at age data, for which the same age-length-key is used in all years. (Given true parameter values were M=1, Q=2500, σ = 0.05, σ_1 = 0.3).

Note however that this is only the case if F_i, which is roughly proportional to the average annual fishing mortality, increases steadily after the year in which the age-length key is assessed. Conversely, F_i may decrease, in which case the biomass estimator may be negatively biassed. Furthermore, variations in recruitment strength change the age structure of the catch irrespective of the annual set of fishing mortalities, and this will impart bias to the abundance estimator (via the constant age-length-key effect under examination here).

Another concern is the adequacy of the catch at age data for biomass trend estimation purposes. From the point of view of the VPA this data is ambiguous, because changes in catch age structure may be due to a number of competing effects (see Appendix 5.1), of which two are changes in fishing mortality and changes in recruitment. Hence the result shown on Fig. 5.1 seems to be the consequence of the fitting procedure incorrectly allocating age structure changes between these effects. It is likely that the use of a single age length key over the complete history of the fishery facilitates this misinterpretation of changes in catches at age.

The case where reasonable effort data are available is now considered, since using this might reduce the bias arising from use of a single age-length key to within an acceptable range.

5.3.2 Effort included

If reasonable effort information E_i is included, under the assumption that effort is proportional to F_i, then the biomass trend is determined more rigidly. It is desirable to see whether this information is able to outweigh the detrimental effect of using a single age-length key for all years. The result of a single simulation (fitting only one set of data) using a single age length key but including effort information (σ_i=0.3) shown in Fig 5.1, is positively biassed biomass estimates.

Again it should be pointed out that the direction of the bias is indeterminable in practice, since changes in recruitment and F_i since 1965-67, the period during which samples for the age length key were obtained, could cause the bias to be either positive or negative, and Fig. 5.1 shows that this error may be greater than 50% for recent years.

5.4 ANCHOVY SPAWNING BIOMASS VARIANCE ESTIMATES

In the Fournier and Archibald (1982) VPA model, it is assumed that the total <u>number</u> of fish caught each year is estimated directly. This leads to a term involving C_i and C_i^{obs} in the likelihood function. However, the South African anchovy catch <u>mass</u> is monitored directly, distinguishing the local situation from that modelled by Fournier and Archibald. From the catch mass recorded, the total number of fish caught are calculated using the observed length frequency distribution and a fairly accurately known length-weight relationship. It is thus more reasonable to invoke lognormality for the estimate of catch mass (${}^{w}C_i^{obs}$), rather than for the estimates of total numbers caught (C_i^{obs}), because the latter are simply a derivation of other observations. This entails the replacement of the sum for C_i^{obs} and C_i in the likelihood function with one for ${}^{w}C_i^{obs}$ and ${}^{w}C_i$. Note that the model-fitted catch mass, ${}^{w}C_i$, in year i is given by:

$$ {}^{w}C_i = C_i \sum_{j=1}^{m} (P_{i,j} w_j) \qquad \ldots \; 5.23 $$

To maintain the link between ${}^{m}C_i^{obs}$ and C_i^{obs} in terms of the model parameters defined so far, C_i^{obs} should be expressed as:

$$C_i^{obs}={}^{m}C_i^{obs}/\sum_{j=1}^{m}(P_{i,j}^{obs}w_j) \qquad \ldots\ 5.24$$

Note that

$$C_i={}^{m}C_i/\sum_{j=1}^{m}(P_{i,j}w_j) \qquad \ldots\ 5.25$$

and therefore

$$[\ln {}^{m}C_i-\ln {}^{m}C_i^{obs}]^2=\ln^2[(C_i\sum_{j=1}^{m}P_{i,j}w_j)/(C_i^{obs}\sum_{j=1}^{m}P_{i,j}^{obs}w_j)]$$

$$=\ln^2[(\sum_{j=1}^{m}(C_iP_{i,j}/C_i^{obs})w_j)/(\sum_{j=1}^{m}P_{i,j}^{obs}w_j)]$$

$$=\ln^2[(\sum_{j=1}^{m}w_j\exp(\beta_{i,j})/(\sum_{j=1}^{m}P_{i,j}^{obs}w_j)] \qquad \ldots\ 5.26$$

[note that $\ln^2(x) = \ln(x)\ln(x)$]

The VPA log-likelihood function, LF, therefore has the following form, if σ and M are considered given and the effort term is not incorporated:

$$LF=\sum_{i=1}^{n}\sum_{j=1}^{m}q_{i,j}[(\beta_{i,j}-\ln[\sum_{j=1}^{m}\exp(\beta_{i,j})]]-$$

$$\sum_{i=1}^{n}\ln^2[(\sum_{j=1}^{m}w_j\exp(\beta_{i,j})/(\sum_{j=1}^{m}P_{i,j}^{obs}w_j)]/(2\sigma^2) \qquad \ldots\ 5.27$$

The effect of this modification is to allow a relatively greater deviation between $P_{i,j}^{obs}$ and $P_{i,j}$ for smaller fish (compared to the previous situation using catch numbers rather than mass), since they affect the total mass less for the same deviation by number.

Table 5.6 shows the beginning of year and mid-season anchovy masses at age which have been used in the VPA calculations. These were calculated using lengths from the anchovy growth curve quoted in DAMREP (1983), and the length weight regression equation reported by Davies et al (1981). In Table 5.7 the catch age structure, catch mass and effort data which were used are presented.

5.4.1 Results obtained using catch at age data only

A typical set of spawning (one years and older) biomass estimates for the South African anchovy are shown in Table 5.8 (both for $Q = 500$ and $Q = 2500$) together with the estimates of variance (shown as CV's) derived from the resampling procedure described in Chapter 4. In this particular case, M was given a value of 1 yrs^{-1}, and the assumed variances of catch mass estimates are equivalent to C.V.'s of 5%. Data from 1965 - 1967 is excluded from the analysis as it is considered unreliable following the changes in mesh size which occurred in the 1960's. The C.V.'s for $P_{i,j}^{obs}$ are equal to $[(1-P_{i,j}^{obs})/(QP_{i,j}^{obs})]$. Thus a five fold reduction in Q results in an increase in the variance of $P_{i,j}^{obs}$ by a factor of 5, but an increase in C.V. of only $5^{1/2}$ (i.e. 2.24 times). The result is an increase in the C.V. on biomass estimates of about 2 times (see Table 5.8).

5.4.2 Results obtained including effort information

Results obtained with effort information included are shown in Table 5.9, and in Fig. 5.2 a number of results are compared. The variance assumed for the F_i proportional to effort (E_i) error

Table 5.6. Beginning of year and mid-year anchovy masses used in the anchovy VPA (in grams). These were calculated using the DAMREP (1983) anchovy growth curve and length-weight regression constants quoted by Davies et al (1981).

Fish age	Beginning of year mass	Mid-year mass
0	0	5.19
1	8.5	11.3
2	13.0	15.9
3	18.1	19.5
4	20.5	21.6

Table 5.7. Catch age structure $P_{i,j}^{obs}$, catch mass ${}^{m}C_i^{obs}$ (thousand metric tons), and effort data E_i (thousands of standard boat days) pertaining to the South African anchovy fishery.

Year	$P_{i,1}^{obs}$	$P_{i,2}^{obs}$	$P_{i,3}^{obs}$	$P_{i,4}^{obs}$	$P_{i,5}^{obs}$	${}^{m}C_i$	E_i
1965	0.394	0.380	0.165	0.055	0.007	171.0	9.6
1966	0.780	0.168	0.039	0.012	0.001	143.9	8.1
1967	0.760	0.194	0.031	0.012	0.002	270.6	12.1
1968	0.698	0.251	0.037	0.012	0.002	138.1	11.1
1969	0.782	0.184	0.026	0.007	0.001	149.2	10.3
1970	0.815	0.150	0.027	0.007	0.001	169.3	11.1
1971	0.640	0.273	0.060	0.024	0.004	157.3	10.1
1972	0.507	0.372	0.081	0.033	0.006	235.6	13.1
1973	0.885	0.110	0.004	0	0	250.9	12.2
1974	0.564	0.369	0.058	0.009	0	349.8	8.3
1975	0.807	0.147	0.038	0.008	0	223.6	10.3
1976	0.681	0.271	0.040	0.007	0.001	218.3	8.6
1977	0.472	0.283	0.174	0.062	0.008	235.5	11.2
1978	0.660	0.294	0.038	0.007	0.001	209.5	13.0
1979	0.384	0.447	0.139	0.028	0.001	291.4	11.0
1980	0.537	0.367	0.079	0.016	0.001	315.5	10.1
1981	0.577	0.328	0.075	0.018	0.001	292.0	10.8
1982	0.730	0.220	0.041	0.007	0	306.1	11.1
1983	0.710	0.225	0.060	0.005	0	240.2	14.6

Table 5.8. South African anchovy spawning biomass estimates and resampling estimates of the variance (% C.V.'s in brackets), using the catch mass and age structure estimates only, and assuming that Q=2500 or 500, $\sigma=0.05$ and $M=1$ yr^{-1}.

YEAR	Q=2500	Q=500
1968	116 (4.9)	132 (8.5)
1969	116 (5.5)	134 (7.7)
1970	151 (5.5)	176 (6.7)
1971	202 (4.9)	241 (7.0)
1972	207 (4.4)	245 (8.1)
1973	125 (6.4)	155 (13.0)
1974	308 (4.2)	379 (8.0)
1975	232 (5.5)	297 (11.0)
1976	403 (4.4)	524 (7.1)
1977	471 (4.3)	595 (7.5)
1978	417 (4.4)	529 (7.3)
1979	553 (3.6)	667 (6.3)
1980	387 (4.8)	458 (9.5)
1981	293 (8.0)	358 (21.0)
1982	220 (21.0)	286 (39.0)
1983	199 (55.0)	312 (89.0)

Table 5.9. South African anchovy spawning biomass estimates and resampling estimates of variance (% C.V.'s in brackets), using the catch mass and age structure estimates together with effort data. The following model parameter values were assumed: Q=2500, $\sigma^2=0.05$, $\sigma_1{}^2=0.09$ and $M=1$ yrs^{-1}.

YEAR	ESTIMATE	% C.V.
1968	118	5.7
1969	120	6.5
1970	157	6.3
1971	201	4.6
1972	207	4.7
1973	130	6.2
1974	307	5.1
1975	234	6.7
1976	394	5.5
1977	442	5.0
1978	412	5.4
1979	530	5.1
1980	391	5.5
1981	297	6.8
1982	235	11.0
1983	246	21.0

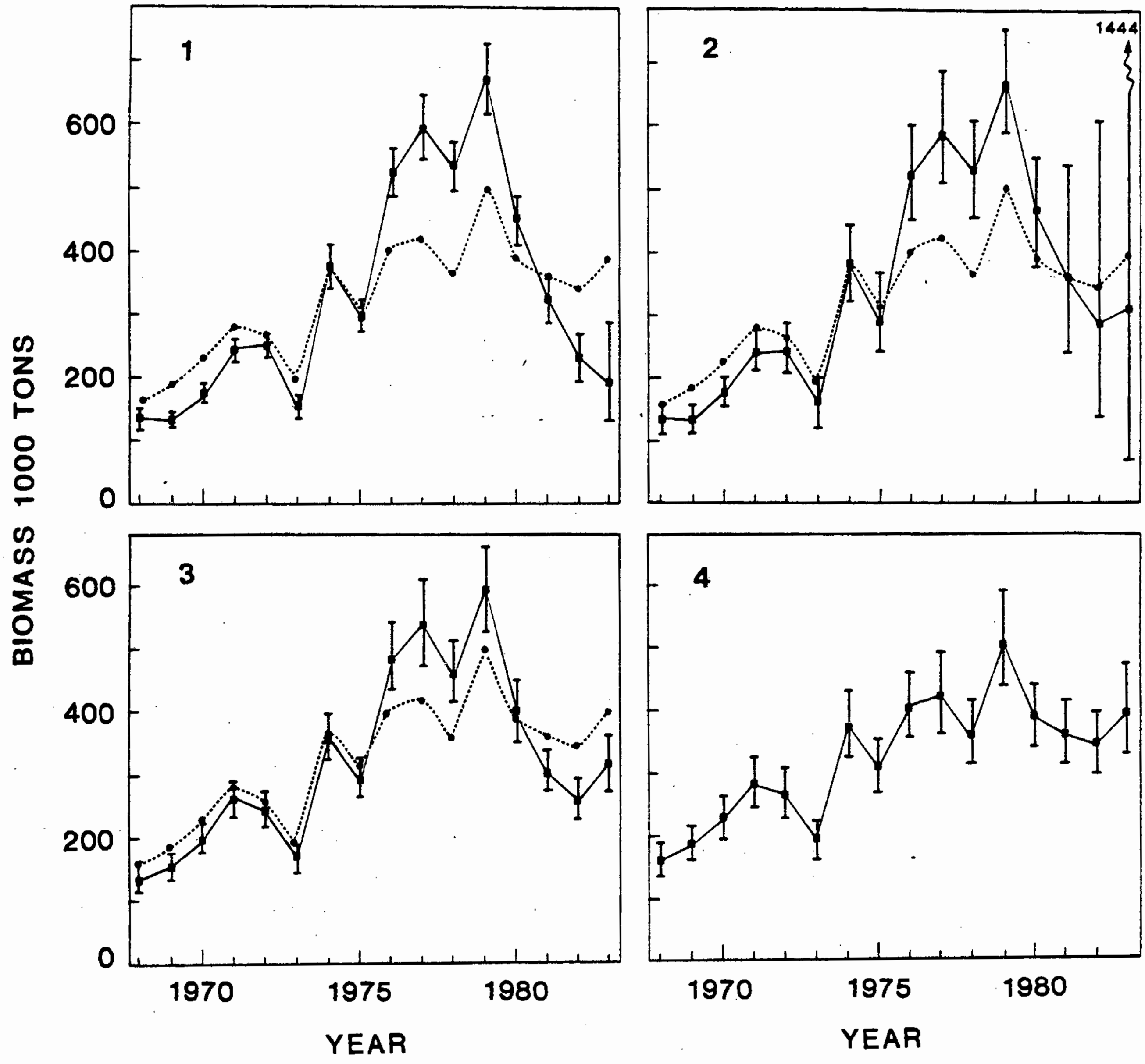


Figure 5.2. 1 - Mean anchovy abundance estimates obtained without effort information, Q=2500, σ=0.05; 2 - without effort information, Q=500, σ=0.05; 3 - with effort information, Q=2500, σ=0.05, σ_1=0.30; 4 - with effort information, Q=500, σ=0.05, σ_1=0.30. The 95% confidence limits are shown for each trajectory in turn. Trajectory 4 is superimposed in dots on plots 1 - 3 for comparison.

is equivalent to a C.V. of about 30%. To take account of the uncertainty regarding appropriate values for these latter variances, model fits were performed for reasonable upper and lower limits for these quantities, these being C.V.'s of between 5 and 15 % for the catch mass values and model error of between 20 and 40 % (C.V.) for the effort relationship. The results (see Table 5.10) suggest that substantial relative differences in biomass estimates only occur in recent years.

5.4.3 The importance of specifying selectivity functional form

The importance of choice of selectivity function for anchovy is illustrated by the spawning biomass estimates set out in Fig 5.3. These results were obtained using:

$$S_j=\alpha_j \quad \text{for all } j \qquad \ldots\ 5.28$$

rather than

$$S_j=j/(j+b)$$

in the model, for which the terminology "relaxing the selectivity function requirement", or "not specifying the selectivity function", is used here. The different biomass trajectories shown in Fig. 5.3 result from fits made to equally probable catch at age data sets generated by the resampling procedure described in Chapter 4. Very different biomass estimates and trends are obtained. The standard deviations of biomass estimates obtained with and without specification of the selectivity function are shown in Table 5.11. Ratios of standard deviations shown there demonstrate the increased dispersion in estimates which results when the selectivity functional form is relaxed. Results are presented for Q = 1000 and Q = 400. At Q = 400 the increase in variance when relaxing the selectivity function requirement is

Table 5.10. Biomass trends for anchovy, for the four possible combinations of upper and lower estimates for σ^2 and $\sigma_1{}^2$ (units: thousand metric tonnes, Q=2500, M=1 yrs^{-1}). The percentage C.V. estimates are shown in brackets.

σ^2	0.0025		0.0223	
$\sigma_1{}^2$	0.039	0.148	0.148	0.039
1968	132 (5)	133 (5)	124 (23)	116 (22)
1969	132 (5)	131 (5)	109 (17)	109 (17)
1970	174 (5)	172 (5)	157 (17)	162 (18)
1971	242 (4)	241 (4)	277 (17)	279 (15)
1972	248 (3)	248 (4)	313 (20)	308 (16)
1973	150 (6)	149 (6)	146 (18)	147 (15)
1974	370 (5)	372 (5)	395 (19)	384 (16)
1975	290 (5)	292 (4)	310 (17)	306 (15)
1976	511 (4)	518 (4)	622 (17)	589 (15)
1977	578 (5)	590 (4)	782 (18)	714 (16)
1978	514 (4)	526 (4)	665 (15)	586 (14)
1979	655 (4)	667 (4)	851 (18)	715 (17)
1980	443 (4)	446 (4)	410 (16)	387 (14)
1981	332 (6)	325 (5)	257 (24)	257 (22)
1982	255 (10)	238 (8)	179 (25)	193 (27)
1983	260 (19)	219 (16)	196 (33)	225 (38)

substantial.

A value of Q = 400 is just beyond the range (on the small side) of Q values suggested by the results shown in Table 5.2. However, in obtaining the values of Q shown in Table 5.2, the only error in catch age structure assessment considered was that due to randomly sampling the catch. Since there are other factors which are likely to increase the imprecision of catch age structure values even further (e.g. errors in assessing age), a value of Q = 400 is probably still quite plausible.

The cause of the wide dispersion in estimated biomass trajectories with the selectivity function unspecified may be due to ambiguities inherent to the catch at age data, and consequently the flatness of the VPA likelihood function with respect to model parameters. For the current formulation of the VPA problem (with selectivity unspecified), there are 3 different possible explanations for high apparent total mortalities in catches-at-age (see Appendix 5.1). These are either persistently increasing recruitment trends, high fishing mortalities (F_i), or selectivities (S_j) decreasing with age. Stock emigration is another possibility; however the model as set up here does not permit fitting of emigration parameters. The fitting procedure can thus fairly arbitrarily interpret the observed catch age structure (and apparent Z level) as being due to combinations of high F_i, S_j decreasing with age, and recruitment increasing with time. Further changes in that structure (a decreasing mean catch age) can also be interpreted either as increasing F_i's, or as an increasing recruitment trend.

Fig. 5.3 shows the wide variety of interpretations possible when the selectivity function requirements are relaxed; each

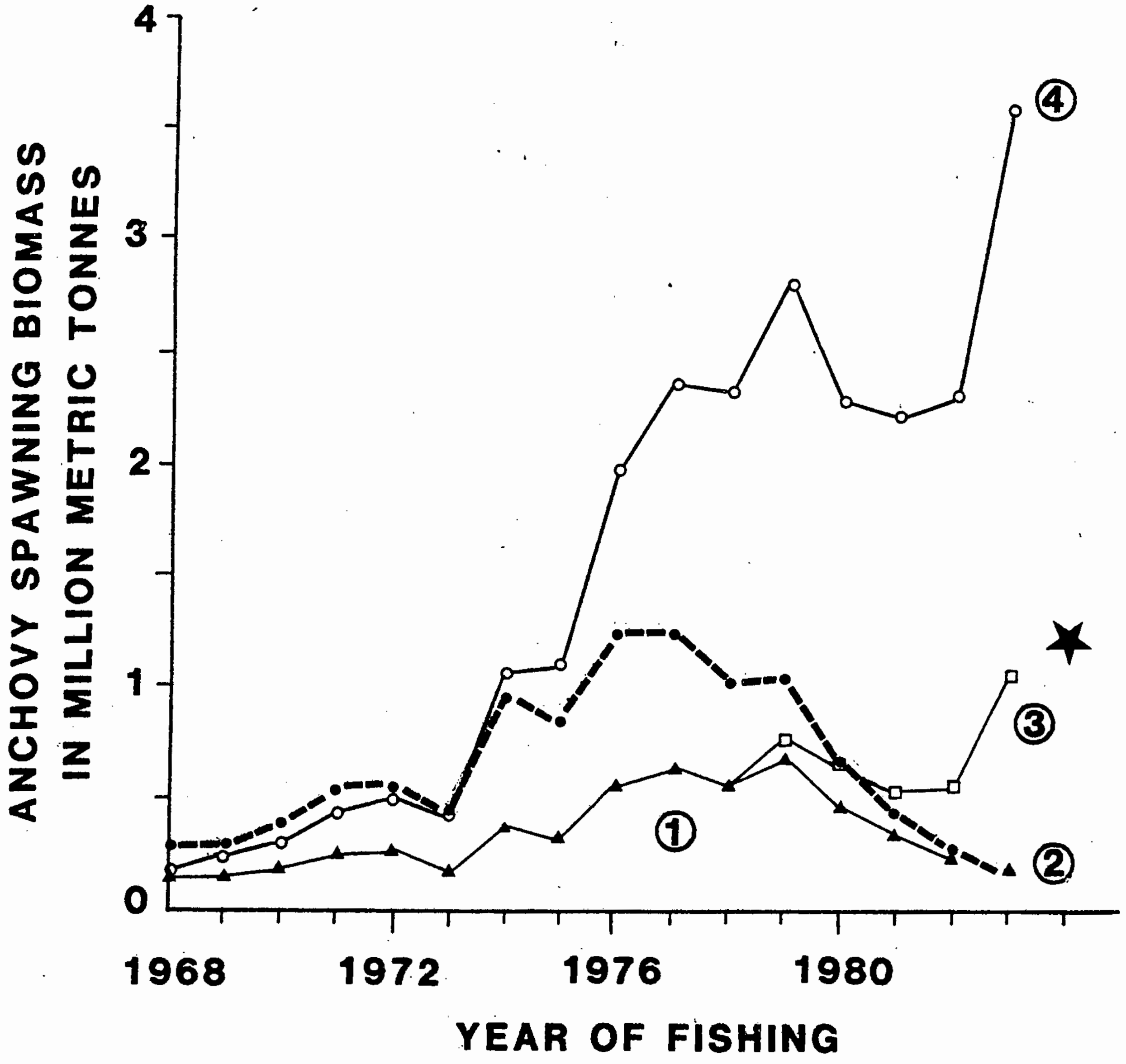


Figure 5.3. A selection of VPA model fits to various resampled data sets obtained when the selectivity function is not specified and without use of effort information. The single model parameter b associated with the selectivity functional form ($S_j = j/(b+j)$) has been replaced by m-1 parameters (m=5 for anchovy), i.e. a selectivity parameter for each age class except one (There is an arbitrary parameter since fishing mortalities are given by the product of F_i and S_j). The star shows the November 1984 direct survey estimate of anchovy spawning biomass.

Table 5.11. Standard deviations and their ratios for biomass estimates obtained with (σ_s) and without (σ_{ns}) specification of selectivity function form for the South African anchovy. Effort information is not included in both cases. Units of the standard deviations are thousand tonnes.

	Q = 1000			Q = 400		
	σ_{ns}	σ_s	(σ_{ns}/σ_s)	σ_{ns}	σ_s	(σ_{ns}/σ_s)
1968	9.7	7.3	1.33	29.4	11.3	2.61
1969	10.4	8.3	1.56	41.5	12.6	3.30
1970	13.8	10.7	1.29	63.6	13.3	4.79
1971	21.7	13.0	1.66	96.5	19.5	4.95
1972	24.1	12.2	1.96	107.8	18.7	5.76
1973	21.7	9.9	2.19	94.4	17.2	5.48
1974	48.6	17.4	2.79	222.5	30.1	7.39
1975	54.2	18.8	2.89	267.8	25.1	10.70
1976	83.4	27.1	3.08	413.5	28.0	14.80
1977	86.0	29.7	2.89	408.7	29.6	80.00
1978	66.4	24.4	2.72	333.2	32.9	10.10
1979	48.5	25.2	1.93	328.6	50.1	6.55
1980	34.5	23.2	1.49	246.6	46.8	5.27
1981	33.5	26.3	1.27	223.8	50.2	4.46
1982	45.7	45.8	1.00	224.5	70.2	3.20
1983	114.2	100.2	1.14	328.6	167.3	1.96

corresponds to different yet equally probable data sets (resampled).

A recent survey estimate (based on acoustics and egg production measurements, Armstrong and Butterworth, 1986) of the anchovy spawning biomass is shown as an asterisk on this figure, and this information can potentially be used to eliminate unlikely biomass trajectories by including a term for this estimate in the VPA likelihood function. However considering the imprecision and probable inaccuracy in the anchovy VPA, this does not seem warranted, and the current practice of using the survey estimates alone seems definitely preferable.

5.5 DISCUSSION AND CONCLUSIONS

Variance calculations for VPA biomass estimates for anchovy show that without effort information, recent biomass estimates are poorly determined, with C.V.'s of as large as some 90% (Table 5.8). Including the (questionable) measure of gross pelagic fishing effort, the most recent biomass estimate is given with a C.V. of some 20% (Table 5.9).

On the other hand, if the fishing selectivity function is not specified (emigration is suspected to occur which would be mimicked by selectivity decreasing with age rather than being asymptotically flat - Appendix 5.1) and effort information is excluded, the biomass estimates are intolerably poorly determined, with certain parameter estimates being confounded with others (Fig. 5.3). The standard deviation of abundance estimates for the case where the selectivity function is unspecified are shown on Table 5.11.

Estimator precision is increased by including effort data in the likelihood function. In doing this, the resultant estimates of stock size depend heavily on the validity of the assumed relationship between F_i and effort E_i. A number of good reasons can be given suggesting $F_i \risingdotseq kE_i$. For example, anchovy form large shoals with densities in the region of 15 kg m^{-2}. A typical fishing day includes the time taken to get to the most favourable region, the time taken to make a number of sets, and the time taken to move from one shoal to the next. It seems that only the last measure, which is some measure of inter-shoal separation, has a direct link with population density; therefore the available fishing time measures (see Table 5.7) are probably not appropriate for the purposes required here.

In addition, there is observational evidence suggesting that the habitat range of the Northern anchovy <u>Engraulis</u> <u>mordax</u> decreases with decreasing population size (MacCall, 1983), which would destroy the linearity of the relationship between stock density and abundance. The relationship between abundance and density for these cases is probably more of the form:

$$\text{fish density} = kB_i^{a} \quad 0 < a < 1 \qquad \ldots 5.29$$

where B_i is the fish abundance by mass in year i. Therefore, still assuming CPUE is proportional to fish density:

$$CPUE_i \propto B_i^{a} \qquad \ldots 5.30$$

or

$$k\ CPUE_i^{1/a} = B_i \qquad \ldots 5.31$$

For the previous model (k CPUE = B) the relative annual change of population biomass, $[(B_{i+1}-B_i)/B_i]$ is equal to the relative change in the catch rate $[(CPUE_{i+1}-CPUE_i)/CPUE_i]$. However for the above model:

$$[(B_{i+1}-B_i)/B_i] = (1/a)[(CPUE_{i+1}-CPUE_i)/CPUE_i] \quad \ldots\ 5.32$$

so that a unit relative change in population biomass is 1/a times greater than a unit relative change in CPUE (1/a > 1 since $0<a<1$). For the South African anchovy, one would therefore expect that the estimates of biomass shown in Table 5.9, obtained assuming E_i and F_i are linearly related (crudely equivalent to CPUE proportional to population biomass), underestimate the true fluctuations in recruitment and stock size.

There are further problems with measures of local pelagic fishing effort, the main one being the fact that effort has not been split between species, so that the available data reflect only the gross undirected effort. These gross effort values are the ones listed in Table 5.7 and used in results including effort data.

5.5.1 The basis for an anchovy VPA

At this time the basis for carrying out a VPA assessment of the South African anchovy is being seriously undermined by the results of recent ageing (Armstrong and Prosch, pers. commn) and direct survey abundance estimates (Armstrong and Butterworth, 1986).

Even if fishing proceeds in an approximately spatially random fashion over the area fished, the appropriateness of VPA is critically dependent on the absence of emigration. This recent research provides strong evidence that substantial emigration is taking place. Daily growth ring studies (Armstrong, Prosch, pers. commn) indicate that the current (and probably also the historic) anchovy catches consist almost exclusively of recruits (greater than 90% by number), compared to previous estimates averaging some 60% (see Table 5.7). This means that the catch age structure estimates, the $P_{i,j}^{obs}$, upon which the VPA calculations reported earlier in this chapter are based, are substantially in error, and it appears that they are in fact primarily artifacts obtained from the application of a single age length key to the length distribution of recruits. This is corroborated by hydroacoustic surveys, which point to a large anchovy spawner stock inhabiting the Agulhas Bank well away from the main fishing grounds.

Anchovy reaching the West Coast fishing grounds appear therefore to be recruits in their first year of life, which subsequently migrate southwards onto the Agulhas Bank as they mature. This is illustrated on Fig 5.4. If this is eventually conclusively shown to be the case, then the VPA estimates discussed earlier in this chapter are purely artifactual. To accommodate the migration pattern into the VPA, it would be necessary to inflate natural mortality to a level at which it included the effect of migration away from the fishing grounds. On the other hand, leaving the selectivity function unspecified gives the model further opportunity to interpret large amounts of emigration from the system as steadily increasing recruitment, since this would help to explain an unrealistically large apparent Z in the age structure of the catch. Trajectory 4 of Fig. 5.3 seems to have opted for this interpretation.

A fundamental problem with the application of VPA to the South African anchovy stock has thus been revealed by a single hydroacaustic survey, which suggests that the stock is distinctly geographically stratified by age. The fishery may thus have been sampling essentially only one year class each year, namely recruits, and this eliminates the VPA as a viable method for stock assessment purposes. The only useful catch statistics data left are the catch masses and the effort data, and the latter is also of questionable relevance in view of the shoaling behaviour of pelagic fish. Thus there are no reliable indirect indices of historic or present population biomass, and previous discussions of anchovy biomass trends presented in the literature and based on VPA must be considered with scepticism.

Currently, sustainable yields for anchovy are being calculated using the very recent catch history and direct survey estimates of population biomass (values are given in Armstrong and Butterworth, 1986).

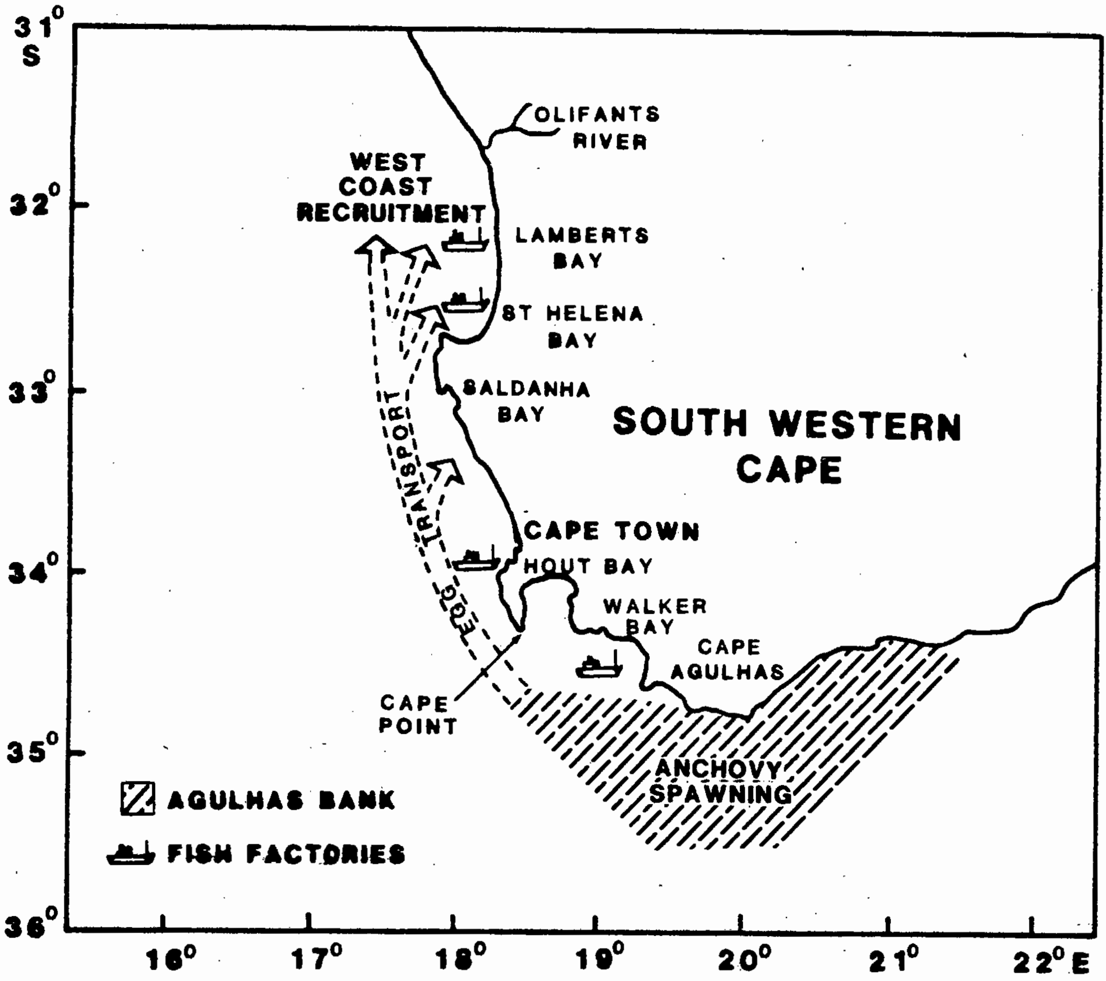


Figure 5.4. Map showing roughly the site of anchovy spawning. Fishing takes place mainly in the vicinity of the fish factories indicated (i.e. mainly on the west coast). Recent surveys have indicated that fish on the west coast (north of Cape Point) are in their first year of life, which is consistent with the model of egg transport and west coast recruitment suggested on the map. Anchovy first spawn at roughly 1 of age, so that the fish which occur on the west coast are continuously migrating southwards to their spawning grounds. Thus fish become less available to the fishery as they age.

6. OBTAINING A PILCHARD RECRUITMENT INDEX FROM COMMERCIAL GUANO RECORDS

6.1 INTRODUCTION

Seabird guano has been collected from islands off the South African and Namibian coasts since at least as far back as the nineteenth century for use as an agricultural fertilizer. The government has controlled this industry since 1896, and guano harvests from each island have been recorded in various annual government reports. The total South African and Namibian guano harvests for the period 1897 -1975 are shown in Fig 6.1. Islands off the coast of Peru are also sites of guano production; however the annual guano harvest there is far more impressive, and annual yields of between 50 and 200 times larger than the South African harvest are common. For example, in 1962 guano harvests were 1.1, 4.6 and 206 thousand metric tonnes in South Africa, Namibia and Peru respectively.

Off the South African west coast, enhanced coastal upwelling, caused by coastal topography and longshore winds, brings nutrient rich water up into the euphotic zone where it is utilised by phytoplankton (Andrews and Hutchings 1980). This vigorous primary production supports large stocks of zooplankton, and both zooplankton and phytoplankton are consumed in large quantities by pilchard and anchovy. Pilchard and anchovy are in turn the main prey of the guano producing seabirds, the Cape Gannet _Sula capensis_, Cape Cormorant _Phalacrocorax capensis_ and Jackass Penguin _Spheniscus demersus_. In this chapter these three species are loosely termed "seabirds". It is likely that seabird population dynamics, and hence guano production, are closely related to the dynamics of their prey, and therefore the guano

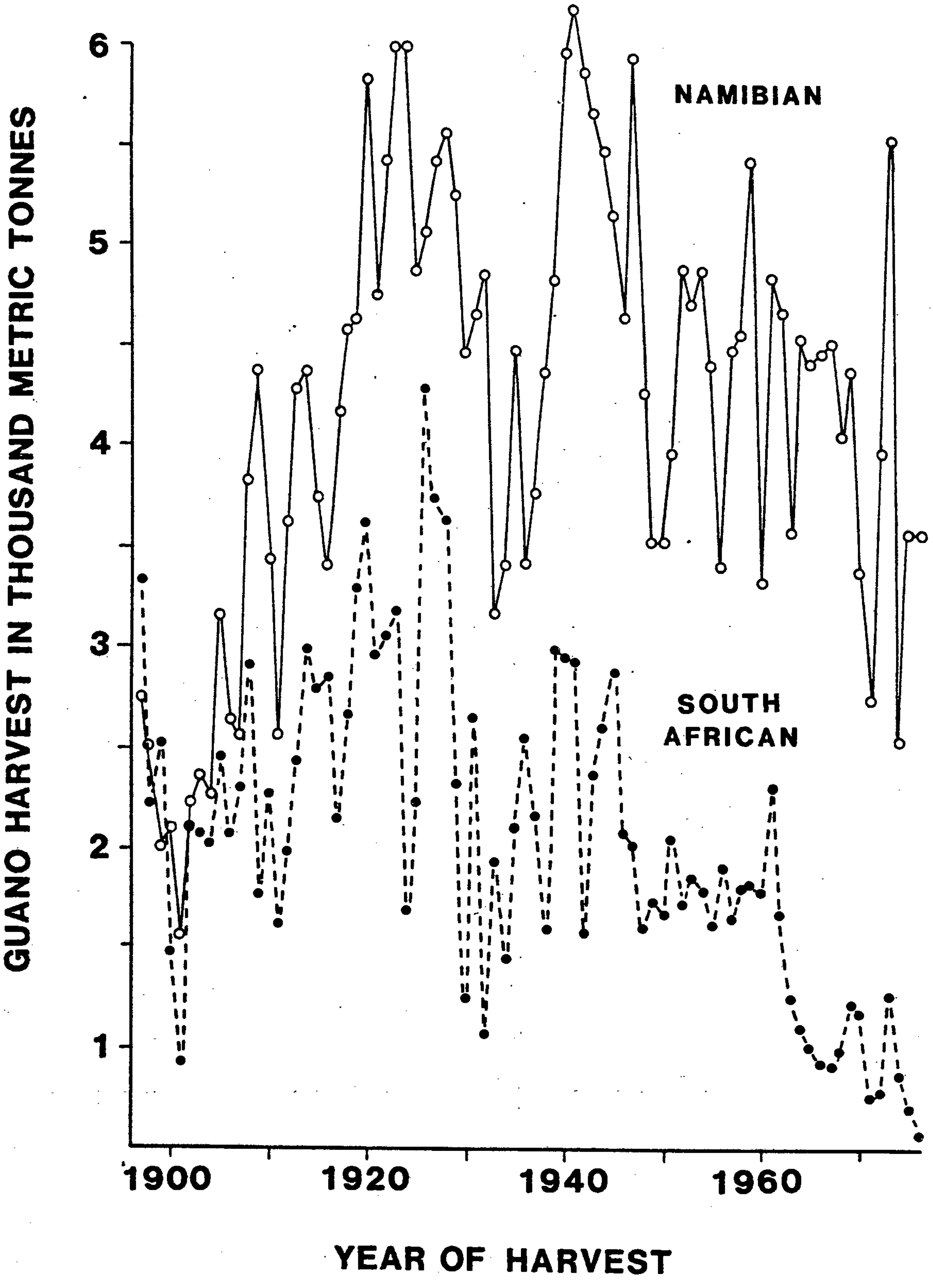


Figure 6.1. Estimates of total guano harvests made at South African and Namibian islands. (The location of various islands is shown in Fig. 6.2.)

data may be useful for fishery management. More specifically, the guano data may contain some relative abundance information or some other index, which could be used in conjunction with catch statistics in the VPA to improve the precision of pelagic fish abundance and trend estimates. This lack of precision is a logical consequence of the inadequacy of the catch-at-age data, as noted in previous chapters.

The work reported here is an attempt to relate guano production to pelagic fish population dynamics - an approach suggested by Crawford and Shelton (1978). However, whereas Crawford and Shelton (1978) concentrated on similarities in trend between the various island guano records and fish populations, the following research goes further in analysing the time dependent behaviour of the guano records, and possible consistencies between guano production at separate islands. The final analysis is treated in the next chapter (7) where the relationships hypothesised and tested here are used to produce pilchard biomass trends which are consistent with models of guano production.

Guano harvest records at 17 islands off the South African and Namibian coasts are subjected to 3 analyses, viz. for linear correlation using Pearson's correlation coefficient, and for periodic behaviour using first univariate and then bivariate spectral analysis techniques. These results indicate certain patterns within the guano records which may be consistent with aspects of pilchard population dynamics. Evaluations concerned with possible complex links between guano yields at certain islands provide additional evidence that a pilchard recruitment index is given by the Lambert's Bay Bird Island guano harvest.

The location of the principal guano producing islands is shown in

Fig. 6.2. Seventeen of these "islands" (one of them represents a combination of the guano harvest from all the wooden man-made platforms off Namibia) were harvested more or less regularly, and can therefore be compared meaningfully. Although South African fish stocks are of primary interest as far as this document is concerned, Namibian guano productions may prove to be important to this analysis. This is because historically the bulk of the Namibian guano harvest was derived from the southern group of Namibian islands, and it is possible that at that time the birds were feeding on a northern limb of the South African pilchard stock. Newman (1970) presents tagging study results showing that Namibian pilchards are practically isolated from the South African stock. It is significant however that the pilchards considered in his study were tagged in the Walvis Bay vicinity (only a few of those tagged were recovered in Table Bay in the South African fishing grounds) which was at the centre of the Namibian pilchard fishery. This is well north of the group of Namibian guano islands (see Fig. 6.2) considered here. Accordingly Newman's results, while indicative of two stocks, do not necessarily fix the inter-stock boundary at the Orange River - it could be further north.

6.2 DIFFERENT ANALYSIS PROCEDURES APPLIED TO GUANO HARVEST TIME SERIES

6.2.1 Correlations between guano harvest time series

The significance levels of elements of the Pearson's correlation matrix of guano records at the 17 islands plus the two trends of total harvest indicated in Fig. 6.1 are shown in Fig. 6.3. It is clear from this that significant correlation exists between a number of islands:

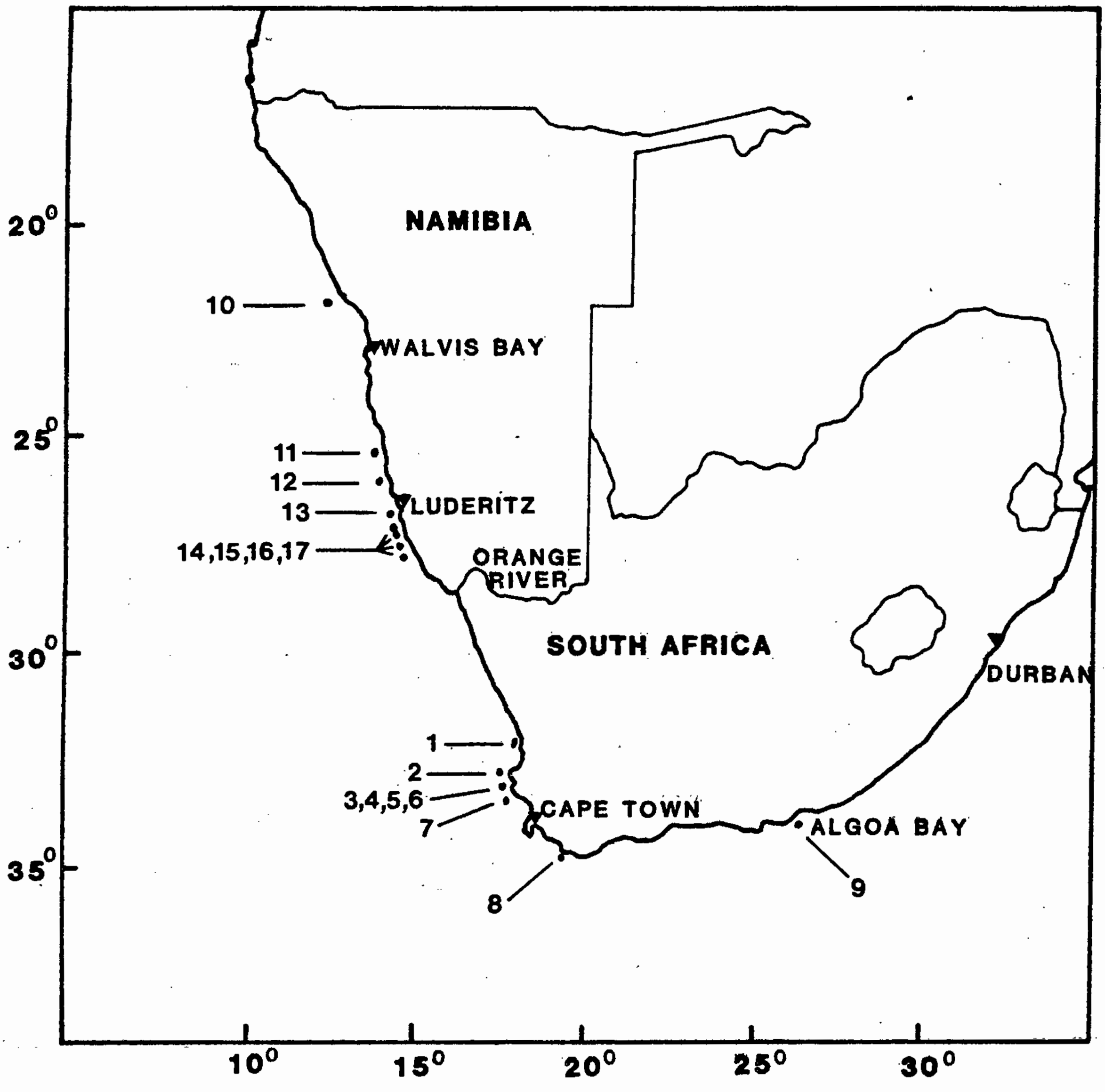


Figure 6.2. Approximate location of 17 guano islands along the South African and Namibian coastlines:

1. Bird Is., Lambert's Bay; 2. Paternoster Rocks; 3. Malgas Is.; 4. Marcus Is.; 5. Jutten Is.; 6. Vondeling Is.; 7. Dassen Is.; 8. Dyer Is.; 9. Bird Is., Algoa Bay; 10. Cape Cross Platforms; 11. Mercury Is.; 12. Ichaboe Is.; 13. Halifax Is.; 14. Possession Is.; 15. Pomona Is.; 16. Plumpudding Is.; 17. Sinclair Is.

		S.A.									NAMIBIAN									
		CG 1	C 2	GC 3	P 4	CP 5	CP 6	PC 7	PC 8	GP 9	C 10	G 11	CP 12	P 13	GP 14	P 15	P 16	P 17	SA 18	NAM 19
S.A.	CG 1																			
	C 2	■																		
	GC 3		▣																	
	P 4			■																
	CP 5			■																
	CP 6				□	■														
	PC 7		▣			■														
	PC 8			■		■														
	GP 9			▣				▣												
NAMIBIAN	C 10																			
	GC 11					▣	□	■	□	■										
	CP 12			■		■		▣	■	▣		■								
	P 13		■	■		▣		◘	■		■	■	■							
	GP 14	■	▣	■		■		■	■	□	■	■	■	■						
	P 15	▣		■		□		■		■		■	▣	■	■					
	P 16			■					■		▣		□	■	■	■				
	P 17		▣	■					■		◘		▣	■	■	▣	■			
	SA 18	■	■	■	▣	■	■	■	■			■	■	■	■	■	◘	◘		
	NAM 19					■		■		■	▣	■	■	▣	■	■	■	■	▣	

■ P < 0.0005 ◘ P < 0.001 ▣ P < 0.005 □ P < 0.01

Figure 6.3. Significance levels of Pearson's correlation coefficients for relationships between all unique pairs of guano harvest series at the 17 islands under consideration, and including the totals for South Africa (18) and Namibia (19) (there are 171 possible unique pairs). The letters C, G and P refer to the dominant seabird species on each island, viz. cormorants, gannets or penguins respectively.

1. The southern Namibian Islands (numbers 11-17) form a well correlated group with coefficients of determination (r^2) as high as 80%, possibly because they are all very close to each other.

2. Some South African islands (eg. numbers 3,5,7,8) are significantly correlated with some Namibian Islands. If guano and fish yields are strongly related, then this indicates that either prey stocks in these two regions are part of a single large stock, or there are two separate stocks responding to the same set of dominating environmental conditions. Note however that the Namibian islands under consideration are located south and away from the main Namibian fishing grounds (which were centred close to Walvis Bay for the period that pilchard dominated that fishery, and have subsequently spread northward rather than southward). Only since about 1950 with the construction of platforms has guano from the vicinity of Walvis Bay, where fishing effort was primarily directed, contributed significantly to the total Namibian guano production.

3. Some South African islands are well correlated with each other though not as many as in Namibia. This is probably due to the South African birds feeding on a mixed diet (Rand, 1960a,b), and the greater distances between the South African islands.

Studies have confirmed that pilchards were at one time important in seabird diets, and this suggests the possibility of some link between pilchard dynamics and guano production. For example, before the stock off Namibia was heavily fished in the late 1960's, pilchard formed more than 85 % of seabird diet. Matthews

(1961) found that for the years 1957 - 1958, pilchard comprised about 87 % by mass of the diet of gannets, 94 % of the diet of penguins and 87 % of the diet of Cape cormorants.

Off South Africa, again before the pilchard was drastically depleted by overfishing in the 1960's, seabird diet contained about 50% pilchard by mass (Rand, 1960a, b). This reflects the multispecies nature of the South African pelagic fishery, which is based on 6 main species, compared to (at that time) the mainly pilchard based Namibian fishery. Apart from its dominance by weight in the diets of the birds, pilchard is also more nutritional, having a higher lipid content and calorific value than other prey species. This is reflected in gannet fledgling growth rates, which are faster on a diet of pilchard than on hake (Batchelor, 1982). On the basis of this, and the dietary evidence mentioned above, it is reasonable to expect that seabird population dynamics and therefore presumably their guano production as well have been strongly related to pilchard population trends - at least prior to the South African pilchard population collapse.

Crawford and Shelton (1978) have demonstrated that guano production at certain islands is significantly related to VPA biomass estimates of pilchard population size and recruitment. However, the VPA estimates they utilised are of dubious value, since there is considerable debate about pilchard growth rates (and hence estimates of the catch age structure), and only catch mass and age structure information was used, which is insufficient for determining stock trends precisely (Chapter 3; Pope and Shepherd, 1982). Furthermore, there are potentially a number of complicating factors, which make it unlikely that pilchard stock size and annual guano harvest are linearly related. Some of these are;

1. Because of their discrete locations, islands may, at best, indicate changes in local fish abundance.

2. Variations in forage time, droppings released at sea, and prey-switching further complicate the link between guano and pilchard.

3. Heavy rainfall can wash away substantial amounts of guano.

4. If the birds are food limited and they have a sufficiently high potential rate of reproduction, then fluctuations in guano may simply be a reflection of only slightly damped or undamped oscillations in the bird population with a relatively constant fish abundance (see subsequent analyses for a detailed explanation).

5. When islands are not scraped clean each year, the commercial yield of guano is unrelated to its biological rate of production. This may have occurred when demand for guano was low (this is not well documented by government publications). For example a decrease in the total South African guano yield in the early 1930's (see Fig. 6.1) may have been due to the economic depression at that time.

6. Massive bird exoduses from islands, due for example to tick infestations, have been recorded off Peru and may have occurred off South Africa.

7. Migration of birds between islands would complicate the interpretation of guano data.

8. Increased guano production caused by the construction of artificial wooden platforms for breeding sites (as has occurred north of Walvis Bay) is unrelated to changes in fish abundance.

9. Inter-annual environmental variation could affect guano

production.

These factors confound attempts at distilling ecologically meaningful information from the guano records. The Pearson's correlation coefficients shown on Fig. 6.3 suggest however that some common factor does influence guano production. Some insight into the time dependence of guano records using spectral analysis techniques may be useful for identifying common factors affecting guano production.

6.2.2 Univariate spectral analyses of commercial guano harvest records

In general, the 2m+1 (=n, i.e. n odd) terms of a time series x(t), t=1,2,...,n can be written as a sum of their harmonic constituents:

$$x(t)=a_0/2+\sum_{k=1}^{m}[a_k\cos(2\pi kt/n)+b_k\sin(2\pi kt/n)] \quad \ldots\ 6.1$$

where the values of the constants a_0, a_k and b_k are given by

$$a_0=(2/n)\sum_{t=1}^{n}x(t) \quad \ldots\ 6.2$$

$$a_k=(2/n)\sum_{t=1}^{n}x(t)\cos(2\pi kt/n) \quad \ldots\ 6.3$$

$$b_k=(2/n)\sum_{t=1}^{n}x(t)\sin(2\pi kt/n) \quad \ldots\ 6.4$$

k=1,2,...,m

The harmonic amplitude, c_k, is defined as

$$c_k=(a_k^2+b_k^2)^{0.5} \quad \ldots\ 6.5$$

The normalised unsmoothed sample spectrum estimate, $f^*(w)$, calculated by the NAG routine G13CBF

$$f^*(w)=(1/(2\pi))\ |\sum_{t=1}^{n} x(t)e^{iwt}|^{2} \qquad \ldots\ 6.6$$

$(i=(-1)^{1/2})$ is simply the normalised quantity g_k (Shimshoni, 1971) (for $w=2\pi k/n$ i.e: $2\pi/n, 4\pi/n, 6\pi/n\ldots (n-1)\pi/n$.

$$g_k=c_k^{2}/(\sum_{j=1}^{m} c_j^{2}) \qquad \ldots\ 6.7$$

$f^*(w)$ can be written as

$$f^*(w)=(1/(2\pi))[(\sum_{t=1}^{n} x(t)\cos(2\pi kt/n))^{2}+(\sum_{t=1}^{n} x(t)\sin(2\pi kt/n))^{2}]$$

$$=(1/(2\pi))[a_k^{2}\cdot n^{2}/4+b_k^{2}\cdot n^{2}/4] \qquad \ldots\ 6.8$$

and the normalised version of this is therefore the same as g_k. Therefore the significance levels for g_k estimated by Shimshoni (1971) can be used to test the significance of the unsmoothed sample spectrum estimate, $f^*(w)$, calculated by the NAG routine G13CBF. Shimshoni (1971) gives estimates of the critical value of the estimated spectral density function ($f^*(w)$) for the r th highest peak. The first ten highest peaks were arbitrarily chosen for purposes of analysis. It was therefore necessary to first rank the ten highest spectrum estimates in order to test for significance. For n=78 years, the critical levels have been roughly interpolated from Table 1(a, b, c and d) in Shimshoni (1971). These are shown on Table 6.1.

The sample spectrum estimates, $f^*(w)$, for the commercial guano records at 17 Namibian and South African islands and the two records of total harvest were calculated for the period 1897-1975 using the NAG routine G13CBF. The significance, at either the 1% or 5% level, of the 10 largest estimates for each time series are shown on Table 6.2. Apart from Bird Island at Lambert's Bay,

Table 6.1. Critical levels for the sample spectrum of a time series of length n=78, for peaks of rank τ. To test the significance of various peaks in the normalised periodogram, the peaks must first be ranked. Significance at either the 1% or 5% levels of the peak ranked τ is then given by comparison with the values quoted here (Shimshoni, 1971).

τ	CRITICAL LEVEL	
	1%	5%
1	0.192	0.157
2	0.125	0.109
3	0.101	0.089
4	0.085	0.076
5	0.073	0.067
6	0.066	0.060
7	0.059	0.055
8	0.054	0.049
9	0.049	0.045
10	0.046	0.043

Table 6.2. Significance levels of sample spectrum estimates for the univariate analysis of estimated guano harvests at all 17 islands considered in the correlation matrix (Fig. 6.3). C, G and P refer to the dominant bird species on each island according to Crawford and Shelton (1978), either cormorants, gannets or penguins respectively. The columns are tabulated according to decreasing period, p, which is given in terms of the integer, L, listed above each column: p = 50/L; 50 is the baseline time interval which is the wavelength of the first harmonic component (i.e. the harmonic analysis is based on a time period of 50 years). No significant peaks were observed for p<50/9.

ISLAND	L:	1	2	3	4	5	6	7	8	9
	p:	50	25	16.7	12.5	10	8.3	7.1	6.3	5.6
1. Bird, Lamberts	CG		5	1	1	1	1		1	1
2. Paternoster	C	1	1							
3. Malgas	GC									
4. Marcus	P	1	5	1						
5. Jutten	CP		1		1					
6. Vondeling	CP	5	1	5						
7. Dassen	PC				5					
8. Dyer	PC	1								
9. Bird, Algoa	GP	1								
10. Platforms	C	1	1							
11. Mercury	GC	1	1							
12. Ichaboe	GP									
13. Halifax	P	1	1	5						
14. Possession	GP	1	5	5						
15. Pomona	P	1	1	1						
16. Plumpudding	P	1	1	1						
17. Sinclair	P	1	1							
18. S.A: Total		1	1							
19. Namibia: Total		1	1	1	1					

1 - a significance level of 1%
5 - a significance level of 5%

there are no significant short period oscillations, and generally only the long period oscillations are statistically significant. These significant fluctuations have periods longer than 15 years, and typically 25 or 50 years (note that the 25 and 50 quoted here are accidents of the baseline time interval chosen for performing the spectral analysis: the most that can be said is that there appear to be statistically significant components with periods in excess of 50/3 or, say, 17 years, the next shortest period representable with a baseline of 50 years). From arguments which will be developed in section 6.3, it seems that these long period fluctuations could only be produced by animal populations with ages to maturity of between 7 and 13 years, which is far too long for seabirds or fish. These periodic features are therefore presumably due to some extraneous effect (economic or biological) other than bird or fish dynamics. The sample spectrum for Bird Island Lambert's Bay is shown in Fig. 6.4, peaking at periods of 25, 8 and 6 years.

It should be noted that, at the 5% significance level, 1 in 20 periodogram peaks are expected to exceed the critical level due to pure chance alone. Therefore, of the k significant levels observed, on average k/20 of these are due to chance. Thus since 19 time series have been considered, results which are significant at the 5% level in one island only are probably chance effects. Furthermore one expects to see one significant spectral peak in 100 time series significant at the 1% level by pure chance. Thus there is a 19% chance of seeing one in 19 time series. The two peaks significant at the 1% level at Lamberts Bay, at periods of 6 and 8 years, therefore have a 19% probability of being chance effects without any link to population dynamics. The bivariate spectral analysis gives a

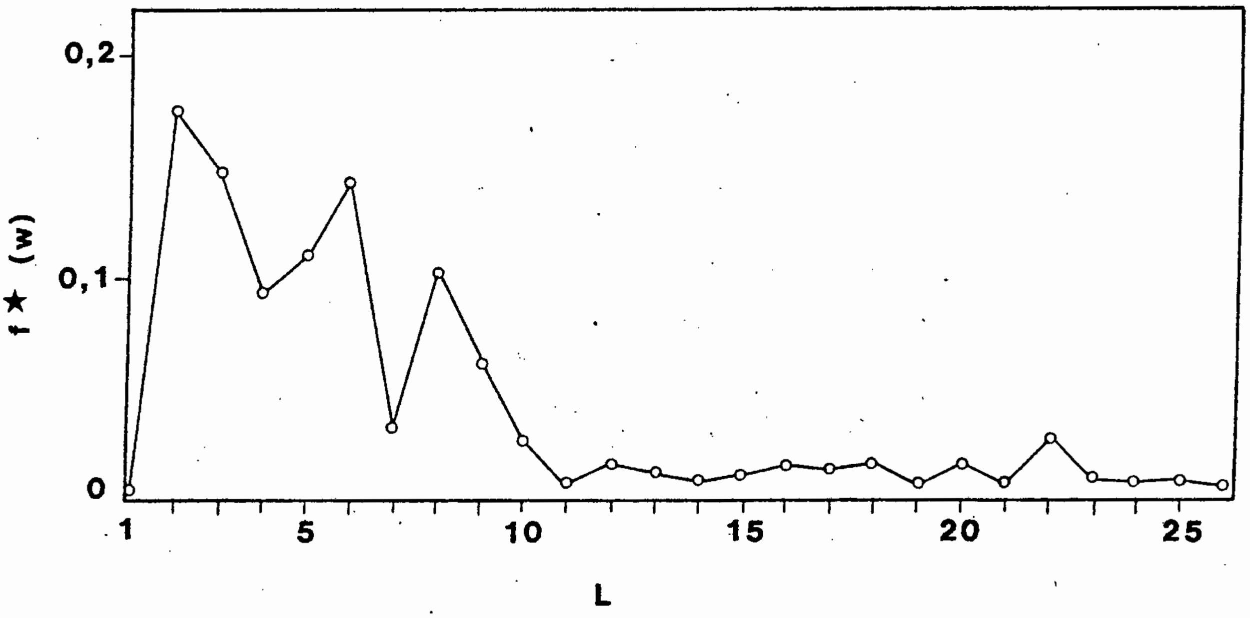


Figure 6.4. The estimated spectral density function, as normalised f*(w), of guano harvests at Bird Island, Lambert's Bay. Periods are calculated as 50/L, where 50 years was the baseline time interval given to the univariate spectral density estimation routine G13CBF of NAG (Numerical Algorithms Group, Oxford, 1977). Significant periodic components at periods of 6, 8 and 25 years are indicated from the critical levels set out on Table 6.1.

stronger result for these short period effects.

6.2.3 Bivariate spectral analysis of commercial guano records

For the situation under discussion, the squared coherency, $W(w)$, is perhaps the most useful quantity provided by the bivariate spectral analysis. This gives some idea of the correlation between two time series at a particular frequency and is defined as (Jenkins and Watts, 1969; Bloomfield, 1976);

$$W(w) = |f_{xy}(w)|^2/[f_{xx}(w)f_{yy}(w)] \qquad \ldots\ 6.9$$

where $f_{xy}(w)$ is the smoothed sample cross-spectrum of the two series x_t and y_t (recall again that $w=2\pi k/n$):

$$f_{xy}(w)=1/(2\pi n)[\sum_{t=1}^{n} y_t \exp(iwt)].[\sum_{t=1}^{n} x_t \exp(-iwt)] \qquad \ldots\ 6.10$$

(note again: $w=2\pi k/n$, k the k th period)

The different ways of smoothing the data distinguish the different types of spectral analysis. The results calculated here were obtained using spectral smoothing by the trapezium frequency (Daniel) window, with a trapezium shape parameter of p=0.5 (NAG Manual; Jenkins and Watts, 1969; Bloomfield, 1976). This gives similar results to other smoothing techniques "with a slight advantage of somewhat less distortion around sharp peaks, but suffering a rather less smooth appearance in fine detail" (NAG Manual).

Fig. 6.5 shows the total number of squared coherencies at all Fourier frequencies and for all possible unique pairs of islands (but excluding the S.A. and Namibian totals since there is some implicit dependence between these and the individual island records, 136 unique pairs altogether) which could not be rejected

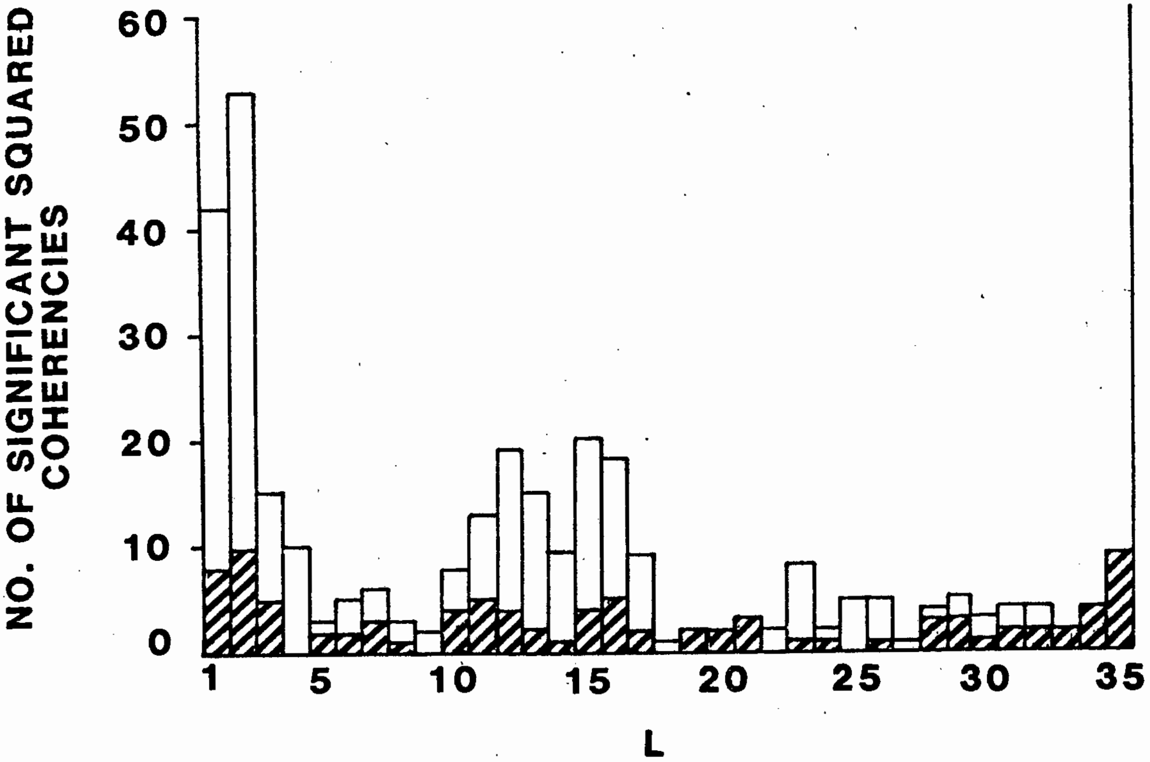


Figure 6.5. The sum of the significant (95% level) squared coherency estimates for all unique pairs of guano harvest time series, excluding the total Namibian and South African harvests. (136 pairs in total). The hatched histogram shows the equivalent result for lognormal noise with an arbitrarily chosen coefficient of variation of 40%. Periods are given as 70/L, where 70 is a baseline time interval given to the bivariate spectral analysis routines. These noise results suggest that smoothing processes have not produced spurious peaks in the analysis of the actual data. The average noise level in the diagram is 3 per 136 (the number of possible unique pairs). It is suggested that the peaks for p (period) > 23 (i.e. L < 3) and 4 < p < 6 (i.e. 11 < L < 17), units all in years, are significant, with the level always above 10 per 136.

at the 5% level by comparing the coherency estimate $W(w)$ with the test statistic T derived from the upper 5% point of the F distribution for (2, d-2) degrees of freedom:

$$T = 2F/(d-2+2F) \qquad \ldots\ 6.11$$

where d is the number of degrees of freedom associated with the univariate spectrum estimate.

There is some suggestion in the literature that smoothing processes can produce spurious periodogram peaks (Grenander and Rosenblatt, 1957). As a rough test of this the significant squared coherencies obtained from 18 series of lognormal random data (of length 78 years) with an arbitrarily chosen CV of 40% were estimated (shown as hatched histograms on Fig. 6.5), and this provides some estimate of the effect of smoothing. Conspicuous above this background effect, there are strong coherencies at periods of 23 years and more, and in addition there is some coherency at periods of 4 to 6 years. It seems most likely that the long term effect is ecologically meaningless. These apparent long term periodic cycles may rather be the reflections of the influences of the growth of the guano industry since the turn of the century, two world wars, the depression years and the pilchard collapses (off South Africa and Namibia) in the last 20 years.

The high frequency component may however be of ecological import. As mentioned earlier, the most obvious cause of these more rapid cycles in guano production is the feedback control mechanism governing seabird population size. Some brief analyses are necessary to throw light on this.

6.3 UNDAMPED OSCILLATIONS IN BIRD POPULATION SIZE

6.3.1 Trends in total guano production

An examination of the total commercial guano production for South Africa (Fig. 6.1) shows that, prior to the start of purse-seining on a large scale (circa 1945), fluctuations in guano, and therefore presumably bird numbers as well, were large. Since then, and particularly following the collapse of the pilchard in the mid-1960's, the fluctuations appear to have been damped (i.e. their amplitude has decreased markedly). The fluctuations are what might result from a system whose dynamics are governed by a compensation mechanism incorporating a feedback delay such as would be caused by the time taken to reach sexual maturity. It is necessary therefore to examine the period and amplitude of the output of some simple dynamic population models, in an attempt either to demonstrate or refute the hypothesis that the fluctuations in guano were due to undamped fluctuations in the seabird population size.

6.3.2 Plausible models of seabird populations

Seabird breeding incidence is generally not continuous, but occurs at a fixed period during the year, and thus the following discrete model is appropriate:

$$V_{i+1} = V_i + rV_{i-T+1}(1-V_{i-T+1}/K) \qquad \ldots\ 6.12$$

where the variables and parameters are

V_i : number of adult seabirds in year i

T : the age to sexual maturity for seabirds

r : the maximum number of mature seabirds produced per annum per capita of mature seabirds

K : the equilibrium population size of adult seabirds

For rT sufficiently large, the continuous version of this model is expected to have an oscillation period of approximately 4T years (May, 1981 obtains just this period for a very similar model), since the rate of change of adult population size changes sign when the number of adults a time T years previously crosses the K threshold. Thus T is the interval on the time axis corresponding to the interval between equilibrium and a peak or a trough; as oscillations are approximately symmetrical about peak/trough positions, the period will be about 4T.

In the discrete case, the period of oscillation is generally less than 4T. This is because peak and trough values can only be preceded by T-1 values above or below K respectively, since thereafter the growth term changes sign and the population size either starts to decrease or increase (Fig. 6.6). Thus one can account for the 2 peak values themselves, 1 trough value, and 2T-2 other values preceding the trough and the right-hand peak in a full peak to peak cycle. A minimum of 2T+1 points must therefore occur from peak to peak inclusive. The remaining points will lie between the left-hand peak and K, and between the trough and K. These are the points descending from peaks to cross K, and ascending from troughs to cross K.

It has been shown that there are only T-1 points to the left of peaks and troughs above and below K. This is not necessarily true of the right hand points (only T-1 points to the right of peaks and troughs above and below K respectively), but simulations suggest the number of right hand points is usually less than T, and often T-1. This latter value would give 4T-1 points from peak to peak inclusive, and hence an effective peak to peak period of 4T-2 years. In addition, simulations showed

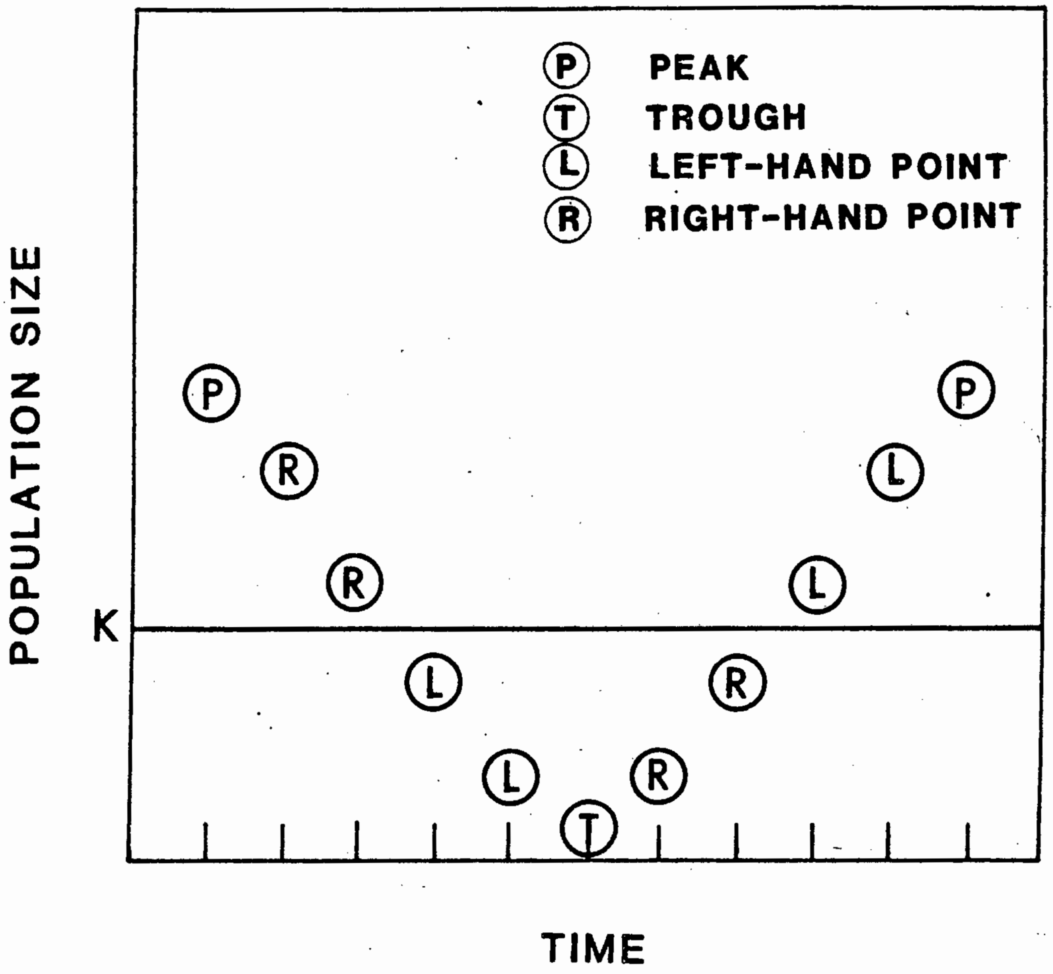


Figure 6.6. A typical population size limit cycle for a species with an age to maturity of 3 years, using the model described by equation 6.2. The 2 peaks marked P(eak), 1 marked T(rough) and 4 marked L(eft-hand) are an unavoidable consequence of the functional form used (equation 6.2). From simulations using parameter values typical of seabirds, there appear to be on average a further 4 points marked R. The peak to peak time interval is therefore given approximately as 10 years. This generalises to a peak to peak interval of 4T-2, for an age to maturity of T years.

that this model is unstable for r greater than about 0.7 years^{-1} (with T=3 years), and the population becomes extinct within a few years.

The problem with the discrete logistic model is that it incorporates birth and death rates in a single term. Thus unrealistically high rates of population decline, and hence extinctions, are possible. It is more reasonable, for adult seabirds, to assume that natural mortality occurs continuously and is density independent (i.e. the effects of food limitation are manifest almost entirely in egg-laying and the first year survival rate). If logistic type density dependence in birth (but not death) rates is permitted, then the following model is more stable and suitable for subsequent analysis than equation 6.12;

$$
\begin{aligned}
V_{t+1} &= V_t \exp(-M_b) + rV_{t-T+1}(1-V_{t-T+1}/K) && : \quad V_{t-T+1}<K \\
V_{t+1} &= V_t \exp(-M_b) && : \quad V_{t-T+1}>K \quad \ldots \; 6.13
\end{aligned}
$$

where M_b is the adult seabird natural mortality constant.

It is useful to examine the behaviour of the adult seabird population size following a perturbation in B_t (a perturbation of 20% of K was used). Note that the peak to peak periods of fluctuation for this model is also expected to be about 4T-2, on the basis of similar arguments to those given earlier.

Simulated solutions to equation 6.13 show that the period of population size fluctuations is critically dependent on T, as expected, and minimally so on r and M_b. It is therefore important to specify the region of parameter space (defined by r, M_b and T) of relevance to seabirds, particularly so for the age to sexual maturity. Various estimates of these quantities are

Table 6.3. Estimates of M (natural mortality) and T (age to sexual maturity) for cormorants, gannets and penguins according to various authors. The logistic equation intrinsic growth rate parameter, r, is estimated as half the mean clutch size. For cormorants the mean clutch size for the range of 1-5 eggs quoted is taken to be 3.

Seabird	T (years)	M_b (yrs^{-1})	Mean clutch size	Clutches/yr	r
Gannets	3-4[1]	0.05-0.08[1,7]	1[1]	1[1]	0.5
		0.02-0.07[9]			
Cormorants	2[2]	0.2[8]	1-5[3]	1[3,2]	1.5
	3[3]				
	3[8]		3[8]	1[8]	
Penguins	2[4]		1.7[4]	2[5]	0.8
	4[5]			3[6]	
	5[6]				
	3[8]	0.13[8]	1.8[8]		0.9

1. Jarvis (1971)
2. Duffy (pers. commn) - "some cormorants may breed at two years"
3. Berry (1976)
4. Jackson et al (1976) - bred in captivity
5. Wilson (pers. commn) - west coast penguins only
6. Randall (1983) - south coast penguins only
7. Nelson (1978)
8. Furness and Cooper (1982)
9. Appendix 6.1

summarised in Table 6.3. The maximum per capita adult production rate, r, is calculated as approximately half the mean clutch size (assuming one pair of birds per nest, 50% males and females at birth, and zero mortality from egg to adult under ideal conditions).

In the differential equation;

$$dV(t)/dt=rV(t)[1-V(t-T)/K], \qquad \ldots \; 6.14$$

the population exhibits stable limit cycles when $rT>1.57$ (May, 1981), and a similar law presumably holds for equation 6.13. Simulations suggest that for M_b values typical of seabirds (Table 6.3), the changeover from a stable equilibrium state (this state includes the possibility of damped oscillation approach to this equilibrium to stable limit cycle behaviour occurs at about $rT=2$. Gannets, with $r=0.5$ and $T=3$ to 4 years are thus not able to maintain limit cycles according to this model, although cormorants and penguins can.

6.3.3 Simulation results

The approximate period and amplitude (half of the difference between peak and trough population sizes, relative to $K=1$) of population size fluctuations generated by equation 6.13 are shown in Tables 6.4 and 6.5. The model predicts that the stable limit cycles for a typical cormorant population may have periods of roughly 5 to 9 years ($T=2$ to 3 years). The period predictions are extremely sensitive to the age at maturity parameter, and minimally dependent on r and M_b. Simulated oscillation amplitudes for the adult seabird population are smaller than suggested by the guano data, for which fluctuations are in the vicinity of half to double the average level (see Table 6.5).

Table 6.4. Approximate simulated periods of seabird population size fluctuations from values of r, T and M_b in the vicinity of those suggested by Table 6.3. Continuous oscillations do not occur for r<1 and T=2, and for parameter combinations indicated below by an asterisk.

		r=1	
	T=2	T=3	T=4
M_b=0.05	*	10	14
M_b=0.10	*	*	13
M_b=0.20	*	*	*
		r=1.5	
	T=2	T=3	T=4
M_b=0.05	6	11	16
M_b=0.10	5	10	16
M_b=0.20	5	9	12
		r=2	
	T=2	T=3	T=4
M_b=0.05	7	12	16
M_b=0.10	6	10	15
M_b=0.20	6	9	12

Table 6.5. Approximate simulated amplitudes of seabird population size fluctuations for values of r, T and M_b in the vicinity of values suggested by Table 6.3. Continuous oscillations do not occur for $r<1$ and $T=2$, and for parameter combinations indicated below by an asterisk. Amplitudes are expressed as proportions of K (K=system carrying capacity).

		r=1	
	T=2	T=3	T=4
M_b=0.05	*	0.1	0.19
M_b=0.10	*	*	0.25
M_b=0.20	*	*	*
		r=1.5	
	T=2	T=3	T=4
M_b=0.05	0.08	0.16	0.29
M_b=0.10	0.11	0.25	0.41
M_b=0.20	0.19	0.32	0.45
		r=2	
	T=2	T=3	T=4
M_b=0.05	0.1	0.21	0.36
M_b=0.10	0.15	0.33	0.54
M_b=0.20	0.22	0.53	0.61

This comparison assumes that guano production is proportional to adult seabird population size, and from these results it appears that fluctuations in guano production do not come from the overcompensation in the seabird population regulation mechanism. However it is not necessarily the case that guano production is solely dependent on adult seabird population size. A reasonable assumption, for example, may be that guano production depends on the total annual bird-days spent on the island (A. Berruti, pers. commn). This factor would increase as breeding effort increased. The result might thus be that guano production is proportional to seabird "recruitment" , for which the relative amplitude of fluctuations depends on the peak levels of the growth term relative to their average value. Note however from expression 6.13 that the value of the growth term reaches 0 under certain conditions. If guano production may be adequately represented by a weighted sum of adult bird population size and seabird "recruitment", instead of just being proportional to V_t, then the expression for guano production might be:

$$G_t = \alpha V_t + \beta X_t$$

where X_t is the production of 0 year old seabirds. If the peak value for X_t is W (remember X_t returns to zero periodically), and if V_t fluctuations are small enough to regard V_t as a constant equal to V, then the amplitude of G_t fluctuations will be approximately $\beta W/2$. Amplitude as a fraction of the mean (approximate) value is therefore just $\beta W/[2(\alpha V+\beta W)]$. Thus the amplitude of fluctuations observed in guano production depends on the weights α and β. These weights are not constrained by 6.13 so that, provided the seabirds are capable of reducing reproductive effort to close to zero (relative to peak levels), fluctuations reaching double the average guano production are compatible with the seabird model.

In summary, under certain assumptions about seabird population dynamics, large fluctuations in guano production occur if the population is perturbed from its equilibrium state. A critical prerequisite for the inducement of continuous oscillations (within the time period over which the model is run, which was 300 years) is that r should be sufficiently large. Gannets cannot produce this effect, since the mean number of eggs produced per adult per year is only 0.5, and rT is less than the critical level of about 2. For cormorants, if survival from laid egg to mature adult is 100%, then model population oscillations do persist. These may be comparable in amplitude to the fluctuations observed in the guano records (eg. Fig. 6.1). Requirements for frequency are more restrictive. Cormorant population oscillation frequencies are only comparable to those detected in guano records if the age to maturity for cormorants can be assumed to be 2 years. Only one of the ornithologists interviewed (D. Duffy, pers. comm.) speculated that "some cormorants may breed at 2 years", referring to the species of importance Phalacrocorax capensis. It is concluded therefore that the high frequency component in the guano records, although perhaps influencing seabird population dynamics in some way, is not caused by seabird population regulation mechanisms. The significant short period component at Bird Island, Lambert's Bay, of about 8 years (see Fig. 6.4) may however be due to cormorant guano fluctuations (Rand, 1963, estimated that there were at one time over 60 000 cormorants on the island, compared to 10 000 gannets), since this period is probably not markedly different to what may be possible for cormorant population size fluctuations.

As a counter to the analysis of probable seabird population oscillation behaviour set out in this section, a brief examination of similar (probable) processes in a typical pilchard population is presented in the following section.

6.3.4 Limit cycle frequencies for typical pilchard recruitment

The recruitment pattern for a wide range of fish stocks is typified by the occurrence of occassional large year classes (Walters, 1984). This behaviour is close to what would be expected if the net survival from the egg stage to recruitment is due to numerous multiplicative effects which are independently either good or bad, which is the scenario envisaged in this thesis for South African pelagic fish stocks. Invoking the central limit theorem, the logarithm of the net environmental effect over the whole period is approximately normally distributed. Recruitment would thus be approximately log-normally distributed. As a first approximation, it can therefore be assumed that pelagic fish recruitment is perturbed from equilibrium by a log-normally distributed net environmental effect. Because of the spawner-recruit-spawner feedback loop, recruitment and stock biomass trends may exhibit certain cyclical features. It is these features which are of interest at this stage.

The spectral densities, $f^*(w)$, (defined by equation 6.8) for the recruitment trend of a fish with a natural mortality equal to 0.5 yrs^{-1}, age to maturity of 2 years (pilchard-like) and log-normal environmental perturbations on recruitment were estimated using the NAG routine G13CBE. Recruitment, R_i, was generated using an asymptotically flat stock-recruit relationship:

$$R_i=[ay_i/(b+y_i)]*\exp(u_i) \qquad \ldots\ 6.15$$

where y_t = spawning biomass, assuming an age to maturity of 2 years, and $u_t' \approx N(0,0.04)$. The generation of R_t and y_t was started at unexploited equilibrium.

6.3.5 Results for simulated pilchard populations

For 18 typical annual recruitment series 78 years in length, and using the fully age-structured model, the highest spectral peak in each case occurred at a period of between 5.5 and 7 years.

In a separate simulation, a spectral analysis of time series generated by equation 6.12 (with $r=1.5$ yrs^{-1}, $T=2$ years) with lognormal noise ($u_t \approx N(0,0.04)$ on the intrinsic growth rate parameter r, resulted in peak values at periods of between 5 and 5.5 years.

Both these models therefore suggest that a typical pilchard population, under random growth perturbations will exhibit continuous oscillations at periods of between 5 and 7 years. It is anticipated that as the variance of the random perturbation is increased beyond some threshold, limit cycles will eventually become undetectable in the spectral analysis; however no attempt to determine this threshold limit was made.

At this stage of the analysis, there seems good reason to believe that some component of pilchard population dynamics (perhaps recruitment) is reflected in the commercial guano harvest records. Further evidence for this can be found in the existence of complex lag effects between certain islands. These are discussed in the next section.

6.4 COMPLEX RELATIONSHIPS BETWEEN GUANO HARVEST TIME SERIES

6.4.1 Crawford and Shelton's findings

There is some evidence that complex relationships exist between certain pairs of islands, for which guano time-series are poorly correlated. In unravelling this link, it is hoped that past pilchard recruitment trends can be identified in the guano harvest data. The best example of this is the relationship between Lambert's Bay Bird Island and Algoa Bay Bird Island guano records.

It is important to note that these islands have not been chosen on the basis of a search for the two islands which give the best results. Rather these have been chosen on the basis of a priori assumptions. Lambert's Bay Bird Island is the main guano island located close to the prime pilchard recruitment grounds, and it is one of two major West Coast gannet colonies in South African waters. It is not clear, just from the records of gross mass of guano harvested, from which bird species the bulk of harvested guano production has resulted. Historical population size censuses suggest that cormorants outnumbered gannets by 6 to 1 (Crawford and Shelton, 1978). However cormorants produce poor quality guano, which is easily blown away (A. Berrutti, pers. commn); gannet guano on the other hand becomes firmly cemented into nests, and is better preserved through to the time when annual guano scraping takes place. In the analysis which follows, it is assumed that guano at Lambert's Bay is produced predominantly by gannets. Additional analyses not reported here indicates that conclusions drawn from this do not change if cormorants are instead assumed to be the dominant guano producers.

The second island, Algoa Bay Bird Island, is the site of a major gannet colony east of the Cape Peninsula, and is isolated from the West Coast colonies. Diet studies (Batchelor, 1982; Batchelor and Ross, 1984) have shown that gannets at Algoa Bay are feeding on adult pilchard, and this suggests that guano production trends at this island may be useful for testing the hypothesis of the Bird Island Lambert's Bay guano trend as a pilchard recruitment index, as first suggested by Crawford and Shelton (1978). The sea route between the two islands is about 1200 km.

The Pearson's correlation coefficient between commercial guano harvests at Bird Island, Algoa Bay and Bird Island, Lambert's Bay is $r = 0.148$, $n=78$, which is not significant at the 5% level. Guano production at these two islands is compared in Fig. 6.7. Crawford and Shelton (1978) have argued that Bird Island, Algoa Bay lags behind Bird Island, Lambert's Bay because the former reflects adult pilchard and the latter recruitment. According to their paper this is because pilchard recruitment occurs in the vicinity of Lambert's Bay and "shoals of adults migrate in an easterly direction each spring and some are located near Algoa Bay in winter" (Crawford and Shelton, 1978). This effect is evident when correlation coefficients (Pearson's) with Algoa Bay lagging Lambert's Bay by successive years are calculated. The results show that the highest correlations are obtained with a lag of 3 years ($r = 0.30$ Fig. 6.8). However, implicit in the argument presented by Crawford and Shelton (1978) is the assumption of direct proportionality for both islands: guano production to pilchard recruitment for Lambert's Bay, and guano production to adult stock size for Algoa Bay. The correct test of this model requires that the Lambert's Bay guano production be

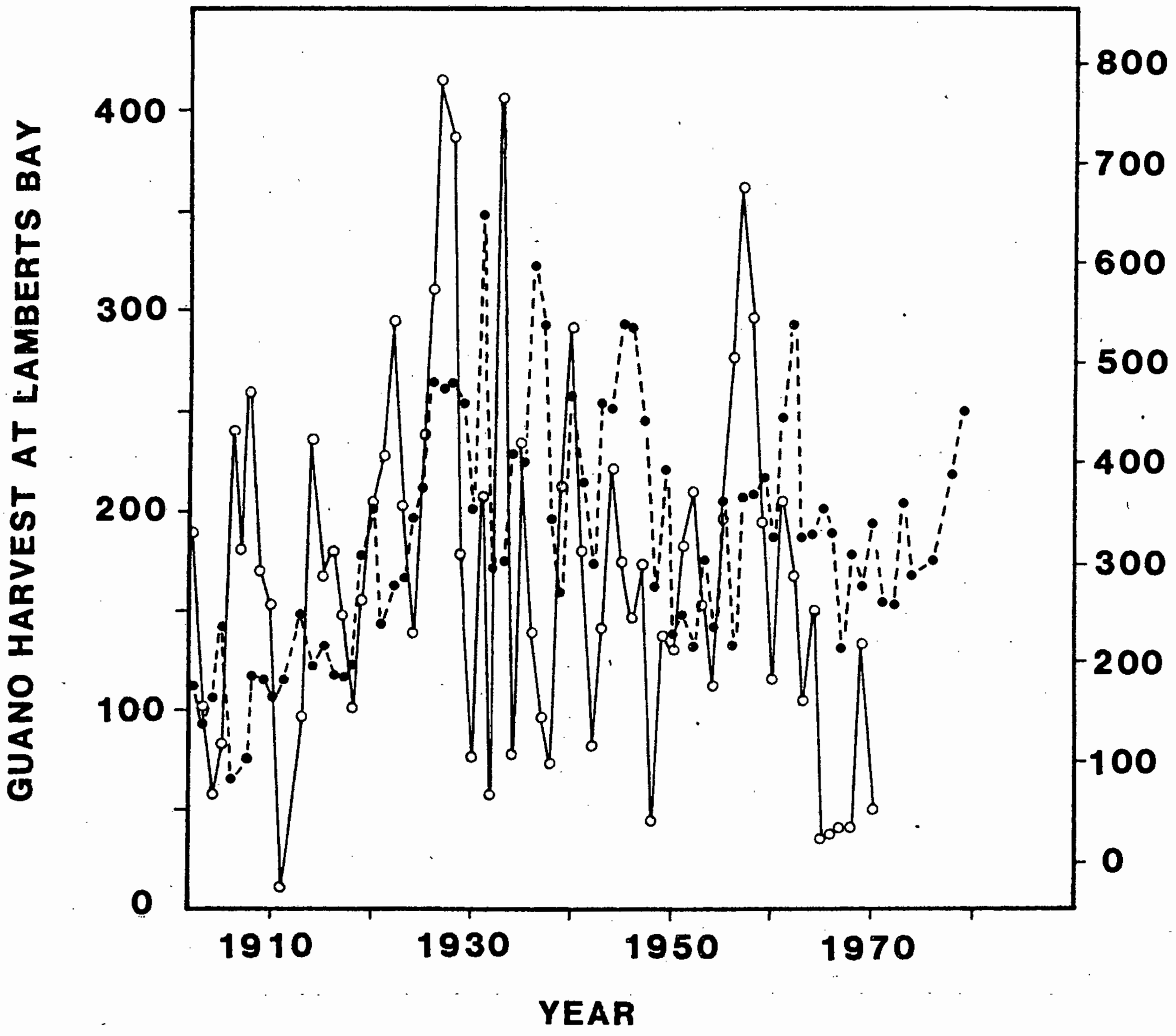


Figure 6.7. Guano harvests at Algoa Bay (dashed) and Lambert's Bay (solid) Bird Islands. From a priori considerations based on diet one might expect a lag between these two time series. This is indicated by the correlation coefficients shown in Figure 6.8. Units: tonnes.

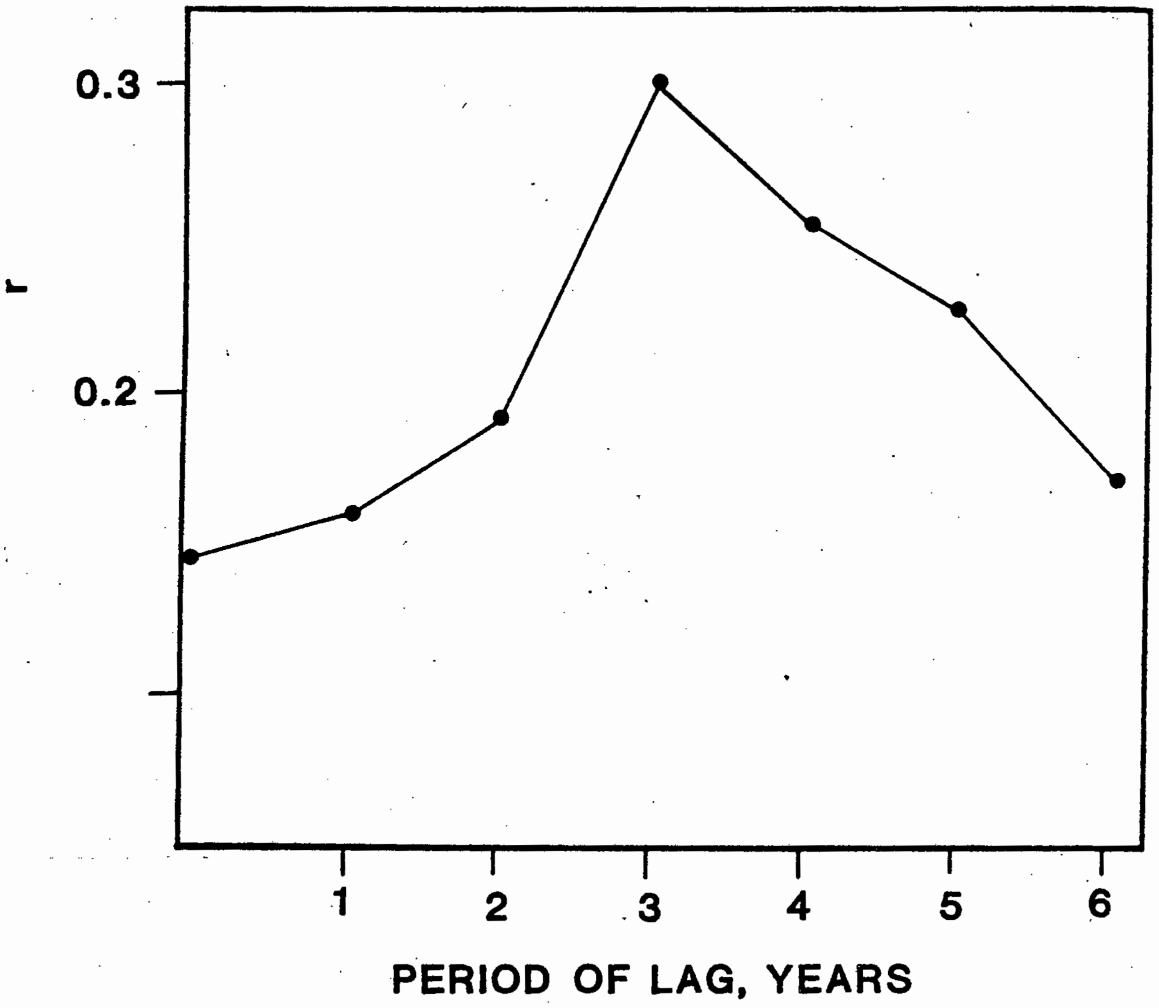


Figure 6.8. Correlation coefficients between guano harvest records for the Bird Islands at Algoa Bay and Lambert's Bay obtained by lagging the Lambert's Bay series through successive years. A peak value of r ≈ 0.3 at a lag of 3 years is indicated.

decayed and averaged over several years in the same way that a large number of year classes contribute to adult fish stock size, before comparison with Bird Island Algoa Bay guano data. The effect of fish growth rate must also be taken into account.

6.4.2 An age structured pilchard population model of the relationship between Lambert's and Algoa Bay guano harvests

A simple model of an age structured pilchard population was developed to test the hypothesis that Lamberts Bay guano harvests reflect (in direct proportion) pilchard recruitment and Algoa Bay harvests reflect the adult populaton biomass. The model assumes a constant (instantaneous) natural mortality rate of 0.5 yrs^{-1}. It allows for fish growth by changing fish mass for each successive year to the new midyear fish mass (pilchard masses at age taken from DAMREP, 1983). Recruitment is assumed proportional to guano harvests at Bird Island, Lambert's Bay. The fish age structure is filled after 9 years of recruitment have been allowed to grow through to adulthood. For each year after this, a hypothetical adult population can be generated by adding up the cohort biomasses of, say, 3-year and older pilchard cohorts in that year. Thus hypothetical adult pilchard biomasses in year i, Y_i, are related to recruitment values, R_k, where recruitment is given in terms of metric tonnes of guano harvested, G_k:

$$Y_i = \sum_{k=i-MAXAGE}^{i-FAGE} w_{i-k} R_k \exp(-(i-k)M)$$

$$= \sum_{k=i-MAXAGE}^{i-FAGE} k\, w_{i-k} G_k \exp(-(i-k)M) \qquad \ldots 6.16$$

w_{i-k} - pilchard mass at the middle of year for age i-k k - a proportionality constant between guano production and pilchard recuitment.

Pilchard FAGE to MAXAGE years old are therefore assumed to occur at Algoa Bay (FAGE = 0 are 0-group pilchard, assumed to be 6 months old).

6.4.3 <u>Results obtained with the age structured pilchard model</u>

The correlation coefficients for the relationship between hypothetical pilchard adults (obtained using the Lambert's Bay guano harvest as recruitment into the model) and the Algoa Bay guano harvest are used on a somewhat (see Appendix 6.2) ad hoc basis for comparing the two time series. No comment about the significance of differences between different correlation coefficients can be made, since these different values are not independent. All possible combinations of FAGE and MAXAGE between 1 and 8 years were estimated. Results are set out on Table 6.6. The largest correlation coefficient, r = 0.55, occurs for FAGE = 3 and MAXAGE = 6, corresponding to 3 to 6 year old pilchards. As FAGE and MAXAGE approach 3 and 6 from other values between 1 and 8, there is a persistent increase to the peak at values of 3 and 6 respectively. This ($r_{fishmodel}$) is greater than the largest r_{lag} (= 0.30) obtained by simply lagging the two time series.

It is possible to go further than this in the analysis, by attempting to remove the effect that seabird population size has on the guano production, hopefully yielding a residual time series containing predominantly pilchard recruitment and adult biomass effects plus some noise. These could then be used with (perhaps) more confidence in the age structured pilchard

Table 6.6. Pearson's correlation coefficients for the relationship between the Algoa Bay Bird Island guano harvest, and a hypothetical pilchard biomass, aged between FAGE and MAXAGE years, derived by using the Lamberts Bay Bird Island guano harvest as recruitment into the hypothetical age structured pilchard population. (n=70 throughout)

		FAGE							
		1	2	3	4	5	6	7	8
MAXAGE	1	0.28							
	2	0.32	0.28						
	3	0.39	0.39	0.43					
	4	0.44	0.44	0.49	0.41				
	5	0.47	0.48	0.53	0.48	0.38			
	6	0.48	0.49	0.55	0.49	0.41	0.28		
	7	0.48	0.49	0.55	0.49	0.40	0.24	0.08	
	8	0.48	0.49	0.55	0.49	0.40	0.23	0.08	0.04

population model for testing the relationship between the two time series. To do this, an appropriate model of the link between guano production and seabird population dynamics is required.

6.5 POSSIBLE RELATIONSHIPS BETWEEN GUANO PRODUCTION, SEABIRD POPULATION SIZE AND RECRUITMENT, AND FISH POPULATION SIZE AND RECRUITMENT

6.5.1 Guano and egg production rates for penguins

Siegfried and Crawford (1978) found that a linear regression between the commercial egg crop and guano yield at Dassen Island for the period 1920 - 1961 produced a significantly positive correlation ($r = 0.436$; $P < 0.01$; $n = 39$). If commercial harvests of both penguin eggs and guano are roughly proportional to the biological egg and guano production rates, it would be reasonable to make the tentative assumption that guano production is proportional to seabird "recruitment". This is really a special case of previous comments made about the dependence of guano production on bird population size and breeding effort, where it was held that guano was some weighted sum of the two;

$$G_t = \alpha V_t + \beta X_t \qquad \ldots 6.17$$

where V_t = adult seabird population size

X_t = production of 0 year old seabirds

It is now being suggested that the weighting for adult birds, V_t, may be essentially zero ($\alpha=0$). This does not mean that guano production is not influenced by adult seabird population size, rather that the influence of V_t is effectively incorporated into X_t (as will be seen). This may require that non-breeding but mature birds do not inhabit the islands to any substantial extent during the breeding season.

Underlying this assumption is the fact that guano production may be an index of the annual number of bird days spent on the island (A. Berruti, pers. commn). However, most guano is produced during the breeding season and the main presence on the islands occurs during the breeding season, so that the two assumptions - proportionality between guano yield and seabird "recruitment", and between guano yield and annual bird days spent on each island - are perfectly compatible. It is assumed from this point onwards that this relationship holds for gannets and cormorants as well as for penguins.

In the following section, a model is suggested which relates guano production, adult size and 0 - group strength for seabirds, and fish population biomass and recruitment.

6.5.2 A consolidated description of how various factors may influence guano production

Three factors of importance to guano production have been identified in the preceeding analyses. These are the strength of the 0 - year - old seabird cohort (X_t), adult seabird population size (V_t), and the biomass of the assemblage of fish age classes (Y_t), consumed by the birds at the island in question. From the work of Siegfried and Crawford (1978) it is suggested that;

$$G_t = kX_t \quad \ldots \ 6.18$$

where k is a proportionality constant

Furthermore, it is suggested that the combined influence of adult seabird population size and fish biomass on seabird "recruitment" (and hence on guano production as well) is represented by the appropriately scaled product of the two

$$G_i = kX_i = sV_i^{\gamma}Y_i \qquad \ldots\ 6.19$$

where s the proportionality constant

The parameter γ is assumed to lie somewhere between 0 and 1. This confers a broader ecological meaning to the seabird population growth relationship, since it permits the derivation of an equilibrium population size from the balance between natural mortality and population recruitment. This is a desirable feature in any model of population dynamics.

The advantage of the seabird model outlined here is that it yields indices of V_i (mature bird population) and Y_i from the single series G_i, and so G_i can be transformed into Y_i. This is achieved by first generating relative estimates of V_i (number of mature breeders) in terms of guano, assuming a constant natural mortality for the seabirds

$$V_i = \sum_{l=T}^{14} [X_{i-l}\exp(M_b(i-l))]$$

$$= (1/k) \sum_{l=T}^{14} [G_{i-l}\exp(M_b(i-l))] \qquad \ldots\ 6.20$$

where T is the appropriate age to maturity for the seabirds found on the island in question. To avoid using all the guano information in generating the initial adult seabird population size, an upper age of 14 years was used for seabirds (If M_b is taken to be 0.05 yrs^{-1}, and T = 3 years, then this amounts to neglecting as much as 50% of the population) . Indices of Y_i are therefore given by

$$Y_i = G_i / V_i^{1/\alpha} = kG_i \left[\sum_{j=T}^{14} G_{i-j} \exp(M_b(i-j)) \right]^{-1/\alpha} \qquad \ldots \; 6.21$$

for all i > 1897 + 14.

6.5.3 Model output

Indices of pilchard recruitment at Lambert's Bay and of pilchard adults at Algoa Bay were calculated assuming that gannets are the dominant contributors to the amount of guano harvested, for which the following model parameters are suggested by the estimates summarised on Table 6.3:

$T = 3$ years

$M_b = 0.05$ yrs^{-1}

The resulting index of pilchard recruitment at Lambert's Bay, Y_i, was then used as input for the age structured pilchard model described in section 6.4.1, producing again the index of pilchard biomass for pilchards aged FAGE to MAXAGE (but now corrected for the influence of seabird population size on guano production) for comparison with the time series of pilchard adults derived from the Algoa Bay guano harvests (also now corrected for the influence of V_i).

The highest correlation coefficient was $r_{bird\ model}=0.61$ for $\tau=1$ (this implies that no density dependence exists, which seems unlikely), FAGE=3, and MAXAGE=6. This ($r_{birdmodel}$) is not substantially larger than the value obtained without accounting for seabird dynamics ($r_{fishmodel}$ - maximum value 0.55).

6.5.4 The discrepancy between model predicted pilchard ages at Algoa Bay, and observed ages

On the basis of the Pearson's correlation coefficient values calculated in the preceeding analysis, the most likely pilchard age classes oocurring at Algoa Bay should be fish between 3 and 6 years old. Diet studies (Batchelor, (1982); Batchelor and Ross, 1984) show instead that gannets at Algoa Bay are consuming pilchards aged predominantly between 1 and 2 years.

Batchelor (1982) found that for the period January 1978 to December 1980 the mean fork length of pilchard consumed by gannets on Bird Island Algoa Bay ranged between 140 mm and 215 mm. Davies (1956) found that for the years 1954 to 1956 the Natal (east of Algoa Bay, further away from the pelagic fishing grounds) sardine run was composed of pilchard of caudal length ranging between 140 mm and 215 mm. Likewise, Baird (1971), reviewing the Natal sardine run for the period 1951 to 1969, found a similar range in pilchard caudal length. According to the best present knowledge of pilchard growth rates (Thomas, 1985) 140 mm to 215 mm corresponds to pilchard of 1 to 2 years old. Furthermore, in all the samples mentioned, there were only a few fish greater (fork or caudal length) than 215 mm in length. From the growth curve information just mentioned, and for an unfished population with $M = 0.5$ yrs^{-1}, one would expect that at least 30 % of fish in a population aged 1 year or greater would measure more than 215 mm caudal length. It is clear therefore that the Algoa Bay gannets are feeding on pilchard aged 1 to 2 years.

However, there are additional considerations which might lead to the observation of gannets at Algoa Bay feeding on 1-2 year old pilchard and a 3 year lag between the guano records in question being compatible. These are concerned with the historic age stratification of the pilchard spawning grounds coupled with restricted flows of egg- containing surface waters (M.J. Armstrong, pers. commn). This explanation is set out in detail in the following section.

6.5.5 A possible two stage explanation for pilchard recruitment near Algoa Bay

Evidence from past egg surveys has shown that prior to the collapse of the pilchard in the mid 1960's, their spawning region spanned the Cape Peninsula (for geographic sites see Fig. 6.9). Older pilchard, greater than 215 mm in caudal length, were traditionally caught north of the Cape Peninsula, in the vicinity of St. Helena Bay on the West Coast. Maturing pilchard with caudal lengths of between 140 mm and 215 mm, aged between 1 and 2 years (according to recent growth curve estimates such as are quoted in DAMREP, 1983), were mainly caught east of the Cape Peninsula, and are thought to have spawned on the Agulhas Bank.

Surface water movement from the West Coast region during a south-east wind is northward, and since this is the prevailing wind direction, the predominant surface water flow in this region is northwards. Thus the dominant surface flow between the Agulhas Bank region and the West Coast region is towards the latter. This is corroborated by the results of drift card studies (Shelton, 1979), and by direct observations of egg and larval transport (Shelton and Hutchings, 1982; Duncan and Nell, 1976; Shelton and Kriel, 1976; Shelton, 1979). However, these studies also suggest that there is considerable surface drift eastwards off the Agulhas Bank, in the direction of Algoa Bay.

A likely scenario for egg transport and subsequent recruitment can therefore be constructed from observed surface water drift. Some of the eggs spawned on the Agulhas Bank may eventually recruit along the immediate inshore region and eastwards thereof towards Algoa Bay, while the rest would be transported onto the West Coast region. These would add to the eggs spawned there, and it is known that pilchard are recruited in the region of St. Helena Bay. Now it is possible that, when the stock was unexploited, the West Coast spawners made by far the greatest contribution to annual recruitment, and that the Agulhas Bank spawners, being only partially mature and losing eggs eastwards, did not make much impact on West Coast recruitment.

For cormorants at Lambert's Bay, the energetically most economic food available may have been pilchard recruits, and it is well known that cormorants are able to exploit small prey such as juvenile anchovies (Rand, 1960b). This is perhaps less likely for gannets which are often found to have consumed large fish (Batchelor, 1982). The reflection of trends in the year class consumed by cormorants, and gannets to a lesser extent, at Lambert's Bay in guano production at Algoa Bay would now be delayed by the oceanographic and geographic spawning age stratification model just sketched. The year class in question would first appear on the Agulhas Bank at about 1 or 2 years old, approaching full maturity, and may make a major contribution to recruitment for the pilchards found at Algoa Bay. However, either because they migrate slowly eastwards, or because they are the optimum size pilchard for gannets to feed on (gannets consume larger fish on average than penguins or cormorants, Rand 1960b; Batchelor, 1982)), they only appear in gannet stomachs at length 140 mm to 215 mm, aged 1 to 2 years. The absence of large pilchard at Algoa Bay and in the Natal sardine run (further east) may be due to a return migration of larger pilchard to the West Coast spawning grounds (Fig. 6.9). Thus an overall lag of about

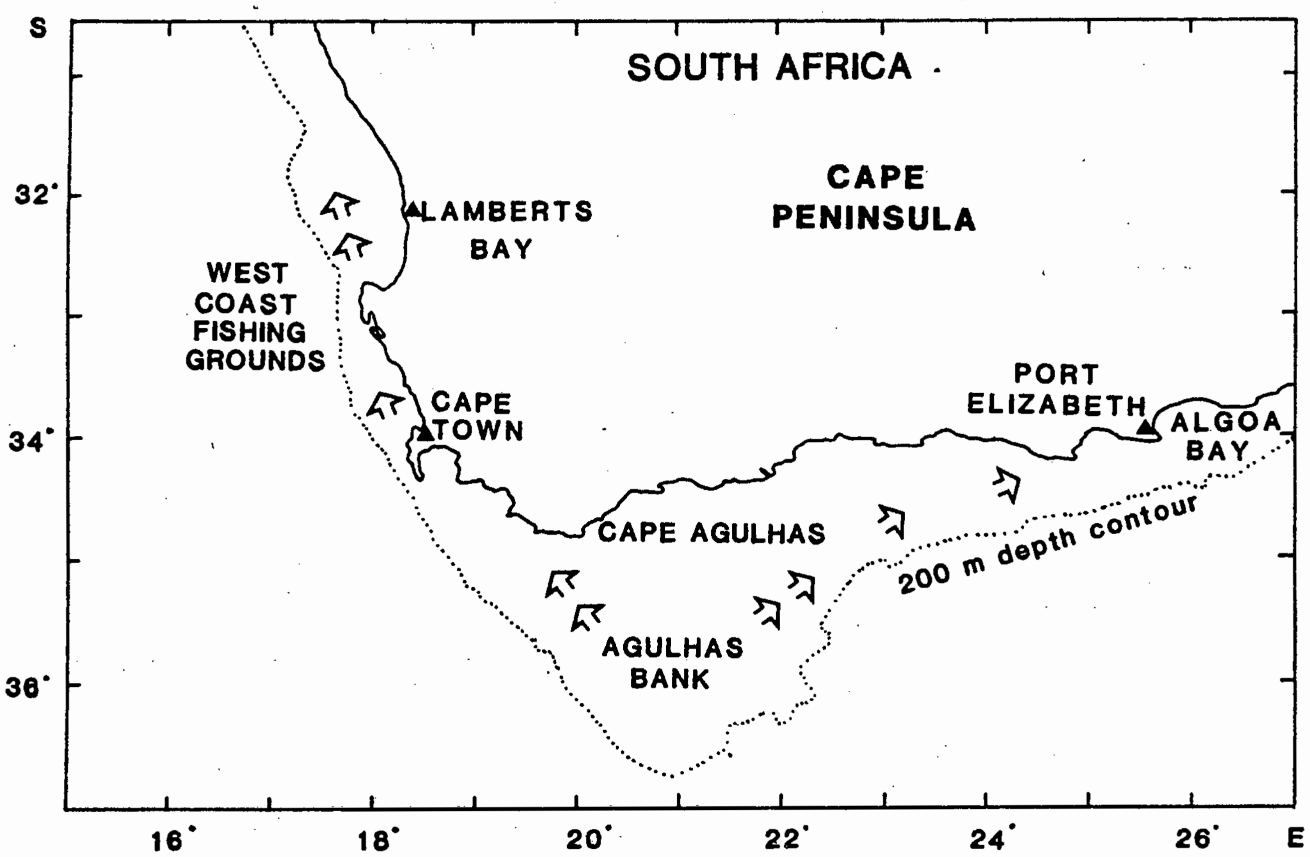


Figure 6.9. Probable pilchard egg transport direction vectors during summer (peak) spawning. Spawners on the Agulhas Bank are at most 2 years of age, those on the West Coast are older than 2 years. Since West Coast spawning may have dominated overall pilchard recruitment, recruitment trends observed at Lambert's Bay take 1 - 2 years to cause the same trends in spawning biomass on the Agulhas Bank, and recruitment resulting from this takes a further 1 - 2 years before appearing in gannet diets at Algoa Bay. The overall lag between guano trends at Lamberts Bay and Algoa Bay would therefore be about 3 years, or about twice the age to maturity for pilchard (M. J. Armstrong, pers. commn)

three years (≈ twice the age to maturity for pilchard) between peak recruitment levels at Lambert's Bay and the biomass of 1 - 2 year old pilchard at Algoa Bay is still feasible.

The correlation coefficients between model predicted and actual guano harvests at Algoa Bay were obtained by lagging the Lambert's Bay guano harvest as a recruitment index for the Algoa Bay pilchard population by either 1 (resultant correlation coefficient = r_1) or 2 (r_2) years. These lagged recruitment series were then used as input into an age structured pilchard population (as done in Section 6.4.2), with additional adjustments for seabird population dynamic effects (as done in Section 6.5.2). Only pilchard aged 1 and 2 years are assumed to occur in gannet diet at Algoa Bay (as found by Batchelor, 1982). The results were $r_1 = 0.34$ (n=69) and $r_2 = 0.44$ (n=68). These values are smaller than the earlier correlation value of 0.55 which was obtained without lag, and assuming that pilchard aged 3 - 6 years occurred in gannet diet at Algoa Bay. However in the case of r_2 the difference seems to be insubstantial (i.e. a difference in terms of explained variance of $1-0.44^2/0.55^2$ or about 26%).

6.6 <u>SUMMARY AND CONCLUSIONS</u>

The original motivation behind much of this chapter was the observation that fluctuations in guano production were apparently damped out following the inception of fishing. Initially it was thought that this was due to a damping out of seabird population size fluctuations as food resources diminished.

Results presented here indicate rather that guano production fluctuations are a direct result of changes in pilchard population size and recruitment. In this context the damping out of guano production fluctuations would merely be a reflection of what was happening to the pilchard population. An explanation for this phenomenon is thus most likely to lie in the influence of fishing on the pilchard population. In an age structured model the effect of fishing is most easily represented by an increase in the total mortality experienced. In terms of a stock recruit plot, the net result is an increase in the slope of the <u>effective</u> replacement line (i.e. including replacement of fish removed by fishing).

The stability criteria for a time delayed age structured population are very complex. However if one ignores the age structuring and the time delay due to age to maturity, then it is clear that the stability of recruitment and spawning biomass is critically dependent on the difference in slope of the recruitment curve and the replacement line at their point of intersection. For points of intersection to the left of the inflexion in the case of a Ricker recruitment curve, this difference is reduced by increments in the slope of the replacement line (i.e. by increasing fishing mortalities). Thus the anticipated effect of fishing is to confer stability on a stock (when instabilities arise in the deterministic sense). This is also likely to be the case for the more complex model incorporating age structuring and time delay. It therefore seems to be plausible that the observed damping of fluctuations in the guano record following the inception of fishing and particularly during and after the pilchard collapse resulted from increasing pilchard population stability (even though the mean population size and hence guano production was decreasing) caused directly by fishing.

The importance of pilchard in the diet of gannets and cormorants suggests that time series of guano harvests may be related to some aspect of pilchard population dynamics.

Time series analyses indicate that, apart from significant low frequency fluctuations, records of guano harvests contain statistically significnt components with periods of between 4 and 6 years. These relatively high frequency features are, at first glance, most likely due to changes in seabird population size, which may exhibit persistent continuous oscillations, such as are possible for typical models of animal populations (May, 1976). For seabirds, some discrete version of the usual continuous form of the logistic equation is suitable for modelling population dynamics, since they breed seasonally. Critical aspects for such models to manifest persistent oscillations are the size of r and T, the maximum specific population growth rate and the age to maturity.

Simulations show that gannets do not have a high enough specific growth rate to describe continuous oscillations (females lay only one egg per year); on the other hand cormorants do, and cormorants may be important producers of guano. For the specific equation chosen and in the neighbourhood of typical parameter values, the period of limit cycles observed in simulations is given roughly by 4T-2. Nearly all of the ornithologists interviewed estimated T for cormorants to be about 3 years (36 months), and this is the figure quoted in the literature (eg. Berry, 1976). One ornitholigist suggested that some cormorants may breed at 24 months, but most at 36 months. This (T=3 years) would produce population size fluctuations with a period of 10 years, and it would therefore seem unlikely that the high frequency component in the guano series is due to seabird dynamics.

Further simulations show that pilchard populations modelled either by a discrete logistic equation or using a spawner-recruit-spawner feedback loop, both with T=2, are quite capable of displaying persistent oscillations of the observed period (shown only within model runs of finite duration), provided that r (logistic model) or a (stock-recruit parameter in age structured model) was taken large enough. The implication therefore is that pilchard population fluctuations are reflected in guano production.

Further evidence for the reflection of pilchard population trends in guano harvests has been suggested by Crawford and Shelton (1978), who argue that the lag between guano production trends at Lambert's Bay and Algoa Bay Bird Islands is caused by the birds at these islands feeding on pilchard recruits and adults respectively. Their hypothesis was tested using some typical models of fish and seabird poulation dynamics. There may be a contradiction between the lag between the two islands in question, and the lag corresponding to the pilchard age structure indicated by Batchelor's (1982) diet studies conducted at Algoa Bay. However, by considering more complex yet plausible models of egg transport in the Benguela current region, the discrepancy can be resolved.

At root, there may be an ecologically meaningful lag of 3 years between Bird Island Lambert's Bay and Bird Island Algoa Bay guano harvest trends. More complex models of the link between the two time series produce better correlation coefficients, however the best of a number of competing models could not be identified. There are some clear cut surface water movement observations which can explain why gannets may be consuming 1 and 2 year old pilchard rather than 3 and more year old pilchard. Furthermore, the guano data appear to contain significant high frequency components which are more likely to be caused by pilchard population size and recruitment fluctuations than by changes in seabird population size. In view of this, the model results which have already been presented, and knowing that Bird Island Lambert's Bay lies at the epicentre of the pilchard recruitment grounds, it seems reasonable to consider using the guano record at Lambert's Bay as a recruitment index in the pilchard VPA likelihood function.

7. PILCHARD STOCK ASSESSMENT USING THE VPA WITH A RECRUITMENT INDEX DERIVED FROM GUANO DATA

7.1 THE BASIS FOR THE PILCHARD RECRUITMENT INDEX

On the strength of results shown in Chapter 6, it appears that records of guano harvests at Bird Island Lambert's Bay may provide some index of pilchard recruitment. Data such as this can be incorporated into the likelihood function for the VPA (Chapter 4), and it is anticipated that they will improve the precision of the resultant VPA stock estimates compared to results obtained using catch mass and age structure estimates alone.

Recruitment indices are however of limited value compared to effort information. This is because, while CPUE (catch per unit effort) is sometimes a reasonable index of stock biomass throughout the period under consideration, the recruitment associated with the index ($N_{i,1}^{ind}$) only gives rise to a stock biomass index m years after the first year of sampling, where m is the number of quantitatively significant year classes in the catch, and only if estimates of the fishing mortalities ($F_{i,j}$) and the natural mortality (M) are obtainable. Additional recruitment indices for years prior to the first year considered in the catch-at-age matrix can however be included to provide an effective stock biomass index starting in the first year considered in the VPA.

The guano data in question extend back to 1897. Diet studies (Rand, 1960a and 1960b; Matthews, 1961) confirm that pilchard used to be the main prey of gannets and cormorants. Recent diet studies show however that cormorants and gannets at Lamberts Bay are feeding predominantly on anchovy (A. Berruti, pers. commn).

It is likely that the switch from a mainly pilchard to a predominantly anchovy diet occurred in the mid-1960's, when the pilchard stock is believed to have suffered a rapid decline. This is reflected in a drop in pilchard predation in the diet of handline caught snoek _Thyrsites atun_, which switched from consuming mainly pilchard to predominantly anchovy over the period 1960 to 1965 (Nepgen, 1979). It may be reasonably argued that a changeover in seabird diet to anchovy occurred at roughly the same time. Therefore it seems quite clear that the current Lambert's Bay guano harvest has ceased to be of any value as a pilchard recruitment index. The latest year of correspondence between the guano series and pilchard recruitment has been chosen as 1971, although this may well be overoptimistic. For the purposes of the VPA, this leaves the period 1972-1983 without any relative index of recruitment. Consequently little confidence should be placed in VPA estimates of abundance and recruitment over that period, considering the limited value of catch mass and age structure information for determining abundances (Chapter 4 and Pope and Shepherd, 1982).

In the analyses which follow, maximum likelihood techniques of abundance estimation (Fournier and Archibald, 1982; Fournier, 1983) are applied to the pilchard catch mass, catch age structure and recruitment index data, each of which are accorded appropriate weights. The pilchard spawning biomass and recruitment data obtained from the aforementioned virtual population analyses are used to estimate the mean stock/recruit function. This in turn is used to estimate the maximum sustainable pilchard yield, the MSY biomass level (i.e. the biomass level which produces MSY, y_{MSY}), and the system carrying capacity for pilchard (K), using a yield per recruit approach. Simulations are carried out to test for bias in the MSY estimation procedure for the specific case analysed. Problems arising because of recruitment fluctuations, which diminish the usefulness of the deterministic MSY concept by increasing the risk of stock collapse, are considered.

7.2 THE PILCHARD LIKELIHOOD FUNCTION

7.2.1 The use of a recruitment index instead of effort data

The VPA likelihood function described in Chapter 3 was used for pilchard (equation 3.37), with some slight modifications. Instead of effort information being included as an index of F_i, the Bird Island Lambert's Bay guano harvest was used as a recruitment index (Table 7.1). This involved replacing the least squares sum:

$$\sum_{i=1}^{n} [\ln k_e E_i - \ln F_i]^2/(2\sigma_i^2) \qquad \ldots\ 7.1$$

by one for recruitment. If the recruitment index is assumed to be lognormally distributed, then for σ_r^2 considered given, the new term is the sum C where:

$$C = \sum_{i=1950}^{1971} [\ln k_r N_{i,1}^{ind} - \ln N_{i,1}]^2/(2\sigma_r^2) \qquad \ldots\ 7.2$$

($N_{i,1}$ and R_i are notationally equivalent, see Appendix 1.1). The quantity k_r is an additional model parameter, being the proportionality constant between $N_{i,1}^{ind}$ and $N_{i,1}$. Note that this sum is only taken from 1950 to 1971, the upper limit of 1971 being based on the most recent year for which it may reasonably

Table 7.1. The pilchard recruitment index (records of guano harvested at Bird Island Lambert's Bay in metric tonnes) for the years 1943 - 1971 (see also Fig 6.7).

Year		Year		Year	
1943	241	1953	253	1963	204
1944	320	1954	211	1964	252
1945	274	1955	295	1965	136
1946	245	1956	377	1966	138
1947	273	1957	461	1967	139
1948	148	1958	395	1968	140
1949	237	1959	295	1969	232
1950	230	1960	215	1970	150
1951	282	1961	306	1971	80
1952	310	1962	269		

be assumed that pilchards dominated seabird diet.

The pilchard catch-at-age matrix contains some genuine zero entries (see age structure matrix, Table 7.2). The way that these are handled is to give the best fit values of catch proportion by age (the $P_{i,j}$) a zero value where $P_{i,j}^{obs} = 0$ (the appropriateness of such a procedure is discussed in section 7.2.3). Thus recruitment ($N_{i,1}$) cannot be calculated using equation 4.26:

$$N_{i,1}=\exp(\beta_{i,1})C_i^{obs}(F_iS_1+M)/[F_iS_1(1-\exp(-F_iS_1-M))] \qquad \ldots\ 4.26$$

This is because the β parameters ($\beta_{i,j}=\ln[P_{i,j}C_i/C_i^{obs}]$) are not defined for $P_{i,j}=0$. Alternatively however, model recruitment can be expressed in terms of other cohort sizes which can be calculated (from non-zero catches at age). Using the link equations for a single cohort, viz. the catch equation (3.9) and the exponential relationship between successive cohort sizes (3.11), the $N_{i,1}$ can eventually be obtained. The computer routine which was used transforms the fundamental β estimate (often at roughly the middle of a cohorts lifespan) into a measure of cohort size, and then calculates earlier and later sizes of the cohort (including $N_{i,1}$) from this using equations 3.9 and 3.11.

The complete pilchard log-likelihood function which has been used is:

$$L(F_i,\ b_1,\ b_2,\ k_r,\ \beta_{fundamental}) = A - B - C \qquad \ldots\ 7.3$$

$$A = \sum_{i=1950}^{1983} \sum_{j=1}^{m} q_{i,j}^{obs}[(\beta_{i,j}-\ln[\sum_{j^*=1}^{m} \exp(\beta_{i,j^*})]]$$

$$B = \sum_{i=1950}^{1983} \ln^2[(\sum_{j=1}^{m} w_j\exp(\beta_{i,j})/(\sum_{j=1}^{m} P_{i,j}^{obs}w_j)]/(2\sigma^2) \ .$$

Table 7.2. Observed catch age structures for the South African pilchard fishery (Source: M. Armstrong, pers. commn). The index j=1 refers to the 0 year class. A zero entry in the table represents a value of $P_{i,j}^{obs}$ of less than 0.0005. Some of the zeros in this table are therefore actually very small values of $P_{i,j}^{obs}$, which are used as such in the actual VPA calculations. However for neatness of presentation they have been given a zero value in the table set out below.

$\underline{P_{i,j}^{obs}}$

Age class:	0	1	2	3	4	5	6	7
Year								
1950	0	0.007	0.101	0.277	0.347	0.229	0.038	0.001
1951	0	0.006	0.060	0.285	0.381	0.230	0.037	0.001
1952	0.003	0.014	0.161	0.342	0.314	0.148	0.017	0
1953	0.004	0.025	0.084	0.324	0.376	0.169	0.018	0
1954	0	0	0.037	0.314	0.432	0.202	0.014	0
1955	0	0	0.025	0.295	0.455	0.211	0.014	0
1956	0.008	0.016	0.079	0.245	0.395	0.237	0.019	0
1957	0.003	0.066	0.185	0.308	0.288	0.131	0.019	0
1958	0	0.053	0.509	0.309	0.098	0.028	0.003	0
1959	0.009	0.127	0.529	0.285	0.048	0.003	0	0
1960	0.004	0.088	0.469	0.355	0.080	0.004	0	0
1961	0.001	0.020	0.220	0.506	0.225	0.027	0	0
1962	0	0.040	0.324	0.410	0.191	0.035	0	0
1963	0.030	0.186	0.274	0.339	0.151	0.020	0	0
1964	0.005	0.101	0.434	0.369	0.082	0.008	0	0
1965	0.147	0.128	0.415	0.242	0.063	0.006	0	0
1966	0.299	0.037	0.273	0.278	0.101	0.011	0	0
1967	0.931	0.023	0.033	0.012	0.001	0	0	0
1968	0.868	0.028	0.062	0.035	0.008	0.001	0	0
1969	0.848	0.060	0.075	0.016	0.002	0	0	0
1970	0.771	0.072	0.134	0.023	0.001	0	0	0
1971	0.903	0.029	0.052	0.015	0.001	0	0	0
1972	0.739	0.075	0.074	0.080	0.028	0.004	0	0
1973	0.885	0.052	0.055	0.007	0	0	0	0
1974	0.257	0.086	0.262	0.278	0.096	0.021	0	0
1975	0.864	0.047	0.051	0.027	0.009	0	0	0
1976	0.172	0.483	0.291	0.049	0.005	0	0	0
1977	0.862	0.025	0.093	0.020	0	0	0	0
1978	0.705	0.184	0.096	0.014	0.001	0	0	0
1979	0.672	0.188	0.127	0.012	0.001	0	0	0
1980	0.449	0.221	0.286	0.044	0	0	0	0
1981	0.552	0.418	0.029	0.001	0	0	0	0
1982	0.709	0.259	0.027	0.004	0.001	0	0	0
1983	0.699	0.241	0.052	0.007	0	0	0	0

$$C = \sum_{i=1950}^{1971} [\ln k_r N_{i,1}^{ind} - \ln N_{i,1}]^2/(2\sigma_r^2)$$

Equation (7.3) assumes that the parameters σ^2 and σ_r^2 are considered given so that the term $-n\ln[\sigma\sigma_r]$ can be left out of the loglikelihood function. The parameters b_1 and b_2 are selectivity function parameters defined in Section 7.2.3.

7.2.2 The use of guano data prior to 1950

An important feature of an age structure matrix such as the one shown in Table 7.2 becomes apparent when one views it in terms of the information pertaining to a specific cohort. The information in question occurs along the diagonal for each cohort in either the catch-at-age matrix or the age structure matrix. Note that there is only one data point on the diagonal corresponding to both the most recent and the oldest cohort. These data points lie at (i,j) = (1983,1) and (i,j) = (1950,8) respectively. For the next most recent and next oldest cohorts there are two data points each, three for each of the following cohorts, and so on until one reaches an inner limit beyond which no further information is available for additional cohorts. The maximum number of data points for the "inner" cohorts is m. If the first year in the rectangular matrix of catch age structures is represented by Y then the birth date for the oldest cohort is Y-m+1. The number of data points for each cohort increases from 1 at birth date Y-m+1 to m at birth date Y. These data decrease again (from m) for cohorts born in year Y+n-m onwards reaching one point for the the most recent cohort born in year Y+n-1.

Thus, for pilchard cohorts born between the years 1943 and 1949 and the years 1976 and 1983, there are fewer data in the VPA than for other pilchard cohorts. On the basis of statistical considerations, one might therefore anticipate that the estimates of cohort size for cohorts making fewer appearances in the catch-at-age matrix than the "inner" cohorts will be less precisely determined than other cohorts sizes. This effect is evident in results obtained for the South African anchovy in Chapter 5, where results (see for example Tables 5.9 and 5.10) show that CV's for biomass estimates increase in the most recent years. This effect is not particularly clear for the very early biomass estimates and CV's shown in these tables.

Additional "constraints" can be imposed on the earlier pilchard cohorts by a more imaginative use of the available recruitment indices (given in term of guano harvests). Specifically, recruitment indices prior to the first year in the catch at age matrix can be used in addition to the available series for 1950 - 1971. It is anticipated that this will reduce imprecision for the very early estimates of pilchard biomass and cohort size (years Y to Y+m-2 and birth dates Y-m+1 to Y-1). The same does not apply to the more recent cohorts for which few data are available since recruitment indices are given for these.

In 1950, the pilchard cohorts born prior to that year appear in the population as the set $N_{1950,j}$, for j=1 to m. The only notable annual pilchard catch made before 1950 is a landing of 17 thousand metric tonnes in 1949 (Butterworth, 1983; see also the pilchard landings set out in Table 7.3). This is very small compared to subsequent landings, so that fishing mortalities for years earlier than 1950 can be safely ignored in the analysis presented here. Cohort size estimates for 1950 are therefore given in terms of their yearclass strength at recruitment (i.e. for j=1) by the following expression:

Table 7.3. South African pilchard landings since 1949 in thousand metric tonnes (Butterworth, 1983; DAMREP, 1983).

1949	17	1961	402	1973	69
1950	85	1962	410	1974	16
1951	102	1963	390	1975	89
1952	170	1964	256	1976	176
1953	133	1965	205	1977	58
1954	88	1966	118	1978	97
1955	122	1967	70	1979	53
1956	77	1968	108	1980	50
1957	110	1969	56	1981	46
1958	194	1970	62	1982	34
1959	260	1971	88	1983	61
1960	318	1972	104	1984	27

$$N_{1950-j+1,1} = N_{1950,j} \exp(-M(j-1)) \qquad \ldots\ 7.4$$

Thus the term C in the log-likelihood function (equation 7.2) can be modified to include the guano data from years 1950-m+1 to 1950-1; for m=8, this corresponds to extending the lower limit of the sum in equation 7.2 to 1943:

$$C' = \sum_{i=1943}^{1971} [\ln k_r N_{i,1}^{ind} - \ln N_{i,1}]^2/(2\sigma_r^2) \qquad \ldots\ 7.5$$

7.2.3 Changes in pilchard catchability with age and over time

The minimum allowable mesh size for purse-seining in South African waters was changed from 38 mm to 12 mm during the mid-1960's (Crawford et al, 1978). This would certainly have caused changes in the relative catchability of pilchard of different ages. The changeover took 2 to 3 years, and most nets were replaced by the beginning of the 1964 season. For the pilchard VPA, this process has been modelled by defining two asymptotic selectivity functions (equation 3.22) for pilchard, one each for the period 1950-1963 and the period 1964-1983:

S_j (1950-1963) = $j/(b_1+j)$

S_j (1964-1983) = $j/(b_2+j)$

This means that the change in net mesh size has been modelled as a step change in selectivity, occurring at the beginning of 1964.

7.2.4 Zero's in the catch-at-age matrix

For pilchard, apart from M, σ^2 and σ_r^2 for which independent estimates are condidered given, the log-likelihood function is maximised by fitting 78 parameters on the basis of 280 pieces of information. There are 55 zeros in the pilchard catch-at-age matrix (Table 7.2 shows extra zeros, but these are small but non-zero values of $P_{i,j}^{obs}$ less than 0.0005). If the zero's are not regarded as input information, then the number of parameters and input information are as shown in Table 7.4)

The zeros referred to in the preceding paragraph occur most frequently in the 5, 6 and 7 year old age classes. Note that in the event of a zero value for $P_{i,j}^{obs}$, the corresponding i,j contribution falls out of the first term in equation 7.3. In this case, whether zeros are regarded as information or not, it amounts to the same thing since either way the associated $P_{i,j}$ parameter is eliminated. The only prior knowledge about $P_{i,j}$ is thus the constraint:

$1 > P_{i,j} \geq 0.$

Non-zero estimates of $P_{i,j}$, even where $P_{i,j}^{obs} = 0$, may however arise through the ß parameters, and these exert some influence in the sum $\Sigma \exp(\beta_{i,j})$ which occurs in both terms A and B of equation 7.3. The procedure adopted here, of assuming $P_{i,j} = 0$ when $P_{i,j}^{obs} = 0$ is strictly incorrect in terms of the $F_{i,j} = F_i S_j$ model. However, from a practical viewpoint it may be preferable. At various points in this thesis reference has been made about the deficiencies in the VPA fitting procedure with regard to process errors. The model naively assumes the best fit $P_{i,j}$ estimates correspond to the actual landed catch proportions, whereas in actual fact these would be the values best conforming to the particular model of selectivity ($F_{i,j} = F_i S_j$ and $S_j = j/(b+j)$) in use. In the event that the process error variance is larger than sampling error variance, the assumption $P_{i,j} = P_{i,j}^{obs}$ is most likely a preferred fitting strategy. The discrepancy between fitted $P_{i,j}$ and the actual landed proportion is most likely to be greatest for $P_{i,j}^{obs} = 0$, since these cases

Table 7.4. A breakdown of the data which are used in the VPA for pilchard, including the Lambert's Bay Bird Island guano harvest as a recruitment index (for years 1943 - 1971) and the model parameters which are estimated (n = 34, m = 8), .

Model Parameters	Quantity	Input Information	Quantity
Fundamental ß's:	n+m-1	Estimates of catch mass:	n
Year dependent fishing mortality effects:	n	Estimates of age-structure	nm-55
Parameters defining the two separate selectivity functions:	2		
The proportionality constant between guano and recruitment	1	Recruitment indices:	29
TOTALS:	2n+m+2=78		n+nm-26=280

exert the least weighting in the likelihood function (they are virtually eliminated, apart from a very indirect appearance in the sum $\Sigma \exp(\beta_{i,j})$). The fitting strategy $P_{i,j} = 0$ when $P_{i,j}^{obs}$ is equal to zero has been adhered to here and in Chapter 8. Admittedly the basis for doing this is not entirely consistent, but it is thought likely to produce less biassed results and therefore to be preferable.

7.3 THE USE OF AGE-LENGTH KEYS FOR THE SOUTH AFRICAN PILCHARD

The age-length keys used for the South African pilchard up to 1976 are tabulated by Crawford et al (1978). For the period 1950 to 1976, only 6 different age-length keys are used, and in particular only one key is used for estimating catch age structures over the period 1950 to 1971. The biases due to the use of a single age-length key, under the assumption of invariant length-at-age distributions in the catches, have been illustrated in Chapter 5.

There are further problems with pilchard aging, which is currently undergoing revision. For example, it is now thought that pilchard previously estimated as 4 years old, were actually 2 years old , and some crude adjustments were made to the various age-length keys reported by Crawford et al (1978) to account for this. These adjustments are reflected in the pilchard catch age structure data used here and which were obtained from the Sea Fisheries Research Institute of South Africa (M. Armstrong, pers. commn) (Table 7.2).

The pilchard catch age structure estimates are therefore extremely imprecise, and VPA calculations are poorly founded. Even so, pilchard VPA results have often been quoted in the literature. The uncertainties associated with such VPA assessments should therefore be quantified, to allow a more meaningful interpretation of the resultant biomass trends (computer time limitations in subsequent calculations were however such that the evaluation of variances by the methods discussed in Chapter 4 proved to be impractical).

7.4 TESTING THE NUMERICAL ESTIMATION PROCEDURE

The optimisation procedure was tested as before, using simulated data without any sampling error, and the software was assumed to be free of errors when true and estimated abundances matched each other to within about 6 significant figures. Double precision variables and optimisation routines were used throughout. The log-likelihood function was scaled by subtracting from it the term $q_{i,j}^{obs} \ln P_{i,j}^{obs}$ summed over all i and j (but excluding terms with $q_{i,j}^{obs} = P_{i,j}^{obs} = 0$). In the absence of any model or sampling errors this would give the loglikelihood function a value of zero for the correct model parameter values. This procedure has been shown in Chapter 4 to produce parameter estimates which are insensitive to starting parameter values used.

7.5 RESULTS

Biomass estimates for pilchard were obtained using the following given values: $M=0.5$ yrs^{-1} (Butterworth, 1983), $\sigma=0.05$, $\sigma_r=0.3$ and $Q=1000$. The chosen value of $\sigma_r=0.3$ has little basis. In Chapter 4 it is suggested that one of the weighting constants is potentially estimable, however it will be seen in Chapter 8 that this may not be feasible (in the sense of producing realistic results) in practical situations. Some value must be used, one that does not eliminate the influence of the recruitment indices.

Ultimately the results will depend critically on the value chosen; the best that can be achieved is to ultimately (not done here) demonstrate the sensitivity of final results to σ_r.

Spawning biomass estimates corresponding to the aforementioned constants are set out in Figure 7.1. The <u>absolute</u> biomass levels obtained prior to 1955 are very large, perhaps unrealistically so, when compared to independent estimates of ecosystem carrying capacity. Shannon and Field (1985) estimate for example that the pelagic system carrying capacity for shoaling fish is in the region of 2 million metric tonnes, suggesting that the earlier VPA biomass values (for 1950-1955) which are in excess of this amount are untenable. Further analyses based on the VPA output whose results are given later in this Chapter suggest however that the value of K, the equilibrium spawning biomass, is about 2.42 million metric tonnes, so that the VPA biomass estimates above this should perhaps be viewed as natural fluctuations above the mean.

The spawning biomass <u>trend</u> shown in Fig. 7.1 is compared to results produced by Centurier-Harris (1977). The essential differences between the results produced here and all earlier estimates (Centurier-Harris's results are typical of all published results preceding those shown here) are clear from Fig. 7.1, viz. whereas the earlier estimates suggest a marked peak in biomass in about 1960 compared to average levels, estimates produced here suggest this peak was only a momentary and relatively inconsequential recovery at about that time (circa 1960) superimposed on a generally declining trend.

It is important now to restate the interpretation accorded these early estimates of pilchard biomass trends. Results such as

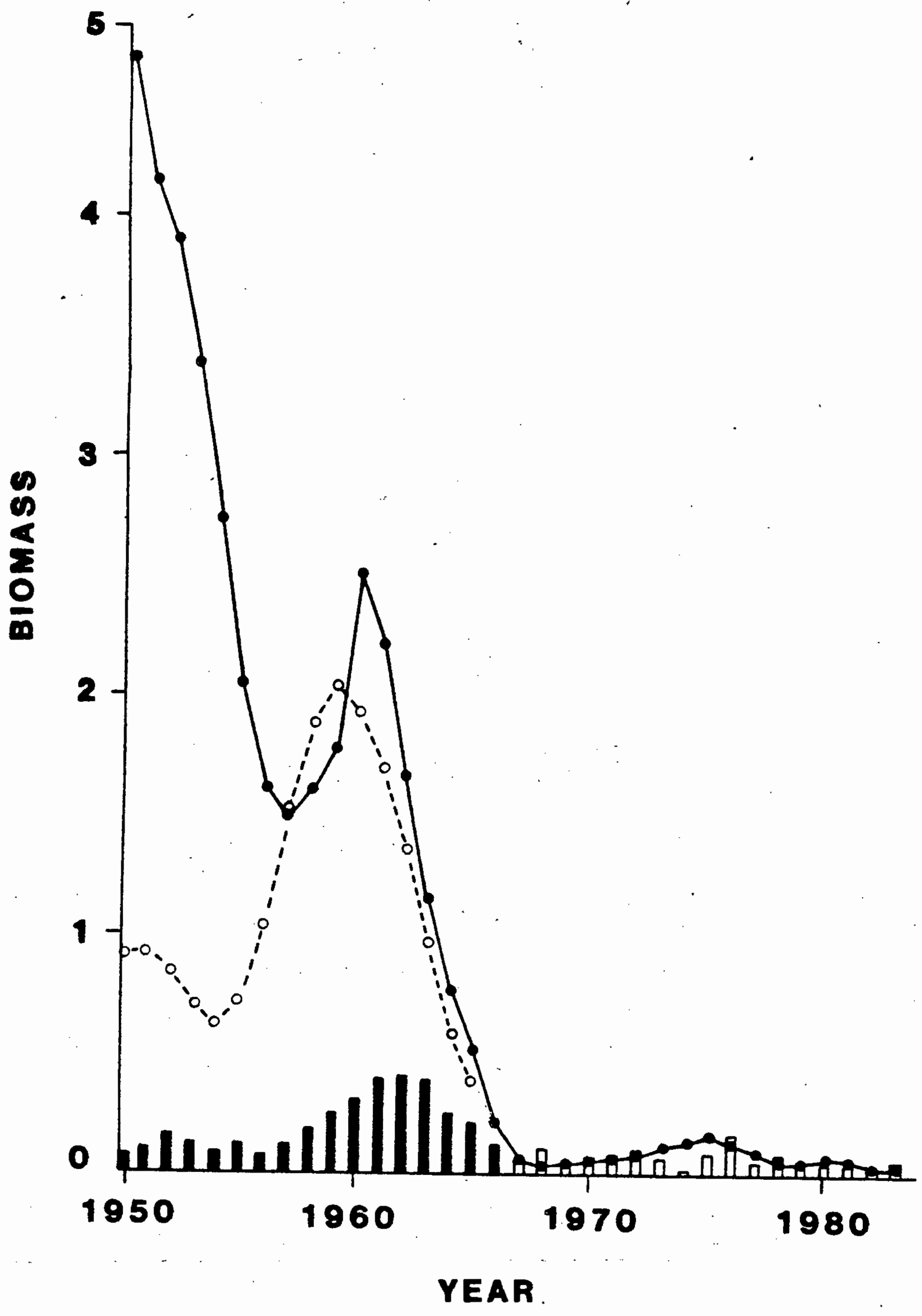


Figure 7.1. The biomass trajectory (units: millions of tonnes) for the best fit of the VPA model to the pilchard catch mass, age structure and guano based recruitment index is shown by the solid line. The dotted line shows the VPA assessment of Centurier-Harris (1977) based only on catch mass and age-structure information. Vertical bars indicate annual catches.

those of Centurier-Harris have led to the interpretation that the growth of pilchard catches in the late 1950's and early 1960's followed the entry of some very strong year classes into the fishery (eg. Crawford et al, 1983), and that subsequently the fishery collapsed following a few years of very poor recruitment. In contrast the main feature of the maximum likelihood estimate of pilchard stock size trends is a decline in biomass from as early as 1950.

From the most recent results produced for this Chapter, it appears that the relative rate of decline in biomass increased markedly in the early 1960's (there is no overall decline from 1954/55 - 1960); from 1960-1965 the overall decline was about 80%; thereafter stock size seems to have remained at a very low level (less than 200 000 tonnes). This suggests that the history of the pilchard fishery may represent little more than a straightforward case of overfishing. Note that just such a reverse S-shaped biomass trend (if one smooths the years 1950-1960) would be anticipated from a population regulated approximately according to logistic growth dynamics, where the fishing effort increases steadily. In particular, the marked decline between 1960 and 1965 would coincide with the population size being reduced below the MSY biomass level. Below that level surplus production declines, in due course precipitating a stock "collapse".

It is of course relevant that the estimates of recruitment for the years 1960 - 1965 all fall below the replacement line (see Figure 7.2), which supports the notion of a recruitment failure contributing to the collapse. However it is quite possible that this is a consequence of inaccuracies in the catch-at-age matrix arising from the use of inappropriate age-length keys (see

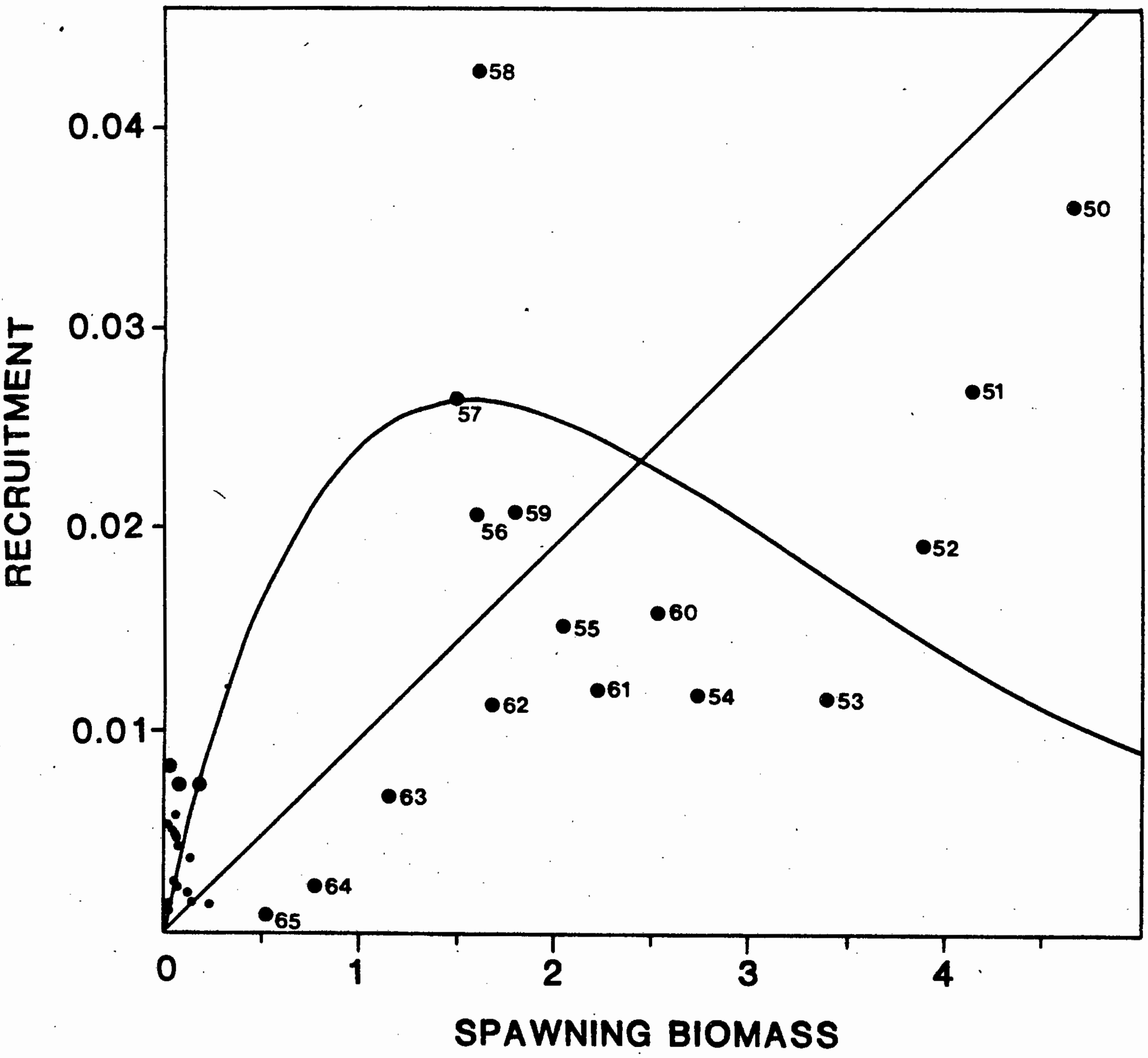


Figure 7.2. Plots of estimated pilchard recruitment and spawning biomasses obtained from the VPA model fit using catch mass, catch age structure, and guano harvest data. The stock-recruit relationship shown is fitted by a method discussed in section 7.6.2. The inclined straight line through the origin is the replacement line (i.e. number of recruits required to maintain the stock in equilibrium at that spawning biomass level). Small dots in the lower lefthand corner represent recruitment and spawning biomass since 1969, in an attempt to achieve better diagrammatic resolution.

Section 7.3). It is also clear however that the substantial biomass decline suggested by VPA for 1950 to 1957 could not have been caused entirely by fishing as catches were relatively low over that period.

One has only very crude data to appeal to for further affirmation of the VPA results; ultimately therefore it may be impossible to ascertain which of recruitment failure or overfishing contributed *most to* the demise of the South African pilchard. For the moment, and the foreseeable (and perhaps indefinite) future, this question must be regarded as unresolved.

7.6 MAXIMUM SUSTAINABLE YIELD ESTIMATES

The maximum sustainable yield for pilchard was calculated using the convexity in the relationship between the spawning biomass and recruitment as a measure of the density dependence of pilchard population dynamics. To do this it is necessary to estimate the mean stock recruit function from the limited data which are provided by the VPA. Allowances for likely error structures must be made. The recruitment and spawning biomass quantities, R_i and Y_i, estimated in the VPA are both subject to measurement error arising from the limitations of the catch mass, age structure and guano data. This error may be modelled as either additive normal or multiplicative log-normal, to name two obvious error models which are easy to implement. If r_i and y_i are the true recruitment and spawning biomass quantities, then either:

$$R_i = r_i + v_i$$

$$Y_i = y_i + w_i \qquad \ldots\ 7.6$$

or

$$R_i = r_i * v_i.$$

$Y_i = y_i * w_i$ ... 7.7

where v_i and w_i represent the estimation errors.

In estimating the recruitment function parameters, a further set of errors due to environmental and biological variation should be considered for recruitment only. For the Ricker stock/recruit relationship, this model prediction error or process error (Walters, 1984) may again either be additive or multiplicative (normal or lognormal for convenience):

$r_i = \alpha y_i \exp(-\beta\, y_i) + u_i$ ... 7.8

or

$r_i = \alpha y_i \exp(-\beta\, y_i) * u_i.$ ... 7.9

where the u_i comprise the process errors.

Walters (1984) notes that random effects in recruitment rates often seem to follow a lognormal pattern (most points near the recruitment curve, but a few very high outliers). He argues that precisely this pattern is expected when there is no small set of major factors that cause the variation, but rather variation is due to many small random effects operating at various life stages of the fish prior to recruitment. There seems little reason to dispute this in the case of the South African pilchard, and it is assumed in what follows that pilchard recruitment is lognormally distributed (homoscedastic errors after log-transformation) about the recruitment function. Furthermore, previous results presented in this thesis (Table 4.6; results using effort information) suggest that the measurement errors for recruitment and spawning stock biomass resulting from the VPA could be multiplicative and very roughly homoscedastic, and these can thus be modelled as lognormal multiplicative errors.

7.6.1 Ludwig and Walters' procedure for estimating the recruitment relationship

The underlying model suggested by Ludwig and Walters (1981) for fitting the Ricker stock/recruit function in the face of the likely error structures mentioned earlier can be summarised as:

$$Y_i = y_i * \exp(w_i)$$

$$R_i = r_i * \exp(v_i)$$

$$r_i = \alpha y_i \exp(-\beta y_i + u_i) \qquad \ldots \; 7.10$$

The errors are all assumed to be homoscedastic and normally distributed:

$$w_i \approx N(0, \sigma_w^2)$$

$$v_i \approx N(0, \sigma_v^2)$$

$$u_i \approx N(0, \sigma_u^2)$$

To proceed with some sort of maximum likelihood estimation procedure it is necessary to be able to assume that u_i, v_i and w_i are independent. This is not strictly the case, since the "observed" spawning biomass as derived from the VPA is functionally related to "observed" recruitment. Thus for example positive errors in recruitment lead to positive errors in spawning biomass in later years. For this reason, estimates produced here, under the assumption of independence of the errors u_i, v_i and w_i may not be entirely reliable.

Ludwig and Walters have considered the estimation procedure outlined so far for the situation where $w_i = v_i$ and the ratio σ_u^2/σ_v^2 is known. Essentially the equality $w_i = v_i$ permits the effects of measurement error and model prediction error to be separated, so that individual r_i can be estimated. Since $w_i = v_i$, the likelihood function with parameters α, β, σ_u, σ_v and the set of v_i's can be set up as a weighted sum of squares of u_i's and v_i's. The general form of the problem of estimating parameters that maximise this function is intractable, and it is necessary to make further assumptions before a solution is possible (Kendall and Stuart, 1967). The critical assumption which must be made is that a value is known for the ratio σ_u^2/σ_v^2, which is the relative weighting for the sum of squares of the v_i's and u_i's. Given this value, the log-likelihood function, with the v_i's, α, β and σ_u^2 as parameters is:

$$L(\alpha,\ \beta,\ \sigma_u^2,\ v_1, \ldots,\ v_n) = -[\sum_{i=1}^{n} u_i^2 + \sum_{i=1}^{n} v_i^2/\eta]/(2\sigma_u^2) - n\log(\eta^{1/2}\sigma_u^2) \qquad \ldots\ 7.11$$

where $\eta = \sigma_v^2/\sigma_u^2$.

Ludwig and Walters suggest that σ_u^2 should be estimated separately (not in the minimisation process) using:

$$\sigma_u^2 = [\Sigma\ u_i^2 + \Sigma\ v_i^2/\eta]/(2n-n-3) \qquad \ldots\ 7.12$$

The denominator (2n-n-3) rather than 2n follows if one applies a degrees of freedom correction for bias of 2n/(2n-n-3) to the maximum likelihood estimate of σ_u^2. There are 2n data, n+3 model parameters (the n v_i's, α, β and σ_u^2), and therefore 2n-n-3 degrees of freedom. This correction assumes that the error model is linear, and that the errors involved all have equal variances (Kendall and Stuart, 1967); this is not the case, and so the correction is really of an ad hoc nature. The estimation procedure proposed by Ludwig and Walters' can be summarised by the following 6 steps:

1. α, β, v_i's estimated for each iteration by a first derivative approximate minimisation algorithm (Newton's method).
2. $y_i = Y_i * \exp(-v_i)$
   $r_i = R_i * \exp(-v_i)$
3. $u_i = \ln(r_i/(\alpha y_i)) + \beta y_i$
4. Sum the term, $u_i^2 + v_i^2/\eta$, over all i.
5. Repeat 1-4, to minimise the sum obtained at stage 4.
6. The stock recruit function is given in terms of the estimated values of α and β.

Ludwig and Walters' basic procedure has been adapted for use in the case of the South African pilchard as shown in the following section.

7.6.2 An adaptation of Ludwig and Walters' (1981) basic model

It is necessary at this stage to consider the fitting procedure for the situation where $w_i \neq v_i$, which is an unavoidable consequence of the VPA estimation procedure (Ludwig and Walters do comment that their procedure can be adjusted to deal with this case). It was noted earlier that the equality $w_i = v_i$ permits the disentanglement of model error and measurement errors, so that an estimate of the effective annual recruitment rate can be obtained. From the point of view of obtaining an estimate of mean maximum sustainable yield, it is only necessary to have some estimate of the mean recruitment function, without specific knowledge about individual annual recruitment rates. The α and β parameters can be fitted without knowledge of the separate effects of the u_i and v_i, and the estimates of α and β lead directly to estimates of MSY.

To use Ludwig and Walters' estimation procedure, note that u_i and v_i must be combined (they cannot be separated at any later may be obtained from:

$$\partial L/\partial \sigma_\varepsilon = \sigma_\varepsilon^{-3} \left[\sum_{i=1}^{n} \varepsilon_i^2 + \sum_{i=1}^{n} w_i^2/\eta\right] - 2n/\sigma_\varepsilon \qquad \ldots\ 7.17$$

Setting this equal to zero produces the estimate:

$$\sigma_\varepsilon^2 = \left[\sum_{i=1}^{n} \varepsilon_i^2 + \sum_{i=1}^{n} w_i^2/\eta\right]/(2n) \qquad \ldots\ 7.18$$

To allow for bias, the (ad hoc) adjustment factor (2n-n-3)/2n (Ludwig and Walters, 1981) is used to give:

$$\sigma_\varepsilon^2 = \left[\sum_{i=1}^{n} \varepsilon_i^2 + \sum_{i=1}^{n} w_i^2/\eta\right]/(2n-n-3) \qquad \ldots\ 7.19$$

A value for σ_ε^2 is needed to obtain a best estimate of the recruitment function $r(\alpha, \beta, y)$. The maximum likelihood procedure estimates $\ln[r(\alpha, \beta, y)]$, and so some correction for skewness in the antilog transformation to get a reasonable estimate of the average recruitment (per se, rather than log-transformed) is required.

The correction for obtaining the mean of x, given an estimate of the mean of ln(x), where ln(x) is normally distributed, is given by Bradu and Mundlak (1970). If the maximum likelihood estimate is an estimate of the expectation of $\ln[r(\alpha, \beta, y)]$ then the expectation of $r(\alpha, \beta, y)$ is, according to Bradu and Mundlak, given by:

$$r_i = \alpha y_i \exp(-\beta y_i)\, g_q(\sigma_u^2/2) \qquad \ldots\ 7.20$$

where:

$$g_q(t) = \sum_{p=0}^{\infty} \frac{q^p(q+2p)}{q(q+2)\ \ldots\ (q+2p)} (q/[q+1])^p\, t^p/p! \qquad \ldots\ 7.21$$

where q is the number of degrees of freedom ($q=n-3$).

However, for typical values of $\sigma_u^2/2$ (<0.5) considered here and for $n>15$, the use of the exponential function instead of g_q produces positive biases of no more than 4% (from data presented by Bradu and Mundlak, 1970). Since σ_u^2 has been assumed to be half of σ_ε^2, the mean corrected recruitment function is therefore estimated to be:

$$r_i = \alpha y_i \exp(-\beta y_i + \sigma_\varepsilon^2/4) \qquad \ldots\ 7.22$$

and the correction can be incorporated as an adjustment to the estimate of α:

$$\alpha \rightarrow \alpha \exp(\sigma_\varepsilon^2/4) \qquad \ldots\ 7.23$$

Model parameters were fitted using the constrained nonlinear optimisation package (E04JAF of the Numerical Algorithms Group Ltd), which has been discussed previously (Chapter 4), rather than the minimisation algorithm suggested by Ludwig and Walters (1981).

In summary, the adapted α and β estimation procedure consists of the following consecutive steps:

1. Specify α, β, and the w_i's.
2. $y_i = Y_i * \exp(-w_i)$
3. $\varepsilon_i = \ln(R_i/(\alpha y_i)) + \beta y_i$
4. Sum the term $(\varepsilon_i^2 + w_i^2/\eta)$ over all i.
5. Repeat 1-4, to minimise the sum obtained at stage 4, using E04JAF.
6. $\sigma_\varepsilon^2 = [\Sigma\ \varepsilon_i^2 + \Sigma\ w_i^2/\eta]/(2n-n-3)$
7. $\alpha \rightarrow \alpha \exp(\sigma_\varepsilon^2/4)$

Following this, MSY, y_{MSY} and K (equilibrium spawning biomass in the absence of fishing) are calculated by the procedure described in Appendix 7.2.

7.6.3 Estimates of MSY and associated variance

The resampling procedure to estimate variances for the VPA has been described in Chapter 4. This produces a set of N possible series of recruitment and spawning biomass estimates, $\{(R_i;Y_i):$ $i=1...n\}$. The computer time necessary for generating these estimates of spawning biomass and recruitment is restrictive, and so to extend the number of possible data sets to allow variance estimation, a further resampling procedure was adopted. The procedure suggested here for estimating $\sigma_{MSY}{}^2$ is validated in subsequent simulations.

For a particular series of $\{(R_i;Y_i):$ $i=1...n\}$ estimated using VPA techniques, further resampled data series were obtained by randomly selecting pairs from the N data sets of $\{(R_i;Y_i):$ $i=1...n\}$. Doing this with replacement, M times, yields a total of MxN sets of $\{(R_i{}^*;Y_i{}^*):$ $i=1...n\}$ (N of these are unique and MxN-N are resampled). In total therefore NxM estimates of MSY are produced (one for each $\{(R_i;Y_i):$ $i=1...n\}$ series), and these can be used to produce a variance estimate for MSY. The method for obtaining the variance estimate for MSY can be summarised as:

1. From the resampling procedure used for the VPA, N possible sets of spawning biomass and recruitment are obtained: $\{(R_i;Y_i):$ $i=1...n\}$.
2. Resample each set of $\{(R_i;Y_i):$ $i=1...n\}$ with replacement M times, to produce the additional (MxN-N) sets $\{(R_i{}^*;Y_i{}^*):$ $i=1...n\}$, a total of MxN data sets of spawning biomass and recruitment series.

3. Obtain α and β estimates for each set using Ludwig and Walters' method, and use these to evaluate MSY.

4. Calculate the variance of the MSY estimates. These are analogous to $(\sigma_{MSY}^2)^c$ discussed in the simulation exercise presented in section 7.8.

In practice this procedure was carried out for N = 1 and M = 30.

7.7 RESULTS FOR THE ACTUAL DATA

The estimated spawning biomass and recruitment data (from VPA) and fitted stock-recruitment function are shown in Fig. 7.2.

Results for MSY were computed using the full set of 34 years of (R_i, Y_i) data obtained from the VPA model fit to the catch mass, age structure and guano data. Selectivities used were those obtained for the period 1964 - 1983 - the best fit suggested no dependence of selectivity on age, i.e. $b_2 = 0$.

The MSY estimate for pilchard was 442 thousand metric tonnes. The C.V. obtained for this estimate was 22 %. The estimated MSY biomass level was 1.02 million tonnes (C.V. = 13 %) and the estimated unexploited population size was 2.42 million tonnes (C.V. = 14 %). These estimates are equivalent to the simulation results obtained in Trial B discussed in the following section. The estimate of σ_u was 0.6, and therefore the calculations were based upon $\sigma_v = \sigma_w = 0.6$ since η was assumed to be 1/2.

7.8 SIMULATIONS FOR DETERMINING THE ACCURACY OF ESTIMATES OF σ_{MSY}^2 OBTAINED USING THE RESAMPLING PROCEDURE

The reliability of the MSY variance estimation procedure (Section

7.6.3) was investigated in a series of simulations. These simulations were also used to illustrate the effect of bias in Ludwig and Walters' estimation procedure for α and β on subsequent MSY estimates.

There are two crucial determinants of the efficiency of the MSY estimator. These are (1) the spawning biomass data contrast and (2) the length of the $\{(R_i;Y_i): i=1...n\}$ time series. Data contrast was regulated by adjusting the total variance of the true (simulated) spawning biomass quantities, $\ln(y_i)$, to be the same as the actual observed total variance of VPA estimates of $\ln(Y_i)$ for the South African pilchard. The modal spawning biomass was set equal to $1/(2\beta)$, with the following stochastic relationship used for generating a typical y_i series:

$$y_i=\exp(x_i)/(2\beta) \qquad \text{... 7.24}$$

where $x_i \approx N(0,\sigma_x^2)$

Bearing in mind that peak recruitment occurs at a spawning biomass of $1/\beta$, and that for pilchard the replacement line cuts the recruitment curve to the right of this point, the situation modelled is that of a heavily exploited stock, and hence is appropriate to the current status and available historic data of the pilchard fishery. The value of σ_x suggested by the total variance of $\ln(Y_i)$ obtained from the VPA was 1.9. A data series of length 34 years was used, which is the same as that available for pilchard. A typical pilchard population was considered in the simulation ($M=0.5$ yrs^{-1}), using model parameters in the vicinity of values estimated for pilchard from previous computations.

Actual recruitment values (r_i) were generated from y_i using a value of $\sigma_u = 0.6$, as suggested by the value of $(\sigma_\varepsilon^2/2)^{1/2}$ estimated by the method described in section 7.6, and typical values of the stock recruit parameters α (0.0378 thousand million recruits per million tonnes of spawning biomass) and β (0.644 per million tonnes of spawning biomass) from fits to the real pilchard data. Observation errors, v_i and w_i, were generated using $\sigma_v=0.6$ and $\sigma_w=0.6$.

In the simulation, 3 different trials were conducted, resulting in the estimates MSY^A, MSY^B, MSY^C, $(\sigma_{MSY}^2)^A$, $(\sigma_{MSY}^2)^B$, and $(\sigma_{MSY}^2)^C$ (the equivalent statistics for y_{MSY} and K were also produced). MSY^A represents the mean estimated MSY in the absence of observation errors, MSY^B is the mean when observation errors for recruitment and spawning biomass are included, and MSY^C is the mean of a set of MSY estimates produced using a resampling procedure. $(\sigma_{MSY}^2)^C$ is an average variance for the MSY resulting from resampling a single set of observed recruitment and spawning biomass data (and not the variance of MSY^C). The difference between MSY^A (this value was used rather than the deterministic MSY because of the non-linear relationship between α and β which almost certainly leads to substantial biases between the deterministic MSY and that derived from a small sample) and MSY^B shows the bias inherent in Ludwig and Walters' estimation procedure for the specific situation under consideration. The difference between $(\sigma_{MSY}^2)^B$ and $(\sigma_{MSY}^2)^C$ indicates the extent to which the resampled estimate of variance over- or under-estimates the true variance of MSY estimates.

The simulation is summarised by the following steps:

1. Simulate a set of true spawning biomass quantities (y_i) with a total variance typical of the South African pilchard, using the

procedure:

$y_i = \exp(x_i)/(2\beta)$ for $i = 1$ to 34.

2. Simulate a set of true recruitment values (r_i) from y_i, including model error: $r_i = \alpha y_i \exp(-\beta y_i + u_i)$. Repeat this a large number of times - 90 repetitions are used here - and store these 90 typical recruitment time series.

3. Split into three different simulations, A, B and C. Each of these produces a mean of MSY estimates and an estimate of σ_{MSY}^2.

TRIAL A

4a. Using all 90 sets (k=1...90) of $\{(r_i; y_i): i=1...n\}_k$ generated in step 2, estimate $MSY_k^{\wedge}$ for each set, using the analytical Ricker model fit described in Appendix 7.1. From these:

$$MSY^{\wedge} = \sum_{k=1}^{90} MSY_k^{\wedge} / 90$$

$$(\sigma_{MSY}^2)^{\wedge} = \sum_{k=1}^{90} [MSY_k^{\wedge} - MSY^{\wedge}]^2 / 89$$

TRIAL B

4b. For each of 90 sets of $\{(r_i; y_i): i=1...n\}_k$ generated in step 2, obtain the equivalent set $\{(R_i; Y_i): i=1...n\}_k$ including observation errors:

$R_i = r_i \exp(w_i)$

$Y_i = y_i \exp(v_i)$

5b. Fit each set $\{(R_i; Y_i): i=1...n\}_k$ using the procedure adapted from Ludwig and Walters (1981), obtaining the 90 MSY estimates, MSY_k^B. From these:

$$MSY^B = \sum_{k=1}^{90} MSY_k^B / 90$$

$$(\sigma_{MSY}^2)^B = \sum_{k=1}^{90} [MSY_k^B - MSY^B]^2 / 89$$

TRIAL C

4c. For each of 90 pairs of $\{(r_i; y_i): i=1...n\}_k$ generated in step 2, obtain the equivalent set $\{(R_i; Y_i): i=1...n\}_k$ including observation errors:

$R_i = r_i \exp(w_i)$

$Y_i = y_i \exp(v_i)$

5c. For each set of $\{(R_i; Y_i): i=1...n\}_k$ obtain a further 29 sets $\{(R_i^*; Y_i^*): i=1...n\}_{k,l}$ $(l=2...30)$ (2700 spawning biomass and recruitment data sets in total, of which only 90 are unique and 2610 are resampled).

6c. Each constituent of the super-set of 29 $\{(R_i^*; Y_i^*): i=1...n\}_{k,l}$ $(l=2...30)$ and its single unique counterpart $\{(R_i^*; Y_i^*): i=1...n\}_{k,l}$ (i.e. designated $l=1$) provides an estimate $MSY_{l,k}^C$. The mean MSY estimate (MSY_k^C) for the k th super-set and the variance of this estimate ($(\sigma_{MSY,k}^2)^C$) is then calculated:

$$MSY_k^C = \sum_{l=1}^{30} MSY_{k,l}^C / 30$$

$$(\sigma_{MSY,k}^2)^C = \sum_{l=1}^{30} [MSY_{k,l}^C - MSY_k^C]^2 / 29$$

7c. MSY^C and $(\sigma_{MSY}^2)^C$ are then given by the means of the values for each of the (90) super-sets:

$$MSY^C = \sum_{k=1}^{90} MSY_k^C / 90$$

$$(\sigma_{MSY}^2)^C = \sum_{k=1}^{90} (\sigma_{MSY,k}^2)^C / 90$$

Note that exactly the same y_i values are used throughout the simulation.

7.8.1 <u>Simulation results</u>

The simulation results are set out in Table 7.5. These do not suggest that MSY^B is positively biassed relative to MSY^A on the basis of a t-test. However the estimated 95% confidence interval on the difference of 44 is [0, 88] suggesting a borderline case. The simulation was repeated using twice as many simulated (R_i, Y_i) data sets. The mean estimates obtained in this case were - MSY^A = 433, MSY^B = 473 - 95% confidence intervals on the difference of 40 are [9, 71]. Therefore a bias correction factor of 433/473 for the actual MSY estimates appears to be warranted. This bias arises from the adapted Ludwig and Walters' (1981) fitting procedure, even when the underlying mathematical relationship and the correct error model are known (i.e. the true value of η is given).

Equivalent estimates for y_{MSY} and K for TRIAL B and C relative to TRIAL A are also shown. Differences evident between trials B or C do not appear to be significant.

It was anticipated that the resampling estimates of variance, the $(\sigma_{MSY}^2)^C$ quantities, would be reduced by a factor of roughly 2 (or $2^{1/2}$ for C.V.'s) relative to $(\sigma_{MSY}^2)^B$ because of the fact that the model degrees of freedom are approximately half the number of

Table 7.5. Simulation results for the estimation procedure (MSY, y_{MSY} and K) used for pilchard. Trials A, B and C are referred to in the text; they represent mean and variance estimates obtained without observation errors on either recruitment or spawning biomass values (A), then including observation errors of equal variance (though the errors in any one year are not necessarily the same) on both recruitment and spawning biomass values (B), and finally using a resampling procedure corresponding to the actual procedure used for estimating the mean and variance of MSY, y_{MSY} and K for pilchard. %C.V.'s are given as $\sigma/x * 100$ where x=MSY, y_{MSY} or K. Units are million metric tonnes. Correction factors for the mean estimates obtained in practice are calculated to be the ratio of the mean estimate obtained in TRIAL A divided by the mean estimate from TRIAL B (when the difference has been demonstrated to be significant).

	TRIAL A	TRIAL B	TRIAL C
MSY	0.434	0.478	0.492
σ_{MSY}	0.065	0.20	0.14
%C.V.	15	42	28
y_{MSY}	0.906	0.955	1.03
σ_{BMSY}	0.070	0.39	0.35
%C.V.	7.7	41	34
K	2.17	2.29	2.37
σ_K	0.176	0.896	1.56
%C.V.	8.1	39	66

input data (this would be the true reduction for a linear model). Reduction in the variance estimates for MSY and y_{MSY} between trials B and C is evident; however the variance of K in TRIAL C is larger than that obtained in TRIAL B (see Table 7.5).

On the basis of these simulations, the estimate of MSY for pilchard therefore becomes 442x433/473 thousand metric tonnes, i.e. 405 thousand tonnes.

The MSY estimate should, in addition to a bias correction factor, be adjusted by an amount to account for the difference between the observed annual catch masses and those given by summation with respect to age in:

$$^{m}C_i^{obs'} = \Sigma\, w_j\, P_{i,j}^{obs}\, C_i^{obs} \qquad \ldots\ 7.25$$

The difference between the annual catch landed, and the equivalent midyear catch $^{m}C_i'$ is due to the difference between integrating for each age class either w(t)N(t)F (catches made) or w(0.5)N(t)F (as actually used for the MSY estimate) over the whole year, where w(t) is instantaneous fish mass, N(t) the number in the cohort, F the fishing mortality and w(0.5) the midyear fish mass for the appropriate age. This is illustrated by the difference between catch masses obtained from summing with respect to age, $\sum_j w_j P_{i,j}^{obs} C_i^{obs}$, and summing according to the breakdown of the catch by length, $\sum_l w_l f_{i,l}^{obs} C_i^{obs}$; since there are more length classes than age classes the summation involving masses at length is expected to give a better approximation to the realised catch mass.

Since in practice C_i^{obs} is derived using $f_{i,l}^{obs}$ and $^{m}C_i^{obs}$ (the catch mass), the term $\sum_l w_l f_{i,l}^{obs} C_i^{obs}$ is equal to the observed catch mass. The equivalent midyear catch masses for the historic time series of catches are plotted versus the realised catches in Fig 7.3. The differences between these two approximations are such that the first term (using numbers caught at age derived from numbers caught at length using the age length key) gives a catch ("C'^{obs} in 1962 of 506 thousand metric tonnes, compared to the observed catch ("C^{obs}") of 410 thousand metric tonnes.

It is not clear that the numbers caught at age in the VPA should be adjusted to force $\Sigma w_j P_{i,j} C_i$ equal to the direct catch mass estimate for all years, since the catch estimate $\Sigma w_j P_{i,j} C_i$ may in fact be the correct equivalent midyear approximation to the annual catch mass. In the latter event it is necessary to investigate the effect of individual fish growth during the year on the MSY estimation procedure. Ideally, the problem should be reformulated to treat the $f_{i,l}$ as model parameters, instead of the $P_{i,j}$ quantities, which eliminates any need for calculating catch masses in terms of the parameters $P_{i,j}$. Appropriate techniques for dealing with this approach have been developed elsewhere (D. Fournier, pers. commn). At the moment however, it seems reasonable to reduce the MSY estimate by about 20%, as indicated by Fig. 7.3.

The final pilchard MSY estimate obtained at this stage is therefore 324 thousand metric tonnes (C.V. is roughly 33%, after correcting the 22% originally obtained by multiplying by a correction factor to allow for bias in the resampling procedure for variance estimation; this factor is taken to be $\sigma_{MSY}{}^{B}/\sigma_{MSY}{}^{C}$, using the numerical results shown in Table 7.5.

7.9 CONSISTENCY CHECKS

The virgin spawning biomass level estimate of 2.4 million metric tonnes agrees well with calculations of the maximum carrying

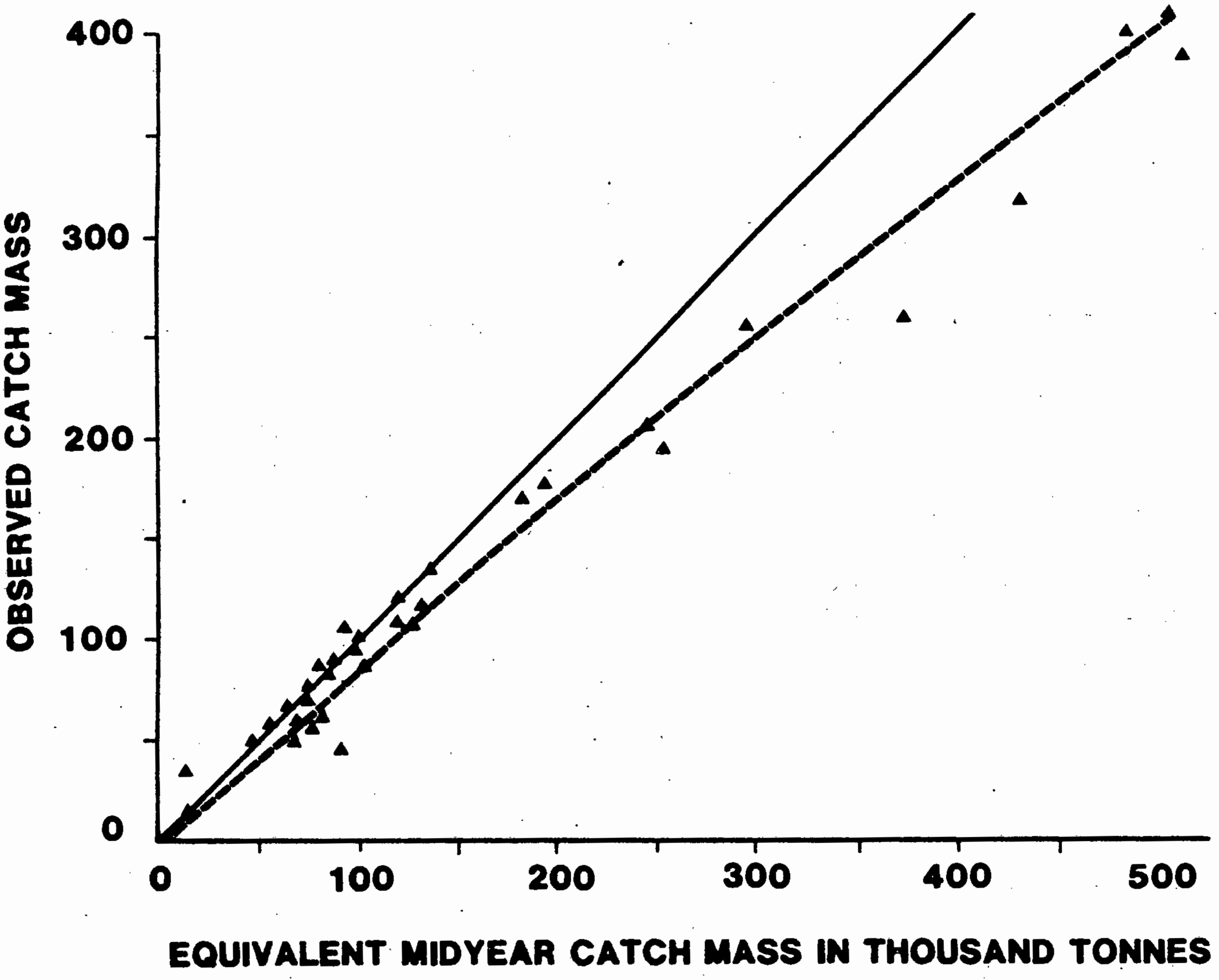


Figure 7.3. A plot of the observed pilchard catch masses ${}^{m}C_{i}^{obs}$ versus the quantities ${}^{m}C_{i}'^{obs}$. The latter results from the following calculation steps:

$$\underset{\sim}{C}^{L}_{i} = C_{i,l}^{obs} = {}^{m}C_{i}^{obs} f_{i,l} / \sum_{l} w_{l} f_{i,l}$$

$\underset{\sim}{C}^{A}_{i} = \underset{\sim}{C}^{L}_{i}\, \underline{a}$ where $\underset{\sim}{C}^{A}_{i} = C_{i,j}^{obs}$

$${}^{m}C_{i}'^{obs} = \sum_{j} w_{j} C_{i,j}^{obs}$$

The solid line is the y=x plot, with a slope of unity. The dashed line is a hand drawn curve with a slope of roughly 0.80.

capacity of the system of 2 million metric tonnes estimated from gross carbon fixing rates achieved by phytoplankton (Shannon and Field, 1985).

In Chapter 6 it was argued that the pilchard population described undamped fluctuations prior to its collapse: the estimates of MSY, y_{MSY} and K obtained so far should be such that they permit this phenomenon to occur.

A deterministic age structured model was developed for examining the potential for oscillations in pilchard recruitment and spawning biomass. Recruitment for the model was given as a function of spawning biomass by the estimated expected recruitment relationship:

$$\begin{aligned} R_i &= \alpha\ Y_i\ \exp(-\beta Y_i + \sigma_u^2/2) \\ &= 0.0465\ Y_i\ \exp(-0.644 Y_i) \qquad \ldots\ 7.26 \end{aligned}$$

where the factor $\exp(\sigma_u^2/2)$ is a rough correction factor which converts the fitted R_i value (estimated as $\ln R_i$) to the expected value of R_i. The units of R_i are 10^9 fish; Y_i is in million tonnes. The slope of the replacement line, γ, in:

$$R_i = \gamma Y_i \qquad \ldots\ 7.27$$

is 0.00976 (units: thousands of million fish per million tonnes of pilchard).

The replacement line cuts the recruitment curve at - $Y_i = -(\ln(\gamma/\alpha)/\beta$; $R_i = -\gamma(\ln(\gamma/\alpha)/\beta$. To test for instabilities, recruitment was given a small perturbation away from the equilibrium point. For the parameter values quoted immediately above for α, β and γ, an M value of 0.5 yr^{-1}, and using 7 pilchard age classes the results show that recruitment and spawning biomass oscillations persist over a 200 year time span. There was little evidence of either an overall increase or decrease in the oscillation amplitude over the 200 year timespan investigated. [The period of oscillation was about 10 years; this conflicts with a 6 year period suggested for pilchard population oscillations in Chapter 6 - however the difference may be due to the density dependent mortality in the logistic model used in Chapter 6 - the age structured model has density independent natural mortality.] This indicates therefore that there are undamped oscillation solutions for the parameter combinations pertinent to the South African pilchard population, for the model type considered here. The estimates of the pilchard stock-recruit relationship produced here therefore leads to continuous recruitment oscillations as suggested in Chapter 6.

7.10 THE EFFECT OF RANDOM RECRUITMENT FLUCTUATIONS ON RISKS OF STOCK COLLAPSE

It is clear that the occurrence of large recruitment fluctuations will produce greater collapse risks than are caused by the catch simply exceeding MSY. In view of this, the deterministic MSY estimates are of limited value from a practical harvesting point of view, and should be adjusted to account for the problem of recruitment variability.

If η is assumed to be about 1/2, the estimated value of σ_u is roughly 0.6 which is equal to a C.V. on recruitment of some 65%. This may appear to be unrealistically large; however the equivalent value for the Californian pilchard <u>Sardinops caerulea</u> is estimated at σ_u = 0.76 (Murphy, 1966 and MacCall, 1979 cited by Beddington and Cooke, 1983). When recruitment fluctuations are that large, it is necessary to relate the MSY concept to probabilities of collapse (see for example Beddington and Cooke, 1983), and it is possible that allowable catches should be held substantially less than the deterministic MSY level in order to reduce the risk of collapse to an acceptable level.

In the brief analysis which follows, pilchard population dynamics is again modelled as an age structured population (as in Section 7.9) with a constant value of natural mortality ($M=0.5\ yr^{-1}$). However, instead of using expected recruitment as input, R_i is now regarded as a random variable so that:

$$R_i = \alpha Y_i \exp(-\beta Y_i + u_i) \qquad \ldots\ 7.28$$

where $u_i \approx N(0, \sigma_u^2)$

Since cohort size is usually calculated from fishing mortalities:

$$N_{i+1,j+1} = N_{i,j} \exp(-F_{i,j}-M), \qquad \ldots\ 7.29$$

a computational problem arises in subjecting the population to a constant catch strategy. It is therefore necessary to solve iteratively for the $F_{i,j}$ for each year using the relationship:

$${}^{*}C_i = \sum_{j=1}^{7} N_{i,j}\ w_j\ F_{i,j}(1-\exp(-F_{i,j}-M))/(F_{i,j}+M) \qquad \ldots\ 7.30$$

The best fit selectivities estimated for the pilchard for 1964 - 1983 were the same for all ages, so that given ${}^{*}C_i$, w_j, M and $N_{i,j}$, there is only one unknown ($F_i = F_{i,j}$) in the above sum. This was solved for each year and catch level using Newton's method.

For each constant catch level considered, 500 possible time series of recruitment and spawning biomass were generated, each 20 years long. It is assumed that the stock status at the start of fishing is y_1 = K/2, since the ideal future scenario for the pilchard may be to allow the stock to return to approximately y_{MSY} from its current depressed state. The situation projected here would therefore follow on from that point in time.

The probability of "collapse" was calculated to be equal to the proportion of 20 year projections in which the spawning biomass dropped below 0.2K (i.e. 480 000 metric tonnes). For these cases the projection was terminated in the year this occurred.

Results are set out in Fig. 7.4. An approximate 10% "collapse" risk over a 20 year period occurs at a catch level of 255 thousand metric tonnes (uncorrected). This value should be reduced by the same 20% factor (correction for bias in catch mass evaluations) and the factor of 433/473 (correction for bias in stock-recruitment fitting procedure) that was applied to the pilchard MSY estimate. The appropriate constant quota strategy for the risk level specified is therefore an annual catch of 187 thousand metric tonnes. This is 47% less than the deterministic MSY of 352 thousand metric tonnes. The reduction is quite substantial, but could perhaps have been anticipated considering the large (65%) C.V. of recruitment fluctuations. There is a very strong subjective element to this result, since the estimate of σ_u=0.6 is a direct result of the value assumed for η. The 47% factor is most likely on the low side, since the corresponding values of σ_v=σ_w=0.6 are much higher than seem reasonable from VPA simulations performed in Chapter 4 (with additional relative abundance information incorporated). Adjustments for this could be made by reducing η, which would increase estimates of σ_u, and hence decrease the constant catch level evaluated above.

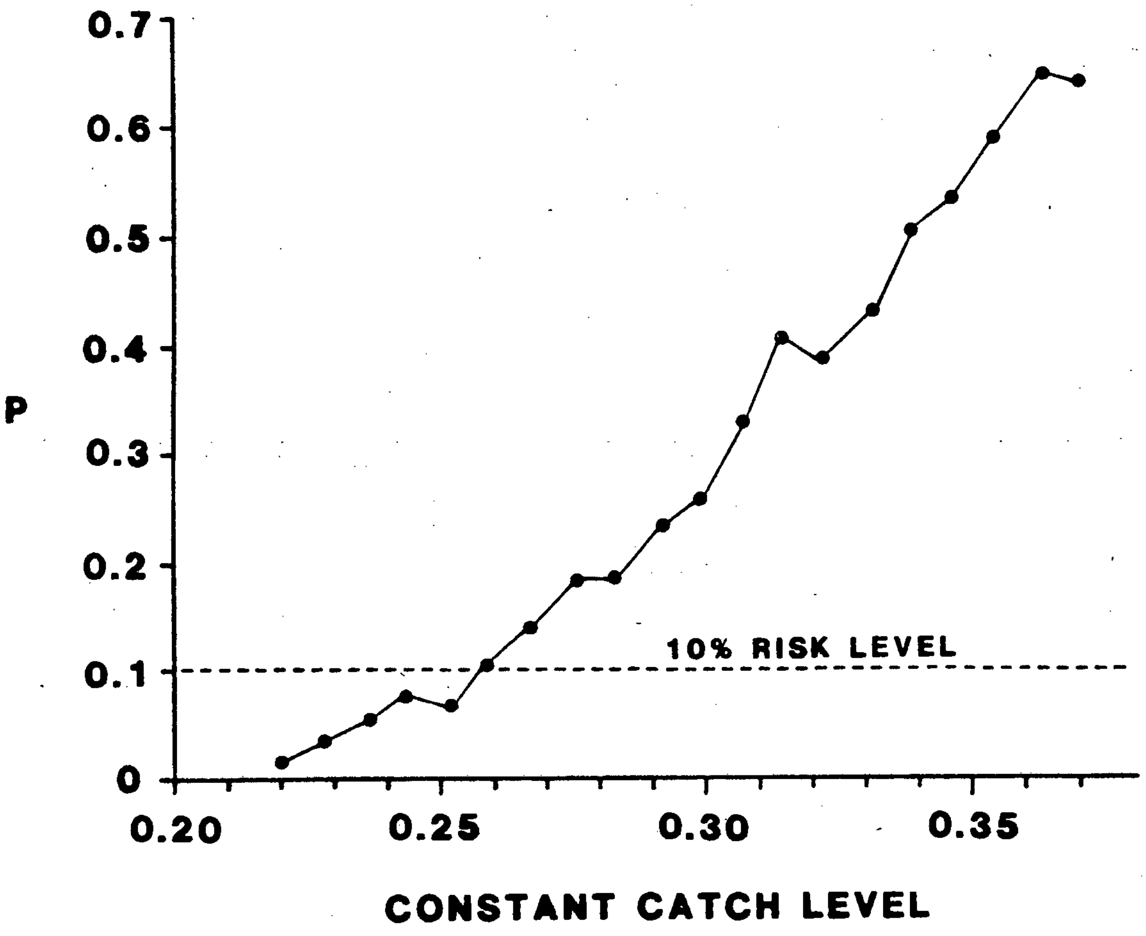


Figure 7.4. A plot of risk of "collapse" (p) of the pilchard stock (defined as spawning biomass falling below 0.2K over a typical 20 year time span of fishing) against the constant annual catch level applied. Projections commence with spawning biomass at 0.5K. Catch levels are uncorrected for the estimation bias of 433/473 and the 20% factor related to the estimation of landed catches (see text). (Catch units: million tonnes)

7.11 SUMMARY

The use of guano harvests to impose additional constraints on the pilchard VPA model parameters has merit as it leads to a considerable improvement in the precision of parameter estimates. Guano data between the years 1943 and 1971, where 1950 is the first year of fishing considered in the VPA, was successfully used to impose additional constraints on the fitting procedure. The overall estimated biomass trajectory suggests a large data contrast, with decreases in population biomass of almost 2 orders of magnitude. It is suspected that the precision of recent pilchard biomass estimates (in years for which guano-based recruitment indices are no longer available) is poor, although this could not be confirmed by using the resampling procedure for variance estimation because of the large amount of computer time needed for obtaining successful VPA model fits (over 4 hours for a single fit).

The 1950 - 1983 pilchard biomass trend estimated in this chapter suggests a different development pattern for the pilchard fishery to biomass estimates evaluated in past years. A biomass decline over the period 1950 to 1968 is the main feature indicated - the biomass peak in 1960 which dominates other analyses is much reduced in the current estimates.

An estimate of pilchard MSY from a yield per recruit calculation procedure was corrected for a discrepancy between observed catches landed (which are by definition consistent with catch-at-length information) and those calculated using the estimated catches-at-age and midseason fish masses. The final MSY estimate is 324 thousand metric tonnes. The estimated pilchard spawning biomass level at MSY is 1.02 million metric tonnes and the system carrying capacity (for pilchard spawning biomass) is K= 2.42 million metric tonnes. This value of K is not inconsistent with the estimate of system carrying capacity of 2 million metric tonnes for all pelagic fish proposed by Shannon and Field (1985). Simulations using the estimated pilchard stock-recruit relationship suggest that the pilchard population is capable of maintaining regular fluctuations, which is consistent with other evidence for this phenomenon presented in Chapter 6.

On the basis of these results an F_{MSY} quota recommendation for pilchard is approximately 32% (MSY/y_{MSY} x 100 %) of the (current) beginning of year spawning biomass level.

Recruitment fluctuations for pilchard were estimated to occur with a C.V. of approximately 65% (σ_u=0.6) about the mean stock/recruit function. The implications of this for enhancing the risks of a stock "collapse" (spawning biomass falling below 0.2 K) occurring were investigated using an age structured population model, with random recruitment fluctuations about the best fitted stock-recruit relationship. It was found that in order to keep the risk of collapse during a typical 20 year fishing period below 10%, the maximum allowable constant catch quota should be about 42% less than the value of 324 thousand metric tonnes mentioned in the second last preceding paragraph. The maximum catch which can be taken continuously consistent with this risk level is therefore about 187 thousand metric tonnes, which is roughly 20% of the spawning biomass at the beginning of the period (as the projections commence with spawning biomass at K/2).

8. ESTIMATES OF HAKE BIOMASS AND SUSTAINABLE YIELD OFF THE WESTERN CAPE

8.1 BACKGROUND

Catches of demersal hakes off the Western Cape consist of two species, Merluccius capensis and Merluccius paradoxus. Merluccius capensis lives closer inshore than does Merluccius paradoxus which is generally caught at depths greater than 200m. However, because of identification difficulties, catch mass and age structure estimates treat all hakes as one species, even though the growth and recruitment rates and probably natural mortalities for the two species differ. Thus the available data cannot be used to carry out two separate VPA's for each species of hake, and any resultant bias in VPA abundance estimates from this cause is practically irremoveable.

Butterworth and Andrew (1984) have noted a serious discrepancy between the results of the Fox and Schaefer dynamic production model fits to the CPUE trends in ICSEAF Division 1.6 (off the Western Cape), and hake biomass trends obtained from their VPA calculations, particularly for recent biomass trends. It has been shown in this thesis (Chapter 4) that the VPA alone (i.e. based on catch mass and age structure estimates) may not produce meaningful estimates of biomass values or trends (see also Pope and Shepherd, 1982). Since Butterworth and Andrew only use the very recent effort data for "tuning" their VPA, the discrepancy could be due to problems similar to those one expects to encounter with a VPA assessment based on catch mass and age structure data only.

More precise biomass estimates are potentially obtainable if the classical VPA data, i.e. the age structure and mass of the catch, are combined with the complete time series of effort data which are available. For example a more objective approach would be to use a likelihood function such as the one suggested by Fournier and Archibald (1982), incorporating all the available effort data. The remainder of this chapter is concerned with the methods used to do this, and some results are obtained. Only the hake stock in ICSEAF Division 1.6 is considered (Fig. 8.1).

8.2 THE HAKE VPA LIKELIHOOD FUNCTION

The basic VPA likelihood function used for anchovy (Fournier and Archibald, 1982 and equation 3.37) has been used for hake, with some slight modifications. It is suspected that the appropriate selectivity function for hake is different to that used by Fournier and Archibald (1982), or to those used in Chapters 5 and 7 for anchovy and pilchard. Although it may be reasonable to use a fixed selectivity function for all years for a stock which is well mixed with respect to age, for hakes off the Western Cape this may not be a valid assumption. The observed length stratification of hake catches and independent data from scientific surveys (A. Payne, pers. commn) suggests, for example, that the hake population is inhomogeneously distributed with respect to age, with younger fish occurring closer inshore. The implication of this is that the age-specific selectivity function varies according to the relative fishing effort which is applied in different areas.

8.2.1 The choice of an appropriate selectivity function

Following the assumptions that natural mortality is constant and fishing mortalities are separable into year and age effects, the choice of a selectivity function becomes a critical determinant

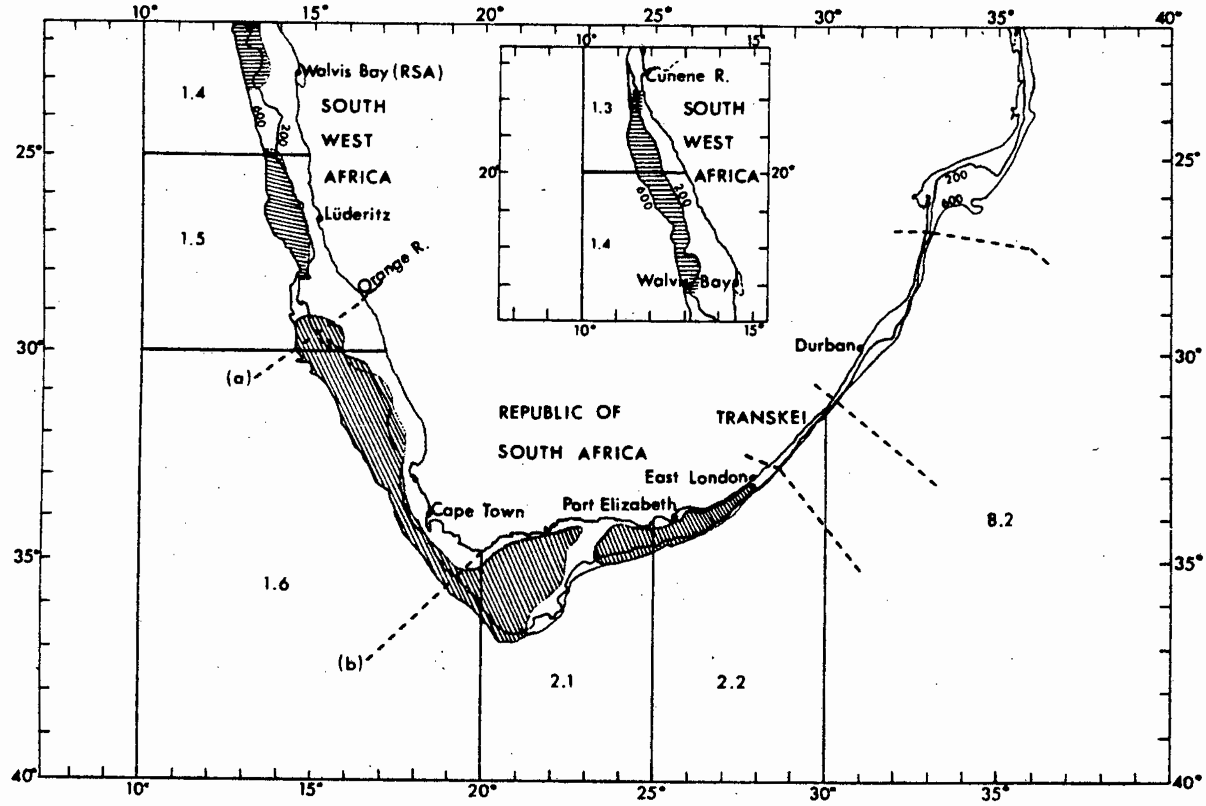


Figure 8.1. Map showing the major trawling grounds (shaded area), the ICSEAF Division boundaries, and the 200 and 600 m. depth contours off southern Africa (based on unpublished SFRI and ICSEAF data). The boundaries marked (a) and (b) refer to the domestically applied boundaries of Division 1.6 and the adjoining Divisions (map and caption taken from Andrew, 1986). Only Division 1.6 is considered in VPA calculations discussed in the text.

Table 8.1. The set of $P_{i,j}$ values used in the hake VPA (R. Leslie pers. commn). The midyear (w_j^*) fish masses used are also shown below (in grams).

UNADJUSTED

AGES: Year	0	1	2	3	4	5	6	7	8	9
1964	0	0	.031	.072	.119	.174	.248	.203	.100	.053
1965	0	0	.031	.058	.110	.198	.248	.205	.100	.051
1966	0	0	0	.127	.214	.302	.193	.121	.035	.009
1967	0	0	0	.181	.278	.278	.154	.082	.022	.005
1968	0	0	.003	.196	.280	.264	.140	.084	.027	.006
1969	0	0	.089	.318	.275	.136	.101	.056	.020	.006
1970	0	.021	.157	.308	.242	.125	.082	.044	.016	.006
1971	0	.125	.254	.355	.153	.048	.037	.018	.008	.003
1972	0	.102	.375	.326	.132	.035	.018	.009	.003	.001
1973	0	.018	.328	.430	.151	.045	.015	.008	.004	.001
1974	0	0	.183	.447	.224	.082	.031	.024	.006	.003
1975	0	0	.150	.264	.208	.179	.102	.066	.024	.007
1976	0	.060	.318	.348	.163	.056	.037	.016	.002	0
1977	0	.137	.535	.214	.071	.033	.008	.002	0	0
1978	0	.072	.716	.152	.039	.016	.005	.001	0	0
1979	0	.114	.545	.215	.063	.046	.013	.003	0	0
1980	0	.056	.472	.289	.112	.048	.017	.005	.002	.001
1981	0	.238	.474	.166	.078	.028	.012	.004	.001	0
1982	0	.351	.465	.110	.040	.022	.009	.003	.001	0
1983	0	.118	.485	.249	.083	.043	.016	.006	.001	0
1984	.004	.092	.466	.264	.105	.044	.017	.005	.002	0

ADJUSTED FOR POSSIBLE SAMPLING BIASES

AGES: Year	0	1	2	3	4	5	6	7	8	9
1964	0	0	0	.075	.123	.179	.255	.209	.103	.055
1965	0	0	0	.059	.113	.205	.255	.211	.103	.053
1966	0	0	0	.127	.214	.302	.193	.121	.035	.009
1967	0	0	0	.181	.278	.278	.154	.082	.022	.005
1968	0	0	0	.196	.281	.264	.141	.084	.027	.006
1969	0	0	0	.349	.302	.149	.111	.062	.021	.006
1970	0	0	0	.374	.294	.152	.100	.053	.020	.007
1971	0	0	0	.571	.246	.078	.059	.029	.013	.004
1972	0	0	0	.622	.252	.067	.034	.017	.005	.002
1973	0	0	0	.658	.231	.069	.023	.013	.006	.002
1974	0	0	0	.547	.274	.100	.038	.029	.008	.004
1975	0	0	0	.311	.245	.211	.120	.077	.029	.008
1976	0	0	0	.560	.263	.089	.059	.026	.003	0
1977	0	0	0	.652	.216	.101	.025	.005	0	0
1978	0	0	0	.715	.183	.075	.023	.004	0	0
1979	0	0	0	.630	.186	.136	.038	.010	.001	0
1980	0	0	0	.611	.237	.101	.036	.011	.003	.001
1981	0	0	0	.574	.269	.096	.041	.014	.005	.001
1982	0	0	0	.597	.215	.121	.048	.014	.003	.001
1983	0	0	0	.627	.209	.107	.039	.015	.003	.001
1984	.004	.092	.466	.264	.105	.044	.017	.005	.002	0
w_j^*:	18	69	208	435	743	1141	1602	2119	2679	3271

of the biomass trend which will be obtained from the VPA. The breakdown of proportion of numbers caught at age (Table 8.1) is revealing in this regard. It is difficult to reconcile the absence of juveniles in the catch over the period 1964 to 1969, followed by their occurrence in the catch since that time, with a model of year invariant fishing selectivities. It seems far more likely, given the stratification of the hake habitat by age, that this has been caused by a gradual shift in trawling strategy to exploit younger fish as the average age of the catch declined (though this effect may also be due to sampling biases which are discussed later).

A tentative 3 parameter selectivity function, which allows for an increasing selectivity of young fish with time, has been used to describe this phenomenon. The selectivity itself has been modelled by a logistic curve with point of inflexion at age j_0 (the age at 50% recruitment/selectivity). The increasing selectivity for younger fish in recent years has been accounted for by allowing j_0 to decrease with time. Initially a complex non-linear relation between j_0 and the year i was attempted:

$S_j=1/(1+\exp(-b_1(j-j_0)))$

$j_0=ib_2/(b_3+i)$ ... 8.1

where b_1, b_2 and b_3 are parameters to be fitted in the maximisation procedure. However, preliminary model fits to the data resulted in very large estimates for b_3, which suggests that an ordinary linear relationship between i and j_0 is adequate. On this basis the following parameterisation was adopted:

$S_j=1/(1+\exp(-b_1(j-j_0)))$

$j_0=b_3-ib_2$ ... 8.2

The structure of the model in terms of model parameters and input data, for the two cases of inclusion or exclusion of effort information, is set out in Table 8.2. In this table the variable θ represents the number of zeros in the catch-at-age matrix. This value has not been specified since it may vary depending on how the data are manipulated (see subsequent discussion).

8.3 RESULTS OBTAINED EXCLUDING EFFORT INFORMATION

Biomass estimates for hake obtained by maximising the likelihood function without effort information are set out in Table 8.3. These same results are plotted on Fig. 8.2. The main feature shown by these results are the unrealistically large biomass estimates for years 1977 to 1984.

It is pertinent to recall earlier reservations (Chapter 5) voiced about the relationship between trends in the catch-at-age matrix and population biomass trends. Appendix 5.1 shows that large apparent total mortalities indicated by the catch-at-age data could be due to at least four possible effects. These are:

1. Large and persistent increases in recruitment.
2. High fishing mortalities.
3. Emigration.
4. Selectivities sharply decreasing with age.

Clearly the model fit upon which the biomass trajectory in Table 8.3 is based ascribes large apparent Z values in the catches-at-age since 1977 to persistently increasing recruitment, with fishing making a continually decreasing impact on the stock. Given the manifold ambiguity of interpretation of the base data (the catches-at-age), and the ecological unacceptability of these results, the maximum likelihood estimates of hake biomass using

Table 8.2. A breakdown of the data which are used in the VPA for hake, and the model parameters which are estimated ($n = 21$, $m = 10$). The variable θ represents the number of zeros in the catch-at-age matrix, which may vary according to the manner in which the data are utilised.

NO EFFORT

Model Parameters	Quantity	Input Information	Quantity
Fundamental Betas:	$n+m-1$	Estimates of catch mass:	n
Year dependent fishing mortality effect:	n	Estimates of age structure of the catch:	$nm-\theta$
Parameters defining the selectivity function:	3		
TOTALS:	$2n+m+2 = 54$		$n+nm-\theta=231-\theta$

EFFORT INCLUDED

Model Parameters	Quantity	Input Information	Quantity
Fundamental Betas:	$n+m-1$	Estimates of catch mass:	n
Year dependent fishing mortality effect:	n	Estimates of age structure of the catch:	$nm-\theta$
Parameters defining the selectivity function:	3		
The proportionality constant between effort and F_1:	1	Effort data:	n
TOTALS:	$2n+m+3=55$		$2n+nm-\theta=252-\theta$

Table 8.3. VPA results for hake using only catch mass and age structure information ($M=0.3$ yrs^{-1}, $\sigma=0.05$, $Q=1000$). Biomass values are given in thousand tonnes.

YEAR	BIOMASS
1964	742
1965	626
1966	491
1967	400
1968	362
1969	352
1970	372
1971	392
1972	360
1973	353
1974	427
1975	704
1976	1 494
1977	3 681
1978	9 604
1979	23 843
1980	58 912
1981	152 600
1982	381 442
1983	878 124
1984	1 458 922

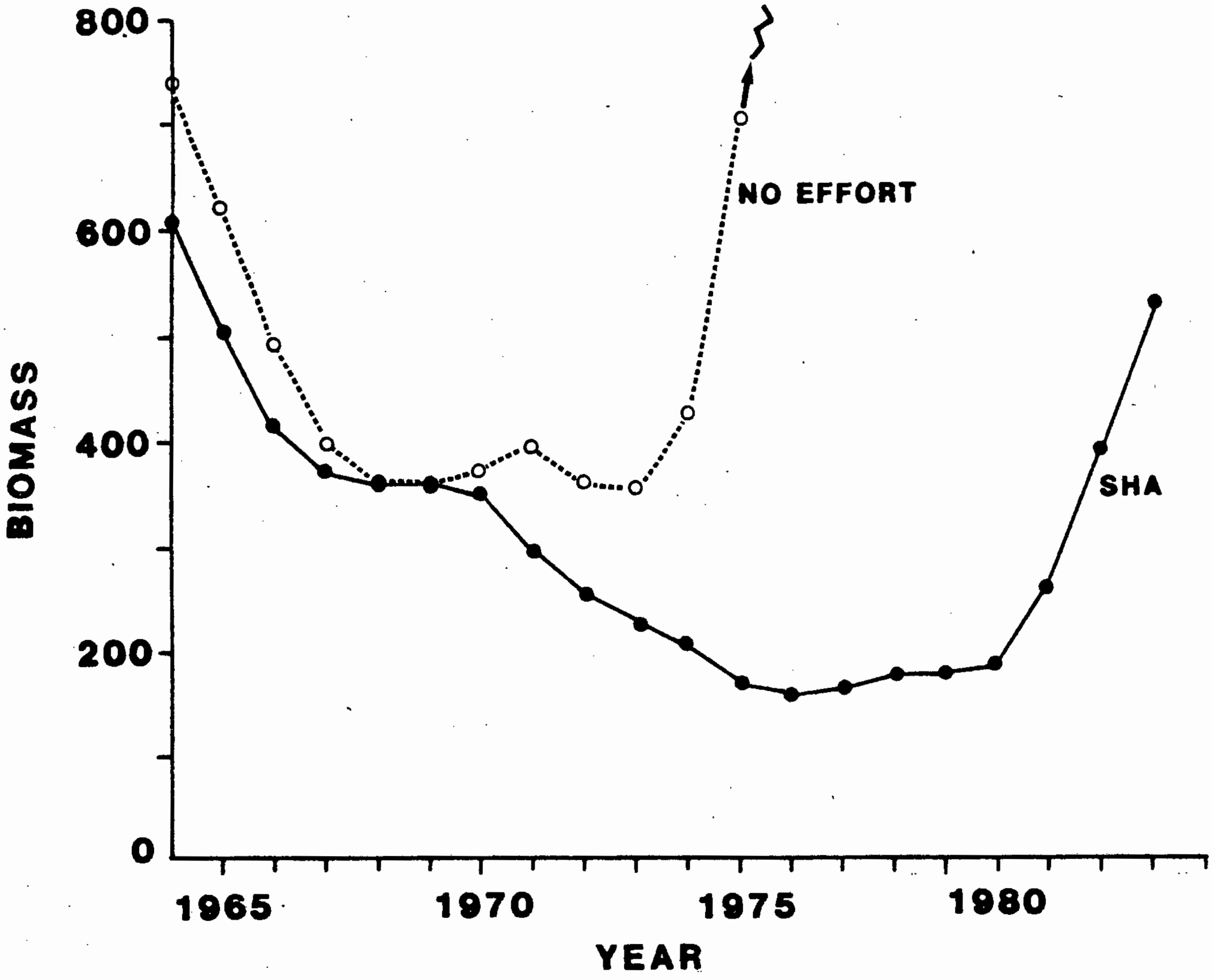


Figure 8.2. A plot of VPA results set out in Table 8.3 obtained without using effort data compared to the result obtained with effort (series SHA in Table 8.5).

catch-at-age information alone are rejected at this point.

Progress in hake stock assessment within this framework can therefore only be made by including further data relevant to the hake fishery in the hake VPA likelihood function. This is provided by the fishing effort values for hake in ICSEAF Division 1.6, which are shown in Table 8.4 (which also lists annual catch masses and catch rates (CPUE)).

8.4 PRELIMINARY RESULTS: EFFORT INCLUDED

The position of the maximum of the objective function was estimated as before using the NAG routine E04JAF (Numerical Algorithms Group Ltd., Oxford) with double precision variables, and the double precision version of the optimisation routine. Hake biomass estimates, corresponding to the estimated position of the likelihood function maximum are shown in Table 8.5 (Series SHA - obtained using Q=1000, σ=0.05 and σ_1=0.1).

The rationale behind the chosen weights (values for Q, σ and σ_1) can only be partly justified. The small value of 0.05 used for σ reflects a belief that the catch mass is fairly precisely determined. This does not include additional errors associated with poorly determined discard rates (a portion of the catch is normally discarded at sea) which is considered as a bias. The value of Q could not be determined in the same manner as for anchovy in Chapter 5 since the base data was inaccessible. The value of Q=1000 was therefore chosen on the assumption that anchovy and hake sampling levels for age determination purposes are likely to have been similar. The value of σ_1=0.1 was chosen on the basis of the magnitude of the residual errors in a dynamic Fox model fit to the catch mass and effort data (Butterworth and

Table 8.4. Measures of fishing effort, E_i ('000 standard boat days), and annual catch mass, ${}^{m}C_i^{obs}$ ('000 metric tonnes), used in the hake VPA. CPUE is the ratio ${}^{m}C_i^{obs}/E_i$.

YEAR	E_i	${}^{m}C_i^{obs}$	CPUE
1964	11.12	162.3	14.60
1965	18.76	203.3	10.84
1966	18.34	195.0	10.63
1967	17.65	176.7	10.01
1968	14.35	143.6	10.01
1969	19.15	165.1	8.62
1970	19.71	142.5	7.23
1971	28.49	202.0	7.09
1972	49.78	243.9	4.90
1973	31.75	157.8	4.97
1974	26.45	123.0	4.65
1975	19.23	89.6	4.66
1976	26.80	143.4	5.35
1977	20.15	97.5	4.84
1978	17.24	101.7	5.90
1979	14.75	90.4	6.13
1980	18.46	101.5	5.50
1981	17.13	99.5	5.81
1982	14.48	84.8	5.86
1983	11.30	73.7	6.52
1984	12.53	83.6	6.67

Table 8.5. Midyear biomass estimates for hake calculated using catch per unit effort plus all the catch age structure information (Series SHA), then excluding the first three years of age structure information in an attempt to correct for sampling biases (Series SHD), and by the methods of Fox and Schaefer (Butterworth et al, 1986: Series SHB and SHC respectively), using the CPUE data only. VPA results are obtained with M=0.3, Q=1000, $\sigma = 0.05$ and $\sigma_1 = 0.1$. The CV's refer to the biomass estimates of series SHD. All the VPA biomass estimates shown here were calculated with effort data.

YEAR	VPA (all ages)	FOX	SCHAEFER	VPA (excluding ages 0, 1, 2)	% C.V.
	SHA	SHB	SHC	SHD	
1964	606	1001	784	668	2.2
1965	506	926	734	532	2.1
1966	418	840	671	414	1.6
1967	370	772	624	351	1.5
1968	359	733	603	336	1.9
1969	352	701	587	338	1.9
1970	359	671	571	362	1.8
1971	350	622	536	364	1.5
1972	295	522	448	309	1.6
1973	251	441	372	270	2.1
1974	224	417	346	252	2.5
1975	210	425	349	247	2.3
1976	168	423	343	198	2.0
1977	159	418	331	187	1.8
1978	168	432	337	182	2.4
1979	179	452	348	201	1.8
1980	178	473	362	202	2.2
1981	188	491	374	210	2.7
1982	260	518	397	272	3.3
1983	397	559	436	387	4.5
1984	533	603	482	540	5.7

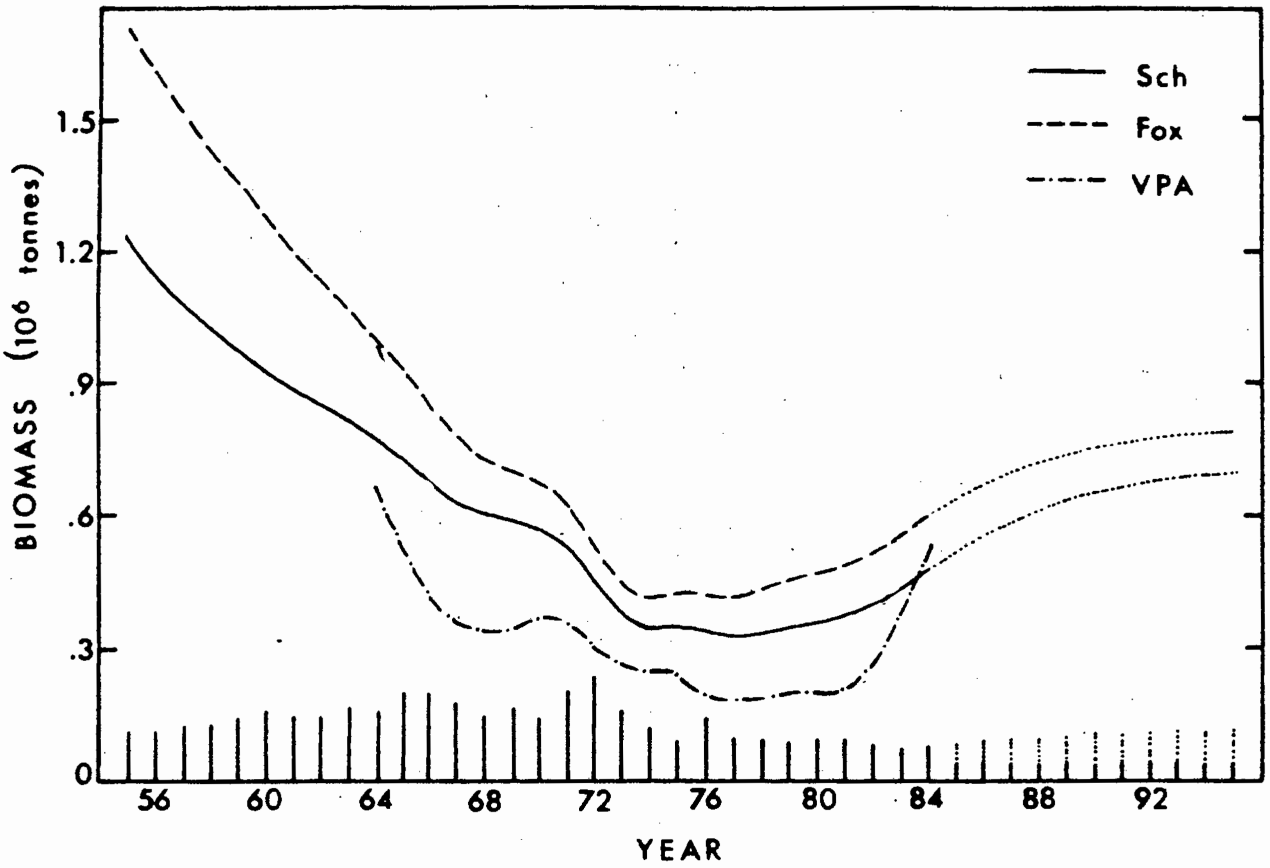


Figure 8.3. Plots comparing some of the biomass estimates presented in Table 8.5. Biomass estimates obtained using the dynamic catch-effort Schaefer model (solid line; Series SHC) are compared with those from the dynamic Fox model (dashed line; Series SHB). Ten year forward biomass projections under an $f_{0.1}$ harvesting strategy are shown for both models. Historical annual catches are indicated by solid vertical bars, with the Fox model $Q_{0.1}$ projections for the next ten years by dotted vertical bars. The dash-dot line shows the VPA solution for a selectivity function whose inflexion point is at an age that decreases linearly with time (Series SHA) (taken from Butterworth et al, 1986).

Andrew, 1984; Butterworth, pers. commn). A value of M = 0.3 yrs^{-1} was used to be consistent with current practice in the ICSEAF Scientific Advisory Council.

The most important difference between the results of the Fox model fit (Butterworth et al, 1986) and this VPA result is that the VPA biomass estimates are smaller than those produced using the dynamic Fox catch-effort production model. This may be because the M value of 0.3 yrs^{-1} used in the VPA is too small; larger values of M would produce larger biomass estimates. There are two other effects which should also be considered. These are (1) cannibalism within the hake community - this was alluded to in Chapter 2, and (2) a combination of sampling biases and variable discarding of juvenile hake.

8.4.1 Cannibalism within the hake community

In Chapter 2 it was suggested that as much as 75% of hake production may be cannibalised and recycled within the hake community. This retained material flow is not realisable as surplus production, and estimates (Chapter 2) indicate that it could be as much as ten times the amount taken by the fishery. It is very likely therefore that age structure and natural mortality for hake are functionally related, and the constant natural mortality assumption is invalid. For the earlier period under discussion (1964 to, say, 1968) when the average age of the stock was presumably higher, the natural mortality for juvenile hake was probably higher than for the latter period, 1968 to 1982, with a smaller proportion of older hake in the population.

Instead of one natural mortality parameter, a set of nm natural mortality parameters, $M_{i,j}$, is necessary to describe this situation. The task of estimating these is a major problem (which is considered by Lleonart et al, 1985) , and it seems probable that available data are insufficient to obtain a useful quantitative result. To define the functional relationship between age structure and age specific natural mortality, some knowledge of the size-spectrum of the cannibalised diet and food consumption rates are needed for all ages of hake. Given this knowledge, the relationships mentioned would permit the formulation of an overdetermined problem by reducing the nm natural mortality parameters to a smaller number. However, comprehensive hake diet data are only available since 1979 (Lleonart et al, 1985), and this is really inadequate for obtaining a satisfactory solution.

It is nevertheless possible to anticipate the qualitative outcome of a study attempting to account for the effect of cannibalism in the VPA. It is generally expected that the mean age of a fish population declines as fishing effort and hence fishing mortalities increase. Thus the rate of cannibalistic consumption of younger hakes by adults should decrease, as the proportion of adults in the population is reduced. This has important implications for the VPA result, series SHA, shown in Table 8.5 and on Fig 8.3. If natural mortality was relatively larger between 1964 and 1968 for younger fish (because a greater proportion of young fish were cannibalised then) than assumed in the VPA, then the true biomass trend for hake (in Division 1.6) would be different to that shown. For example, one might expect that the biomass decrease which occurs over most of the period of fishing since 1964 is in fact larger than suggested by the VPA result (SHA) shown in Fig 8.3, coincident with the postulated decrease in (age-averaged) M because of the backcalculation procedures used in the VPA (equations 3.9 and 3.11). Note however that this might increase the discrepancy between the CVPA (i.e. VPA incorporating cannibalism) generated biomass trend, and that suggested by the dynamic catch-effort models; a comprehensive analysis would have to consider the CPUE data as input to the CVPA, and weight it appropriately.

A steeper hake biomass decline is thus the anticipated effect of changes in total natural mortality (resulting from cannibalistic consumption). Further it is suspected that M for juveniles is much higher than previously thought (from the results presented in Chapter 2) and the net effect of this would be to shift the VPA estimated biomass trend to a larger level, irrespective of between year variations in M.

Alternatively, in the absence of any direct measurements of the age dependence of hake natural mortality coefficients, one could argue that hake natural mortality is independent of cannibalistic effects. This may be possible if an appreciable amount of non-cannibalistic mortality occurs, which is flexible enough to allow M to remain constant for hake, irrespective of changes in population age structure and hence cannibalism. [This may be the case for smaller hake which are extensively preyed upon by snoek; however there are no identified major predators of larger hake apart from hake themselves.] In this situation M may, for example, be adequately estimated from the von Bertalanffy growth curve parameters and a measure of habitat temperature, as suggested by Pauly (1980).

At the current low level of understanding of the dynamics of natural mortality for South African hake, the hypothesis that cannibalism causes changes in hake natural mortality to be functionally related to population age structure cannot be rejected, and this is a sufficient reason for continuing investigations of the implications of cannibalism for the VPA and CPUE assessment methods.

8.4.2 Possible sampling biases and variable discard rates

About 25 000 fish are measured each year to estimate the length frequency distribution of the annual hake catch. Some differences in this sampling scheme have occurred in the last 5 years. Before 1977, all measurements of length of fish were made on shore from the landed catch. It is known that a considerable portion of the trawled catch was discarded in the past, notably juvenile hake. These discard *rates* are largely unquantified, but could have an important effect on estimates of current stock size and historic biomass trends. This is due to the fact that the discarding has involved young fish, coupled with recent increases in the marketability of juveniles. The net effect for recent years would be for estimated age structures to reflect increasing catches of juveniles. This effect, which is evident in the catch age structure (Table 8.1) and trends in mean catch age (Fig. 8.4) could therefore be an artifact of the changes in the discard rate for younger fish, and the fact that sampling occurs after the discard (though starting in 1977 some onboard sampling has been instituted in an attempt to surmount this problem).

It is conceivable that the influence of these postulated sampling biases can be substantially reduced by excluding data pertaining to the affected age classes from the analysis. If the full annual catch mass is attributed to the remaining age classes, then the effective $P_{i,j}^{obs}$ matrix is the adjusted version set out in Table 8.1 (zeros are dealt with in the same manner described in Chapter 7). Table 8.5 shows the biomass trend (VPA - series SHD) which is obtained when age structure information for ages 0 to 2 are left out of the VPA likelihood function, and the annual catch mass is attributed to ages 3 and above. [Note however that the age structure estimates for 1984 were not adjusted, to maintain some constraint on recruitment estimates for 1982, 1983 and 1984]. The resampled estimates of variance (as percentage C.V.) for these estimates are shown in Table 8.5. The result of the adjustments is that 16 of the 21 biomass estimates in series SHD are now larger than the preliminary estimates in series SHA; nevertheless all the values in the "bias corrected" series, SHD, are still smaller than the Fox model results (see also Fig. 8.5 for a comparison between Fox model results and series SHA and SHD).

8.4.3. Sensitivity to M

The sensitivity of biomass estimates to M is illustrated in Fig. 8.6, which shows the (total) biomass estimates obtained using either M=0.3 yrs^{-1} or M=0.4 yrs^{-1}. As anticipated, biomass results obtained with M=0.4 yrs^{-1} are slightly larger than those produced using M=0.3 yrs^{-1}. In a later section (8.5) the sensitivity of surplus production parameters to the value of M is investigated.

8.4.4. Fitting σ_1

The problem of fitting σ_1 using the VPA likelihood function is treated briefly in Chapter 4. Results presented there indicate that σ_1 may be reasonably precisely estimated in this way, provided the correct values of Q and σ are given.

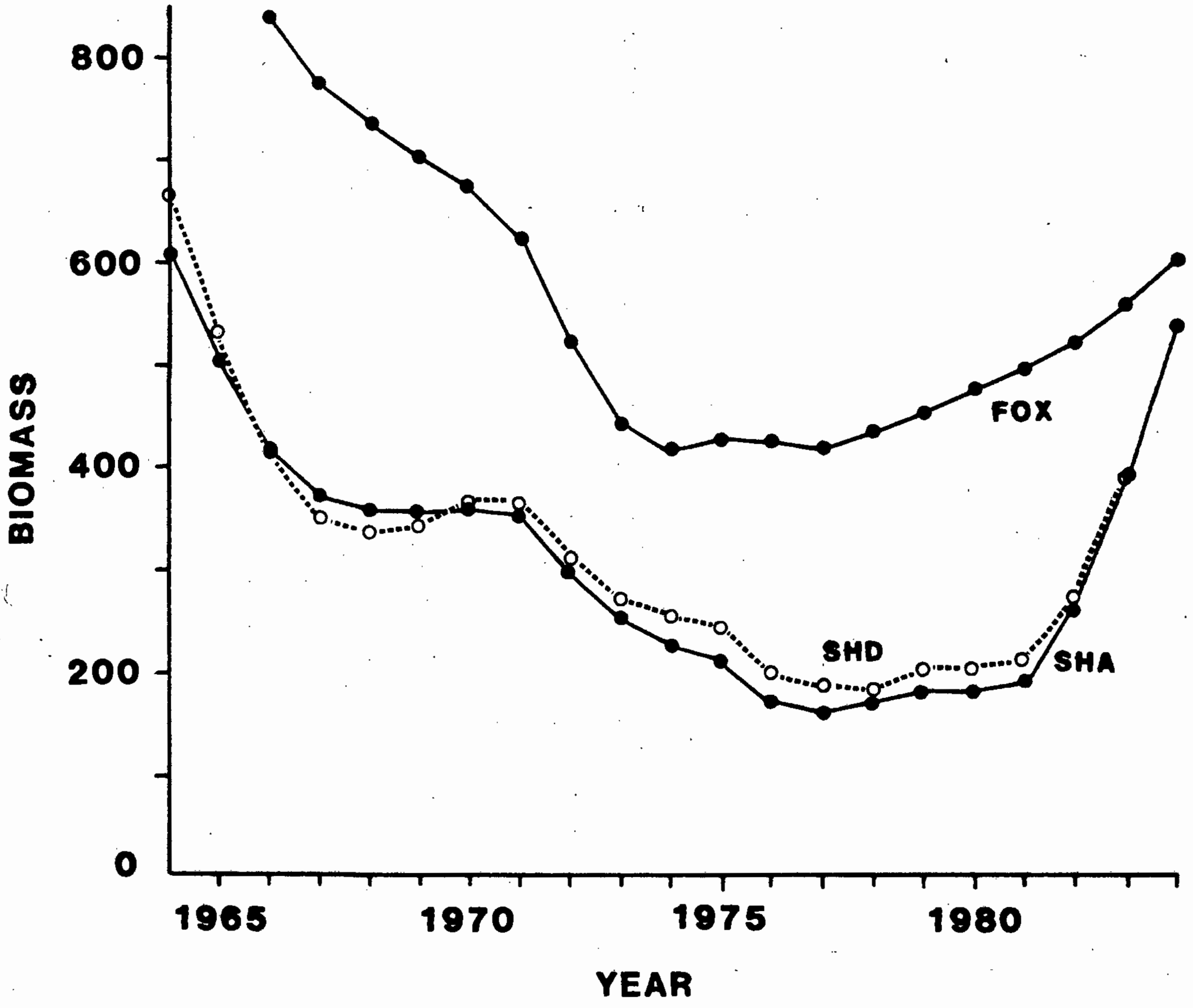


Figure 8.5. Biomass estimates obtained using the Fox model (FOX), and those produced using the VPA with effort; SHA and SHD refer to results presented in Table 8.5 - they are the trends obtained with and without adjustment of the catch-at-age matrix for possible sampling bias. (Biomass units: thousand tonnes).

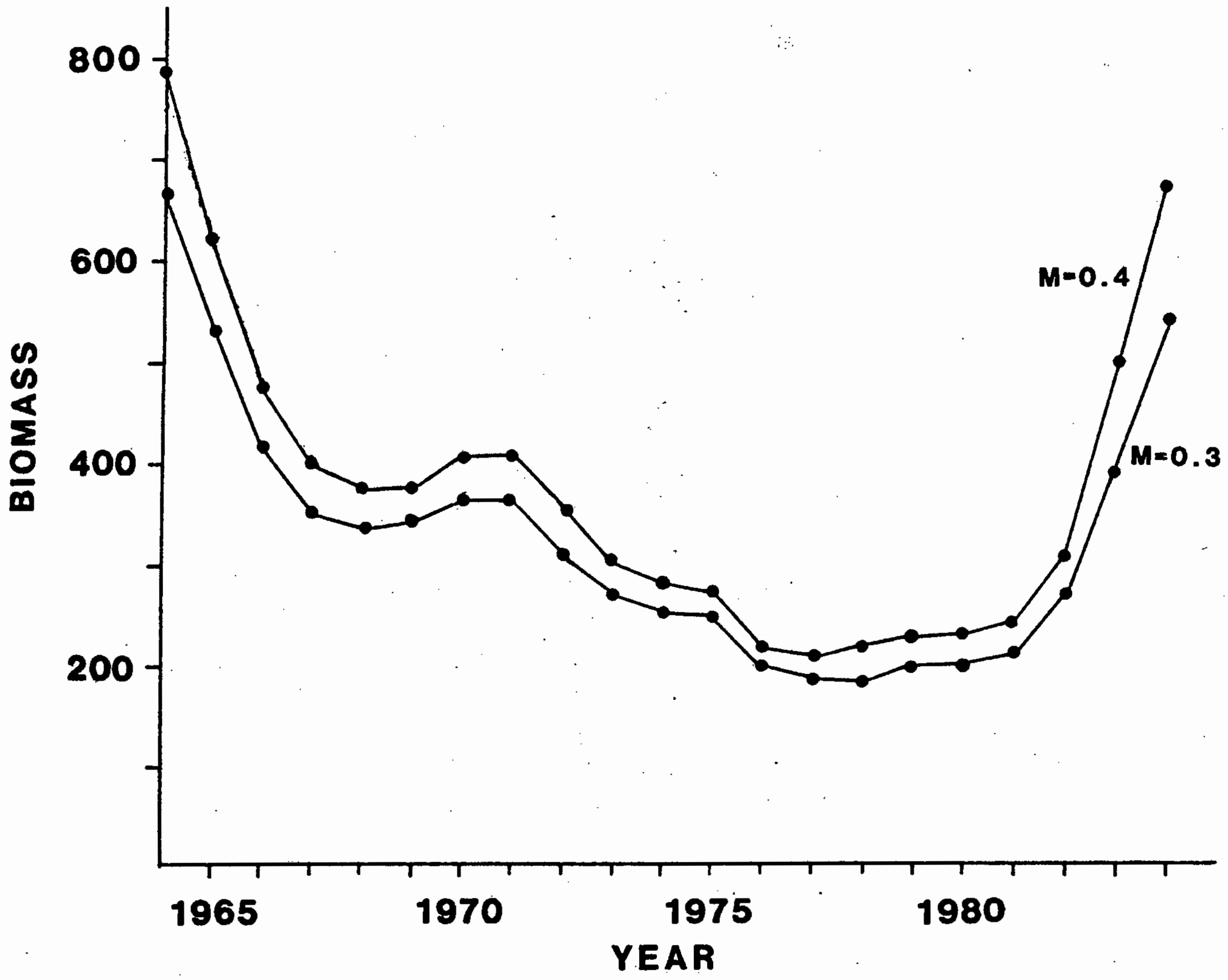


Figure 8.6. Biomass estimates obtained using VPA with effort information for two different values of natural mortality. The result for M=0.3 yrs^{-1} is series SHA in Table 8.5. The result for M=0.4 yrs^{-1} is the equivalent of SHA with a different value of M. (Biomass units: thousand tonnes).

An attempt at fitting σ_1 was made, with the values of Q and σ given as 1000 and 0.05 respectively. As discussed in Chapter 4, the term $[-n\ln\sigma_1]$ must be incorporated in the usual log-likelihood function.

A rather surprising estimate of $\sigma_1 = 4.19$ was obtained from the best fit of the VPA model to the hake data. The weight applied to the effort term $(1/\sigma_1{}^2)$ in this fit is therefore 0.057, compared to the value of 100 for the previously prescribed value of $\sigma_1 = 0.1$ (this latter value was based on the results of separate catch-effort production models).

The biomass trajectory associated with this result is - not surprisingly - practically identical (i.e. to within less than 1%) to the earlier results obtained without using effort information (see Fig. 8.2 and Table 8.3). These show unrealistically large biomass values from 1977 - 1984.

The underlying reason for this result is almost certainly a basic conflict between the catch-at-age information, the observed CPUE trend, and the selectivity functional form assumed. Two reservations about the calculation procedures used should be noted:

i) Variations in the $P_{i,j}{}^{obs}$ data have been entirely attributed to multinomial sampling error arising from ageing a finite-sized (Q) sample of the catch. However the selectivities that determine the relative proportions of different age-classes harvested each year would be expected to show annual fluctuations (arising from variations in environmental factors and fleet distribution). Neglecting this "process" error means that the multinomial approach will under-represent the degree of variation to be expected in the $P_{i,j}^{obs}$ data.

ii) The selectivity functional form assumed, despite allowing for some variation with time, may well be inadequate to account for the major change in catch age structure (particularly for the younger ages) evident in Table 8.1. (Although this may be primarily related to inappropriate allowance for changes in discard patterns as discussed elsewhere in this Chapter).

Fixing Q relatively large and keeping σ small only allows for small differences between the $P_{i,j}$ and $P_{i,j}^{obs}$, this lack of flexibility means that biomasses derived from the $P_{i,j}$ have little chance of good agreement with the CPUE trends, and so given the chance of an unspecified σ_1, the fitting procedure reacts by downweighting the contribution of the effort term in the log-likelihood function.

Two approaches to improve matters should perhaps be considered in future. The first is an ad hoc decrease in the value of Q to allow for the selectivity "process" errors referred to in i) above. However the appropriate magnitude of such a reduction is not known a priori.

The second is to attempt fits using alternative forms of the selectivity function which allow for more flexibility in year-dependence (even perhaps different functions over different periods). Unless, however, independent procedures can be developed upon which to base estimates of such selectivity changes (e.g. some relation to geographical changes in the fishing fleets operating pattern), this approach carries the penalty of needing to estimate yet more parameters from a likelihood function already exhibiting relatively poor contrast.

A critical component of the search for more appropriate representation of the selectivity should be inspection of the residuals, $\varepsilon_{i,j}$, where

$$\varepsilon_{i,j} = P_{i,j}^{obs} - P_{i,j}; \qquad \ldots\ 8.3$$

analyses reported here (and in previous chapters) have not examined this aspect of age-structure fits, but it is suspected that the fits for hake may show markedly non-random patterns of residuals.

These considerations probably have a bearing on the variance estimates obtained for the hake biomasses (see Table 8.5) which appear unrealistically small. The re-sampling procedure used to calculate these is not strictly of the "bootstrap" type, because a particular parametric distribution function (the multinomial) has been used to generate the pseudo-data sets $P_{i,j}^*$.

Provided that σ and σ_i can be fixed independently, and a selectivity form found for which the $\varepsilon_{i,j}$ are random, it may be possible to develop a resampling procedure more in line with the fundamental bootstrap concept. This would need to use the $\varepsilon_{i,j}$ residuals themselves to generate the pseudo-data sets $P_{i,j}^*$, rather than rely on the multinomial distribution. In this manner selectivity "process" error could be better taken into account. Such analyses have yet to be attempted, but it is anticipated that they would provide more realistic variance estimates for VPA assessments of hake biomasses in Division 1.6.

8.5 CALCULATION OF MAXIMUM SUSTAINABLE YIELD

Estimates of maximum sustainable yield for hake in Division 1.6

were obtained using the same approach as adopted for pilchard (Chapter 7 and Appendix 7.1). As in that case, further sets of spawning biomass and recruitment time series were generated by resampling the data, for the purpose of estimating variances for MSY and associated parameters.

8.5.1 Results

The stock-recruit relationship fitted to spawning biomass and recruitment estimates obtained using the VPA procedure (SHA - Table 8.5) is plotted in Figure 8.7.

Using an age to maturity of 4 years and natural mortality M=0.3 yrs^{-1}, the MSY estimate based on this stock-recruitment curve fit was 142 thousand metric tonnes. The estimated CV was 5.8%. The corresponding estimates of K and y_{MSY} (in terms of spawning biomass) were K = 530 thousand metric tonnes (CV = 7.5%) and y_{MSY} = 259 thousand metric tonnes (CV = 11.6%).

Butterworth et al (1986) produce dynamic catch-effort model based MSY estimates of 123.37 (CV = 5.5%) and 139.20 (CV = 7%) thousand metric tonnes, using Fox and Schaefer models respectively, and considering the catch and effort information only. The VPA based estimate of MSY is therefore not "significantly" different to either result (strictly these estimates are not completely independent, so that the standard statistical test cannot be applied). Their estimate of the 1984 stock status, expressed by the ratio of the current <u>total</u> biomass to the MSY biomass level, was:

Fox model - 0.661
Schaefer model - 0.712

VPA based results produced here suggest an equivalent value

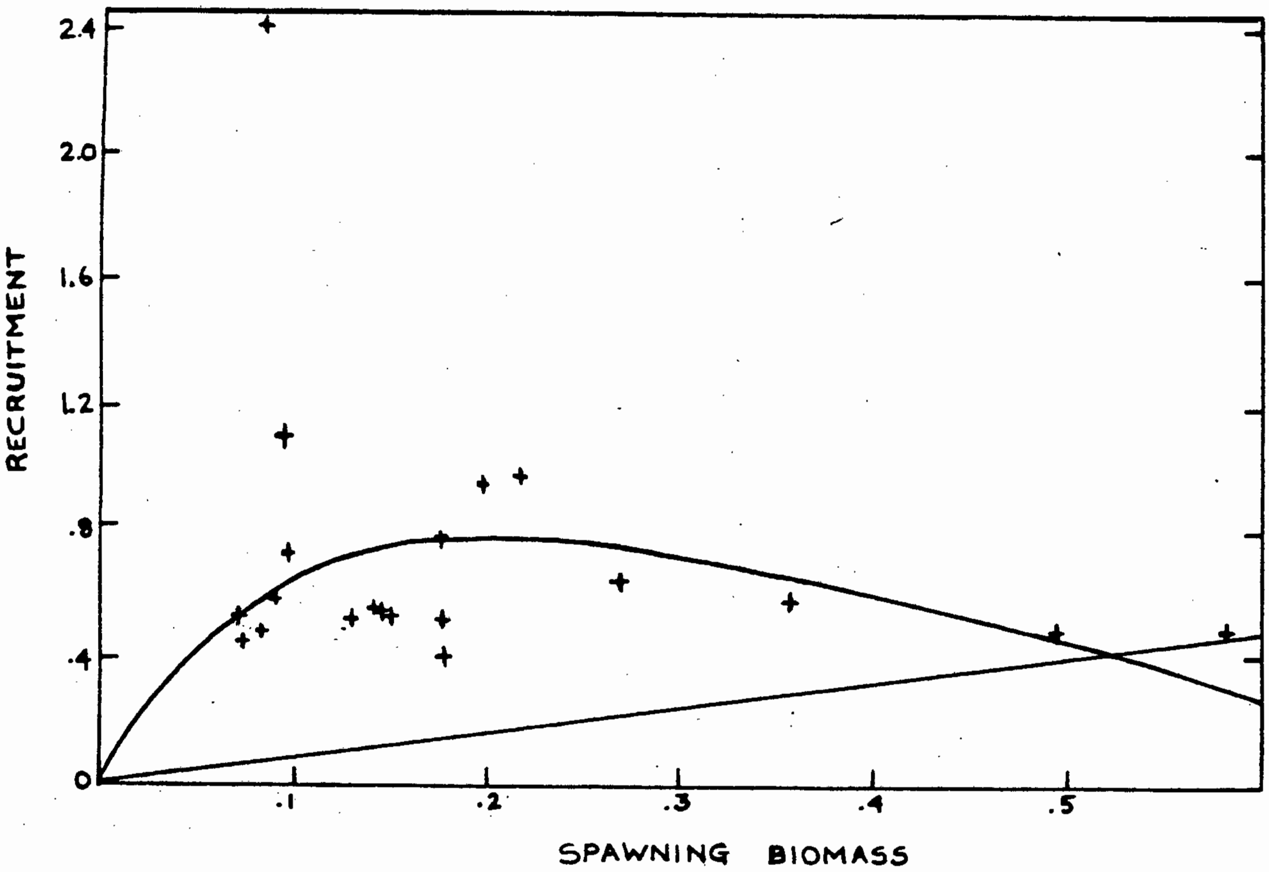


Figure 8.7. Estimates of recruitment (0-group year class strength in billions (10^9) and spawning biomass in million tonnes obtained from VPA assessments using a selectivity function varying with time are shown (+). The straight line through the origin is the replacement line - the number of recruits required at that spawning biomass to maintain the stock at that level. The curved line is the best fitted Ricker stock-recruit function for the spawning biomass and recruitment estimates shown (the VPA used $M = 0.3$ yrs^{-1}; $Q=1000$, $\sigma = 0.05$, $\sigma_1 = 0.1$, and included effort information, i.e. this corresponds to series SHA - Table 8.5.)

(y_{1984}/y_{MSY}) in terms of spawning biomass of 0.56, roughly 16% less than the Schaefer model result.

Thus although no serious disagreement between MSY estimates produced here and by Butterworth et al (1986) is evident, the quota recommendation based on a harvesting strategy such as f_{MSY} would differ because of the different estimates of the ratio of current to MSY level biomass in each case. Results were also produced using a value of M=0.4 yrs^{-1}. The estimates were:

K = 545

y_{MSY} = 256

MSY = 142

y_{1984}/y_{MSY} = 0.68

This suggests that the estimate of y_{1984}/y_{MSY} resulting from the combined VPA and stock-recruit fitting procedure is unacceptably sensitive to the value of M chosen, and the catch-effort production model based estimates are probably superior, since they circumvent the problem of specifying a value for M. [The larger value of M in fact produces results closer to Butterworth et al's; however this cannot necessarily be considered indicative of 0.4 being closer to the true value of M because of the large uncertainty in the current year biomass estimate obtained from the VPA.]

8.5.2 Possibility of biases

The potential impact of observation errors on the efficiency of the MSY estimation algorithm was quantified in a simulation study, using the same methods used in Chapter 7 for pilchard MSY estimates. As before, the suitability of the resampling procedure was investigated using parameters typical of the hake stock considered here. Some typical values from the VPA and recruitment function regressions, assuming that $\eta=1/2$, are:

σ_x = 0.8 (standard deviation of the logarithm of the time series of spawning biomass estimates)

$\sigma_u = \sigma_v = \sigma_w = 0.3$

$M = 0.3$ yrs $^{-1}$

$\alpha = 0.013$ (units: 10^9 recruits per thousand metric tonnes)

$\beta = 0.0054$ (units: per thousand metric tonnes)

Results suggest that the differences between mean estimates of MSY without observation error, with observation error, and using the resampling procedure (Table 8.6) are insignificant.

8.5.3 <u>A generalised recruitment function</u>

Before calculating the $f_{0.1}$ quota recommendation arising from these estimates, it is worthwhile re-examining their basis.

Consider the functional form chosen for fitting the stock-recruit data, to isolate just one feature of the model. There are two a priori statements which might be reasonable to make about the relationship between stock and recruitment. These are:

1. The function should pass through the origin.
2. The recruitment curve and the replacement line must intersect in the first quadrant.

There are two commonly used recruitment functions which satisfy both 1 and 2:

(a) $r = \alpha y\, e^{-\beta y}$ ... 8.4

(b) $r = \alpha y/(\beta+y)$ ... 8.5

The question arises: "Which is the better model?" This problem has been discussed by Schnutte (1985), who suggests that a more

Table 8.6. Simulation results for the MSY estimation procedure used for hake. Trials A, B and C are referred to in Chapter 7; they represent mean and variance estimates obtained without observation errors on either recruitment or spawning biomass values (A), then including observation errors of equal variance (not equal errors) on both recruitment and spawning biomass values (B) and finally using a resampling procedure corresponding to the actual procedure used for estimating the mean and variance of MSY for hake (C). %C.V.'s are given as σ/MSY * 100. Units are thousand metric tonnes. Correction factors for the mean estimates obtained in practice are calculated to be the ratio of the mean estimate obtained in TRIAL A divided by the mean estimate from TRIAL B (when the difference is significant).

	TRIAL A	TRIAL B	TRIAL C
MSY	98	102	101
σ_{MSY}	7.5	15	16
%C.V.	7.7	15	16

generalised form, which encompasses both (a) and (b), should be defined, (i.e. (a) and (b) will be special cases of this form). To describe a general form of the sort envisaged, a further parameter is introduced. An appropriate form in this case [originally due to Deriso (1980)] is:

$$r = \alpha y(1-\gamma\beta y)^{1/\gamma} \qquad \ldots\ 8.6$$

(a) and (b) are the special cases $\gamma \rightarrow 0$ and $\gamma = -1$ respectively. [Apart from these special cases, the continuum $\gamma = -1 \rightarrow -\infty$ provides recruitment curves which are unbounded and show asymptotic behaviour increasingly close to that of a linear trend. For $0 \geq \gamma > -1$, recruitment tends asymptotically to zero at a declining rate as γ decreases towards -1.] Now to identify the better of two models (a) or (b), it is necessary to be able to say something about the value of γ. If γ is very poorly determined, then (a) and (b) are indistinguishable.

Consider again the fitting procedure set out by Ludwig and Walters (1981). This can very easily be extended to fit a generalised non-linear recruitment function, say $f_r(y)$. The model would then be:

$$r_i = f_r(y_i)\exp(u_i) \qquad \ldots\ 8.7$$

where the notation is the same as used in Chapter 7 (i.e. u_i = process error, r_i = true recruitment, y_i = true spawning biomass). Following (8.7):

$$R_i = r_i \exp(v_i) = f_r(y_i) \exp(u_i+v_i)$$

$$= f_r(y_i) \exp(\varepsilon_i) \qquad \ldots\ 8.8$$

Also as before:

$$Y_i = y_i \exp(w_i) \qquad \ldots\ 8.9$$

(R_i = observed recruitment; Y_i = observed spawning biomass; v_i and w_i are the multiplicative observation errors for recruitment and spawning biomass respectively, expressed as natural logarithms of the multiplicative factors.)

Again the n+3 model parameters can be estimated by minimising the sum of squared residuals, the w_i and ε_i, assuming that they are normally distributed. The difference of the particular case for $f_\gamma(y) = \alpha y(1-\gamma\beta y)^{1/\gamma}$ compared to that considered in the previous chapter is just the additional parameter γ.

Attempts at fitting the generalised model to the hake stock recruit data were successful (with a $\gamma < 0$ constraint). The best fitted value of γ was -0.68, with a CV of 54% (using the resampling procedure outlined in Chapter 7). Results for MSY, y_{MSY} and K were:

$$K = 612 \; ; \; y_{MSY} = 259 \; ; \; MSY = 137$$

Thus the estimates of MSY and y_{MSY} are little changed, although K shows some substantial difference compared to the value of 530 000 tonnes obtained before when fitting the Ricker model ($M=0.3$ yrs^{-1} in both cases). The data do not therefore appear to support a distinction between the two specific types of asymptotic behaviour for recruitment, $\gamma = 0$ or $\gamma = -1$. In addition, freeing the "distinction" parameter γ does not meaningfully alter estimates of sustainable yield and MSY biomass levels.

8.5.4 An $f_{0.1}$ quota strategy

Decisions about quotas are frequently based on the idea of marginal yields (to take cognisance of economic factors such as the costs of applying additional fishing effort). The marginal yield of a fishery can be expressed as the ratio of the change in sustainable yield to the small change in fishing effort, δ_E required to effect that yield change - in the limit $\delta_E \to 0$ this becomes the derivative of sustainable yield with respect to effort.

A common management objective in terms of this marginal yield concept is to bring the stock to and maintain it at a size for which the marginal yield is 1/10 of the marginal yield realised for the pristine stock size. At this point the terminology commonly adopted is:

stock size: $y_{0.1}$
effort level: $E_{0.1}$
surplus production: $G(y_{0.1})$

To achieve the $y_{0.1}$ biomass level it is necessary to have a quota strategy which will result in equilibrium at this biomass level (at least conceptually) - the equilibrium catch at this point would be $G(y_{0.1})$. One such strategy is to take a quota which is the same fraction of the current biomass level as $G(y_{0.1})/y_{0.1}$. This quota is conveniently represented as $Q_{0.1}$ - i.e.:

$$Q_{0.1} = y_{current}G(y_{0.1})/y_{0.1}$$

In general the yield per recruit based analysis produces skewed effective production curves (i.e. $y_{MSY} \neq K/2$). Allowance for skewness can be made by incorporating a further parameter in the usual logistic equation:

$$G(y) = ry(1-y/K)^{\tau}$$

If catch is proportional to effort times stock biomass, $C=qEy$, then the equilibrium catch level in terms of effort is:

$C = EKq[1-(Eq/r)^{1/\tau}]$

From this one can show that:

$dC/dE_{E=0} = Kq$

Also:

$dC/dE = qy - Kq(1-y/K)/\tau$

The value of $y_{0.1}$ is obtained by solving:

$$dC/dE\Big|_{E_{0.1}} = 0.1\ dC/dE\Big|_{E=0}$$

Solution yields:

$y_{0.1} = \{(1+\tau/10)/(1+\tau)\}K$

A deterministic estimate of τ is given by:

$\tau = K/y_{MSY} - 1$

For the values pertinent to the stock under consideration (obtained using the Ricker stock/recruit fit), viz. K=530 and y_{MSY}=259, the value of τ is 1.046 which is very close to 1; skewness of the surplus production curve is apparently not a serious problem in this case.

The $f_{0.1}$ strategy quota recommendation is therefore 70 thousand metric tonnes ($y_{current}$ - i.e. y_{1984} - is given in spawning biomass units as 145 thousand metric tonnes). For M=0.4 yrs^{-1} in all calculations $Q_{0.1}$ = 85 thousand tonnes. These values are to be compared with values of 89.6 and 93.6 obtained from dynamic Fox and Schaefer models reported in Butterworth et al, 1986.

8.6 CONCLUSIONS

In terms of results presented here, there is no substantial difference between the quota recommendations arising from use of either the dynamic catch-effort production model or the VPA estimates. The only quantity for which some discrepancy initially seemed to exist was the ratio of the current stock size to the MSY biomass level. This estimate was subsequently shown to be unreliable, since it is sensitive to the value of M chosen, for which the a priori basis is rather poor.

Superficially the VPA results for sustainable yields and quota recommendations appear to provide independent confirmation of the values obtained by Butterworth and Andrew (1984) and Butterworth et al (1986) which are based on dynamic catch-effort models. This implication is however almost certainly grossly misleading. A quick comparison of VPA results produced without use of effort information (Table 8.3 and Fig 8.2) with those including effort (Fig. 8.2; Fig 8.3) is enough to see that the final results are critically dependent on the use of the effort data, and hence on the weighting given to the effort data. The conclusion that the VPA based MSY estimates are little more than a re-representation of the dynamic catch-effort modelling estimates seems to be fairly clear.

It is necessary at this stage to consider what additional insight (if any) regarding the hake stock has been gained from the VPA assessment. The VPA without effort does admittedly yield sensible results over the period 1964 - 1974, which are in accordance with the estimates using catch-effort data only. After that time increasingly unrealistic biomass estimates are produced. In terms of the discussion in Appendix 5.1 an inappropriate description for changes in catch age structure is chosen, viz. a decreasing mean age of the catch (see Fig. 8.4) is put down to very large and persistent increases in recruitment. Even with effort data included this effect is noticeable for the years 1982 to 1984 (Figs 8.2 and 8.3). Alternatively to suggestions made in Appendix 5.1 this tendency may be due to the (suspected) reduction in discard rates for smaller hake. Note in this regard that a reduction in mean catch age has been observed for hake stocks in ICSEAF Divisions 1.3+1.4, 1.5, and less so in 1.6, over the last approximately 10 years (Butterworth et al, 1986). If there has indeed been a substantial change in the hake discard pattern, this would definitely add to the problem of resolution faced by the VPA, perhaps facilitating the complete misinterpetation shown in Fig. 8.2.

The possibility that discard rates have decreased in recent years raises some disturbing questions about the validity of even the dynamic catch-effort production model results. Note that if the discard rates have declined steadily in recent years, then recorded catch rates would be artificially elevated relative to a "true" CPUE trend which <u>is</u> proportional to stock size. One of the most striking features of the catch rates for the hake stock considered here (Division 1.6) and others (Divisions 1.3+1.4 and 1.5) is the recent increase in CPUE (catch per unit effort), following many years of decline (Butterworth et al, 1986). In fact this recent upturn in CPUE trends is central to the precision of MSY estimates; Andrew (1986) notes that it is only following a changeover from decreasing to increasing catch rates that estimates of the surplus production parameters become at all reliable. However, an increase in CPUE, and a concommittant decrease in mean catch age is qualitatively consistent with a reduction in the discarding of smaller hake, without necessarily any real change in stock biomass. The potential impact of this effect depends on the extent of discarding; the CPUE for Division. 1.6 has increased by 43% since 1974 (see Table 8.4). For the true CPUE trend to have remained constant, with the increase being due to a change in discard practices alone requires that past discards constituted at least some 43% of the annual catch. Adjustments to shorebased estimates of catches before 1972 have been made to attempt to correct for this - the pre-1972 catches are all elevated by 39% (Andrew, 1986). The main concern voiced here is however about the levelling off followed by a steady increase in CPUE since 1972 which this adjustment does not affect.

The practice of discarding economically unattractive portions of the catch at sea is illegal in South Africa, and so it is understandably difficult to obtain even reliable comments about the extent of discarding both past and present. The possibility of modelling the discarding process is therefore remote. However in view of the distorted impression which can be created in the catch statistics record by discarding, it is not clear that the industry would be protecting even its own interests by possible clandestine dumping of fish at sea (if indeed they have in the past!). Long term investment decisions which are based on results such as are reported here are valid only insofar as the reporting of past catches has been sufficiently accurate.

The potential implications of this discard problem, and the difficulty of policing an industry which enjoys such vast autonomy in its operations, viz. its activities at sea, raises an interesting yet disturbing possibility. It is conceivable that fishermen can learn the link between favourable quotas awarded, and data submissions. Thus for example reporting favourable catch rates would in terms of the models set out here produce favourable quota allocations. The possible existence of this maxim in the minds of fishermen is worth investigating.

As far as improvements to the current stock assessment practices are concerned, VPA must on the basis of results produced here be seen as an expendable appendage, which is only justifiable if real benefits in terms of parameter precision [or insights other than those directly related to stock size and recruitment trends such as have been developed here with regard to the discrepancy between VPA and dynamic catch-effort model biomass estimates] can be demonstrated. VPA is potentially useful in detecting particularly good or bad year classes (this is only useful if management can respond to such information) - however the success of such calculations depends on whether independent estimates of changes in selectivity can be developed. The most pressing issue which must be addressed is the question of whether recent (over the last 10 years or so) increases in catch rates are real or artifactual.

The scenarios sketched above are perhaps far too pessimistic - unfortunately on the other hand use of the results of mathematical models without appropriate consideration of the validity of the underlying assumptions is potentially misleading, perhaps leading to excessive optimism. The statistical calculations reported have been carried out in the most optimistic and liberal manner. For example, the assumption of independence of errors is the bedrock of all the fitting procedures used. Strictly this should have been confirmed by testing the residuals obtained from the fitting procedures, whenever these were employed. Nevertheless it has been noted

9. GENERAL DISCUSSION AND CONCLUSIONS

9.1 WHOLE SYSTEM STUDY

A crude evaluation of material flows and standing stocks in the Benguela system, including lesser studied species (inaccessible to research, or commercially unimportant), was made. Results suggest that the productivity of juvenile (demersal) hake (3 million metric tonnes per year or more), squid (>0.5 million metric tonnes per year), and midwater shoaling fish (together > 0.5 million metric tonnes per year) is substantial compared to the total annual landing of marine fish off South Africa which is in the region of 0.5 million metric tonnes. Accordingly the biomass of these species is also substantial compared to these landings. The consumption of anchovy and other epi-pelagic shoaling fish is dominated by piscivorous fish, particularly snoek (snoek may consume as much as 0.4 million metric tonnes of anchovy per year, out of a total anchovy production of in the region of 2 million metric tonnes wet mass per year). Anchovy are the main fish consumers of plankton, with an estimated annual consumption of more than 20 million metric tonnes (mainly wet zooplankton).

Total plankton production is in excess of 800 million metric tonnes wet mass per year. About 32% of this enters higher trophic levels in the epi-pelagic region (mainly through grazing by zooplankton). It is suspected that the bulk of production in the zooplankton and demersal communities (the latter is dominated by hake) is recycled by omnivorous feeding in the case of zooplankton and cannibalism in the case of hake. For hake, the amount of material recycled annually is in the region of 10 times the amount taken each year by trawlers. The biomass and production and consumption rates of apex predators such as seals, whales, dolphins and seabirds (sizes of the order of metres to 10's of metres) is quantitatively unimportant by comparison with species consisting of smaller individuals (sizes of the order of 10's of centimetres).

Some of these findings are important to present stock assessment procedures. For example, the phenomenon of hake cannibalism has some very important implications for hake assessment using VPA. Overall, however, this study demonstrated that material flows in the ecosystem are known only very crudely, and this suggests that the development of complex multispecies models is somewhat unrealistic at present.

9.2 VIRTUAL POPULATION ANALYSIS

The bulk of this thesis has been directed at the practical application of cohort analysis techniques. The classical application of VPA, as propounded for example by Gulland (1965), is restricted to the use of catch mass and age structure information only. In the more generalised approach of Fournier and Archibald (1982), based on maximum likelihood estimation, these data may be wedded with any other abundance-related information concerning the fishery. The main reservation voiced with regard to Fournier and Archibald's model concerns their failure to adequately distinguish between measurement errors (in the catch-at-age matrix), and process errors (also termed model prediction errors).

This problem of making the appropriate distinction between process error and measurement error is perhaps of most importance in regard to the techniques used for estimating variances. The bootstrap-like procedure developed here produces a reasonable impression of the errors in the VPA subject to sampling errors on estimates of $P_{i,j}$ and C_i. However the additional model error which exists will almost certainly cause the variance estimates shown for the anchovy and hake biomass estimates to be underestimates (for convenience of expression we refer to "underestimates", though more correct terminology would be "negatively biassed").

Simulation results which were obtained with the VPA model described by Fournier and Archibald (1982) confirmed Pope and Shepherd's (1982) findings regarding the inadequacy of the catch-at-age data alone. Results produced using effort information for which CPUE was approximately proportional to stock biomass showed marked improvements in biomass estimate precision.

Further investigations into the precision of estimates of σ_1 and M indicate that whereas σ_1 might be reasonably precisely estimated from the fitting procedure, M is not given with very good precision. The simultaneous estimation of σ and σ_1 proved to be unsuccessful, as was suspected by analogy with linear models.

An overall recommendation arising from the simulation results was therefore that VPA should not be attempted in the absence of additional relative abundance information. Analyses of actual data for some South African stocks was commenced with this in mind.

9.2.1 Anchovy

Minimum estimates (i.e. negatively biassed by an unquantified amount due to unaccounted model error) of the variance of South

African anchovy biomass values evaluated using VPA were obtained. If one assumes that;

1. anchovy ages have been correctly determined,
2. the available CPUE information is linearly proportional to abundance (i.e. no systematic bias effects such as catchability inversely related to abundance),
3. the value of M, the natural mortality (assumed constant) is known,
4. the weighting constants for the catch mass and effort data in the VPA likelihood function are known,
5. samples from which age information is obtained are chosen at random, and
6. age specific selectivity is invariant over the history of the fishery,

and that the statistical assumptions discussed in Chapter 3 are valid, then the coefficient of variation for the abundance estimate for the most recent year is in the region of 20% (for the chosen values Q, σ, σ_1). In a separate simulation it was shown that, because anchovy ages have been determined using the same age-length key estimated from samples collected from 1965 to 1967 instead of from a key determined each year, the current year abundance estimate may be in error by about 50%.

There is however little cogent reason to expect proportionality between fishing mortality (F_i) and effort expended (E_i) for anchovy, since their shoaling behaviour probably destroys the possibility that population density and catch per unit effort rates (if effort is some gross fishing time estimate) are linearly related. Further recent direct survey results show that there must be considerable emigration from the main fishing grounds with increasing age, which could be mimicked by a corresponding decrease in selectivity (see Appendix 5.1). These two considerations suggest that the VPA parameter fit should be attempted without use of effort information or specification of the asymptotic slope of the selectivity function. This less defined model yields a much wider range of possible biomass trajectories. These include trajectories showing a consistent and overall 10 fold increase in stock abundance and recruitment between 1968 and 1983 (see Fig. 5.3).

This feature could be an artifact of failing to take emigration of anchovy into account in the model. The effect of emigration would be to cause additional falloff of numbers caught at age with increasing age, over and above the effects of fishing and natural mortality. Assuming that the VPA model was approximately correct, except for the emigration effect, the fitted parameters may, for some data sets, opt for an artificially increasing recruitment trend which would continually elevate apparent Z (total mortalities) values for the catch for each year.

Perhaps the most significant recent development for VPA are recent results concerning the anchovy growth rate. On the basis of growth rate studies it is now thought that the anchovy catch consists predominantly of recruits, whereas previous results suggested a substantial contribution from older age groups. This virtually rules out VPA, since the basic age structure input hardly exists anymore. It is therefore concluded that anchovy biomass trajectories arising from the VPA based only on catch mass and age structure information are artifacts of measurement and model inadequacies, and are of no value whatsoever for future management of the anchovy resource.

9.2.2 Analysis of guano records

In the case of the local pilchard stock, the only other data available for augmenting catch-at-age information are commercial records of guano yields (effort data are affected by the same problems as detailed for anchovy). Two important features of the guano harvest time series were investigated. These were the existence of significant periodicities with periods of between 4 and 6 years, and a 3 year lag between guano harvest trends at two islands, which may be related to birds at the two islands feeding on pilchard differing in age by 3 years.

From the properties of simple difference equations and considering plausible ranges for seabird population dynamic parameters, it was argued that the guano producers, the seabirds, are incapable of producing periodic effects with such high frequencies, and these periodicities are more likely to be due to fluctuations in pilchard population size and recruitment which subsequently affects guano production.

The case for pilchard dynamics being reflected (at a fine scale) in guano records is strengthened by evidence of a 3 - year lag between trends in guano harvest at two islands. The leading island, Bird Island Lamberts Bay, is situated in the prime pilchard recruitment grounds near St. Helena Bay, while dietary evidence shows that gannets at the second island (Bird Island Algoa Bay) are, and probably were historically, feeding on large pilchard. The analysis of this phenomenon was primarily concerned with the possibility of a more complex link between the two time series, other than the obvious pilchard juvenile - adult lag during migration.

A major flaw in the argument that the lag between guano production trends at the two islands is simply the difference in age between fish consumed by the two different gannet colonies is the fact that the gannets at Algoa Bay are consuming mainly 1.5 to 2 year old - and not 3 year old - pilchards. This may however be due to a more complex model of recruitment on the West Coast and how this eventually influences the food source of gannets at Algoa Bay. The existence of restricted surface water flows may prevent recruitment peaks which develop in the vicinity of Lambert's Bay from reaching Algoa Bay in the same year. Rather, the young adults (at first maturity) of the aforementioned recruitment peak would first migrate onto the the Agulhas Bank. Some of the recruitment from <u>their</u> spawning activity would then be able to reach Algoa Bay, since the Agulhas Bank straddles the oceanic divergence zone restricting the eastward flow of eggs spawned on the West Coast. Gannets would then feed on these pilchard recruits once they reach an age of about 1.5 years. Hence an overall lag of 3 years is still feasible, comprising the 1.5 years for adults to reach the Agulhas Bank, and another 1.5 years for their recruits to become attractive as prey for gannets at Algoa Bay.

In view of these results, the use of the Lambert's Bay guano harvests as pilchard recruitment indices in the VPA likelihood function is advocated.

9.2.3 Pilchard

VPA model fits confirmed that the South African pilchard population biomass collapsed by more than an order of magnitude in the mid 1960's, and this particular observation is still in general agreement with previous work (Centurier-Harris, 1977; Crawford et al, 1983). The details of the estimated trend are however contrary to the conventional scenario developed on the basis of this earlier work. These earlier estimates all suggested a marked biomass peak in 1960 - they indicated in fact that the fishery expanded in the late 1950's and early 1960's concurrently with the entry of some very strong year classes into the fishery. The pilchard collapse subsequent to 1960 has in the past often been attributed to catches climbing too high over this period of enhanced recruitment together with relative recruitment failures throughout 1960-1965.

The results produced here suggest a far weaker biomass peak in 1960, with the essential feature between 1950 and 1965 being a decline in biomass. This tends towards a more classical model of overfishing and subsequent collapse as the MSY was exceeded and the stock size fell below the MSY biomass level.

The pilchard MSY was estimated from the convexity in the relationship between (VPA generated) recruitment and spawning biomass. After applying a correction factor for the discrepancy between observed catch masses and those consistent with the catches at ages and a further bias correction factor for bias arising in the stock-recruitment fitting procedure, a MSY of roughly 324 000 metric tonnes was suggested, with an estimated C.V. of about 33%. Estimates of the MSY spawning biomass level of about 1.0 million metric tonnes and an unexploited spawning biomass (K) of 2.4 million metric tonnes were obtained.

In further analysis it was shown that the occurrence of excessive recruitment variability means that even if the MSY is taken regularly with the stock at K/2 ($\approx y_{MSY}$), the probability of the stock dropping below 20% of K in a typical 20 year fishing period is about 60%. To keep the "collapse" risk ("collapse" is defined here as the stock dropping below 20% of K) less than 10% in a typical 20 year period, the maximum allowable constant catch quota is only about 187 000 metric tonnes. This is most likely an optimistic result, since additional uncertainty about the estimate of MSY which would lead to additional "collapse" risks has not been accounted for.

9.2.4 **Hake**

The VPA fit for the hake stock in ICSEAF Division 1.6 was attempted without using any effort information. Sensible results (see Table 8.3 and Fig. 8.2) were only obtained for the period 1964 - 1974; thereafter the best-fit biomass trajectory increased to grossly unrealistic levels. This phenomenon is understandable in terms of the basic ambiguity of the catch-at-age data with respect to stock sizes (see Appendix 5.1). Clearly the VPA without effort data is of no particular value on its own; effort information must be included in the likelihood function.

With additional effort information in the VPA likelihood function, the resultant MSY estimate is essentially the same as that produced by the dynamic catch effort model for which results are reported by Butterworth and Andrew (1984) and Butterworth et al (1986). One is therefore led to question the value of incorporating the information about catch age structure into the analyses (viewed from the point of view of the dynamic production model analysis based on CPUE data; Butterworth and Andrew, 1984), considering first the meaninglessness of VPA without effort or relative abundance information, and secondly the large number of additional and questionable assumptions which must be made before the age data can be taken into account:

1. A value must be assumed for natural mortality (M), normally taken to be age independent and constant over the catch history.

2. Appropriate weights between the age structure (Q), catch mass (σ^2) and effort ($\sigma_1{}^2$) data must be ascribed.

3. A model must be chosen to represent changes in selectivity with age, and possibly also changes of this age-pattern with time.

Another concern is the matter of discarding at sea. Conclusions that this stock has made a substantial recovery since the mid-1970's, and estimates of MSY and TAC's, depend critically on the recent upward trend in biomass. However this optimistic picture seems contradicted by the decreasing trend of the mean age of the catch. Both could be real effects - brought about, for example, by geographical changes in fishing strategy and associated selectivity changes together perhaps with enhanced recruitment trends. However, if discards of younger ages were frequent in the the past, and have recently been much reduced (economic reasons such as the increased marketability of smaller hake could account for this) then some disturbing possibilities become evident. For example the current CPUE increase, with decreasing catch age, may be caused by a reduction in discarding; at the same time the actual stock size could well be decreasing. All the current estimates of hake sustainable yield would be invalidated by such an effect if substantial. The increased CPUE trend could also be an artifact of undetected efficiency increases in the fishing operation - vessel power factors are only crudely determined. These problems warrant further investigation.

9.3 FINAL COMMENTS

In summary, there are two obstacles which prevent the achievement of reliable population size estimates using catch-at-age data only. First, the database in question is equivocal in its specification of abundance trends. A particular catch age structure may be generated by one or more of at least five factors. These are fish recruitment trends, fishing mortality (and/or selectivity), natural mortality and emigration. It may be reasonable to set natural mortality constant over all years and ages and to specify its value, so that data input to the VPA likelihood function exceeds the number of model parameters. Even so, the remaining factors have to be distinguished in the analysis. The model developed here frequently failed at this point, and this manifested itself in the extensive imprecision of anchovy biomass estimates and also the severe problems demonstrated for the hake VPA in the absence of effort data.

This problem of the flatness in parameter space of the likelihood function given only catch and age information is appreciably reduced when additional relative abundance information is incorporated in the likelihood function, and if the functional form of the age specific fishing selectivities is specified. In this context, the usefulness of information concerning the age structure of the catch (added to the catch mass and abundance index data) is at best to define q, the proportionality constant between a biomass index (such as CPUE) and absolute abundance. The efficiency of VPA in giving some indication of the value of q is critically dependent on the natural mortality constant (M) assumed; the basis for the M value chosen is frequently poor - therefore even q seems not to be given reliably by the age structure data.

The fundamental question is really one of whether this additional age structure data does in reality provide more information on the biomass trajectory than is given by the catch mass and CPUE trends alone. The example of hake presented earlier suggests that the VPA does not substantially enhance the precision of biomass estimates - though admittedly a more rigorously statistical analysis is necessary to confirm this. The incorporation of the additional age structure data requires a number of extra parameters, which leads to further estimation problems. This is so even ignoring certain parameters which must be specified in order to obtain results with acceptable precision. These include a value for natural mortality and relative weighting for effort and age structure information. Simulation results suggest that these, and mainly the weighting factors and the natural mortality, cannot be satisfactorily estimated by the fitting procedure (Chapter 3). The overall impression is that age structure data is unlikely to improve precision of estimates of management related variables for the major South African fisheries.

A recent local development has been the use of direct surveys for obtaining absolute abundance estimates for anchovy. The techniques being used are hydroacoustic survey and egg production estimates. Preliminary biomass estimates have been made public, placing the November 1984 anchovy spawning biomass at roughly 1.3 million metric tonnes (Armstrong and Butterworth, 1986). Potentially these estimates offer the first opportunity to make real progress in pelagic fishery management, and have already revolutionised theories about stock habitat, migration and dynamics. However, skepticism about their short term value has already been voiced (Butterworth, 1985), and it seems that, even when the simplest management models, based on direct surveys, are used, gross uncertainty is unavoidable because, inter alia, of the high variances and potentially large systematic errors for these direct survey assessments. This uncertainty is emerging as an inherent feature of scientific knowledge about marine ecosystems, dispelling the myth of an eventual, precise understanding of ecodynamics.

In an attempt to cope with this uncertainty, models are being developed which unite the complete ambit of dynamic and statistical effects and limitations concerning the fishery. It is now realised that appreciable data contrast is necessary for estimating the surplus production function with reasonable precision, and that this data contrast may have to be the result of conscious decisions rather than accidental mismanagement (Smith and Walters, 1981; Ludwig and Hilborn, 1983; Walters, 1984). Achieving this data contrast by varying quotas has limited merit because of the high cost involved (R. Bross, pers. commn), and other imaginative options involving shifting effort between a number of contiguous stocks with the same overall quota allocation (Ludwig and Hilborn, 1983) are perhaps more appealing (this may be feasible for the various ICSEAF Divisions for hakes in Namibian or South African waters).

The essential implications of the analyses presented in this document are that the apparent accuracy offered by earlier deterministic cohort analysis techniques was (and continues to be) misleading; estimated biomass trajectories and subsequent MSY estimates must be interpreted with appreciation of their statistical precision. The necessity for direct surveys is fully endorsed by this work, which indicates that adequate assessment of some major South African fisheries on the basis of catch statistics alone is at best open to reservation (e.g. hake) and at worst probably impossible (e.g. anchovy and pilchard).

ACKNOWLEDGEMENTS

The research reported in this thesis was supported by the Benguela Ecology Programme (Council for Scientific and Industrial Research, South Africa).

I am indebted to Dr. D.S. Butterworth and Prof. J.G. Field in their official capacity as supervisors, and to Dr A.D. MacCall for suggesting the main research theme to me.

The guano and catch data used here were provided by the Sea Fisheries Research Institute of South Africa, via personal communications with the following individuals:

Dr M.J. Armstrong Mr P.A. Shelton:	Pelagic catch statistics
Dr R.J.M. Crawford:	Guano records
Mr R. Leslie:	Hake catch statistics

All illustrations which are not specifically credited otherwise were drawn by D. Gianakouras.

REFERENCES

Agger, P., Boetius, I., Lassen, H. (1971). On errors in the virtual population analysis. ICES C.M. 1971/H:16 (mimeo).

Alant, T.G., Baird, D., Crawford, R.J.M., Dreyer, T.P., Du Plessis, J.J., Erasmus, T., Field, J.G., Shelton, P.A., Siegfried, W.R. (1983). Final report of the scientific committee of enquiry into the exploitation of pelagic fish resources of South Africa and South West Africa. Cape Town, Republic of South Africa, 171 p.

Andrew. P.A. (1986). Dynamic catch-effort models for the southern African hake populations. M.Sc. thesis, Univ. Cape Town, S. Africa.

Andrews, W.R.H., Hutchings, L. (1980). Upwelling in the Southern Benguela Current. Prog. Oceanog. 9: 1-81.

Anon, (1977). Line fish catch of kob, geelbek, yellowtail, silverfish, hottentot, steenbras, and red fish estimated to be between 15 000 and 20 000 mt. Rep. Sea Fish. Brch S. Afr. 45: 8-10.

Anon, (1978). Line fish catch of kob, geelbek, yellowtail, silverfish, hottentot, steenbras, elf and red fish estimated to be between 15 000 and 20 000 mt. Rep. Sea Fish. Brch S. Afr. 46: p 8.

Anon, (1979). Line fish catch of kob, geelbek, yellowtail, silverfish, hottentot, steenbras, elf and red fish estimated to be between 15 000 and 20 000 mt. Rep. Sea Fish. Brch S. Afr. 47: p 9.

Anon, (1980). Line fish catch of kob, geelbek, yellowtail, silverfish, hottentot, steenbras, elf and red fish estimated to be more than 10 000 mt. Rep. Sea Fish. Brch S. Afr. 48: p 9.

Armstrong, M.J., Butterworth, D.S. (1986). Estimates of yield for the South African anchovy resource in ICSEAF Division 1.6. Colln scient. Pap. int. Commn SE. Atl. Fish. (submitted).

Armstrong, M.J., Shelton, P.A., Prosch, R.M., Grant, W.S. (1983). Stock assessment and population dynamics of anchovy and pilchard in ICSEAF Division 1.6 in 1982. Colln scient. Pap. int. Commn SE. Atl. Fish. 10: 7-25.

Assorov, V.V., Kalinina, M.I. (1979). Some peculiarities of the feeding habits of Cape hake and South African deep-water hake. Colln scient. Pap. int. Commn SE. Atl. Fish. 6:219-227.

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A., Thingstad, F. (1983). The ecological role of water-column microbes in the sea. Mar. Ecol. Prog. Ser. 10:257-263.

Bailey, G.W. (1985). Distribution and cycling of nutrients at four sites in the Benguela system. Int. Symp. Upw. W Afr., Inst. Inv. Pesq., Barcelona 1:305-317.

Baird, D. (1971). Seasonal occurrence of the pilchard Sardinops ocellata on the east coast of South Africa. Investl Rep. Div. Fish. S. Afr. 96:1-19.

Baird, D. (1978a). Catch composition and population structure of the commercially exploited mackerel Scomber japonicus, 1954-1975. Fish. Bull. S. Afr. 10: 50-61.

Baird, D. (1978b). Food of the mackerel Scomber japonicus from Western Cape waters. Fish. Bull. S. Afr. 10: 62-68.

Batchelor, A.L. (1982). The diet of the Cape Gannet, Sula capensis, breeding on Bird Island, Algoa Bay. M. Sc. thesis, Univ. of Port Elizabeth, S. Africa.

Batchelor, A.L., Ross, G.J.B. (1984). The diet and implications of dietary change of Cape Gannets on Bird Island, Algoa Bay. Ostrich, 55:45-63.

Beddington, J.R., Cooke, J.G. (1983). The potential yield of fish stocks. FAO Fish. Tech. Pap. 242: 47p.

Berry, H.H. (1976). Physiological and behavioural ecology of the Cape Cormorant, Phalacrocorax capensis. Madoqua, 9(4):5 - 55.

Best, P.B., Butterworth, D.S., Rickett, L.H. (1984). An assessmment cruise for the South African stock of Bryde's whales (Balaenoptera edeni). Rep. Int. Whal. Commn 34:403-423.

Bloomfield, P. (1976). Fourier analysis of time series: an introduction. New York: John Wiley.

Botha, L. (1980). The biology of the Cape hakes Merluccius capensis Cast. and M. paradoxus Franca in the Cape of Good Hope area. Ph. D. thesis, Univ. of Stellenbosch, S. Africa.

Boyd, A.J., Cruickshank, R.A. (1983). An environmental basin model for west coast pelagic fish distribution. S. Afr. J. Sci. 79: 150-151.

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

Brown, P.C. (1983a). Phytoplankton development in newly upwelled water. Proceedings 5th National Oceanographic Symposium, Grahamstown, S. Afr., 24-28 January 1983. Abstract in S. Afr. J. Sci. 79: 144.

Brown, P.C. (1983b). First look at phytoplankton production on the Agulhas Bank. 1st Ann. Congr. Phycol. Soc. southern Africa, Johannesburg, Jan. 1983. Abstract in S. Afr. J. Sci. 79: 382.

Butterworth, D. S. (1983). Assessment and management of pelagic stocks in the southern Benguela region. In: Sharp, G. D., Csirke, J. (ed.) Proceedings of FAO-IOC Expert consultation to examine changes in abundance and species composition of neritic fish stocks. San Jose, Costa Rica, 18-29 April 1983. FAO Fish. Rep. 291: 329-405.

Butterworth, D. S., Andrew, P. A. (1984). Dynamic catch-effort models for the hake stocks in ICSEAF Divisions 1.3-2.2. Colln scient. Pap. int. Commn SE. Atl. Fish. 11(II): 29-58.

Butterworth, D. S. (1985). Future prospects for South African anchovy assessment and management with associated implications for research planning. Unpublished document, Univ. Cape Town, S. Africa.

Butterworth, D. S., Bergh, M. O., Andrew, P. A. (1986). A comparison of dynamic catch-effort model and VPA assessments for the hake stocks in ICSEAF Divisions 1.3 - 1.6. Colln scient. Pap. int. Commn SE. Atl. Fish.(submitted).

Caddy, J. F. (1985). Species interactions and stock assessment - some ideas and approaches. In: Bas, C., Margalef, R., Rubies, P. (eds.). International Symposium on the most important Upwelling Areas off Western Africa (Cape Blanco and Benguela). Instituto de Investigaciones Pesqueras, Barcelona (in press).

Carter, R. A. (1982). Phytoplankton biomass and production in a southern Benguela kelp bed system. Mar. Ecol. Prog. Ser. 8: 9-14.

Carter, R. A. (1983). The role of phytoplankton and micronekton in carbon flow through a southern Benguela kelp bed. Ph. D. thesis, Univ. Cape Town, S. Africa.

Centurier-Harris, O. M. (1977). Estimates of size and interaction of the South African anchovy and pilchard populations. M. Sc. thesis, Univ. of Cape Town, S. Africa.

Chłapowski, K. (1976). Food and feeding of hake in Southwest African seas. Colln scient. Pap. int. Commn SE. Atl. Fish. 4: 115-120.

Clark, C. W. (1973). The economics of overexploitation. Science, 181: 630-634.

Clark, C. W. (1985). Bioeconomic modelling and fisheries management. New York: John Wiley.

Clark, W. G. (1981). Restricted least-squares estimates of age-composition from length composition. Can. J. Fish. Aquat. Sci. 38: 297-307.

Conover, R. J. (1978). Transformation of Organic Matter. In: Kinne, O. (ed.) Marine ecology. New York: John Wiley, p. 221-500.

Crawford, R. J. M. (1979). Implications of recruitment, distribution and availability of stocks for management of South Africa's western Cape purse-seine fishery. Ph. D. thesis, Univ. Cape Town, S. Africa.

Crawford, R. J. M. (1981a). Distribution, availability and movements of pilchard Sardinops ocellata off South Africa, 1964-1976. Fish. Bull. S. Afr. 14: 1-46.

Crawford, R. J. M. (1981b). Distribution, availability and movements of anchovy Engraulis capensis off South Africa, 1964-1970. Fish. Bull. S. Afr. 14: 51-94.

Crawford, R. J. M., Centurier-Harris, O. M., Wingate, G. H. L., Kriedemann, B. D. (1978). Revision of species and age composition of landings in the South African purse-seine fishery, 1950-1976. Fish. Bull. S. Afr. 10: 69-88.

Crawford, R. J. M., Shelton, P. A. (1978). Pelagic fish and seabird interrelationships off the coasts of South West and South Africa. Biol. Conserv. 14: 85-109.

Crawford, R. J. M., Shelton P. A., Hutchings, L. (1983). Aspects of variability of some neritic stocks in the southern Benguela system. In: Sharp, G. D., Csirke, J. (ed.) Proceedings of FAO-IOC Expert consultation to examine changes in abundance and species composition of neritic fish stocks. San Jose, Costa Rica, 18-29 April 1983. FAO Fish. Rep. 291: 407-448.

Crawford, R. J. M., de Villiers, G. (1984). Snoek and their prey-interrelationships in the Benguela upwelling system. S. Afr. J. Sci. 81: 91-97.

Cushing, D. H. (1969). Upwelling and fish production. FAO Fish. Tech. Pap. 84: 40 p.

DAMREP, (1983). Data and management report for the South African pelagic fish resource in 1983. Compiled by the pelagic section of the Sea Fisheries Research Institute, Cape Town, South Africa.

Davies, D. H. (1956). The South African pilchard Sardinops ocellata, Migration, 1950-1955. Investl Rep. Div. Sea Fish. S. Afr. 24: 1-52.

Davies, S. L., Newman, G. G., Shelton, P. A. (1981). A review of the South African multispecies pelagic fishery in ICSEAF division 1.6 in 1980. Colln scient. Pap. int. Commn SE. Atl. Fish. II: 51 - 74.

Deriso, R. B. (1980) Harvesting strategies and parameter estimation for an age-structured model. Can. J. Fish. Aquat. Sci. 37: 268-282.

Doubleday, W. G. (1976). A least squares approach to analysing catch at age data. Res. Bull. int. Commn NW Atlant. Fish. 12: 69-81.

Duncan, C.P., Nell, J.H. (1976). Surface currents off the Cape coast. Investl Rep. Div. Sea Fish. S. Afr. 76:1-19.

Dupont, W.D. (1983). A stochastic catch-effort method for estimating animal abundance. Biometrics, 39:1021-1033.

Efron, B. (1981). Nonparametric estimates of standard error: The jacknife, the bootstrap and other methods. Biometrika, 3:589-599.

Everson, I. (1977). The Southern Ocean: the living resources of the Southern Ocean. FAO Rep., GLO/SO/77/1. Southern Ocean Fisheries Survey Programme, FAO, Rome. 156pp.

Finney, D.J. (1951). On the distribution of a variate whose logarithm is normally distributed. Suppl. J. R. Statist. Soc. 7: 155-161.

Fournier, D. (1983). An analysis of the Hecate Strait Pacific cod fishery using an age-structured model incorporating density dependent effects. Can. J. Fish. Aquat. Sci. 40:1233-1243.

Fournier, D., Archibald, C.P. (1982). A general theory for analysing catch at age data. Can. J. Fish. Aquat. Sci. 39:1195-1207.

Furness, R.W., Cooper, J. (1982). Interactions between breeding seabird and pelagic fish populations in the southern Benguela region. Mar. Ecol. Prog. Ser. 8:243-250.

Gordon, H.S. (1954). The economic theory of a common property resource: the fishery. J. Polit. Econ. 62:124-142.

Gray, D.F. (1977). An iterative derivation of fishing and natural mortality from catch and effort data giving measurements of goodness of fit. ICES, C.M. 1977/F:33 Demersal Fish Comm. 13 pp. (mimeo).

Grenander, U., Rosenblatt, M. (1957). Statistical analysis of stationary time series. New York: John Wiley. p 91-94.

Greene, K. (1978) Ecosystem description of the California current. Final report, Marine Mammal Commission Contract No. MM7AC-026.

Gulland, J.A. (1965). Estimation of mortality rates. Annex to Arctic Fisheries Working Group Report. ICES, C.M., Doc. No. 3 (mimeo).

Hardin, G. (1968). The tragedy of the commons. Science, 162: 1243-1247.

Hatanaka, M., Sekino, K., Takahashi, M., Ichimura, T. (1957). Growth and food consumption in young mackerel, Pneumatophorus japonicus (Houttuyn). Tohoku J. Agric. Res. 7: 351-368.

Hatanaka, M., Takahashi, M. (1960). Studies on the amounts of the anchovy consumed by the mackerel. Tohoku. J. Agric. Res. 11(1): 83-100.

Idyll, C.P. (1973). The anchovy crisis. Sci. Am. 228(6): 22-29.

Jackson, F., Siegfried, W.R., Cooper, J. (1976). A simulation model for the population dynamics of the Jackass penguin. Trans. Roy. Soc. S. Afr. 42: 11-21.

Jarvis, M.J.F. (1971). Ethology and ecology of the South African gannet, *Sula capensis*. Ph. D. thesis, Univ. Cape Town, S. Africa.

Jenkins, G.M., Watts, G.W. (1969). Spéctral analysis and its applications. San Fransisco: Holden-Day.

Kendall, M.G., Stuart, A. (1967). The advanced theory of statistics: vol 2: inference and relationship. 2nd edition, London: Charles Griffin.

King, D.P.F. (1977). Distribution and relative abundance of eggs of the South West African pilchard *Sardinops ocellata* and anchovy *Engraulis capensis*, 1971/1972. Fish. Bull. S. Afr. 9: 23-31.

Lane, E.D., Kingsley, M.C.S., Thornton, D.E. (1979). Daily feeding and food conversion efficiency of the diamond turbot: an analysis based on field data. Trans. Am. Fish. Soc. 108: 530-535.

Le Clus, F. (1983). Observations on the egg production of pilchard and anchovy off South West Africa in 1981/1982. Colln scient. Pap. int. Commn SE. Atl. Fish. 10: 139-145.

Lindley, D.V. (1947). Regression lines and the linear function relationship. Suppl. J. R. Statist. Soc. 9: 218.

Lleonart, J., Salat, J., Macpherson, E. (1985). CVPA, an expanded VPA with cannibalism. Applications to a hake population. Fish. Res. 3:61-79.

Ludwig, D., Hilborn, R. (1983). Adaptive probing strategies for age-structured fish stocks. Can. J. Fish. Aquat. Sci. 40:559-569.

Ludwig, D., Walters, C.J. (1981). Measurement errors and uncertainty in parameter estimates for stock and recruitment. Can. J. Fish. Aquat. Sci. 38:711-720.

MacCall, A.D. (1979). Population estimates for the waning years of the Pacific sardine fishery. Rep. CCOFI 20: 72-82.

MacCall, A.D. (1983). Population models of habitat selection, with application to the northern anchovy. Ph. D. Dissertation, University of California at San Diego.

Magnuson, J. J. (1969). Digestion and food consumption by skipjack tuna Katsuwonus pelamis. Trans. Am. Fish. Soc. 98: 379-392.

Mann, K. H. (1982). Ecology of coastal waters: a systems approach. Oxford: Blackwell.

Matthews, J. P. (1961). The pilchard of South West Africa Sardinops ocellata and the maasbanker Trachurus trachurus. Bird predators, 1957 - 1958. Investl Rep. mar. Res. Lab. S. W. Afr. 3: 1-35.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. Nature, 261: 459-467.

May, R. M. (1981). Models for single populations. In: May, R. M. (ed.). Theoretical ecology, principles and applications. 2nd Edition, Boston: Blackwell Scientific Publications, p5-29.

Mills, E. L. and Fournier, R. O. (1979). Fish production and the marine ecosystems of the Scotian Shelf, eastern Canada. Mar. Biol. 54: 101-108.

Murphy, G. I. (1966). Population biology of the Pacific sardine Sardinops caerulea. Proceedings of the California Academy of Science, Fourth Series. 34 (1): 1-84.

NAG; Mathematical Software (1977). Numerical Algorithms Group Ltd. Oxford.

Nelson, J. B. (1966). Population dynamics of the gannet (Sula bassana) at the Bass Rock, with comparative information from other Sulidae. Jn. of Animal Ecology, 35(3): 443-470.

Nelson, J. B. (1978). The sulidae; gannets and boobies. London: Oxford University Press.

Nelson, G., Hutchings, L. (1983). The Benguela upwelling area. Prog. Oceanogr. 12: 333-356.

Nepgen, C. S. de V. (1979). The food of the snoek Thyrsites atun. Fish. Bull. S. Afr. 11: 39-42.

Nepgen, C. S. de V. (1982). Diet of predatory and reef fish in False Bay and possible effects of pelagic purse-seining on their food supply. Fish. Bull. S. Afr. 16: 75-93.

Newell, R. C., Linley, E. A. S. (1984). Significance of microheterotrophs in the decomposition of phytoplankton: estimates of carbon and nitrogen flow based on the biomass of plankton communities. Mar. Ecol. Prog. Ser. 16: 105-119.

Newman, G. G. (1970). Migration of the pilchard Sardinops ocellata in southern Africa. Investl Rep. Div. Sea Fish. S. Afr. 86: 1-6.

Newman, G.G., Crawford, R.J.M., Centurier-Harris, O.M. (1974). Assessment of fishing effort, stock abundance and yield for the South African multi-species pelagic fishery. Colln scient. Pap. int. Commn SE Atl. Fish. 1:163-170.

Olivieri, E.T., Hutchings, L. (1983). Zooplankton grazing in the southern Benguela. Proc 5th National Oceanographic Symposium, Grahamstown, S. Afr., January 1983. Abstract in S. Afr. J. Sci. 79: 145.

Pauly, D., (1980). On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. J. Cons. int. Explor. Mer. 39:175-192.

Pella, J.J., Tomlinson, P.K. (1969). A generalized stock production model. Bull. Inter-Amer. Trop. Tuna Commn 13:421-496.

Platt, T., Irwin, B. (1973). Caloric content of phytoplankton. Limnol. Oceanogr. 18: 306-309.

Platt, T., Mann, K.H., Ulanowicz, R.E. (1981). Mathematical models in biological oceanography. Paris: UNESCO Press.

Pope, J.G. (1972). An investigation of the accuracy of virtual population analysis using cohort analysis. Res. Bull. Int. Comm. Northw. Atl. Fish. 9:65-74.

Pope, J.G. (1977). Estimation of fishing mortality, its precision and implications for the management of fisheries. In: Steele, J.H. (ed.). Fisheries mathematics. New York: Academic Press.

Pope J.G., Shepherd, J.G. (1982). A simple method for the consistent interpretation of catch-at-age data. J. Cons. int. Explor. Mer. 40: 176-184.

Rand, R.W. (1960a). The biology of guano-producing seabirds. The distribution, abundance and feeding habits of the Cape penguin, Spheniscus demersus, off the southwestern coast of the Cape Province. Investl Rep. Div. Fish. Un. S. Afr. 41: 1-28.

Rand, R.W. (1960b). The biology of guano producing seabirds. 3. The distribution, abundance and feeding habits of the cormorants Phalacrocoracidae off the south-western coast of the Cape Province. Investl Rep. Div. Fish. Un. S. Afr. 42: 1-32.

Rand, R.W. (1963). The biology of guano-producing seabirds. 4. Composition of colonies on the Cape Islands. Investl Rep. Div. Fish. Un. S. Afr. 43: 1-32.

Randall, R.M. (1983). Biology of the Jackass Penguin Spheniscus demersus (L.) at St. Croix Island, South Africa. M. Sc. thesis, Univ. Port Elizabeth, South Africa.

Ricker, W.E. (1973). Linear regression in fishery research. J. Fish. Res. Board Can. 30: 409-434.

Sampson, D.B. (1984). Variance estimators for virtual population analysis. M.Sc. thesis, Univ. Washington, Seattle, USA.

Schnutte, J. (1985). A general theory for analysis of catch and effort data. Can. J. Fish. Aquat. Sci. 42: 44-429.

Shannon, L.V., Mostert, S.A., Walters, N.M., Anderson, F.P. (1983). Chlorophyl concentrations in the southern Benguela Current regions as determined by satellite (Nimbus-7 Coastal zone colour scanner). J. Plankton Res. 5: 565-583.

Shannon, L.V., Field, J.G. (1985). Are fish stocks food limited in the southern Benguela pelagic ecosystem? Mar. Ecol. Prog. Ser. 22: 7-19.

Sheldon, R.W., Prakash, A., Sutcliffe, W.H., Jr. (1972). The size distribution of particles in the ocean. Limnol. Oceanogr. 17: 327-340.

Shelton, P.A. (1979). The effect of environmental factors on spawning and recruitment of the anchovy, Engraulis capensis Gilchrist; compared with the lightfish, Maurolicus muelleri Gmelin. M. Sc. thesis, Univ. of Cape Town, S. Africa.

Shelton, P.A., Kriel, F. (1976). Surface drift and the distribution of pelagic-fish eggs and larvae off the south-east coast of South Africa, November and December 1976. Fish. Bull. S. Afr. 13: 107-109.

Shelton, P.A., Hutchings, L. (1982). Transport of anchovy, Engraulis capensis Gilchrist, eggs and early larvae by a frontal jet current. J. Cons. int. Explor. Mer. 40: 185-198.

Shelton, P.A., Armstrong, M.J. (1983). Variations in parent stock and recruitment of pilchard and anchovy populations in the southern Benguela system. In: Sharp, G.D., Csirke, J. (ed.) Proceedings of FAO-IOC Expert consultation to examine changes in abundance and species composition of neritic fish stocks. San Jose, Costa Rica, 18-29 April 1983. FAO Fish. Rep. 291: 1113-1133.

Shepherd, J.G. (1982). A family of general production curves for exploited populations. Math. Biosci. 59: 79-93

Shimshoni, M. (1971). On Fisher's test of significance in harmonic analysis. Geophys. J. R. Astr. Soc. 23: 373-377.

Shushkina, E.A., Vinogradov M.YE., Sorokin YU.I., Lebedeva L.P., Mikheyev, V.N. (1978). Functional characteristics of planktonic communities in the Peruvian upwelling region. Marine Biology, 18(5): 579-589.

Siegfried, W.R., Crawford, R.J.M. (1978). Jackass penguins, eggs and guano: Diminishing resources at Dassen Island. S. Afr. J. Sci. 74: 389-390.

Siegfried, W.R., Field, J.G. (eds.) (1981). A description of the Benguela Ecology Programme. S. Afr. National Scientific Programmes Report, 54: 1-39.

Sims, S.E. (1982). The effect of unevenly distributed catches on stock-size estimates using Virtual Population Analysis (Cohort Analysis). J. Cons. int. Explor. Mer. 40: 47-52.

Smith, A.D.M., Walters, C.J. (1981). Adaptive management of stock-recruitment systems. Can. J. Fish. Aquat. Sci. 38: 690-703.

Sorokin, YU.I., Mikheyev, V.N. (1979). On characteristics of the Peruvian upwelling ecosystem. Hydrobiologia, 62(2): 165-189.

Stander, G.H., le Roux, P.J. (1968). Notes on fluctuations of the commercial catch of the South African pilchard (Sardinops ocellata), 1950-1965. Investl Rep. Div. Sea Fish. S. Afr. 65: 1-14.

Stander, G.H., de Decker, A.H.B., (1969). Some physical and biological aspects of an oceanographic anomaly off South West Africa. Investl Rep. Div. Sea Fish. S. Afr. 81: 1-46.

Thomas, R.M. (1985). Age studies on pelagic fish in the south-east Atlantic, with particular reference to the South West African pilchard, Sardinops ocellata. Ph. D. thesis, Univ. Cape Town, S. Africa.

Ulltang, O. (1977). Sources of errors in and limitations of Virtual Population Analysis (Cohort Analysis). J. Cons. int. Explor. Mer. 37: 249-260.

Walters, C.J. (1984). Managing fisheries under biological uncertainty. In: May, R.M. Exploitation of Marine Communities. Dahlem Konferenzen 1984. Berlin, Heidelberg, New York, Tokyo: Springer-Verlag. p 263-274.

Wonnacott, T.H., Wonnacott, R.J. (1977). Introductory statistics. New York; Chichester; Brisbane; Toronto: John Wiley.

Appendix 1.1

NOTATION

Consider A as representing any of the variables or parameters under consideration in this work; then additional sub-, super- and prescripts are used to alter its sense. These are usually;

A ; Refers to the true value of a variable or parameter - this can also represent an estimate of A, when defined as such in its context.

A^{obs} ; This is the observed value of A.

A^{ind} ; An index of A.

$A_{i,j,l}$; Any possible combination of i, j or l may occur, where i refers to year, j to age and l to length class. For example, C_i is the total number of fish caught in year i, while $C_{i,j}$ is the total number of fish caught in year i which belong to age class j.

${}^{w}A$; This defines the units of A as mass rather than number (of fish).

$A(t)$; A continuous variable, dependent on time, t (l is sometimes used, where l=fish caudal length).

Only a limited number of variations of the base symbol A are defined in the table presented below. The interpretation of other symbols follows from the definitions already shown.

ENGLISH SYMBOLS

a_0 ; the constant term in the expression for a time series in terms of its Fourier components (see equation 6.2).

a_k ; the coefficent of the <u>real part</u> of the kth Fourier component of a time series (see equation 6.3).

$a_{i,l,j}$; the proportion of fish (in year i) aged j which also fall in the length class l ("age length key").

$\underline{a}$; matrix representation of the age length key, $a_{i,j,l}$.

$A_{j,l}$; The number of fish caught which are aged j and occur within length class l (in a particular year).

B_i ; The biomass of a fish population at the beginning of year i .

$B_{i,k}$; The k th estimate of the biomass of a fish population (from for example a resampling exercise).

b ; The parameter of the functional form used to describe S_j, the age specific selectivity (see equation 3.29).

b_1, b_2, b_3; parameters used in the hake selectivity function (see equation 8.1 and 8.2).

b_k ; the coefficent of the <u>imaginary part</u> of the kth Fourier component of a time series (see equation 6.4).

B(t) ; instantaneous fish population biomass.

C_i ; The number of fish caught in year i (just C if no year reference is intended).

${}^{\bullet}C_i$; The mass of fish caught in year i (or just ${}^{\bullet}C$ as above).

$C_i(t)$; The instantaneous catch rate in year i, where the time variable refers to fish age.

C.V. or CV ; Coefficient of variation, defined as the square root of the variance (or standard deviation) divided by the mean.

${}^{14}C$; Carbon isotope with atomic weight 14 - this symbol refers to a particular method of estimating carbon fixing rates attained by phytoplankton.

c_k ; The harmonic amplitude of a time series at the k th Fourier frequency (see equation 6.5).

$D_{i,j}$; the natural logarithm of the ratio of the catch of a cohort during the interval i, j to the catch during the interval i+1, j+1 (catches by number) (see equation 3.17).

E_i ; The fishing effort expended in year i (just E if no year reference is implied).

F_i ; A measure of average annual fishing mortality, corresponding to the model where $F_{i,j} = F_i S_j$.

$F_{i,j}$; The instantaneous catch rate per fish for a cohort in year i where age is j .

$f_{i,l}$; the proportion of fish in the catch in year i which fall in length class l .

$f^*(w)$; The normalised sample spectrum estimate of a time series, defined by equation 6.8

$f_r(..)$; A generalised recruitment function (see equation 8.6 and 8.7).

G_i ; Annual commercial guano harvests at various seabird island breeding sites.

$\underline{g}$; matrix representation of the normalised length distribution of the various age classes in the catch (see equation 5.11 and section 5.3).

$g_{j,l}$; the normalised length distribution of the various age classes in the catch, subscript variable representation of $\underline{g}$.

$g_q(t)$; a function (Bradu and Mundlak, 1970) used in lognormal estimation.

g_k ; the normalised periodogram of a time series for the various k Fourier components (see equation 6.7).

H ; The number of fish which are sampled in order to obtain $f_{i,l}^{obs}$.

h_l ; the number of fish sampled and aged in length class l for purposes of building up an age-length-key.

i ; A subscript variable which defines the year of fishing being referred to. Thus i=1981 refers to the period 1 January 1981 to 31 December 1981.

j ; A subscript variable which defines the fish age being referred to. Thus j=3 refers to fish aged between exactly 3 years old and exactly 4 years old. Birthdays are assumed to occur on the 1 st January.

j_o ; defines the point of inflexion in the hake selectivity function (see equation 8.1 and 8.2).

K ; The carrying capacity of an ecosystem for a species.

K_1 ; The gross growth efficiency of an individual animal.

k_e ; proportionality constant between fishing effort (E_i) and fishing mortality (F_i).

k_r ; proportionality constant between recruitment index and recruitment.

k_s ; proportionality constant between biomass index and biomass.

lmax ; The total number of length classes considered in reporting the length frequency distribution of the catch.

l ; A subscript variable which defines the fish length class to which reference is made. These vary for different fish, but the length increment between successive length classes is normally 0.5 cm. of caudal length (alternatively it may be used as a continuous variable for fish length, eg. in C(l)).

$M_{i,j}$; The natural mortality (instantaneous death rate per fish) for a cohort aged j in year i .

M ; The natural mortality (instantaneous death rate per fish) assumed constant for all cohorts at all ages.

MSY^A, MSY^B, MSY^B; various variables representing maximum sustainable yield - defined in section 7.8.

m ; the maximum number of age classes which are considered in the VPA.

n ; The number of years of fishing considered in the VPA.

$N_{i,j}$; The number of fish in the population on 1 January in year i aged exactly j years; this usually refers to numbers as estimated by the VPA.

O_2 ; symbol for molecular oxygen - this symbol refers to a particular method of estimating carbon fixing rates attained by phytoplankton.

$P_{i,j}$; The proportion of fish in the catch in year i which fall into age class j (i.e. aged j to j+1 years).

Q ; The number of fish which are aged each year to determine $P_{i,j}^{obs}$ from a random sample of the annual catch.

q ; The catchability coefficient for the relationship between CPUE and population biomass.

$q_{i,j}$; The number of sampled fish which are found to lie in age class j in year i.

r ; The maximum per capita population growth rate for logistic growth dynamics (also used for Pearson's correlation coefficient).

R_i ; The estimated annual recruitment into a fishery in year i.

r_i ; The true annual recruitment into a fishery.

$r(\alpha, \beta, y)$; recruitment as a function of the two parameters, α and β, and spawning biomass, y.

S_j ; A weighting factor, dependent on age, j, which is equal to selectivity ($F_{i,j}$ modelled by $F_i S_j$).

T ; The age to maturity for various seabird species (and sometimes for pilchards).

t ; the continuous time variable.

U_1, U_2 ; Uniform U(0,1) random variates.

u_i ; the model error of the logarithm of recruitment in year i (see equation 7.10).

v_i ; Observation error of the logarithm of spawning biomass (see equation 7.10).

V_i ; The size of the adult seabird population in year i.

$W(w)$; The squared coherency of a time series at frequency w (see equation 6.9).

w_i ; Observation error of the logarithm of recruitment (see equation 7.10).

w_l ; The mass of an individual fish halfway between the length limits of length class l (normally at length l+0.25 cm).

w_j ; The mass of an individual fish in age class j, aged exactly j+1/2.

X_i ; The number of newborn seabirds produced in year i.

Y_i ; The spawning biomass (estimated) of a fish population at the beginning of year i.

y_i ; The true spawning biomass of a fish population at the beginning of year i.

Z, Z_1, Z_2, Z_i ; Standard normal N(0,1) random variates.

Z ; The total mortality coefficient.

GREEK SYMBOLS

α ; parameter in the Ricker recruitment function which is the maximum per capita (of spawner biomass) recruitment rate (see equation 7.10).

β ; parameter in the Ricker recruitment function. Maximum recruitment occurs when the spawning biomass is equal to $1/\beta$ (see equation 7.10).

η ; the ratio of model error variance to observation error variance used in the likelihood function set up for estimating the stock recruit parameters (see equation 7.11).

σ^2 ; the variance of $\ln(C_i^{obs})$ about $\ln(C_i)$ - equation 3.31.

σ_1^2 ; the model error variance for the relationship between E_i and F_i (variance of the log of the residuals) (see equation 4.5).

σ_r^2 ; the model error variance for the relationship between $N_{i,1}^{ind}$ (recruitment index) and $N_{i,1}$ (true recruitment) (see equation 4.19).

σ_s^2 ; the model error variance for the relationship between the spawning biomass index and the true spawning biomass (see equation 4.23).

σ_u^2 ; model error variance for recruitment (see section 7.6.1).

σ_v^2 ; observation error variance for the logarithm of spawning biomass (see section 7.6.1).

σ_w^2 ; observation error variance for the logarithm of recruitment (see section 7.6.1).

σ_ε ; $= \sigma_u^2 + \sigma_v^2$

σ_x^2 ; total variance for the logarithm of spawning biomass (equation 7.24).

γ^i ; As defined on page 72.

$\Phi^{i,j}$; As defined on page 72.

τ ; Statistical rank, also used as a generalised exponent for skew surplus production curves.

γ ; Generalised recruitment function parameter in equation 8.6 and 8.7 - also used as the slope of the replacement line for stock and recruitment relationships.

MATHEMATICAL AND STATISTICAL CONVENTIONS

$\exp(x) = e^x$ where $e = 2.718281828$

$\ln^2(x) = \ln(x)\ \ln(x)$ (i.e. not $\ln(\ln(x))$)

$\ln(x) = \log(x) = \log_e(x)$

E(X) ; the expected value of the random variable X, also termed the first moment or the expectation of X.

Var(X) ; the variance, or second moment of the random variable X

i ; $(-1)^{1/2}$, where applicable

$x \approx N(0, y)$; x is a random variable with true mean of 0 and variance of y

ACRONYMS

BEP ; Benguela Ecology Programme

CPUE ; Catch Per Unit Effort

CZCS ; Coastal Zone Colour Scanner

ICSEAF ; International Commission for the Southeast Atlantic Fisheries

MSY ; Maximum Sustainable Yield

SFRI ; Sea Fisheries Research Institute (Department of Environment Affairs; South Africa)

TAC ; Total Allowable Catch

VPA ; Virtual Population Analysis

Appendix 4.1. Generating observed catch mass and age structure data sets from the parameters of the underlying distributions.

Age structure

Multinomial sampling for estimating the catch age structure was simulated by dividing U(0,1), a uniform distribution on the interval (0,1), into j segments such that the width of each segment equalled $P_{i,j}$. Q random numbers from U(0,1) were then generated, and the number falling within the range of each segment was thus the estimate $q_{i,j}^{obs}$. The $P_{i,j}^{obs}$ estimates are given by;

$$P_{i,j}^{obs} = q_{i,j}^{obs}/Q$$

Total catch

The generation of C_i^{obs} was simulated by use of the relationship;

$$\ln C_i^{obs} = \ln C_i + Z_i \sigma$$

where $Z_i \approx N(0,1)$

For each year of fishing, a Z_i value was generated by making two selections from U(0,1), U_1 and U_2, and transforming them;

$$Z_i = (-2\ln U_1)^{0.5} \sin(2\pi U_2)$$

from which Z_i can be shown to be normally distributed with a mean of 0 and variance of 1 (Appendix 4.2). σ was calculated from the relationship;

$$\sigma = [\ln(CV^2+1)]^{0.5}$$

where CV is the desired coefficient of variation of C_i^{obs} about C_i;

$$CV = (\mathrm{var}(C_i^{obs}))^{0.5}/C_i$$

Appendix 4.2. Demonstrating that if U_1 and U_2 are from uniform distributions on [0,1], the random variable $(-2\ln U_1)^{1/2}\sin(2\pi U_2)$ is exactly normally distributed with mean 0 and variance 1.

For U_1 and $U_2 \approx U(0,1)$, the probability density function, $f(Z_1,Z_2)$, is sought, where;

$$Z_1=(-2\ln U_1)^{1/2}\sin(2\pi U_2)$$

$$Z_2=(-2\ln U_1)^{1/2}\cos(2\pi U_2)$$

Solving for U_1 and U_2 from this yields;

$$U_2=\tan^{-1}(Z_1/Z_2)/2\pi$$

$$U_1=\exp(-(Z_1^2+Z_2^2)/2)$$

The Jacobian is;

$$J=\begin{vmatrix} \partial U_1/dZ_1 & \partial U_2/dZ_1 \\ \partial U_1/dZ_2 & \partial U_2/dZ_2 \end{vmatrix}$$

$$=\begin{vmatrix} -Z_1\exp(-(Z_1^2+Z_2^2)/2) & 1/Z_2/(1+Z_1^2/Z_2^2)/2\pi \\ -Z_2\exp(-(Z_1^2+Z_2^2)/2) & -Z_1/Z_2^2/(1+Z_1^2/Z_2^2)/2\pi \end{vmatrix}$$

$$= \exp(-(Z_1^2+Z_2^2)/2)\ [(Z_1^2/Z_2^2+1)/(1+Z_1^2/Z_2^2)]/2\pi$$

Therefore;

$$f(Z_1,Z_2) = \exp(-Z_1^2/2)/(2\pi)^{1/2}\ .\ \exp(-Z_2^2/2)/(2\pi)^{1/2}$$

Therefore;

$$f(Z_1) = \exp(-Z_1^2/2)/(2\pi)^{1/2}$$

so that $Z_1 \approx N(0,1)$.

Appendix 5.1. Ambiguities inherent in associating the total mortalities seen in the catch-at-age matrix with those in the population.

The VPA regression discussed in this thesis can be partly viewed as an attempt to relate the mortalities apparent in the catch-at-age matrix to the true total mortality of the population.

Let $Z_{i,j}$ represent the sum of fishing mortality and natural mortality (i.e. specific instantaneous mortality rates) for the actual population. The mortality apparent from the catch-at-age matrix, $Z_{i,j}^{app}$, is

$$Z_{i,j}^{app} = \ln[C_{i,j}/C_{i+1,j+1}]$$

The analysis is substantially simplified if:

$$C_{i,j}/C_{i+1,j+1} \approx F_i S_j \; N_{i,j} \; / \; (F_{i+1} S_{j+1} \; N_{i+1,j+1})$$

[From the catch equation this is true if:

$$\{Z_{i+1,j+1}[1-\exp(-Z_{i,j})]\}/\{Z_{i,j}[1-\exp(-Z_{i+1,j+1})]\} \approx 1$$

i.e. this requires that Z does not fluctuate too much, which would require that in the equation $Z=FS+M$, either M is constant and dominant, or FS does not vary too much]

If recruitment is constant, and if there is no net emi- or immigration across the stock habitat boundaries, then

$$\ln[N_{i,j}/N_{i+1,j+1}] = Z_{i,j}$$

With the additional possibility of net loss or gain by migration, and given that recruitment varies,

$$\ln[N_{i,j}/N_{i+1,j+1}] = Z_{i,j} + R + E$$

where R and E are the exponential rate constants representing recruitment fluctuations and emi/immi-gration. Thus, if $F_i = F_{i+1}$, the logarithm of the catch ratio, the apparent total mortality is given approximately by:

$$Z_{i,j}^{app} = \ln[C_{i,j}/C_{i+1,j+1}] = F_i S_j + M + q + R + E$$

where $q = \ln(S_j/S_{j+1})$

The naive approach of equating the log of catch ratios to $Z_{i,j}$ therefore ignores changes in selectivity with age, recruitment with time or the possibility of emigration. Large apparent values of mortality manifest in the catch-at-age matrix may therefore be ascribed to :

i) large F - large fishing mortalities
ii) large E - large emigration rates
iii) large q - large decreases in selectivity with increasing age
iv) large R - large changes in recruitment.

There are no further constraints within the catch-at-age data alone which will indicate which of the effects i) to iv) are dominant.

Appendix 6.1. Estimation of gannet natural mortality on the basis of observed frequencies of 3 year olds in the population

For gannets it is possible to estimate the natural mortality, M_b, from the observed frequencies of three year olds in the population, which are in turn estimated on the basis of the distinctive plumage of three year old gannets. A crude estimate for Malgas Island, the site of the other West Coast gannet colony, is that between 1:15 to 1:50 adults can be identified as three year olds (A. Berruti, pers. commn). If it is assumed that only birds of three years and older occur on the island and that there is no age dependent absence from the island at and after that age, then the adult population of three years and older, $A_{3\to\infty}$ can be expressed in terms of the number of three year olds, A_3, and the annual survival, s, as

$$A_{3\to\infty} = A_3+sA_3+s^2A_3+\ldots+s^nA_3. \qquad \ldots\ 6.22$$

As n tends to infinity the sum tends to $A_3/(1-s)$. The proportion of three year olds in the population is thus 1-s. Therefore the range in observed proportion of three year olds of 1:15 to 1:50 corresponds to estimates of s of between 93% and 98%. This is equivalent to natural mortalities for gannets, M_b, of between 0,07 yrs^{-1} and 0,02 yrs^{-1}, and this agrees well with Jarvis's (1971) estimate of 0,05 - 0,08 yrs^{-1}, and for estimates made by Nelson (1966) for <u>Sula bassana</u>.

Appendix 6.2 Sum of squares versus covariance as a measure of correspondence between Algoa Bay and transformed Lambert's Bay guano harvests.

Let A_i and B_i represent the observed guano harvests at Bird Island Lamberts Bay and Bird Island Algoa Bay respectively. Consider that a_i and b_i are hypothetical time series which are related in some complex manner:

$$b_i = f(a_i) + u_i$$

where f is the transformation applied to a_i at time i. It is assumed that u_i is a random variable with mean 0 and variance σ_u^2. $f(a_i)$ is more conveniently written as the series c_i:

$$b_i = c_i + u_i$$

A_i and B_i are corrupted versions of a_i and b_i:

$$A_i = a_i + v_i$$

$$B_i = b_i + w_i$$

where v_i and w_i are random variables with means 0 and variances σ_v^2 and σ_w^2 respectively.

The object is to determine the best representation of b_i by tranforming a_i by means of the function f.

CRITERION A

A suitable criterion might be to choose f by minimising the sum of squares:

$$S = \Sigma (b_i - c_i)^2$$

However as there is no measure of b_i or c_i, the best that can be attempted is to minimise:

$S^* = \Sigma (B_i - C_i)^2$

where now $C_i = f(A_i)$

If it is reasonable to assume that $C_i = c_i + x_i$ where x_i is random with mean zero and variance σ_x^2, then S^* can be rewritten as:

$S^* = \Sigma (b_i + w_i - c_i - x_i)^2$

which reduces to:

$S^* = S + n(\sigma_w^2 + \sigma_x^2)$ + (covariance terms)

where n is the length of the time series.

The covariance terms have expectations of zero. The main concern is that the desired criterion S is poorly reflected in S^* since the two terms σ_w^2 and σ_x^2 are involved in S^*. Changes in S due to different choices of f may therefore be only weakly reflected in S^* (depending on the size of σ_w^2 and σ_x^2). An alternative criterion, the covariance between B_i and C_i, is considered below.

CRITERION B

An alternative criterion for choosing f is to maximise the covariance (cov) between b_i and c_i:

$\text{cov} = \Sigma (\bar{b} - b_i)(\bar{c} - c_i)$

Again, only cov* can be measured:

$\text{cov}^* = \Sigma(\bar{B} - B_i)(\bar{C} - C_i)$

Assuming that

$\bar{B} = \bar{b}$

$\bar{C} = \bar{c}$

The cov^* can be rewritten in the form:

$$cov^* = \Sigma\,(\bar{b} - b_i - w_i)(\bar{c} - c_i - x_i)$$

Let $y_i = \bar{c} - c_i$

$z_i = \bar{b} - b_i$

Then

$$cov^* = cov + (\sigma_{wx} + \sigma_{wy} + \sigma_{xz})n$$

Recalling that:

w_i	are the errors	between	B_i and b_i
x_i	"	"	c_i and b_i
y_i	"	"	c and c_i
z_i	"	"	b and b_i

it therefore seems plausible that

$$\sigma_{wx} + \sigma_{wy} + \sigma_{xz} \approx 0$$

Note however that cov^* can be made arbitrarily large by scaling C_i upwards. It seems therefore that cov^* should be normalised to eliminate this effect. An obvious normalisation for cov^* is given by the standard correlation coefficient:

$$r_{BC} = cov_{BC}/(\sigma_B \sigma_C) \approx cov_{bc}/(\sigma_B \sigma_C)$$

$$r_{BC} \approx cov_{bc}/[(\sigma_b{}^2+\sigma_w{}^2)^{0.5}(\sigma_c{}^2+\sigma_x{}^2)^{0.5}]$$

recall that $S^* = S + n(\sigma_w{}^2 + \sigma_x{}^2)$

In assessing the suitability of using either r_{BC} or S^* as a criterion for choice of best f, a major concern is that improvements in r_{BC} are masked by compensatory changes in $\sigma_x{}^2$ and

σ_c^2, or changes in only σ_x^2 in the case of S^*. The effect of changes in σ_x^2 on either r_{BC} or S^* can be compared by contrasting their first derivatives with respect to σ_x^2. These are:

$$|dr_{BC}/d\sigma_x^2| = |r_{BC}/(2(\sigma_c^2+\sigma_x^2))|$$

$$dS^*/d\sigma_x^2 = n$$

Since

$\sigma_c^2 + \sigma_x^2 >> 1$,

and

$r_{BC} < 1$,

it is likely that

$|dr_{BC}/d\sigma_x^2| << 0.5$,

so that compensatory changes in σ_x^2 masking increases in r_{BC} seem less of a problem than for S^*. Therefore the correlation coefficient has been preferred as the criterion for choosing the best transformation function f.

Appendix 7.1. Fitting the stock recruit curve (Ricker version) in the absence of observation errors on either recruitment or spawning biomass estimates.

The spawner biomass - recruitment relationship is fitted according to the model;

$$R_i = \alpha Y_i . \exp(-\beta Y_i + u_i)$$

$$u_i \approx N(0, \sigma_u^2)$$

Assuming that

$$R_i = r_i$$

$$Y_i = y_i$$

then the maximum likelihood of α and β is found by minimising the sum of squares of residuals, $\sum_{i=1}^{n} u_i^2$, which is given by;

$$\sum_{i=1}^{n} [\ln R_i - \ln(\alpha Y_i) + \beta Y_i]$$

An analytical result for this can be found by finding the roots of the derivatives of this sum with respect to α and β. The result is that, summing everywhere for i=1 to n;

$$\beta = [\Sigma Y_i \ln R_i / \Sigma Y_i - \Sigma Y_i \ln Y_i / \Sigma Y_i - \Sigma \ln R_i / n + \Sigma \ln Y_i / n] / [\quad]$$

$$[\quad] = \Sigma Y_i / n - \Sigma Y_i^2 / \Sigma Y_i$$

$$\alpha = \exp[(\Sigma \ln R_i - \Sigma \ln Y_i + \beta \Sigma Y_i) / n]$$

The α and β values calculated in this way give estimates of the expected value of $\ln R_i$, $E[\ln R_i]$, as a function of Y_i;

$$E[\ln R_i] = \ln(\alpha Y_i) - \beta Y_i$$

From the point of view of MSY estimation, one is really interested in the expected value of R_i, $E[R_i]$, since this will be the average recruitment produced by the stock, and this should be used to obtain the average MSY of the stock. For a log-normally distributed random variable this is given by

$$E[R_i]=\exp(E[\ln(R_i)]+\sigma_u^2/2)$$

for sample sizes approaching infinity.

σ_u^2 can be estimated from the residuals obtained from the least squares fitting procedure, correcting for the reduced degrees of freedom;

$$\sigma_u^2=\sum_{i=1}^{n} u_i^2/(n-3)$$

An unbiassed estimate of the expectation of α is;

$$\alpha = \alpha g_q(\sigma_u^2/2)$$

where q are the degrees of freedom and the function g_q was first derived by Finney (1951);

$$g_q(t) = \sum_{p=0}^{\infty} \frac{q^p(q+2p)}{q(q+2)\ldots(q+2p)} [q/(q+1)]^p t^p/p!$$

For ease of implementation, the asymptotic correction factor was used as an ad hoc correction for α, and not Finney's (1951) result; the estimated value of σ_u^2 was used to obtain an estimate of the expectation of α;

$$\alpha \rightarrow \alpha\exp(\sigma_u^2/2).$$

For sample sizes greater than n=15, and σ^2 less than 1, this correction factor is positively biassed by at most 4% (Bradu and Mundlak, 1970).

Appendix 7.2. The computation of MSY, Y_{MSY} and K given the parameters defining the stock - recruit function

Given α and β, the equilibrium spawner biomass, Y_{eq}, is found by solving

$$Y_{eq} = R_{eq} \sum_{j=a}^{m} w_j \exp(-M.(j-1))$$

w_j is the beginning of year individual fish mass at age j

and

$$R_{eq} = \alpha Y_{eq} \exp(-\beta Y_{eq})$$

for Y_{eq}, where a is the value of j at sexual maturity; j=1 represents fish aged 0 to 1 years - m is the oldest age class, therefore m is the oldest effective age. The result is

$$Y_{eq} = \ln[\alpha/(\sum_{j=a}^{m} w_j \exp(-M.(j-1)))]/\beta$$

For a given value of F_i, which is the same for all years, the equilibrium spawner biomass for that particular level of fishing is given by a similar relationship;

$$Y_{eq} = \ln[\alpha/(**)]/\beta$$

where

$$(**) = \sum_{j=a}^{m} w_j \exp(-M(j-1) - \sum_{k=1}^{j-1} FS_k)$$

S_k is the selectivity at age k. It follows that equilibrium recruitment for $F_i = F$ for all years is given by

$$R_{eq} = \ln[\alpha(**)]/\beta/(**)$$

Catch masses are calculated by the catch equation;

$$C = \sum_{j=1}^{m} w_j N_j FS_j (1-\exp(-FS_j-M))/(FS_j+M)$$

where $N_1=R_{eq}$

$N_j=N_{j-1}.\exp(-M-FS_{j-1})$

w_j are midseason fish masses .

To estimate MSY, F is increased incrementally until the maximum catch is reached at F_{MSY}. The spawning biomass at F_{MSY} is y_{MSY}, and K is the spawning biomass for F=0.