Spring 2001

In Support of a Rationally Managed Fishery: Age and Growth in Patagonian Toothfish (*Dissostichus eleginoides*)

Julian R. Ashford
Old Dominion University

Recommended Citation
Ashford, Julian R.. "In Support of a Rationally Managed Fishery: Age and Growth in Patagonian Toothfish (*Dissostichus eleginoides*)" (2001). Doctor of Philosophy (PhD), dissertation, Biological Sciences, Old Dominion University, DOI: 10.25777/7r1x-x719
https://digitalcommons.odu.edu/biology_etds/45

IN SUPPORT OF A RATIONALLY MANAGED FISHERY: AGE AND GROWTH IN PATAGONIAN TOOTHFISH (*Dissostichus eleginoides*)

by

Julian R. Ashford
B.Sc. May 1983, University of Durham (United Kingdom)
M.Sc. May 1993, University of Durham (United Kingdom)

A Dissertation submitted to the Faculty of
Old Dominion University
in Partial Fulfillment of the Requirement for the Degree of

DOCTOR OF PHILOSOPHY
ECOLOGICAL SCIENCES
OLD DOMINION UNIVERSITY
May 2001

Approved by:

Cynthia Jones (Director)

Mark Butler (Member)

Eileen Hofmann (Member)

Inigo Everson (Member)

ABSTRACT

IN SUPPORT OF A RATIONALLY MANAGED FISHERY: AGE AND GROWTH IN PATAGONIAN TOOTHFISH (*Dissostichus eleginoides*).

Julian R. Ashford
Old Dominion University, 2001
Director: Dr. Cynthia Jones

Patagonian toothfish (*Dissostichus eleginoides*) occur on the continental shelves and shelf breaks of southern South America and the Southern Ocean. Stock structure, critical to good fisheries management, can be inferred from growth differences between areas, but available growth data are compromised by inconsistencies in age estimation methods, sampling and sample sizes, and techniques used to derive estimates. I asked the scientific question: how is growth in Patagonian toothfish structured spatially within the Southern Ocean? I developed a multi-stage randomized design to sample fish caught by commercial longline, and an age estimation methodology. Because toothfish are difficult to age, I developed an ANOVA model for estimating precision and accuracy of age data relative to a standard, as the basis for a protocol for quality control of age data. The methodology was used to obtain age data from toothfish taken from the Falkland Islands and South Georgia in the South Atlantic, and the Kerguelen Islands and Heard Island in the southern Indian Ocean. I estimated von Bertalanffy growth parameters for each area, constructed models to describe rival hypotheses of stock mixing and separation between areas, and selected between the models using normal likelihood methods. The abundance of the captured population varied at a scale of c500 m (76% of variance), and between fishing days (24%). Most variation in length composition was captured at scales less than 500 m (79%). I calculated that sampling 16 10-coil lengths of line/day on 36 days of a voyage of 60 days hauling would be the optimal sampling strategy. Significant bias in age estimation was found between readers and between readings by one reader but, once accounted for, precision of age estimation remained similar between sexes; however, a validation test of the accuracy of the age estimation methodology was inconclusive.

Growth data supported the hypothesis of stock separation between the Falkland Islands and South Georgia, but not between South Georgia and Kerguelen.

ACKNOWLEDGEMENTS

The work described in this dissertation was part of a collaboration between research groups from several nations belonging to the Convention for the Conservation of Antarctic Marine Living Resources (CCAMLR). As a result, my thanks are due to a large number of people.

I would first like to thank Prof. Guy Duhamel of the Museum National d'Histoire Naturel, Prof. Carlos Moreno and Pedro Rubilar of the Universidad Austral de Chile, Zaida Young of the Chilean Instituto de Fomento Pesquero, Dr. Dick Williams of the Australian Antarctic Division, Martin Purves of the South African Sea Fisheries Research Institute, and Dr. Karl-Hermann Kock of the Bundesforschungsanstalt für Fischerei for providing large numbers of samples, and extending their generous support to the project. Dr. Peter Horn of the New Zealand National Institute of Water and Atmospheric Research kindly read otoliths for the precision study.

At the British Antarctic Survey, Dr. Inigo Everson has been a key collaborator and a considerable source of encouragement and insight, and I would like to express my special appreciation for this. Thanks also to Dr. Tony North for providing photographs, data and advice, and Ken Robinson for his expertise with the Scanning Electron Microscope. My old supervisor, Martin White, was instrumental in originally bringing the BAS and the Center for Quantitative Fisheries Ecology (CQFE) at Old Dominion University together; as all who worked with him will know, he was constantly encouraging and full of information - his loss will be felt by the many who shared time with him in the Antarctic.

At Old Dominion University, a debt of gratitude is due to the other members of the CQFE: notably Steve Bobko, Steve Wischniowski, Jolene McDowell, Emanis Dorval and Brian Wells. Dr. Eileen Hoffman of the Center for Coastal Physical Oceanography, Dr. Eric Anderson of the Economics Department, and Dr. Dayanand Naik and Dr. John Morgan of the Department of Mathematics and Statistics gave me considerably of their

time and advice. Thanks also to Dr. Mark Butler, Dr. Frank Day, and Dr. Alan Savitsky of the Biology Department for their roles both as teachers and members of my student committees.

Those near to a Ph.D. candidate probably suffer the most in the degree process, and I would like to record my deepest gratitude to Lisa Bell, who was closest to the ongoing activity, and my family, who at least had an ocean as a buffer.

Finally, I would like to thank Dr. Cynthia Jones who, as my principal teacher through the peaks and troughs of the last seven years, has never ceased to challenge, inspire and mould my way of thinking.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1. Objectives

The development of longlining techniques for fishing Patagonian toothfish (*Dissostichus eleginoides*) led to a rapid increase in exploitation rate during the early 1990s, which progressively affected fisheries around southern South America and the Southern Ocean. The level of illegal fishing was high yet the fish were considered to be long-lived and slow-growing, indicators of vulnerability to rapid over-exploitation. Efforts toward rational management of the stock by national fishing authorities and the Commission of the Convention for the Conservation of Antarctic Marine Living Resources (CCAMLR) were impeded by lack of knowledge of the basic biology of toothfish, precluding the use of an array of powerful population dynamical techniques commonly used by fisheries managers.

Rigorous analyses of growth, mortality and population age structure can be achieved if accurate age data are available. With age-based information, the effect of management choices can be simulated using dynamic pool and age-structured assessment (ASA) models. Cohort life tables and matrix models can be constructed, allowing life-history patterns and density-dependent effects in population regulation to be assessed. Age-structured models can be used to examine the effect of age-specific vital rates, allowing managers to ascertain which stages of the life history cycle are vulnerable to over-exploitation under a particular set of conditions. Moreover, age-based data can aid modelling of community interactions, allowing impacts on other species to be assessed as part of an ecosystem approach to fisheries management.

However, the effectiveness of these models depends fundamentally on the quality of the age data and fulfillment of model assumptions. Error can be introduced through unrepresentative sampling practices, and through inaccurate or imprecise age estimation.

The journal model is taken from *CCAMLR Science*

It can also be introduced by excessive immigration and emigration if the model domain does not reflect the boundaries of the population; inaccurate growth rates will in turn bias estimates of productivity used to set catch levels.

In this project, I focus on a general research question on growth in Patagonian toothfish as a vehicle for addressing the basic building blocks of a rationally managed fishery: representative sampling techniques; age estimation methodologies with estimates of error in the age data produced; delineation of population structure; and estimates of population growth that reflect spatial and within-population variation.

Research Question

How is growth of Patagonian Toothfish structured spatially within the Southern Ocean?

Goals of this dissertation

1. To establish a sampling methodology that will allow representative sampling of the catch for Patagonian toothfish and calculation of population parameters with estimates of reliability.
2. To develop a methodology for estimating age of Patagonian toothfish using otoliths, including criteria for interpreting age from otolith microstructure.
3. To develop a quantitative method for estimating and monitoring the sources of error in age data.
4. To validate age estimation criteria.
5. To select the model that best describes growth in Patagonian toothfish and its variation in space.
6. To use the model of growth to infer stock structure, and potential linkages between fishing areas.

These objectives are addressed in four sections. Questions these sections address are:

1. How can a representative sample of the catch population be obtained?

2. How can age of individual fish be estimated, and how precise are these estimates?

3. Are estimates of age accurate?

4. Does growth vary between fishing areas, and is the pattern of growth consistent with current hypotheses of stock structure for Patagonian toothfish?

1.2. Patagonian Toothfish

The Patagonian toothfish (*Dissostichus eleginoides*) is a large-semi-pelagic predator, belonging to the family *Nototheniidae* (Figure 1.1). It occurs on the shelf and shelf-slope off islands and banks in the Southern Atlantic, Indian and Pacific Oceans, notably within the influence of the Antarctic Circumpolar Current (Yukhov 1972, Skora and Sozinski 1983, SC-CAMLR-XIV Annex 5). In the South Atlantic, it occurs as far south as the South Sandwich Islands (Ashford 1993); its congenor, *Dissostichus mawsoni*, occurs further south. *D. eleginoides* also occurs off the South American coast from Peru to Cape Horn and north off Argentine Patagonia.

Fishing began in the 1970s off northern Chile and in the South Atlantic and South Indian Oceans. Initially, only bottom trawling was used over continental shelves, and fishing rates by catch were low and mostly incidental with some directed fishing. Long-lining for toothfish was developed in northern Chile during the 1980s; long-lines were probably first used at South Georgia in 1986 and off the Kerguelen and Crozet Islands in the southern Indian Ocean in 1992 (SC- CAMLR XI). Longlining allowed access to deeper waters along the shelf-break and gave better returns: exploitation of toothfish grew rapidly. New fisheries were developed off southern Chile, Argentina and the Falkland Islands; and off Heard, Macquarie and the Prince Edward Islands (Figure 1.2). The main catching nations have been Chile and Russia, and more recently Argentina and Norway, but vessels registered in diverse countries such as Panama and Vanuatu have been involved (SC-CAMLR-XVI/4). The toothfish is highly prized in Japan, selling at US$6.00/kg in mid-1995, and is now being targetted at the United States market (Seafood

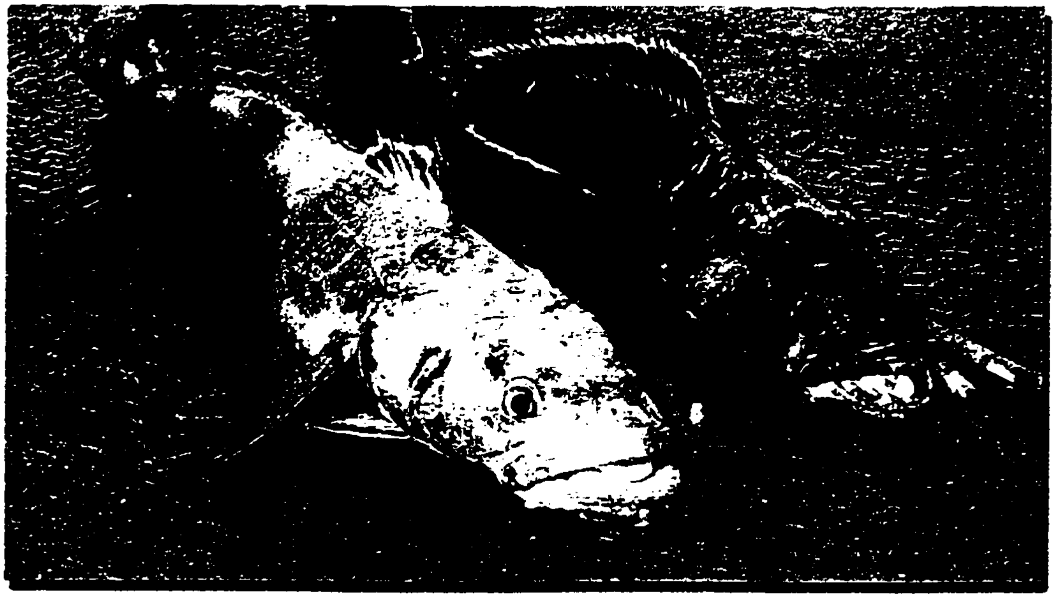


Figure 1.1: Patagonian toothfish (*Dissostichus eleginoides*)

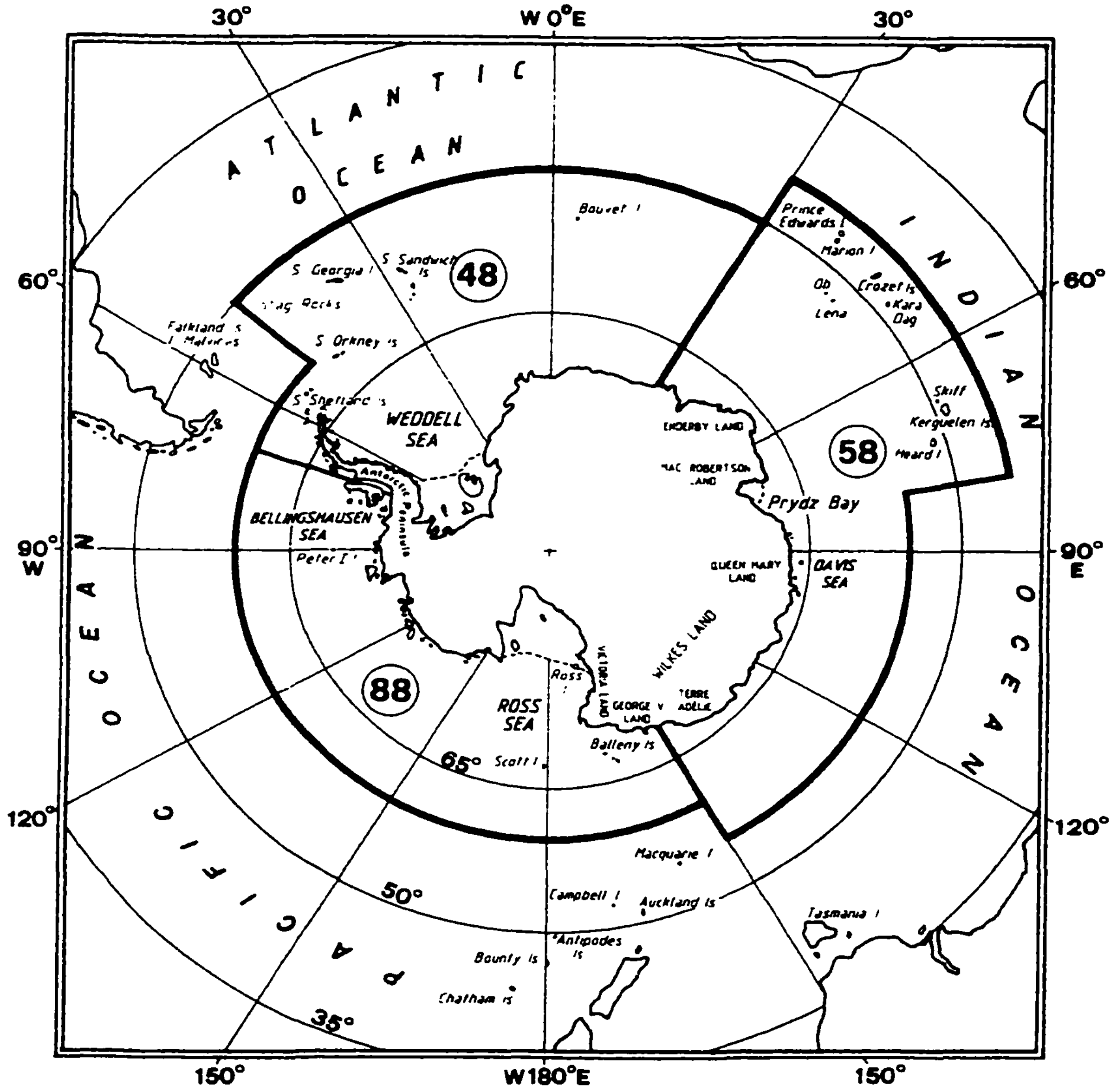


Figure 1.2: Map of the Southern Ocean showing the peri-Antarctic islands in the Atlantic (48), Indian (58), and Pacific (88) sectors. From Fischer and Hureau (1985).

International - September 1995). Total catch in 1995 around South Georgia and the adjacent Rhine and North Banks alone was estimated at 6171.1 tonnes (SC-CAMLR XIV, Annex 5). In 1997, the catch for the CCAMLR area was estimated at 107,000-115,000 tonnes, of which 70% was estimated to have been caught illegally (SC-CAMLR-XVI/4).

Patagonian toothfish grow to over 2m long, and are thought to live upwards of 20 years (Hureau and Ozouf-Costaz 1980). Absolute fecundity is low at between 48,000 - 528,900 eggs/individual over the life history, although relatively high for a notothenid (Chikov and Melnikov 1990, Kock et al.1985). Spawning is annual, occurring between June and August (SC-CAMLR XI). Size of first spawning is ca. 80 cm for males and ca. 100 cm for females. Little is known of their early life history: eggs are considered to be pelagic and have been found occasionally in the water column (SC-CAMLR-XIV, Annex 5). Larvae and post-larvae have been found only infrequently, and generally within the upper 50 m of the water column (Yefremenko 1979). Juveniles are found on the continental shelf at ca.500 m, with increasingly large individuals along the continental slope in deeper waters. Maximum depth at which fish have been caught is ca. 2900 m (SC-CAMLR-XIV, Annex 5).

Patagonian toothfish display considerable flexibility in feeding patterns. Diet varies between regions and with life stage and depth. Juveniles largely eat krill (*Euphausia superba*) in Antarctic regions, while adults feed on cephalopods and other fish (e.g. *Champsocephalus gunnari*) feeding on krill (Konforkin and Kozlov 1992), suggesting an ability to move large distances from the continental shelf. This evidence is corroborated by their regular occurrence in sperm whale stomachs in pelagic waters of the Southern Ocean (Yukhov 1972). Diet off Chile is almost exclusively of fish, while off South Georgia benthic crustaceans are present in a large proportion of full stomachs from fish taken in water deeper than 1000m (Zhivov and Krivoruchko 1990). The digestive tract is adapted for ingesting large items and large amounts of food at a time (Korovina et al 1988) while the proportion of full stomachs varies between regions and with depth.

Little is known of population structure. Variations in published growth parameters suggest a population boundary between the southern South America - Scotia Arc region and the islands of the Southern Indian Ocean (SC-CAMLR-XI). This evidence is supported by increasing levels of a disease causing degeneration of protein (jelly meat) off southern Chile, which has also appeared off South Georgia but has not been recorded in fish from the Indian Ocean or northern Chile. Parasite loadings indicate some isolation between fish populations found along the north and south coasts of Chile but also between the Patagonian shelf and South Georgia (SC-CAMLR-XIV Annex 5). This evidence may however be related more to differences in the environment encountered by older life stages than to different sources of population recruitment. Mark- recapture studies have commenced off the Falkland Islands but returns have yet to start in large numbers (J. Barton, pers. comm.); mark-recapture studies along the Scotia Arc are problematic due to the poor condition of fish captured by long-line (Ashford 1993, 1994), although studies have been initiated using fish taken by trawl. Genetic studies have been attempted but have either not been successful due to the high lipid content of the tissue (R. Williams, pers. comm.), or have not yet been published. Stock depletion experiments have indicated that movement of fish between areas may be large over periods as short as 10 days (SC-CAMLR XIII). Length frequency data indicate five consistent peaks in the early life history which may correspond to year cohorts; however, little age-based data exists and no validation studies of aging techniques have yet been published in the literature.

1.3. Area of interest

The Southern Ocean is a circumpolar ocean, connecting the southern Atlantic, Indian and Pacific Oceans. It is a major site of production of several water masses, including Antarctic Surface Water, Antarctic Intermediate Water and Antarctic Bottom Water, which spread out of the Antarctic regions to lower latitudes. Mixing with North Atlantic Deep Water creates Circumpolar Deep Water, which may form the source of the Common Water of the Indian and Pacific Oceans (Gordon 1988). The Southern Ocean is thus marked by considerable heterogeneity of water masses both vertically and horizontally, with clear differentiation of water types so close to their origin.

Within the Southern Ocean, the Antarctic Circumpolar Current (ACC) moves eastward around the Antarctic continent (Gordon 1988, Anon 1989). After the ACC passes through the Drake Passage, the current moves north along the Patagonian shelf to meet the Brazil Current, after which it travels eastward across the Atlantic. In the Indian Ocean, the current passes along the Crozet and Kerguelen Island groups, to meet the Macquarie Ridge before splitting into two off the Campbell Plateau south of New Zealand. Extensive mixing with Pacific Waters occurs until the current returns to the Drake Passage (Gordon 1988). Much of the transport occurs within three fronts (polar, Antarctic and continental) which mark large changes in temperature and salinity (Figure 1.3). These fronts shift over time and have large implications for organisms living in the Southern Ocean (e.g. Hofmann et al. 1998).

1.4. Stock assessment and management

In the Southern Ocean, fishing has been largely within the area managed by the Commission of the Convention for the Conservation of Antarctic Marine Living Resources (CCAMLR). The Commission is composed of delegations from signatory nations to the Convention, and its headquarters are located in Hobart, Australia. Data from the toothfish fishery are collected, and analyzed each year by the Working Group for Fish Stock Assessment in preparation for the annual meeting of the Commission and its Scientific Committee. Decisions taken at the meetings control Antarctic fishing and promote marine research. The objective of the Convention is the conservation of Antarctic fisheries including rational use based on an ecosystem approach to management.

Stock assessment is concerned with advising decision-makers on the biological effects of different possible harvest actions (Gulland 1983). The role of stock assessment

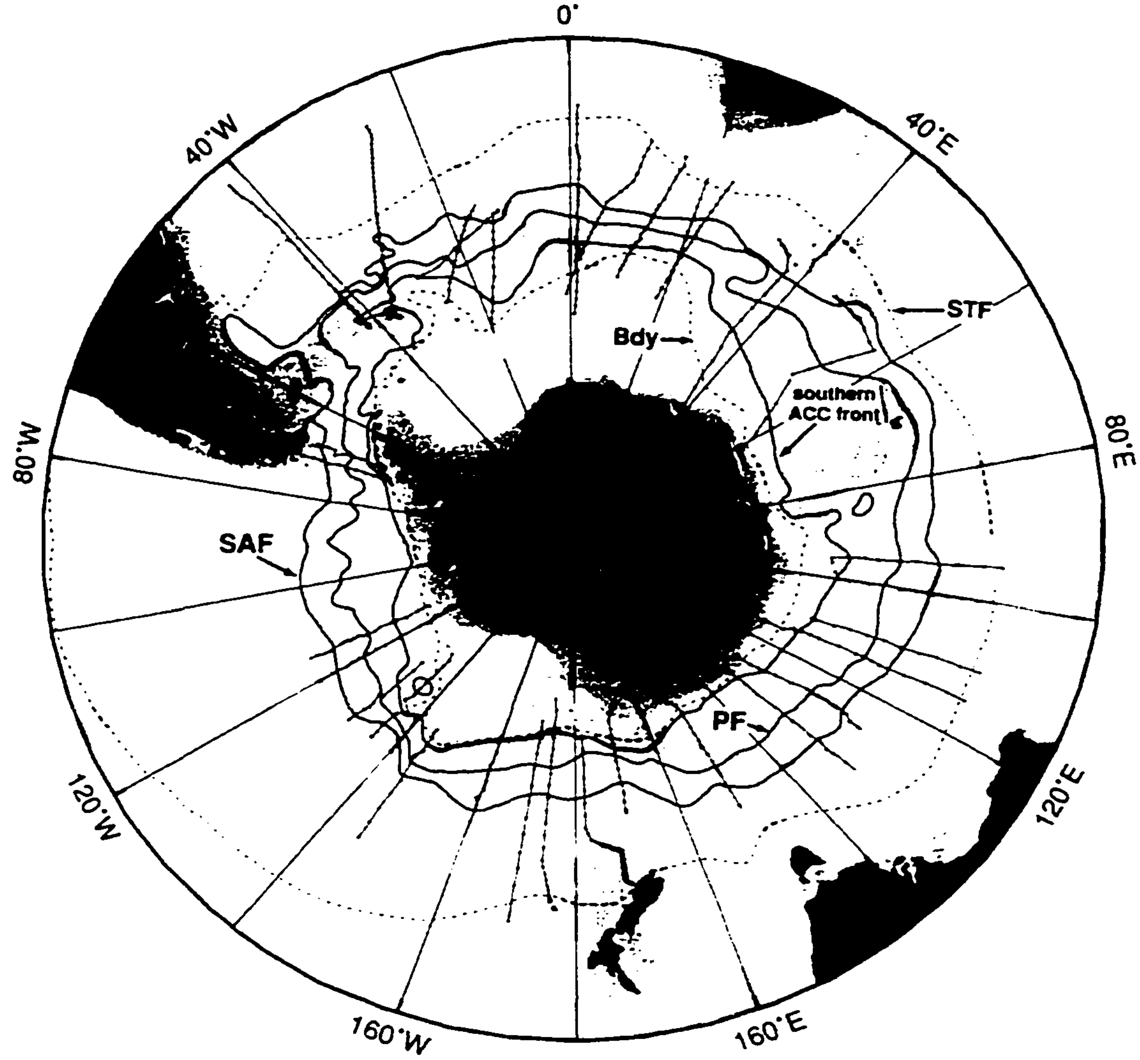


Figure 1.3: Map of the Southern Ocean showing the position of fronts in the Antarctic Circumpolar Current. SAF = Subantarctic Front, STF = Subtropical Front, PF = Polar Front, southern boundary of the ACC is shown as a dashed line (Orsi et al. 1995).

biologists in fisheries management is to provide structured analyses of available data in comparing management choices (Hilborn and Walters 1992), based on the dynamics of the exploited populations. However, their responsibility is rarely to decide between options; instead, practical decisions are taken by fisheries managers after balancing biological, social, economic and political factors (Gulland 1983). Stock assessment is therefore a special, applied case of the general problem of predation and interaction between populations wherein humans are the predators (Beverton and Holt 1954).

Beverton and Holt (1957) developed a simple population model as a starting point for their quantitative approach to stock assessment. They considered a closed population put forward by Russell (1931) in which weight increments, in the form of recruitment and growth, were balanced by weight decrements, in the form of capture through fishing and natural death:

$$S_{i+1} = S_i + (A + G) - (M + C) \qquad (1)$$

where S_i is the total weight of the exploited phase of the population at time $i = 1$ and 2, A is recruitment, G is growth, C is fishing capture, and M is natural death over the period. They eliminated the interdependence between the four primary factors by defining each in terms of parameters assumed to be independent of population density, and proceeded to construct a simple analytical population model:

$$Y / R = Fe^{-M(t_c - t_r)} W_\infty \frac{\sum U_n e^{-nk(t_c - t_0)}}{F + M + nK} \qquad (2)$$

where Y/R = yield-per-recruit; F = instantaneous fishing mortality coefficient; M = instantaneous natural mortality coefficient; W_∞ = asymptotic weight; U_n = summation parameter; t_c = mean age at first capture; t_r = mean age at recruitment to the fishing area; t_0 = hypothetical age when length is zero; K = Brody growth coefficient. Similar dynamic pool models were derived by Thompson and Bell (1934) and Ricker (1958), but using different methods to estimate W_∞ and mortality, and calculating yields over small time intervals which were then summed. Using these methods, the fishing mortality can be calculated at which the maximum sustainable yield (MSY) is achieved, and the effect of different management scenarios simulated.

An alternative approach is to use the logistic growth equation to model the exploited population. Assuming catch to be proportional to fishing effort and stock size, Schaefer (1954) proposed that the change in biomass would equal population production less exploitation:

$$\frac{dB}{dt} = rB(1 - \frac{B}{k}) - qEB \tag{3}$$

where B is the biomass of the stock, r is the intrinsic rate of population growth, E is the fishing effort, and q is a parameter describing the efficiency of the fishing gear. The term $r(1\text{-}B/K)$ is therefore the rate of surplus production, or yield, which under equilibrium conditions equals the rate of exploitation. The model can be shown to follow the form of a parabola, so that Y_{MSY} and B_{MSY} occurs at the point where $dY/dB = 0$.

A third method is to use age-structured stock assessment (ASA) models to simulate the dynamics of individual cohorts through time. ASA models now form the basis for management advice in many world fisheries (Megrey 1989), linking data on age composition to catch rates, and allowing reconstruction of population dynamics. Age-specific vital rates can be estimated, as well as abundances of individual cohorts. An example is Gulland's (1965) sequential population assessment (SPA) model, in which he used a backward solution to link successive age-groups, combining the Baranov (1918) catch equation:

$$C_i = \frac{F_i}{F_i + M} N_i (1 - e^{-(F_i + M)}) \tag{4}$$

where C_i = catch of fish age t, and N_i = the population at the beginning of the year, and the exponential survival model:

$$N_{i+1} = N_i e^{-(F_i+M)} \qquad (5)$$

ASA models can be applied without knowing fishing effort, catchability, or gear selectivity, and avoid the problems of catch per unit effort (CPUE) as an index of population abundance (Megrey 1989). ASA models provide estimates of absolute population abundances and can predict the contribution of individual cohorts to the fishery, so are particularly appropriate where quotas are used to control fishing effort and year-class strength is variable between years.

To use the models, data are collected which can be fed directly into the models, or from which the terms can be calculated. Mortality can be calculated using the catch curve, based on the exponential survival model above where the total mortality (Z) is equal to the sum of fishing (F) and natural mortality (M). Alternatively, survival can be calculated using maximum age methods (Royce 1972, Hoenig 1983), or removal methods (Lesley and Davis 1939, Delury 1947). Recruitment to the fishery can also be calculated using fomulations of the model, back-calculating to find N_0, the total number of a cohort when it enters the exploited population. Growth can be modelled using the von Bertalanffy model (Beverton and Holt 1957). Thus, length at time t is :

$$l_t = L_\infty(1 - e^{-K(t-t_0)}) \qquad (6)$$

where L_∞ is the mean maximum size achieved by a fish, K is the rate at which L_∞ is achieved, and t_0 is the hypothetical age at which length is 0. Other models are also available to simulate growth (e.g. Schnute 1981).

However, with the exception of surplus production models, these methods all rely on accurate and representative information on age. Furthermore, mean age at first capture and at recruitment to the exploited population are normally assessed directly using age data. The models also assume a single, closed population with no emigration or immigration. Dynamic pool and surplus production models assume equilibrium conditions, while most ASA models are deterministic and are not well structured to assess stochastic effects. Yet these conditions are seldom met in practise: biometric data is subject to varying levels of measurement and sampling error, particularly for age data; movement within and between populations, and environmental influences, lead to variability at different temporal and spatial scales.

Assessment of this variability in biological systems complicates the process of reaching valid conclusions. Fisher (eg. 1922) and others developed several statistical frameworks for estimating parameters and their variability or error in populations. They further developed procedures to make inferences, including measures of belief in estimates using confidence intervals and tests of hypotheses based on these estimates, which form a useful tool for biologists and fisheries managers. In one of these approaches, a scientific hypothesis is advanced, for which models are developed which represent null and alternate statistical hypotheses. The model parameters are estimated by least squares and an experiment is designed with a treatment that will test between the hypotheses. The null (or reduced) model describes the situation where the treatment results in no detectable difference between treated and untreated populations; the alternate (or full) model where the treatment results in a detectable difference. Random allocation of experimental units to treatments allows a population of possible experiments to be generated under the null hypothesis, of which one has actually occurred. A probability analogy can therefore be employed (Edwards 1992): the value for the test statistic for the actual experiment can be compared against those for the population in a probabilistic randomization test (Keuhl 1995). Alternatively, significance tests based on normal distribution theory are equally good provided the normal distribution assumption is valid. The method gives a measure of the probability that the result would happen under the null hypothesis, and therefore whether the latter should be rejected or not rejected.

Fisher (1921) also developed the concept of likelihood as an alternative to probability as a relative measure of belief between rival hypotheses. In the probability approach, the results are considered to be the variable and the hypothesis is fixed or constant; in the likelihood approach, the data are fixed and the likelihoods of the hypotheses are the variable. Likelihood depends on a distribution function rather than the probability analogy, and has very different properties. The likelihood for each hypothesis is calculated, and the ratio between the two determines which of the hypotheses is selected. Thus, likelihood allows different types of information to be included in a single framework, and the relative merits of rival hypotheses can be assessed in the light of observational or experimental data that bear upon them (Edwards 1992).

In fish populations, age error can result from age estimation techniques (eg. Beamish and McFarlane 1987); it can also result from unrepresentative sampling. The latter can be minimized by using probabilistic sampling designs with good spatial coverage (e.g. Dorval 1998). Residual error can be assessed to prevent age biases, to estimate random variability using a quality control program, and to evaluate the reliability of model results and population parameter estimates (Kimura and Lyons 1991).

Error can also result if the model domain does not reflect the boundaries of the population and the population does not behave homogeneously within the boundaries. Inaccurate growth rates will bias estimates of productivity used to set catch levels. Linkages between populations can profoundly influence their dynamics, altering parameter values and even the persistence of populations in time and space under different conditions (e.g. Pulliam 1988, Hanski and Thomas 1994). Environment is also likely to contribute to spatial and temporal variation, structuring the space in which a population exists, linkages between populations, and contributing stochastic effects at different scales. For example, temperature and resource availability influence growth, and may vary widely within and between populations of a single species. As a result, hypotheses on movement and mixing need to be constructed and assessed, and growth estimated at an appropriate spatial scale.

With accurate information on age, growth and population structure, managers can construct stock assessment models to estimate yield, ascertain the life history pattern of a species, and pinpoint which age classes are vulnerable to over-exploitation under a particular set of environmental conditions. As importantly, models can be assessed for how well they handle the error detected in the data and how assumptions conform to the characteristics of the population. This can be balanced against data availability, cost and logistical demands to select the most appropriate model to simulate management choices in the fishery.

As a result of the increase in fishing activity within the CCAMLR area during the early 1990s, a workshop was convened in 1995 to review methods used for the assessment of *D. eleginodes*. The workshop considered the current state of knowledge of toothfish biology and demography; stock identity; abundance; and yield. It recommended that research programs be developed to determine absolute estimates of abundance based on both fisheries dependent and independent data; to estimate total removals; and improve estimates of biological and demographic parameters. For the latter, the workshop considered the following to have high priority: determination of age distribution, including validation of age estimations from otoliths and scales, and the magnitude of age biases. Secondly, stock identity and mobility should be determined; and thirdly, estimates of von Bertalanffy growth parameters appropriate for yield calculations given size selectivity of different fishing methods.

1.5. Method: hypothesis-testing and its conceptual context.

In this dissertation, I follow a hypothesis-testing approach using probability-based and maximum likelihood techniques to estimate parameters and make inferences. The logical structure used in a line of investigation needs to be clear to allow evaluation of the influence of the conceptual context on the conclusions reached (Underwood 1997). I therefore review the conceptual background underpinning my approach.

In the traditional model of science, scientific development is viewed as a piecemeal process through which the constellation of facts, theories, and methods are added to the sum of scientific knowledge (Kuhn 1962). The slow, incremental gathering of empirical observations increases rational understanding, 'uncluttered and unbiased by theoretical preferences rooted in social constraints' (Gould 2000). The process 'begins in superstitious ignorance and moves toward final truth by successive accumulation of facts' (Gould 1977).

This model of science as an objective process was articulated by Bacon in 1620 and expressed more recently in the logical positivist school of thought with its emphasis on verification to prove hypotheses (Magee 1997). Pearson (1900) advocated a statistical methodology and view of nature which he saw as consistent with this model. He attempted to clarify how a scientist uses statistical techniques to construct a world view from the raw material of the senses. Facts are classified and expressed as scientific laws which are then tested against more facts to make sure they are correct. By successive approximations, scientists approach the fundamental laws of nature (Kingsland 1995).

However, the traditional model of science has not stood up to close scrutiny over time. Observations are governed by previously held cognitive structures, which limit the observer in ways that are largely unrecognised (Underwood 1997). Studies in psychology have shown the importance of prior experience and context in perception, for example in the perception of illusions between cultures, and in altering the physiology underlying visual perception (e.g. Gregory 1977). Gestalt psychologists have shown that the same information can be constructed in different ways, for example in seeing the same diagram of a cube in more than one orientation.

Evidence from other disciplines is inconsistent with the model. In anthropology and ethnobotany, a wide variety of incompatible knowledge systems have been documented which nevertheless allow different societies to interact successfully with the physical world (eg. Cox 2000). Historians of science have shown that contemporary social and political forces play a large role in influencing the scientific process, and that contemporary theory influences the interpretation of observations (Gould 1977). Furthermore, science has included beliefs incompatible with those held today, and inconsistent with a view of science as a process of accretion (Kuhn 1962). Indeed, the realist view was linked to the emergence of European science from its Renaissance and religious precursors as a framework for understanding the physical world. The realist emphasis on the power of objective observation was understandable in countering belief in textual authority, yet Gould (2000) showed that Bacon himself recognized the mental and social obstacles to an objective form of scholarship.

Popper (1959) also attacked the logic of the model. Proof requires that every possible observation is available; otherwise, it requires the inductive argument that unavailable observations can be inferred from the cases actually available. This is usually not possible or practicable: contrary observations may simply not yet have been sampled and the truth of the proposition cannot therefore be regarded as proven (Magee 1997, Underwood 1997). Similarly, reports from the past do not provide a reason to suppose that any one event will occur in the future rather than any other event (Miller 1999).

On the other hand, a social constructionist view of science as one belief system among others ignores the technical efficacy of scientific achievement (Gould 2000). Instead, Popper (1959) argued that hypotheses cannot be proven, but only disproved. A hypothesis is tested repeatedly: if it stands up to the attempts at disproof, it acquires a degree of respect, always with the reservation that it might later be disproved. Hilborn and Mangel (1997) have pointed out that, in contrast to Pearson's ideas, the classical statistical approaches developed by Fisher and others provide a quantitative underpinning for Popper's system of falsification. Popper further argued the importance of falsification in choosing between competing theories (Magee 1997), and Edwards (1992) noted the consistency of this approach with likelihood methods. A similar view of science as a process of competition was developed by Kuhn (1962): in this, the collection of data occurs within a paradigm, or view of how the world works. During periods of normal science, inconsistencies mount between the data and predictions from the incumbent paradigm. The growing crisis weakens the paradigm, generating the conditions whereby a challenger can displace the incumbent in a scientific revolution, or paradigm shift (Ashford 1999). Knowledge is then re-structured in a gestalt manner. The importance of competition between rival hypotheses in ecology has been emphasized by Hilborn and Mangel (1997), who argued that likelihood methods provide a statistical basis for evaluating this competition in confrontation with data.

In contrast to the realist view, Popper and Kuhn's ideas demand that scientific knowledge be regarded sceptically, with the possibility any hypothesis may later be disproved, no matter how well tested. Popper conceptualizes a 'third kingdom' between those of the subjective and objective: information is created but can be treated independently from its authors through the material products of publication and criticism (Magee 1973). Knowledge attains a physical form which can accumulate through archiving, but follows conceptual and logical rules distinct from the objective world it describes. The sceptical approach is distinct from relativism, which denies an objective world about which true and false statements can be made (Miller 1999). Scepticism denies only that we can ever know for certain that a statement is absolutely true or false. Therefore the claim of science on rationality lies in the critical search for errors when constructing hypotheses, and not from any justification or proof of hypotheses. The approach therefore emphasizes how the development of scientific knowledge is constrained by the process of empirical criticism by which hypotheses are tested.

The history of population ecology has been marked by controversy between quantitative theoreticians and empirical field biologists, notably in the debate over population regulation at the Cold Stream Harbor Symposium in 1957. Kingsland (1995) argued that this can be characterized as an Hegelian dialectic, with a synthesis in the subsequent work of Hutchinson and MacArthur. Thus the quantitative approach developed by MacArthur depended on a dialectical relationship between experiment and the formal construction of theory, in which 'insights obtained from the field should be extended through mathematical analysis in order to generate new predictions and therefore new field research.' Hilborn and Mangel (1997) characterized this process of epistemological development in quantitative ecology as a competitive confrontation between hypotheses and the data. In their view, generalized models are used to generate quantitative hypotheses that can be tested. Models are used to define and represent hypotheses that are a general statement about the natural world. The models are then used to evaluate the hypotheses against each other, subjecting them to tests through which one of the competing models can be selected.

1.6. This study

This dissertation is a response to the recommendations of the workshop on methods for the assessment of *Dissostichus eleginoides* to examine age distributions including validation and analysis of error in age estimation, to determine population structure, and to estimate growth parameters. A proposal for the project was reviewed by an *ad hoc* group of scientists from CCAMLR nations involved in *Dissostichus* research during the 1996 meeting of the CCAMLR Working Group on Fish Stock Assessment (WG-FSA). Response to the proposal was favorable and supportive, the project being viewed as 'ambitious and very well planned'. In-kind support in the form of otolith collections were recommended to be made available from the sampling programs at Heard Island, Kerguelen, Marion and Prince Edward Islands, and South Georgia. The group welcomed the initiative, suggesting that the project provided a valuable focus for identifying researchers and exchanging information on the availability of material. The project was recommended to the WG-FSA as being 'well organised and highly relevant to the needs of the working group'.

I present results from the project in this dissertation. In line with Underwood's (1997) appeal for clarity in the logical structure underpinning a line of investigation, I explicitly adopt a sceptical view of rationality and scientific knowledge, viewing the scientific process as a formalized procedure for hypothesis-testing by creating scientific and statistical hypotheses and attempting to disprove them in the light of data, either individually or in a process of competition between rival hypotheses. I emphasize the relationship between theoretical development and empirical test by reviewing the existing body of knowledge, using this to generate hypotheses capable of disproof, and using the results of each chapter to develop theory further and suggest new questions that can be examined.

I follow this procedure to examine how growth in Patagonian toothfish is structured through the Southern Ocean and off the southern South American shelf, and how variability in growth may be used to infer population structure. In addressing these questions, growth should be modelled using representative age data. Consequently, I develop a representative sampling design and a method for estimating age and its error which can be used for quality control; I also attempt a test of a conceptual model of annulus deposition in otoliths. As well as addressing a fundamental question on spatial variation in growth, I therefore address the basic methodological prerequisites for simulating the population dynamics of Patagonian toothfish, needs specifically identified by the Scientific Committee of CCAMLR.

CHAPTER II

SAMPLING THE LONGLINE CATCH POPULATION

2.1. Introduction

2.1.1. Sampling the longline: the problem

Longlining is a passive method of fishing, relying on active movement by fish to the gear. The fish are attracted by chemicals from the bait carried to them by the current and diffusion; response is by rheotaxis, approaching the gear by swimming upstream. The chemicals form a plume downstream of the longline: the active space of the gear is the volume of water influenced by the plume (Bjordal and Lokkeborg 1996). This is determined by physical variables such as strength and direction of the current, turbulence, physical topography creating a barrier or impeding flow through friction, and the rates of release and diffusion of chemicals from the bait (Olsen and Laevestu 1983). These variables change over time, affecting the active space: for instance, Lokkeborg (1990) demonstrated the rate of release of amino acids from mackerel bait decreased dramatically in the first two hours of soak time.

Physiological, ecological and behavioural factors specific to the target species are also important: the olfactory threshold of the fish limits the effective size of the active space, and the response rate is affected by feeding history and foraging strategy (Bjordal and Lokkeborg 1996). Population density and structure may also influence response rates. For instance, large fish are able to swim further and faster than small, but are likely to gauge their response to chemical information on the size of the bait and its distance, the likelihood of the bait remaining available if population density is high, and the availability of alternative food sources.

Studies from areas outside the Antarctic indicate that changes between lines or along a line may determine which fish are taken. Matsuoka *et al.* (1992, 1995) found longer submersion times can influence the catch through higher rates of fish loss or saturation of baited hooks. Somerton and Kikkawa (1995) found an abundant target species may saturate the gear, or other species, if in abundance, may exclude the target species, leading to differences in selectivity where the distribution is clumped. Lokkeborg and Bjordal (1992) found behavioural mechanisms may be involved in size selectivity: larger specimens may have larger feeding ranges or territories, and be more successful in competing for baits. Depletion studies have indicated that toothfish move widely within a fishing area and distribution may be clumped (Rubilar et al. 1994).

Fishing success is therefore dependent on a suite of physical and biological variables that are likely to vary considerably over time and space. In the toothfish fishery, setting usually takes 1-2 hrs for a line of 10,200 hooks (eg. Ashford 1993), and hauling 10-28 hrs; a single line can cover 10 km of the continental shelf-break. Different points of a longline are likely to be exposed to very different sets of physical variables over different windows of time; soak time will increase greatly for the last part of the line hauled compared to the first. If not carefully conducted, samples may not be representative of the catch population, but give estimates that are biased or, if variability is large, estimates that are unbiased but can result in a single sample that differs considerably from the true catch population parameter. The CCAMLR Working Group for Fish Stock Assessment has requested samples be taken representatively (SC-CAMLR-XV/4, Anon 1996b). However, no standardised method has been developed, and most sampling has been undertaken haphazardly or systematically, despite the warning by Kock and Selling (1996) over the interpretation of data collected for longlines in this manner.

2.1.2. The longline system of fishing

Longlines and their terminology vary widely between vessels and fisheries, but essentially consist of a long main line (also called groundline or motherline) to which are attached shorter branch lines (secondary line, gangline or snood) that carry a baited hook at the free end. Secondary lines are spaced at regular intervals specific to the fishery. The practical unit in rigging and deploying longlines is the skate (coil or basket), a length of mainline rigged with a standard number of secondary lines. Skates are linked to make a fleet of longline gear, which is deployed, or set, from the stern of the vessel as a single longline. The line is left to soak before being hauled using a hydraulic winch, usually into the wind and over the starboard side forward of the centre of the vessel. During hauling, the fish are removed from the hooks, bled, dressed and iced or frozen; skates are repaired, coiled in tubs (or baskets) and the hooks rebaited in preparation for the next set (Bjordal and Lokkeborg 1996).

Longlines can be deployed demersally, pelagically or semi-pelagically. In demersal longlines, the ends of the line are usually secured by anchors and marked by surface buoys carrying lights and radar reflectors, and the line is weighted between skates. Intermediate buoys may be deployed so that, if a line breaks, the vessel can resume hauling at the next buoy: this saves time in steaming to the far end of the line and allows the vessel to continue hauling into the wind.

The toothfish fishery currently uses three types of demersal longline gear. The Spanish system (Figure 2.1a) uses a series of several hundred coils of line joined together to form one continuous mother line. The mother line is thin and a strong retaining line is deployed in parallel and anchored at both ends: the two lines are linked by connecting-lines at the end of each coil. The Japanese system (Figure 2.1b) does not use a retaining line, only a stronger mother-line with a series of two or more surface buoys along the length of the line. The Norwegian system is automated and uses a continuous mother-line with hooks spaced at regular intervals all along (Figure 2.1c). The line is divided into magazines consisting of a number of coils, and weights are placed at the end of each section of line corresponding to a magazine (Ashford et al. 1997).

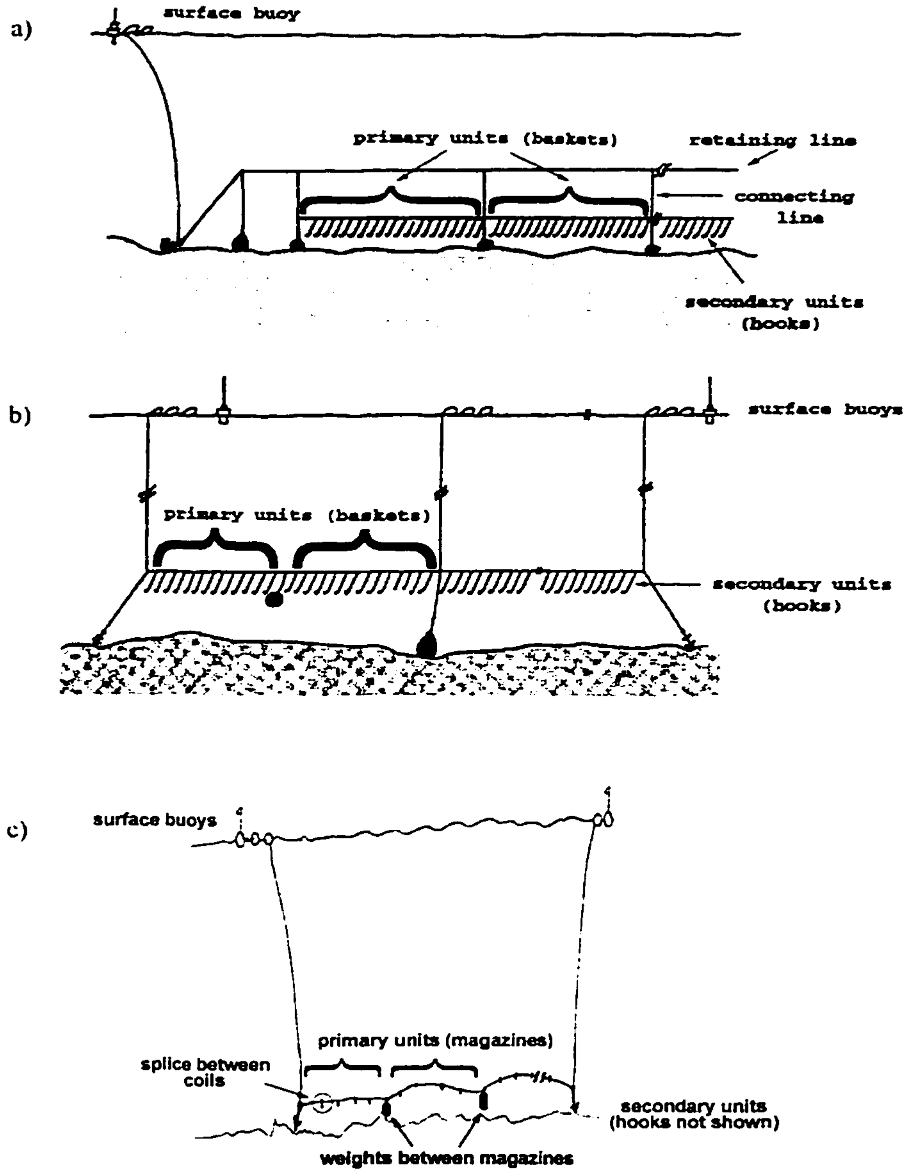


Figure 2.1: Demersal longline equipment used in the fishery for Patagonian toothfish: a) the Spanish manual system, b) the Japanese manual system, c) the b) Norwegian automated system.

2.1.3. How to sample representatively?

To obtain a representative sample of a population, data can be taken by selecting sampling units randomly from a known list of the sampling units making up the population (the sampling frame), using a simple random sampling design. Population parameters can be estimated with estimates of their variance. Thus, for a given population of N units, the population mean μ is the average of the y-values in the whole population (Thompson 1992):

$$\mu = \frac{1}{N}\Sigma^{N} y_i \tag{7}$$

The finite population variance is given by:

$$\sigma^2 = \frac{1}{N-1}\Sigma^{N}(y_i - \mu)^2 \tag{8}$$

For a sample taken by simple random sampling, the sample mean is the average of the y-values in the sample:

$$\bar{y} = \frac{1}{n}\Sigma^{n} y_i \tag{9}$$

The sample variance s^2 is defined as:

$$s^2 = \frac{1}{n-1}\Sigma^{n}(y_i - \bar{y})^2 \tag{10}$$

If all the possible samples of the population were taken, under simple random sampling, the mean of all the sample means will equal the real mean of the population: that is, a sample mean will be a design-unbiased estimate of the population mean. The same applies for the sample variance. According to the Central Limit Theorem, the estimated means from the samples will follow a normal distribution around the true mean, with variance:

$$\mathrm{var}(\mu) = (\frac{N-n}{N})\frac{\sigma^2}{n} \qquad (11)$$

where *(N-n/N)* is the finite population correction factor. An unbiased estimator of this variance is:

$$\hat{\mathrm{var}}(\bar{y}) = (\frac{N-n}{N})\frac{s^2}{n} \qquad (12)$$

This allows inferences to be made based on the well-understood properties of a normal model, even when the *y*-values are not normally distributed. One inference that can be made is the construction of confidence intervals as a measure of the reliability of each estimate. This measures the range of values within which the population parameter is likely to lie at a given level of probability (usually 95%). Thus, for a normally distributed, unbiased estimator of a population parameter θ, a $(1-\alpha)$ confidence interval for θ is given by:

$$\hat{\theta} \pm z\sqrt{\mathrm{var}(\hat{\theta})} \qquad (13)$$

where Z is the upper $\alpha/2$ point of the normal distribution, and α is the allowable probability of error.

Larger sample size reduces the population correction factor, reducing the variance of the mean and the width of a confidence interval, indicating the estimate is more reliable. There is therefore a trade-off between the effort needed to record more data, and the increase in reliability of an estimate. If the level of reliability desired can be specified, the optimal number of samples can be calculated, ensuring that time and effort are not wasted collecting extra data which will do little to increase confidence, or too little data are collected to allow reliable conclusions. Thus, if N is large relative to n, so that the finite population correction factor can be ignored, the necessary sample size is:

$$n_0 = \frac{z^2\sigma^2}{d^2} \tag{14}$$

where d is the maximum allowable difference desired.

Probability designs like simple random sampling provide unbiased estimates of population parameters and their variability, without relying on any assumption about the population itself. They guard against regularities in the population and remove recognized and unrecognized human sources of bias - especially desirable when the results are to be used in a political context with conflicting interests, like fisheries management. Randomization also implies a finite array of designs based on the mix of the sampling units that could have been chosen. In contrast, haphazard, unrandomised systematic, and pseudo-randomised designs are susceptible to conscious or unconscious assumptions about the population by the sampler, give estimates with unknown bias, do not give estimates with known properties and a measure of reliability, and do not conform to the probability analogy. In a situation with real conflicts, conclusions derived from either method are vulnerable to attack by a dissenting view (Thompson 1992).

Simple random sampling (SRS) has the added advantage that each data point can be considered independent and identically distributed (Fisher 1922), without any auto-correlation. This fulfills a basic assumption of most statistical analyses, including those involved in Analysis of Variance (ANOVA). However, identifying a suitable frame has often proved a problem in fisheries-based sampling. An SRS solution for small catches would be to collect caught fish and label each one, creating a frame from which fish can be selected by label. However, this quickly becomes impracticable as the vessel and catch gets bigger. In the case of toothfish, hauling occurs over 10-28 hrs, and fish are processed concurrently. Simple random sampling is clearly not feasible logistically.

A hierarchical randomised sampling design is a practical alternative, and one that can reduce the variance of an estimator when the variation of the underlying population is distributed in certain ways. A frame of M sampling units is divided into N primary sampling units (PSUs). This allows a variety of approaches: in cluster sampling, n PSUs are sampled randomly from N, and all m_i sampling units in each selected PSU are sampled; in multi-stage sampling, a sub-sample is randomly selected from the the m_i units as well. In randomised systematic sampling, n PSUs are selected from a frame of N, but the m_i units within each PSU are distributed systematically through the population. Once a sampling frame is identified, a hierarchical sampling design will allow an observer to sample the catch population during hauling of a line. Furthermore, multi-stage or cluster sampling will allow the catch population to be divided into PSUs of a practicable size for sampling during sampling sessions.

However, toothfish are not captured predictably, so identifying a hierarchical frame to sample based on the catch population is not possible. Instead, the frame can be based on the line itself, divided into coils, or groups of coils, which can be randomly selected. However, the selected m units are not independent but are organized hierarchically: the properties of the sampling design model are different from that for simple random sampling, and population parameters, their variances, confidence intervals and finite population variance have to be estimated differently. In this chapter, in response to the request by the CCAMLR WG-FSA that samples be taken representatively (SC-CAMLR-XV/4), I present a design for randomly sampling longlines which can be used to assess variability at different spatial scales. The results of a field trial of the design are also presented. I used two nested linear models with random effects to analyse the variance components of frequency and total length for the catch population, with the objective of determining the distribution of variation in frequency and length composition at the various levels of the sampling design.

2.2. The sampling design

For all three fishing methods, all the line set during a night forms the sampling frame, whether this was on a single longline or divided between several longlines. For the Spanish and Japanese methods, the sampling frame can be divided into sequences of coils (eg. a longline with 300 coils has 10 sequences of 30 consecutive coils). These sequences can be used as PSUs, which can be further sub-divided into sampling units consisting of shorter sequences to examine variability at different spatial scales. The y-value measured is frequency of fish. For the Norwegian system, the analogous PSU is the sequence of coils within a single magazine, or sequence of magazines depending on the size of the magazine (e.g. a longline of 20 magazines could have 20 PSUs of one magazine, or 10 PSUs of two magazines, etc). Otherwise the design would be the same.

2.2.1. To estimate frequency parameters

Each line is sampled using a two-stage randomised design, with simple random sampling at each stage. Fish abundance is used as an example parameter to illustrate the method. Thus, for the frame for each day's hauling, an unbiased estimator of the total number of fish in the ith primary unit is:

$$\hat{y}_i = \frac{M_i}{m_i} \Sigma^{m_i} y_{ij} \tag{15}$$

where M_i denotes the number of secondary units in the ith primary unit, of which m_i are sampled. Since simple random sampling is used at the first stage, an unbiased estimator of fish abundance for a line (τ_h) is:

$$\tau_h = \frac{N}{n} \Sigma^n \hat{y}_i \tag{16}$$

where N denotes the number of primary units in the frame, of which n are sampled. An unbiased estimator of the mean fish abundance per primary unit is:

$$\bar{y} = \hat{\tau}_h / N \tag{17}$$

and an unbiased estimator of the mean fish abundance per secondary unit is:

$$\bar{y}_i = \hat{\tau}_h / M \tag{18}$$

Sampling units are independent within each PSU but, because the design is hierarchical, they are not independent and identically distributed throughout the sampled catch population and there will be auto-correlation between sampling units within a PSU. Variances are therefore more complicated to estimate than in SRS. Thus, the variance of the estimate of τ_h is:

$$\mathrm{var}(\hat{\tau}_h) = N(N-n)\frac{\sigma_u^2}{n} + \frac{N}{n}\Sigma^N M_i(M_i - m_i)\frac{\sigma_i^2}{m_i} \tag{19}$$

where σ_u^2 is the finite population variance of the total numbers of fish in the PSUs:

$$\sigma_u^2 = \frac{1}{N-1}\Sigma^N (y_i - \mu_1)^2 \tag{20}$$

and σ_i^2 is the population variance within the ith PSU. Thus, for i=1,....,N,

$$\sigma_i^2 = (\frac{1}{M_i - 1})\Sigma^{M_i}(y_{ij} - \mu_i)^2 \tag{21}$$

The first term represents the variance that would be obtained if every sampling unit in a selected primary unit were observed, while the second term is the variance due to estimating the primary unit values y_i from sub-samples of sampling units. The two terms can be estimated by replacing the population variances with sample variances, and parameters with estimates. Thus,

$$\hat{\mathrm{var}}(\hat{\tau}_h) = N(N-n)\frac{s_u^2}{n} + \frac{N}{n}\Sigma^n M_i(M_i - m_i)\frac{s_i^2}{m_i} \qquad (22)$$

where

$$s_u^2 = \frac{1}{n-1}\Sigma^n(\bar{y}_i - \bar{y})^2 \qquad (23)$$

and, for i=1,.....,n,

$$s_i^2 = (\frac{1}{m_i - 1})\Sigma^{m_i}(y_{ij} - \bar{y}_i)^2 \qquad (24)$$

An unbiased estimator for the variance of mean fish abundance per PSU and sampling unit can be calculated by dividing var(τ_h) by N^2 and M^2 respectively. The finite population variance for a line σ_h^2 is the sum $\sigma_h^2 = \sigma_u^2 + \sigma_i^2$.

To estimate the total abundance of the toothfish population sampled during the trial, the data for each line can be pooled in an *a posteriori* fashion by treating each day's hauling as a single stratum within a stratified design. Thus, an unbiased estimator of the overall population total τ_{st} is obtained by adding together the stratum estimators:

$$\hat{\tau}_{st} = \Sigma^P \hat{\tau}_h \qquad (25)$$

where P is the number of days hauling. The variance of the stratified estimator, because of the independence of sampling unit selection between days, is the sum of the individual stratum variances:

$$\mathrm{var}(\hat{\tau}_{st}) = \Sigma^P \mathrm{var}(\hat{\tau}_h) \qquad (26)$$

which can be estimated by summing the variance estimators for each day's hauling. The stratified estimator for μ_{st} is:

$$\mu_{st} = \hat{\tau}_{st} / P \tag{27}$$

and the variance of the estimator is:

$$\hat{\mathrm{var}}(\hat{\mu}_{st}) = \frac{1}{P^2}\hat{\mathrm{var}}(\hat{\tau}_{st}) \tag{28}$$

2.2.2. To estimate length parameters

To illustrate the method, total length (TL) is used as a proxy for other biological composition data, and mean TL as a parameter example. To estimate mean TL and its variance, the line is divided into N PSU's and M sampling units as above, which are selected using SRS at each level. However, the y-value is TL for every fish within a sampling unit. This gives a three-stage cluster design with PSUs as above, but the sampling units are now each fish and the shorter sequences of line used above are secondary sampling units (SSUs). Thus, mean TL for a line is:

$$\bar{x}_{ij} = \frac{1}{L}\hat{\mathrm{T}}_h \tag{29}$$

where L is the number of fish caught in a day's hauling, and

$$\hat{\mathrm{T}}_h = \frac{N}{n}\Sigma^n \hat{x}_i \tag{30}$$

where

$$\hat{x}_i = \frac{M_i}{m_i}\Sigma^{m_i} \hat{x}_{ij} \tag{31}$$

and

$$\hat{x}_{ij} = \Sigma^{L_{ij}} x_{ijk} \tag{32}$$

where x_{ijk} is the TL of an individual fish.

Since all fish within a cluster are sampled, the variance of mean TL is:

$$\hat{\text{var}}(\bar{x}_{ij}) = \frac{1}{L^2}\hat{\text{var}}(\hat{\text{T}}_h) = \frac{1}{L^2}\left[N(N-n)s_u{}^2 + \frac{N}{n}\Sigma^n M_i(M_i - m_i)\frac{s_i{}^2}{m_i}\right] \tag{33}$$

where

$$s_u{}^2 = \frac{1}{n-1}\Sigma^n(\bar{x}_i - \bar{x})^2 \tag{34}$$

and

$$s_i{}^2 = (\frac{1}{m_i - 1})\Sigma^{m_i}(x_{ij} - \bar{x}_i)^2 \tag{35}$$

However, the number of fish caught in a day's hauling is often not known, and the number of fish is unequal between SSUs, so unbiased means cannot be estimated simply by dividing T_h by L, or variances by dividing $var(T_h)$ by L^2. Instead, the total population frequency for a line estimated in 2.2.1 may be used as an estimate for L, so that:

$$\bar{x}_{ij} = \frac{\hat{\text{T}}_h}{\hat{L}} = \frac{\hat{\text{T}}_h}{M\bar{y}_i} \tag{36}$$

and

$$\hat{\text{var}}(\bar{x}_{ij}) = \frac{1}{L^2}\hat{\text{var}}(\hat{\text{T}}_h) = \frac{1}{M\bar{y}_i}\hat{\text{var}}(\hat{\text{T}}_h) \tag{37}$$

Alternatively, if fish are pooled into a single sample population, the mean and variance of the sample population are often used as estimates of the catch population mean and variance.

2.3. Methods and Materials

Within the CCAMLR Observer Scheme, the field trial was undertaken between 7 April - 23 April 1997 aboard the B/F *Cisne Verde*, to test whether a rigorous system was operable using the Spanish double-line system. A total of 12 days hauling was sampled. Every tenth connecting line was already marked with a spliced-in piece of coloured rope, which was used to monitor the amount of line remaining to be hauled. Every coil carried a weight of ca. 6-7 kg at its end and an extra weight half-way along; it also carried 52 hooks, each on a secondary line attached to the mother line. Hooks were Ancora recto size 5, and were baited with a single thawed sardine of ca.15 cm SL, the hook placed through its head. Surface buoys were placed at either end of the longline. All setting operations occurred at night and a similar number of hooks were laid each night over either one or two lines. The Fishing Master determined the number of coils laid during the set, depending on the nature of the seafloor, weather conditions and time available before dawn.

Hauling operations began generally ca. 8 hrs after setting. Hauling was carried out from a platform on the starboard side of the vessel, below and directly aft of the bridge, supervised by the Fishing Master or his assistant from a work station on the aft starboard wing of the bridge. Two winches were used to bring in the retaining and mother line simultaneously. Generally the winch for the mother-line was attended by one of two winchmen working a watch of ca. 3-4 hrs. Fish caught were gaffed as they came to the surface by the winchman and two men handling gaff-hooks with handles ca. 6m long stationed on the upper deck above and aft of the winch. They then passed along a conveyor belt to the stern, falling through a port in the deck to land in a holding bay. Fish were processed using a circular saw to cut off the heads and tails, before evisceration by hand and cleaning automatically.

Fish were sampled on the factory deck. The known number of coils laid during a night's setting operations provided the sampling frame. The size of the primary sampling units was set at sections of 30 coils, secondary units consisted of three sections of ten coils within each primary unit. At the start of hauling, random number tables were used to select four sampling units, of which two were randomly assigned to sampling fish for biometric data on the factory deck, and two to sampling by-catch at the hauling point.

Catch rates from Rubilar et al. (1994) were used to calculate an expected number of ca.1.5 fish caught/coil, from which forty coils/day's hauling were calculated as needed to achieve the CCAMLR target of 60 fish/day. It was therefore necessary to select randomly two SSUs within each PSU. Fish caught on each 10-coil length were placed in a holding bay, and processed after the coils had been sampled. All fish sampled were measured for total length to the nearest centimeter; sex was also recorded.

Components of variance

To examine the sources of variation in the sampling design used to estimate fish abundances, I adapted the method used by Chittenden (1989), employing a nested linear model with random effects:

$$y_{hij} = \mu + a_h + b_{i(h)} + e_{j(hi)} \qquad (38)$$

where $h = 1,2,....12$; $I = 1,2$; and $j = 1,2$. The mean frequency is μ, a_h are the random effects due to strata (days hauling), $b_{i(h)}$ are the random effects due to PSUs within stratum, and $e_{j(hi)}$ are the random effects for sampling units within PSU. The effects a_h, $b_{i(h)}$ and $e_{j(hi)}$ are assumed to be independent of one another. The random effects for days hauling a_h are assumed to be a random sample from a population with a mean of zero and variance σ^2_a. The null hypothesis, that all effects due to days hauling are equal, is H_0: σ^2_a

= 0; the alternative hypothesis, that there is variability among days, is H_a: $\sigma^2_a>0$. Similarly for effects $b_{i(h)}$ due to PSUs within a day's hauling, with variance $\sigma^2_{b(a)}$. The finite population variance σ^2_y is the sum of the variances $\sigma^2_y=\sigma^2_a+\sigma^2_{b(a)}+\sigma^2_{e(ba)}$.

To examine the sources of variation in the sampling design used to estimate mean TL, a similar model was used but with an extra level, corresponding to the lengths measured within each SSU:

$$y_{hijk} = \mu + a_h + b_{i(h)} + c_{j(hi)} + d_{k(hij)} \qquad (39)$$

where $h=1,2,\ldots 12$; $i=1,2$; $j=1,2$; and $k=1,2,\ldots n_{ij}$. $c_{j(hi)}$ is the effect of SSU within PSU, and $d_{k(hij)}$ is the effect of individual fish within SSU. As above, all effects are assumed to be independent and random with means 0 and variances σ^2_a, $\sigma^2_{b(a)}$, $\sigma^2_{c(b)}$, and $\sigma^2_{d(c)}$ respectively.

For both models, the estimates of the variance components were found using analysis of variance in which the expected mean squares are estimated using the observed mean squares, and solved for each variance (Kuehl 1994). The expected mean squares are shown in Table 2.1. For the first model, the estimators for the components of variance are:

$$\sigma_{e(ba)}^2 = MSE \qquad (40)$$

$$\sigma_{b(a)}^2 = \frac{[MS(B/A) - MSE]}{2} \qquad (41)$$

$$\sigma_a^2 = \frac{[MSA - MS(B/A)]}{4} \qquad (42)$$

Table 2.1: Expected mean squares for one-way random effects nested ANOVA used to examine the components of variance due to different levels of the sampling design in estimating a) toothfish catch abundances and b) toothfish total length.

	Source of variation	*Expected mean square*
a)		
	day's hauling	$\sigma_{e(ba)}^2 + 2\sigma_{b(a)}^2 + 4\sigma_a^2$
	PSU	$\sigma_{e(ba)}^2 + 2\sigma_{b(a)}^2$
	Sampling unit	$\sigma_{e(ba)}^2$
b)		
	day's hauling	$\sigma_{d(c)}^2 + n\sigma_{c(b)}^2 + 2n\sigma_{b(a)}^2 + 4n\sigma_a^2$
	PSU	$\sigma_{d(c)}^2 + n\sigma_{c(b)}^2 + 2n\sigma_{b(a)}^2$
	SSU	$\sigma_{d(c)}^2 + n\sigma_{c(b)}^2$
	Fish	$\sigma_{d(c)}^2$

For the second model,

$$\sigma_{d(c)}{}^{2} = MS(D/C) \tag{43}$$

$$\sigma_{c(b)}{}^{2} = \frac{[MS(C/B) - MS(D/C)]}{n_{ij}} \tag{44}$$

$$\sigma_{b(a)}{}^{2} = \frac{[MS(B/A) - MS(C/B]}{2n_{ij}} \tag{45}$$

$$\sigma_{a}{}^{2} = \frac{[MSA - MS(B/A)]}{4n_{ij}} \tag{46}$$

where n_{ij} = the number of fish within a SSU.

2.3. Results

A sample of 1090 fish were taken from 48 SSUs sampled over twelve days. The frequencies, mean TL and standard deviation are shown in Table 2.2 for each SSU nested within PSU within line. The TL distribution is shown in Fig. 2.2. There was little evidence of any trend in TL over a day's hauling (Fig 2.3).

For the abundance data, the catch per sampling unit varied from 3-97 fish, with mean = 22.6 fish/unit and variance of the mean = 1.99. The ANOVA was performed on untransformed and log-transformed data: residuals using untransformed data were not normally distributed (Shapiro-Wilks test, α=0.05), so transformed data were used for the analysis. The spread of residual variances was high but was reduced by transformation. The results of the ANOVA are shown in Table 2.3. Variation between days hauling was not significant at $\alpha = 0.05$. However, the result was marginal with Pr > F = 0.0533, and a low power of $1\text{-}\beta = \sim 0.40$. Variation between PSUs within days hauling was not significant at $\alpha = 0.05$, so $H_0{:}\sigma^2_{b(a)} = 0$ cannot be rejected. Furthermore, the variance component for $\sigma^2_{b(a)}$ was negative: this was accepted as evidence of a true value of zero

Table 2.2: Frequencies, mean TL in cm and standard deviation (SD) for Patagonian toothfish sampled from 10-coil lengths of line during hauling by the B/F *Cisne Verde* between 7-23 April 1997 off South Georgia.

Line	*Freq*	*Mean TL*	*SD*		*Line*	*Freq*	*Mean TL*	*SD*
1	3	98.0	3.61		7	25	101.4	11.0
1	18	99.5	9.64		7	36	92.6	13.5
1	21	91.2	12.39		7	11	88.9	12.0
1	19	87.5	10.40		7	13	91.5	14.4
2	12	103.2	14.1		8	35	98.6	10.9
2	13	93.1	14.91		8	26	97.1	15.4
2	11	89.5	13.77		8	13	105.4	18.5
2	18	90.4	16.01		8	20	110.9	11.0
3	22	93.3	12.86		9	9	93.8	8.2
3	27	101.0	12.77		9	10	96.1	17.6
3	17	98.4	8.34		9	10	98.1	12.6
3	30	94.7	10.79		9	36	93.7	12.5
4	9	91.4	14.70		10	12	91.2	9.4
4	21	88.9	10.62		10	11	95.9	15.8
4	28	92.8	15.01		10	26	95.0	15.1
4	29	87.3	11.62		10	4	110.5	10.1

Table 2.2: continued.

Line	*Freq*	*Mean TL*	*SD*		*Line*	*Freq*	*Mean TL*	*SD*
5	18	104.7	9.47		11	26	91.2	12.3
5	18	103.8	8.99		11	48	86.8	8.0
5	10	103.8	11.90		11	26	84.7	6.8
5	24	103.4	9.92		11	35	87.5	9.8
6	20	109.7	25.6		12	60	85.2	7.7
6	16	93.2	11.4		12	32	95.7	19.1
6	30	104.8	14.2		12	20	86.2	6.8
6	30	104.8	14.2		12	97	92.3	10.8

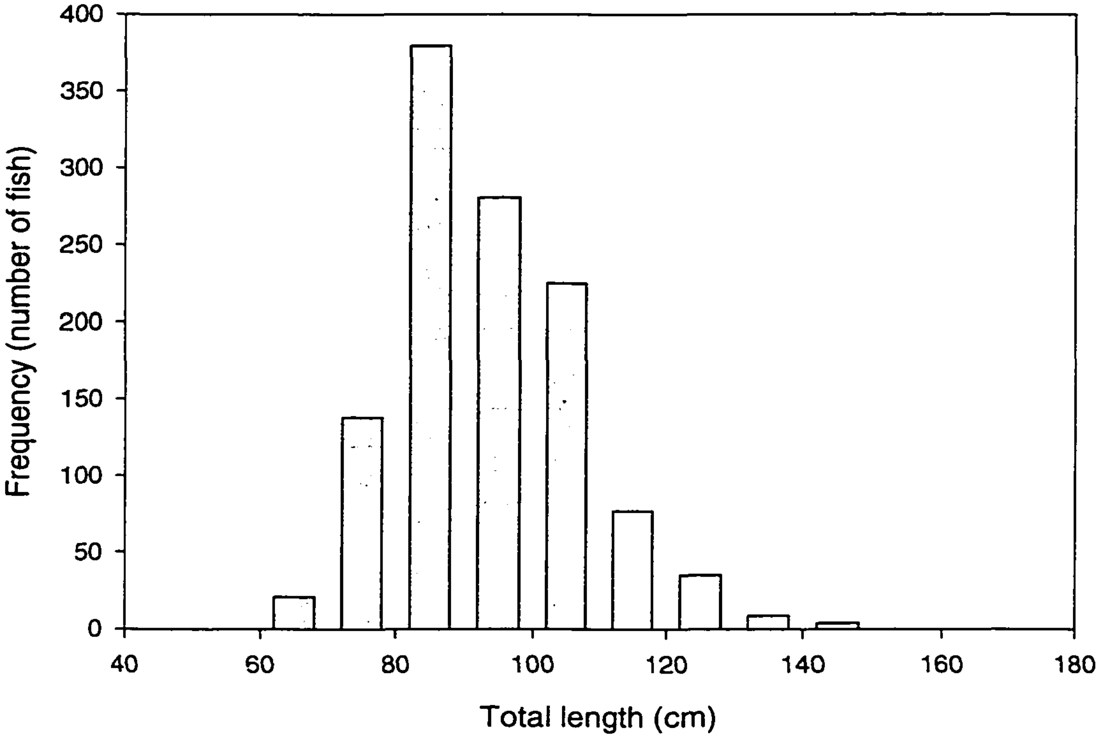


Figure 2.2: Frequency distribution of total lengths for Patagonian toothfish sampled from the catch taken by the B/F *Cisne Verde* between 7-23 April 1997 off South Georgia.

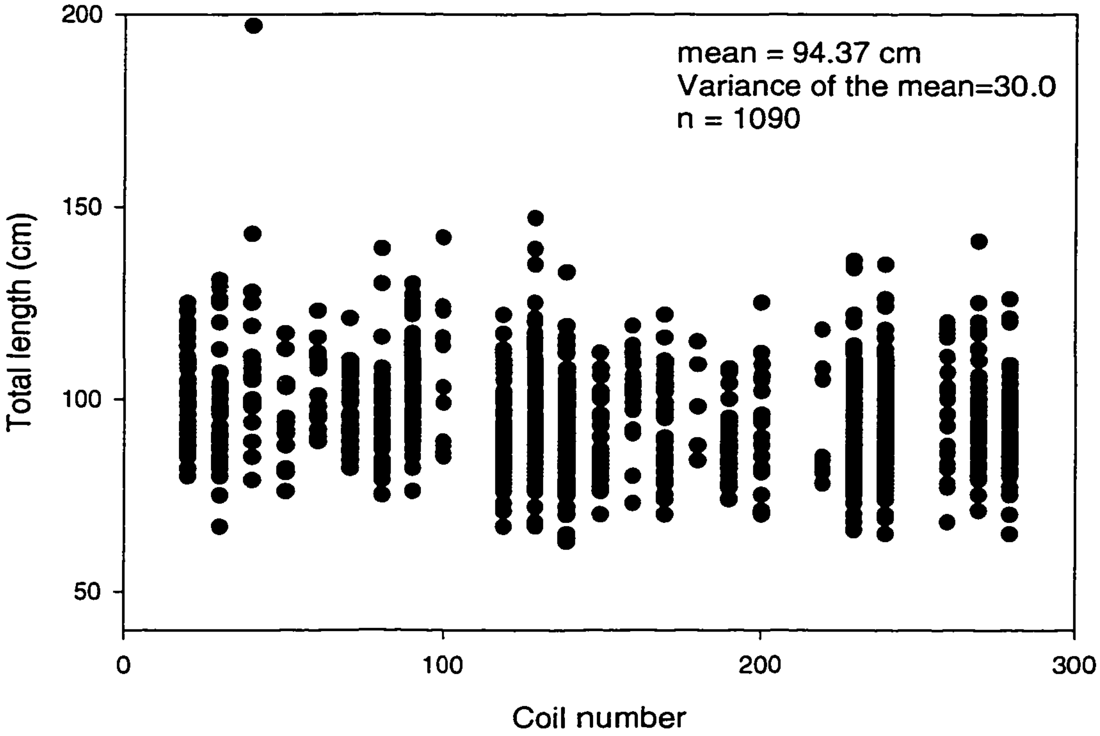


Figure 2.3: Scatter plots showing total length of fish caught versus line loci.

Table 2.3: Results from 1-way ANOVA analysing components of variance for toothfish abundances at three spatial scales: line, 30-coil sampling units (PSU), 10-coil sampling units.

Source	*df*	*Sum of squares*	*Mean square*	*F*	*p*
Total	47				
line	11	7.4025	0.67295	2.18	0.053
PSU	12	3.3608	0.28007	0.91	ns
SSU (error)	24	7.4001	0.30834		

$\sigma_e^2 = 0.30834$ 76%

$\sigma_b^2 = -0.01414$ 0%

$\sigma_a^2 = 0.09822$ 24%

(Kuehl 1994), so that the variance among lines accounted for 24% of the variation in toothfish abundances and the variance among SSUs accounted for 76%.

The practical advantage of a multi-stage design is that it is often less expensive to observe many secondary units when grouped than to observe the same number of units randomly spread over the population. Sampling effort can be allocated depending on the relative costs of sampling the various units (Thompson 1992). Thus, using observer hours as the unit of cost, the optimum number of sampling units/ day can be calculated using:

$$m = \sqrt{\frac{c_1\sigma_e^2}{c_2\sigma_a^2}} = 9.7 \qquad (47)$$

where c_1 = cost of sampling a selected day (15 hrs), and c_2 = the cost of sampling the SSU's (0.5 hrs). Using the cost function:

$$C = c_0 + c_1 n + c_2 nm \qquad (48)$$

where C = total observer cost/voyage of 60 days (60 days=1440 hrs), c_0= fixed overhead (12 hrs/day rest periods = 720 hrs), and solving for n, a total of n = 36 days should be sampled. Thus, rather than sampling every day, a better allocation of effort based on these results is for the observer to randomly select 36 fishing days over the voyage, and randomly sample 10 sampling units within that day, leaving other days free for other tasks.

For the biometric data, the maximum TL sampled was 197 cm; the minimum was 63 cm. The ratio estimators gave mean TL = 94 cm and variance of the mean = 30; whereas pooling all fish and assuming independence gave mean TL = 94 cm and variance of the mean = 0.18. The ANOVA was performed on log-transformed data, as residuals for untransformed data were not normally distributed and the spread of variances was larger than for transformed data. The results of the ANOVA are shown in Table 2.4. As the number of fish were not equal between SSUs, the squares of the mean

Table 2.4: Results from 1-way ANOVA analysing components of variance for toothfish log (TL) at four spatial scales: line, 30-coil sampling units (PSU), 10-coil sampling units (SSU), within SSU. Approximate F_0 statistics are shown.

Source	*df*	*Sum of squares*	*Mean square*	*F*	*p*
Total					
Line	11	3.3473	0.30430	5.140	0.0051
PSU	12	0.6675	0.05563	1.106	ns
SSU	24	1.0601	0.04417	2.734	0.0001
Fish (error)	1042	16.8332	0.01615		

$\sigma_d^2 = 0.01615$	79%
$\sigma_c^2 = 0.00141$	6.9%
$\sigma_b^2 = 0.00012$	0.6%
$\sigma_a^2 = 0.00275$	13%

deviations in the Sum-of-Squares for SSU were weighted by the number of fish in each SSU, and coefficients for the variance components in the expected mean squares were calculated as detailed in Kuehl (1994). Approximate F_0 statistics were then calculated to test the null hypotheses of no effects at the various levels of the model (Table 2.4). Degrees of freedom were calculated using the Satterthwaite Approximation. Variation between lines was significant at $\alpha = 0.05$, so that $H_0{:}\sigma^2_a = 0$ can be rejected. It was not significant between PSUs, but was significant between SSUs, so that $H_0{:}\sigma^2_{c(b)}=0$ can be rejected. Using the coefficients for the variance components calculated above, variance among lines accounted for 13% of the variation, and variation between fish within a SSU for 79%; variation between SSUs was only 6.9%, and between PSUs almost negligible at 0.6%.

Allocating observer effort between day and SSU using the cost functions above, the optimum number of SSU's/day to sample is:

$$m = \sqrt{\frac{c_1\sigma_c^2}{c_2\sigma_a^2}} = 15.38 \qquad (49)$$

and n = 31.3 days should be sampled.

2.4. Discussion

2.4.1. The sample design

Using this design, catch can be sampled at the same time as hauling with little impact on commercial processing. Therefore, this approach will be invaluable when using observers to obtain data on catch and by-catch. The design proved easy to put into practice: SSUs of 10 coil sequences were easier to sample than individual coils, and PSUs of 30 coil sequences meant fewer, longer sampling sessions which were easier to co-ordinate with crew members. The method also proved highly adaptable, allowing sampling on the factory deck to be integrated with rest periods, sampling tasks at other stations (e.g. Ashford et al. 1994, 1995, 1996a), and other observer tasks in one sampling schedule on a day-by-day basis (e.g. Ashford and Croxall 1998).

The sampling method assumes independence between strata in the design. Obvious structures to use for the strata are the individual lines. However, fishing days were used because setting of lines at night divided fishing into clear daily cycles, with a reasonably consistent number of hooks set each night, over either one or two lines. If more than one line was placed on one night, they were generally close to each other, frequently laid within an hour of each other, and were considerably shorter than single lines laid in a night. Furthermore, most travelling occurred immediately after hauling, before the next setting operations, so using days hauling conformed better to the principle of stratification, that the population be divided in such a way that units within a stratum are as similar as possible.

The design for estimating abundances was balanced and the number of sampling units known. It proceeded in a straightforward manner with unbiased estimates of the mean and its variance. However, the design for estimating length composition was unbalanced with an uncertain number of fish caught on the line segment making up a SSU. Population parameters are often estimated by pooling all fish sampled, but this assumes the fish are independent and identically distributed as under a SRS design; this is not the case under a hierarchical design unless fish are randomly distributed in the catch. The pooled estimates are likely to be biased. Estimating population parameters using an estimate of L, the number of fish on the line, indicates that the pooled estimator may substantially under-estimate variance, giving unrealistically narrow confidence intervals. The pooled estimator may therefore be misleading: tests of hypotheses based on pooled variance may result in unjustified rejection of the null hypothesis, leading to erroneous conclusions.

However, the estimators using L will only be unbiased if L is known. An estimate of L will be a random variable: the estimate of mean length and its variance will therefore be a ratio of random variables. Ratio estimators have different properties, may not follow a normal distribution, and are not unbiased (Sokal and Rohlf 1995). But the ratio estimators reflect more faithfully the method of sampling, give a more precautionary variance than the pooled estimate, and the estimate of L has a low variance in any case: the number of PSUs and SSUs are known, and only the number of fish in a SSU is unknown and has to be estimated by mean frequency/SSU. Thompson (1992) pointed out that a biased estimator can be better than an unbiased estimator, if the bias is small and the mean square error of the biased estimator is much smaller than for the unbiased one. Estimates are then more likely to fall near the real value. Similar criteria can be used to select the better of the two biased estimators in this study. Further work is needed to identify the properties of the ratio estimators, and the size of the biases and mean square errors for both methods.

2.4.2. Components of variance

This study demonstrates that abundances of the captured population vary at the spatial scale corresponding to ten coils of line (ca. 500m). Abundances also vary at the higher scale of fishing days. For length composition, most variation is captured within SSU at scales less than 500m, but some is also present at higher scales. Both ANOVAs demonstrated that variance between PSUs at the scale corresponding to 30 coils (1500m) was negligible.

For the ANOVA examining toothfish abundances, the estimates of variance components were not unbiased within the model design because the estimate for between-PSU variance was taken as zero when in fact it was negative. For the ANOVA examining length composition, 79% of the variability was attributable to within SSU variance, 13% to between day and 6.9% to between SSU variance. No variance estimates were negative, so the estimates were unbiased within the model design. Typically, the variance within clusters should be as great as possible to obtain the most precise estimators: these results indicate that this is substantially true for this sampling design, although account needs to be taken of variance at the day and SSU levels.

This study identifies several scales at which it is important to sample toothfish catches, in estimating catch population parameters and constructing tests of hypotheses. Taking these results together, a two-stage cluster design would be appropriate with day at the first stage and 10-coil sequences at the second stage, within which all fish would be sampled. Thirty-coil sequences are unnecessary in the design and need only be used for convenience. Observer effort allocated to sampling 32 days out of 60 days hauling on a voyage, and 16 SSUs within a day, would give the best trade-off between cost and estimation of parameters describing abundance and biometrics. However, variation at other spatial scales are beyond the inferential limits of the study, and further work is needed to examine variation at higher temporal scales and between fishing areas.

Additionally, inferences on the wild population are beyond the scope of this study. In active fishing methods like trawling, the catch population can be linked strongly to the underlying wild population, allowing reasonable extrapolations between the two. But in a passive method like longlining, the captured population selects itself and its structure may be considerably different to that of the wild population, depending on the variables governing choices made by the fish. These variables are poorly understood for toothfish, and more work is needed to determine the relationship between the wild and captured populations, and how these variables and their fluctuations mediate the relationship. Comparative studies between catches in time and space are likely to be more reliable; but differences may reflect changes in the mediating variables rather than the wild population.

2.4.3. Implications of the study for toothfish and other fisheries

The hierarchical sampling design presented is largely driven by considerations of convenience. However, the results of the ANOVAs indicate that for toothfish, most variation is captured at lower levels in a way that makes a clustered design efficient in providing estimates with low variance compared to alternative designs. The ratio estimates, although biased, may be considerably more realistic for examining length and frequency compositions than the common alternative of pooled estimates, especially once observer effort is optimally allocated.

In contrast, a haphazard sampling design will not provide estimates in which the properties can be fully described, or avert bias through unconscious assumptions. It will also not give the information to allow these issues to be addressed. The advantage of convenience is cursory, as the randomized method is easy to implement once understood by the observer; the few minutes saved each day by not selecting sampling units randomly will quickly be outweighed by the much greater precision in confidence and savings in unnecessary time spent sampling, or in generating results that are too unreliable to justify conclusions (Thompson 1992). Furthermore, observers sampling haphazardly allocate their time in a number of discrete periods during a day: the haphazard methodology is hierarchical and incurs similar disadvantages to the randomized design, but without the means to assess these problems.

The line itself may not always be practical as a frame: proxies for the line, such as time, can provide a useful alternative. However, time does not allow the frame to be identified beforehand, except as a predicted mean time. The frame will therefore be a random variable, resulting in some unpredictable statistical properties. On a slow day's hauling, the end of the line will overshoot the frame, so that the later coils will have no chance of being sampled; on quick days, the frame will undershoot and some selected sampling units will not be sampled. Variations in the rate of hauling will be confounded with distribution of frequencies: the number of fish sampled per time unit will be a function of hauling rate and the distribution of the catch. When a line is snagged, no or few fish will be sampled. The variation in numbers of fish caught/unit will be magnified along one tail of the distribution, skewing it artificially, introducing possible artificial outliers, and further unbalancing the number of fish/SSU. Where possible therefore, the line is a much better frame than a proxy.

Hierarchical designs are used to sample biometric data in other fisheries when a fishery's structure is complex. However, managers frequently improvise a sampling design, leading to assumptions that may not be justified. For instance, catch from individual trawls is frequently sampled by scooping a bucket through the topmost layer of fish in the trawlers' hold. This assumes that the fish are mixed randomly in that layer, and there are no sorting effects either in the population captured by the net, or in their fall into the hold. The distribution of trawls in the first sampling stage is often unpredictable, so managers improvise with a set of rules (such as 'sample the last trawl each day'), and make the assumption that the selected PSU's are a random sample from a larger population in time and space. This assumption is then justified by comparing the last trawl with other trawls in the day to look for discrepancies in selected estimates and their variances (Chester 1984). Parameters are estimated by weighing the means of the SSUs selected. However, significant error can arise if the conclusions from these comparisons are applied outside their inferential limits, or if coverage within the frame was insufficient. Furthermore, the estimators do not weigh correctly for the full sampling frame and so are biassed: the number of fish (L) in the frame remains unknown.

Similarly, Chittenden (1989) sampled Atlantic croaker and weakfish from commercial boxes taken from pound-nets and haul-seines in Chesapeake Bay. All fish in a box were sampled in a cluster-sampling design with the box as PSU, and fish as sampling units. He used a 1-way ANOVA to examine the variance components contributed by among-box and within-box variation, using TL as a proxy for other biological composition data (weight, age, sex): 98% of variation was attributable to within-box variation, so he concluded that the strategy in box-selection was not a major problem, providing a justification for pooling fish for subsequent analysis (e.g. Piner 1999), thereby circumventing the problem of an unknown L.

However, pooling fish gives biased estimates where significant between-PSU variability exists; and this may be the case on seasonal time-scales outside the inferential limits of Chittenden's (1989) study. If TL is insensitive as a proxy for other biological data, for instance masking shifts in sex ratio, the problems of extrapolating beyond the limits of inference may be compounded. Analysis of variance components can therefore be useful for allocating sampling effort, and can indicate at what level sampling may go most awry when it is not properly randomised. But, as in the previous example, alternatives to proper randomisation at each level of the design may lead to biases when assumptions are not justified, and can only be regarded as a fallback when a properly randomised hierarchical design cannot be implemented. Biometric estimators will be unbiassed only when weighed for a known number of fish in the frame.

The problem of an estimated L encountered in this study therefore has wider applicability than simply to longline sampling. It will apply where: the fishery structure dictates a hierarchical design; fish do not arrive at the sampling station in a predictable manner so that higher order sampling units are based on gear or other features of the fishery not directly related to the fish population; and the number of fish sampled in each higher order sampling unit is unequal. The bias due to the ratio estimate, and its mean square error, will depend on the particular fishery, but the properties of the ratio estimator are likely to have widespread implications and deserve further work.

2.5. Conclusion

A preliminary version of the sampling design was presented at the 1996 WG-FSA meeting (Ashford and Duhamel 1996), at the same time as a randomised design for shore-based sampling of the artesanal fishery off northern Chile (Ashford et al 1996b). The Working Group discussed the design and requested a set of protocols for shipboard sampling of longlines, which was presented at the 1997 meeting (Ashford and Duhamel 1997). In the same season, the design was field-tested and simplified to the version presented in this chapter. The simplified design was presented at the 1997 meeting (Ashford and Everson 1997), and a full protocol presented in the following year (Ashford et al. 1998). Due to a lack of consensus in the working group, the design was not recommended as a standard for use within the CCAMLR area, but the simplified version has been used subsequently in several management areas for training observers.

CHAPTER III

ESTIMATING AGE

3.1. Introduction

3.1.1. The importance of age data

The age composition of fish stocks is fundamental to understanding population dynamics and stock productivity. Age data allow population age structure to be modelled, so that growth, mortality and recruitment rates can be estimated (Jones 1992), and used to provide management advice. With age data, individual cohorts can be followed through time. Cohort life-tables can be constructed, from which age-specific survival and mortality rates can be calculated. Life-tables also allow calculation of standardised killing power, allowing comparisons between separate studies; and age-specific fecundity, using data on egg production within each age class. These age-specific vital rates can be used to elucidate life-history strategies.

Age data can also aid in understanding population fluctuations due to intrinsic and extrinsic factors: age-specific mortality over time can be used to examine age-specific density-dependent effects in population regulation, similar to key-factor analysis (Begon and Mortimer 1986). Life-history patterns can be examined for combinations of high fecundity and time-lags, or overcompensation in density-dependent effects which may generate variability in abundance (Begon and Mortimer 1986, May 1975). Population structure may tend toward equilibrium (Hastings 1997), but extrinsic environmental perturbations can induce substantial variation in vital rates, which differ in their effects on different age classes. Monitoring age composition over time can allow the effect of environmental variability to be examined on different time scales. Finally, anomalous fluctuations in age structure through time can alert managers to irregularities of scale in sampling programs (Underwood 1996, Anon 2000), allowing correction.

Age data therefore allow managers to ascertain the life history pattern of a stock species, and which age classes are vulnerable to over-exploitation under a particular set of environmental conditions. Similar life history patterns have common properties, so that management of one stock can benefit from the lessons learned in managing another with a similar pattern. As importantly, age data allow managers to use age-structured stock assessment (ASA) models, providing estimation of age-specific vital rates, and predicting abundances of individual cohorts (Megrey 1989).

3.1.2. Error in Age Data

Good age estimation is critical for management of fisheries. Inaccuracies in aging estimates result in significant non-random bias, producing erroneous output from models (Beamish and McFarlane 1983, Dorval 1998, Coggan et al. 1990), leading to inappropriate management strategies. Yet microstructural features occur in otoliths in a hierarchy of scales corresponding to regular cycles and events in a fish's life. The structures corresponding to an annual period need to be selected from the others, yet interpretation demands a degree of subjectivity which will contribute error to all age determinations (Campana et al. 1995).

Error can be of two forms which are not necessarily linked: accuracy in estimating the true age, and precision (or reproducibility) of measurements by readers (Campana et al. 1995). Inaccuracies can result in bias typically toward younger or occasionally toward older individuals (Beamish and McFarlane 1983). Error due to precision can mask year-class strength: Fournier and Archibald (1982) pointed out that the expected percentage of a small cohort in a population would have a 140% error if aging was correct 80% of the time and recruitment varied widely from year to year. Estimates of mortality for a small cohort would be under-estimated initially and over-estimated in later years; whereas large recruitments would appear smaller and small recruitments larger. Bradford (1990) found that aging errors can halve estimates of recruitment variance, and introduce bias and autocorrelations in recruitment time series estimated from sequential population analysis. Lai and Gunderson (1987) found that biases in von Bertalanffy growth parameters due to aging error resulted in significant increases in the optimum fishing mortality (F*) in a yield-per-recruit model; under-aging using scales resulted in estimates of F* and t_c* (mean age at first capture) that would lead to over-fishing.

Reader error occurs through the incorrect or inconsistent use of aging criteria, influenced by light artefacts in microscopy, discrepancies in preparing samples or poor preparation methodology, and the complexity of the aging material itself (Neilson 1992). Faulty sampling design can also contribute substantially to error in age data. Dorval (1998) used an ASA model (Pope and Shepherd 1982) to simulate the menhaden fishery of the east coast of the United States, and found that undercoverage of areas resulted in average errors in estimating recruitment and fishing mortality of up to 116%. He concluded estimates were unreliable when spatial sampling errors were present, and efficient management requires proper sampling frame coverage and implementation of probability-based sampling designs.

To generate reliable age data, there are three essential requirements (Morison et al. 1998): the catch should be sampled representatively to cover the fishery fully and reduce sampling error (if an age-length key is used, the lengths sampled must be representative of the population). Secondly, the technique for age determination should be validated across the population life history, so that the correct criteria are used for age estimation. Thirdly, quality control is needed to ensure age data are estimated consistently for a given methodology, across the catch and over time (Morison et al. 1998, Anon 2000). Furthermore, quality control should minimize sources of error, and provide a quantitative presentation of variability in age determinations should be provided to incorporate in decisions on age-based analysis and modelling (Kimura and Lyons, 1991). These requirements are particularly important in species for which it is difficult to estimate age, to ensure the best quality age data possible and to gauge whether age data are precise enough to justify the use of ASA techniques. If age data are highly variable and inconsistent, management would better resort to less rigorous approaches that avoid their use.

3.1.3. Estimating age

Fish are usually aged using microstructural features in otoliths and scales. Scales are easier to obtain and prepare than otoliths, and can be sampled several times during the life of a fish. However, scales may be reabsorbed and regenerated, or their development delayed for several months after hatching, or annuli in older fish obscured by compression of the circuli as growth slows with age. True age can then be under-estimated especially in long-lived species in which growth is concentrated in the early life history (Beamish and McFarlane 1987, White 1991). Otoliths are not reabsorbed or metabolically re-worked (Campana and Neilsen 1985), and contain a more complete record of growth from hatching to capture. They also contain trace elements taken up from the water column, which can be used to place a fish retroactively in space and time (eg. Thorrold et al. 1999). Otoliths therefore offer advantages over scales for long-lived fish.

Several techniques have been employed for preparing otoliths for estimating age of fish. Translucent otoliths in which internal aging features can be seen, are read whole. Alternatively, otoliths can be broken to reveal the otolith microstructure, and the exposed surface viewed with reflected light. If age features remain obscure, burning the exposed surface in an alcohol flame can improve definition (Christensen 1964, Chilton and Beamish 1982); baking can have similar results, but with greater control and evenness in applying heat. The method is simple to use and demands little equipment. However, otoliths may crumble or shatter; breaks may not pass cleanly through the nucleus; and the exposed surface is frequently rough, scattering reflected light. Moreover, handling whole or broken otoliths is awkward, slowing age reading.

A third option is to cut thin sections in large numbers by mounting otoliths in parallel rows in resin blocks, and cutting using a diamond saw (Bedford 1983). The plane to be read can be aligned more precisely than by breaking, and viewed using transmitted or reflected light. The sections are mounted on slides which are convenient to handle and store. The method can be adapted for fine scale work using Scanning Electron Microscopy (Ashford et al. 1993). However, the precision of the cut depends on placing otoliths with their nuclei aligned along the saw's cutting track: in many species, this can be difficult because the nucleus is small or cannot be identified accurately from surface marks.

3.1.4. Assessing error due to reader precision

Kimura et al. (1979) originally examined precision error using analysis of variance and variance components. Ages were determined for a single sample by two readers on two occasions, giving four readings. Total variance was partitioned into between-reader (s_A^2) and within reader components (s_e^2) under the assumptions of the random effects model:

$$y_{ij} = \mu + a_i + e_{ij} \qquad (50)$$

where μ = the mean, a_i = the between-reader effect, and e_{ij} = the within-reader effect. Unbiased estimates of precision variability were obtained using the error mean square, and the consistency of within-reader variance could be tested (H_{01}: $s_{e1}^2 = s_{e2}^2$, $F_{v1,v2}$). If no significant difference, the hypothesis that the between reader variance was zero (H_{02}: $s_A^2 = 0$) could be tested, and the variance component s_A^2 estimated. The two null hypotheses together imply there is no difference in the statistical properties of age determinations by the two readers. The approach accounts for the sources of variation and distinguishes between between-reader and within-reader variability, allowing unbiased estimates of each and calculation of the overall variance of an age determination ($V(y_{ij}) = s_A^2+s_e^2$). However, estimates showed a trend of increasing variability with age, and data were analysed by nominal age category (where nominal age was the mean of the four readings taken for each fish), confounding the response variable of observed age (Hoenig et al. 1995). The repeat readings by each reader were also assumed to be replicates, but bias may occur within reader making this assumption questionable without a prior test; however, H_{01} does not specifically test for this. If within-reader bias occurred, it would inflate the estimate of imprecision by a given reader and may artificially weaken the test of differences in age estimates between readers (H_{02}: $s_A^2 = 0$). The calculations for sum-of-squares were also improvised and their mathematical properties are not fully clear. Finally, H_{02} is not a direct test for bias between readers, and the procedure does not allow bias to be estimated.

Another approach is to measure the percentage of repeat readings in agreement, but agreement is likely to fall with age, so that comparisons between older and younger sample populations will show differences that are due to age discrepancies as well as reading error (Beamish and McFarlane 1987). Instead, Beamish and Fournier (1981) proposed using average percent error (APE) as an overall index to allow objective comparisons between species and populatons:

$$APE_j = 100 \times \frac{1}{R} \Sigma_{i=1} \frac{|x_{ij} - x_j|}{x_j} \tag{51}$$

where X_{ij} is the *i*th age determination of the *j*th fish, X_j is the mean age of the *j*th fish, and *R* is the number of times each fish is aged. Chang (1982) developed this approach, using standard deviation rather than absolute deviation to produce an estimate of the average coefficient of variation (CV) that was more statistically rigorous than APE (Campana et al. 1995):

$$CV_j = 100 \times \frac{\sqrt{\Sigma_{i=1} \frac{(x_{ij} - x_j)^2}{R-1}}}{x_j} \tag{52}$$

Kimura and Lyons (1991) found age determinations to be normally distributed with constant CV over wide age ranges, supporting the use of CV for measuring variability in age precision studies. However, the usefulness of APE and CV as overall indices of precision depend on the distribution of error with age: for instance, if error is chiefly due to the difficulty of interpreting the first annulus or early life history, error will remain constant with age, but CV will decrease, so that population age structure will again influence estimation of variability. Similarly, variation in precision among ages may occur due to changes in the pattern of annuli laid down at different life history stages, and will make comparisons sensitive to age distribution (Hoenig et al. 1995). Variability is averaged across observations without accounting properly for the different sources of variation between experimental designs (eg. within and between fish, within and between readers) (Hoenig et al. 1995), and the measures make no distinction between reader bias and variability: any bias will inflate the estimate, so to measure precision variability, bias must be discounted beforehand (Campana et al. 1995).

Bias can occur systematically across the age range of a population, or change with age: one reader may under-age fish at one end of the range and over-age at the other, or bias change non-linearly between years with changes in the structure of the otolith. Regression will detect the first two types of bias and matched-pair tests (parametric and non-parametric) can detect systematic bias, but neither can detect non-linear bias around a 1:1 relationship. Instead, age-bias plots can be used where the age readings of reader Y are presented as the mean age and 95% confidence intervals corresponding to each of the age categories reported by reader X (Campana et al. 1995). Age bias plots allow visual detection of all three forms of bias, but do not allow a quantitative estimate or test of significance (Campana et al. 1995).

Alternatively, a symmetry test can be used where there are changes in bias with age (Bowker 1948, Hoenig 1995). This tests for the hypothesis that an $m \times m$ contingency table consisting of two classifications of a sample into categories (ages given by two readers, or repeat readings by one reader) is symmetric about the main diagonal. The test statistic is:

$$\chi^2 = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \qquad (53)$$

where n_{ij} = the observed frequency in the *i*th row and *j*th column, and n_{ji} = the observed frequency in the *j*th row and *i*th column, and is distributed as a chi-square variable. However, the statistic is sensitive to stochastic effects so that its value will be influenced by large precision variability, and this effect will be complicated further by age structure if variability changes with age. The test also does not account properly for the different sources of variability which may lead to bias; bias may be obscured if it only occurs over a part of the life-history; and the types of bias cannot be easily distinguished or their sources identified.

These techniques can be used to estimate error in programs for assuring the quality of age data. Kimura and Lyons (1991) reported on a program undertaken by the Ageing Unit at the Alaska Fisheries Science Center: 20% of all routine age readings were independently re-aged by another age reader to estimate bias and precision for six species over their life history. Estimates of CV were used to compare ‘ageability’ of species, and age classes within species. Between-reader biases and variability could be monitored for potential effects on later analysis and modelling. Morison et al. (1998) repeated readings for 25% of samples to check for consistency, using APE as a measure of precision. They expected APE to be less than 5%, and used bootstrap techniques (Efron and Tibshirani 1993) to calculate bias-corrected means and confidence intervals. Many laboratories maintain reference collections of age determination structures with known age or estimated precision: subsamples are inserted randomly into regular production samples to ensure that age readings do not ‘drift’ over time (Campana et al. 1995). Reference collections can also be compared between laboratories to assess inter-agency differences (eg. Boehlert and Yoklavich 1984).

3.1.5. Patagonian toothfish

Hureau and Ozouf-Costaz (1980) used polarized light to estimate age from scales of Patagonian Toothfish, and gave brief criteria for age estimation using scales and otoliths. Ages read using otoliths and scales from the same fish were considered to give good agreement, but no quantitative assessment was attempted.

Young *et al* (1995) compared scales and otoliths from toothfish caught off southern Chile. They used thin sections cut transversely through the otolith, and found the nucleus and annuli were difficult to interpret. No criteria were given for interpreting scales: instead, the authors relied on independent readings from scale readers experienced with other species. They found that scales gave significantly lower estimates for age than otoliths in older fish. They estimated mean CV=11.3 for age readings by two observers using scales and otoliths.

Cassia (1998) compared age readings from scales and otolith sections from toothfish captured off South Georgia. She found complete agreement between otoliths and scales in 43.8% of cases for which she could obtain readings; in the other 56% of cases, the discrepancies were found were not significantly different. Otolith sections were frequently too opaque to read, but the percentage of unreadable otoliths was not given nor full criteria for interpretation of otoliths. For scales, she followed the criteria of Hureau and Ozouf-Costaz (1980), using polarized light to detect the first ring.

Both Young *et al* (1995) and Cassia (1998) found difficulties in locating the nucleus and interpreting internal features due to the opacity of the otolith matrix. Ashford and Wischniowski (1998) used a technique involving baking and breaking otoliths to examine toothfish otoliths taken from South Georgia. This gave good readable images, but the otolith pattern was difficult to interpret. Two different sets of criteria were used: C1 developed from the criteria outlined by Hureau and Ozouf-Costaz (1980), and C2 based on generalized criteria used by the Pacific Biological Station (MacLellan 1997). Within and between reader precision was estimated using CV, and was found to be similar. For criteria C1, mean CV_{within}= 8.0 and $CV_{between}$= 9.05.

Kalish and Timmiss (1998) noted that the appearance of opaque and translucent zones is highly variable among specimens from a single locality, and there may be differences in the ‘readability’ of otoliths from different localities. The CVs found above corroborate this observation, but there has been no published attempt to extend the scope of these findings or analyse differences between laboratories. Age estimation may vary significantly at each level, affecting comparisons between estimates of growth and age distribution used in models for stock management.

3.1.6. This study

The recent evidence that age estimations are difficult for toothfish otoliths, introducing variability into age data, means that estimation and management of error is critical if age data are to be used in estimating growth and population vital rates through management models like VPA.

Conventionally, a first step to reducing age estimation error is to develop a consistent preparation methodology, and a conceptual model of how the hierarchical internal structure of an otolith corresponds to the required time-scale. From the model, a clear set of age criteria can be predicted for use by all readers. This model is frequently inferred from related species, or species with structural affinities, that have a known relationship between otolith structure and time. Alternatively, where the structure is simple, the strongest recurring pattern is used, so long as it appears at a reasonable scale conforming to what is known about the species longevity and biology.

At this stage, although the model may be based on more-or-less strong evidence, it cannot be considered to have been tested, or validated. Instead, it provides a working hypothetical standard which can be tested in a validation. Precision error in the data produced by the model can be divided into two forms: bias between readings, and residual random variability. Even without a validation, both forms of precision error can be assessed, and individual readers trained to interpret the criteria without bias and within certain bounds of variability.

In this chapter, I describe the development of a new methodology for preparing toothfish otoliths for age estimation. I present a conceptual model of the correspondence between otolith structure and time, and a set of criteria to aid interpretation. This is followed by an analysis of precision error for two sample sets, and readers within and between laboratories. However, to account for the different sources of precision error properly, analysis of variance is needed (Hoenig et al. 1995): I therefore present an ANOVA model that incorporates an estimate of random precision variability somewhat analogous to Chang's (1982) averaged CV, tests of bias within and between readers, and estimates of any significant bias.

I use the ANOVA model to correct for bias in readings, and look for differences in readability by sex. I also construct a reference collection of otoliths, each with estimated age and variance; instead of relying on the conceptual model of otolith structure with subjective interpretation of criteria, I use the collection as a numerical paradigm for reading otoliths, allowing training, re-familiarization and monitoring of reading consistency over time. Finally, I discuss ways in which the ANOVA model can be generalized for use with error distributions that depart from normal.

3.2. Methods and materials

3.2.1. Preparation methodology

Trial sections through sagittal, transverse and longitudinal planes of otoliths were prepared initially using a diamond saw and viewed using reflected and transmitted light. These indicated that the transverse plane gave the best surface for reading with a single nucleus and no secondary primordia. Sections were then prepared for viewing using SEM (Ashford and White 1993) by etching with EDTA, in an attempt to increase resolution and avoid the effects of opacity and light artefacts. However, sections frequently missed the nucleus and the SEM images of the etched sections were still difficult to interpret. The break and burn technique was also tried (Ashford and Wischniowski 1997): baking otoliths before breaking gave more consistent results than burning, and internal features were more clearly defined, but otoliths tended to break along crenellations where incremental structures are likely to be more compressed and difficult to interpret (Everson 1980).

Finally, a new technique was developed using a grinder to prepare thick sections for viewing by reflected light. One of each pair of otoliths was selected randomly and baked at 375°C for approx. 3.5 minutes, or until a light brown. The otolith was then ground by holding the anterior side against the grinding wheel of a Hillquist Thin Section Machine until an internal mark was revealed which was found to lie consistently just anterior to the nucleus. The ground face was then mounted on a glass slide using Krazy-Glu, left to dry, and ground from the posterior side to form a thick transverse section incorporating the nucleus and avoiding crenellations. The section was finally polished using Mark V Laboratory 3M aluminium oxide polishing paper, covered with Flo-Texx, and viewed using reflected light under a Leica MZ8 binocular microscope at a magnification of x25.

3.2.2. Conceptual model and criteria for estimating age

A conceptual model of the relationship between otolith structure and chronology was developed from brief criteria outlined by Hureau and Ozouf-Costaz (1980). More detailed criteria were elucidated in order to discriminate better between yearly annuli and checks found in different regions within the otolith cross-section.

Hureau and Ozouf-Costaz (1980) divided the section into three regions: a nucleus, a region of concentric large annuli immediately surrounding, followed by a region of narrower regular annuli. Working out from the nucleus, the annuli were largest in the dorsal axis, and compressed on the medial and proximal sides (Fig 3.1). The dorsal axis then became compressed, and the annuli widest on the proximal sides. In the regular region, the narrowest annuli on the proximal side were considered to be annual. The yearly annuli tended to diminish in width to the edge of the otolith, although exhibiting

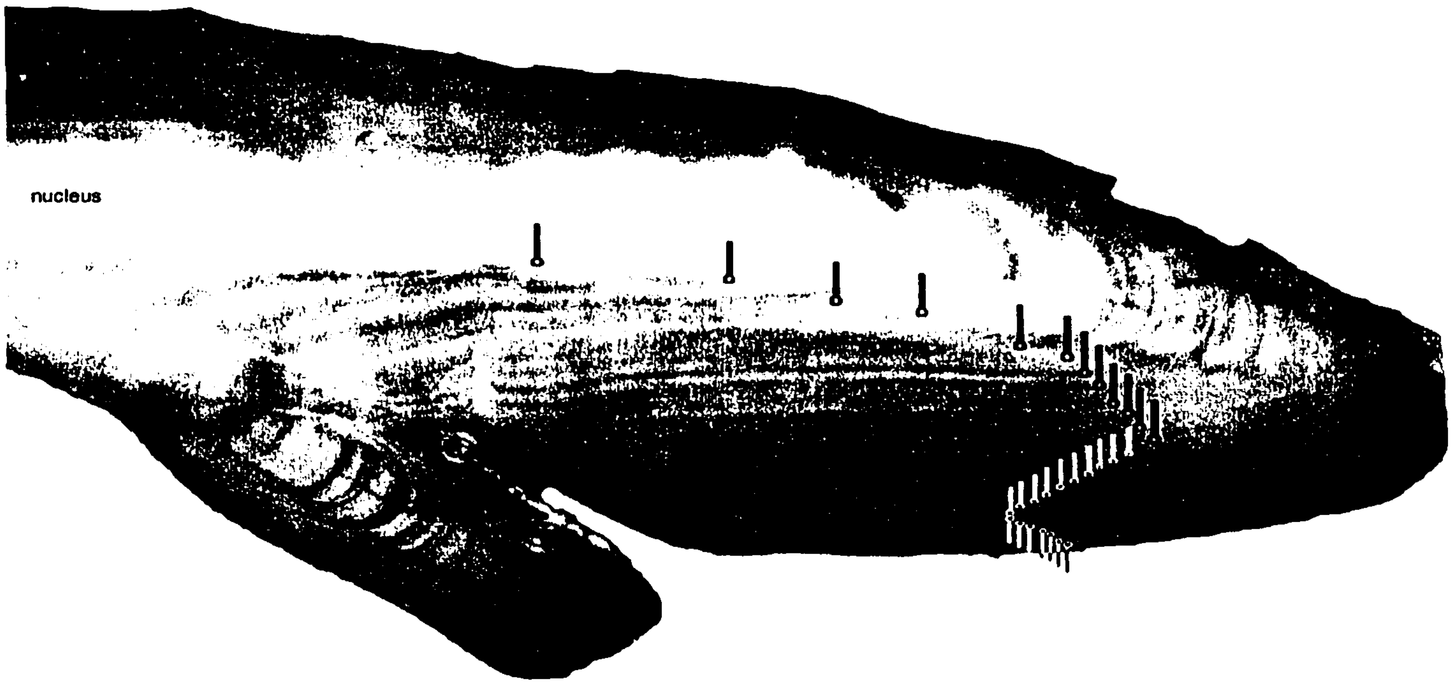


Figure 3.1. Example of age count for transverse section of otolith from *D. eleginoides* using criteria given by Hureau and Ozouf-Costaz (1980). The edge of each yearly annulus is marked by the open circle of each symbol; the first symbol marks the outer edge of the first annulus. The change from dark to light symbols marks the boundary between the region of large clear yearly annuli and the region of regular yearly annuli. Age = 29 years.

some variation in width. In the region of large annuli, heavily calcified zones were interspersed with narrower zones consisting of bundles of narrow micro-increments: these were considered respectively to be the opaque and hyaline zones of yearly annuli. They tended to occur at decreasing intervals but were very variable in contrast along the count path and between fish. The nucleus consisted of a central core strongly marked by regular micro-increments, surrounded by a region with less defined micro-increments forming a dorsal protrusion.

Criteria are summarized in Table 3.1. The count path followed the large annuli along the dorsal axis, moving to the regular annuli along the proximal dorsal axis as the dorsal axis became compressed (Figure 3.1). Structures occurred at different scales in all regions: in the regular region, the narrowest annuli were considered annual as long as they persisted clearly either side of the count path. Marks or structures that did not persist far to either side of the count path or occurred irregularly at a lower scale were considered false checks. In the region of large annuli, distinguishing between annuli and checks was more difficult: annuli were considered to be larger, have stronger contrast between opaque and hyaline zones, and to persist either side of the count path notably into the compressed medial region. Checks tended to be confined to one region or vary considerably in clarity between regions. In the nucleus, a discontinuity was observed running diagonally between the core and the dorsal protrusion. The edge of the nucleus was defined as the inner border of the first hyaline zone, which was typically clearer than the succeeding hyaline zones. As the hatch date of *D. eleginoides* is not known, the nucleus may not represent a full year's growth, so the outer edge of the nucleus was considered as Year 0. The birthday of all fish was taken to be 1 July, so that the outer annulus was counted if the fish was taken after 1 July but not if taken before.

3.2.3. A model for estimating precision error in toothfish age estimations

A randomized complete blocks design was used with a single replicate of each treatment per block. The blocking factor was individual fish, considered randomly drawn

Table 3.1: Criteria for estimating age in Patagonian toothfish.

Region	*Criteria*
1. Nucleus	Consists of: central core marked by regular micro-increments, and a surrounding region developed dorsally. Edge of nucleus defined as the inner border of first annual hyaline zone
2. Large clear annuli	Read along dorsal axis. Large yearly annuli consisting of alternating hyaline and opaque zones: opaque zones heavily calcified, sub-annual checks occur frequently in opaque zones, and narrower hyaline zones often made up of bundles of closely spaced micro-increments. Annuli spaced at regular intervals but trend to decrease in width dorsally.
3. Narrow regular annuli	Read along proximo-dorsal axis. Yearly annuli consist of fine-scale alternating hyaline and opaque zones, often of similar width, which persist along proximo-dorsal edge (sub-annual checks sometimes present but these do not persist).

from the wider population. Each reading was considered a separate fixed treatment, with a single replicate in each cell. The treatment and block effects are assumed to be additive, with no interaction. Let y_{ij} be the *i*th reading on the otolith from the *j*th fish. Under the assumptions of the mixed effects model:

$$y_{ij} = \mu + \tau_i + b_j + \varepsilon_{ij} \tag{54}$$

$$i = 1,2...t; j = 1,2...r.$$

where μ = the general mean

τ_i = the effect of the ith level of the factor reading

b_j = the effect of the jth level of the blocking factor fish

e_{ijk} = the experimental error

The random effect blocking factor b_j (fish) and random error e_{ij} are assumed to be independent normally distributed random errors with mean 0 and variances σ_b^2 and σ^2 respectively. The blocking factor b_j partitions the variance due to individual fish, including effects due to the hierarchical ordering of variation from the sampling design, leaving the variance due to age estimation T_i as the factor of interest. The Sum of Squares error is calculated based on the identity:

$$(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$$

and MSE is an unbiased estimate for σ^2, the overall precision variability. T_i is an estimate of the bias for each reading: where it is found significant using the conventional *F* test, the individual means can be tested using pairwise treatment comparisons like the Tukey Honestly Significant Difference and the Student-Newman-Keuls Multiple Range Test. Bias can then be estimated using the difference between the estimated general mean and estimated treatment mean $(y_{..}-y_{i.})$. A nonadditivity test (Tukey 1949) can be used to detect for interactions between factors, i. e. nonadditivity of the form $\lambda T_i b_j$. The test can be used as an indicator if relative treatment responses differ between blocks containing very different conditions. The experimental design could also be described as a split-plot arrangement with a single replication for whole plot and sub-plot units, or a confounded factorial in which main effects due to fish were confounded with blocks. Thus, the whole plot blocking factor was individual fish, and subplot units were readings.

Furthermore, if two readings were taken by each reader, the treatment term of the model could be further divided into two crossed factors of reader and reading, so that the interaction term obtained could be used as an indicator of imprecision between readings by one reader. Let y_{ij} be the jth reading by the ith reader on the otolith from the kth fish. Under the assumptions of the mixed effects model:

$$y_{ijk} = \mu + a_i + \beta_j + (\alpha\beta)_{ij} + \rho_k + \varepsilon_{ijk} \qquad (55)$$

$i = 1,2 \ldots a;\ j = 1,2 \ldots b;\ k = 1,2, \ldots r$

where μ = the general mean

a_i = the effect of the ith level of the factor reader

b_j = the effect of the jth level of the factor reading

$(ab)_{ij}$ = the interaction effect between reader and reading

ρ_k = the effect of the blocking or whole-plot factor fish

e_{ijk} = the subplot random error

Again, the random effects blocking factor ρ_k and subplot random error e_{ijk} are assumed to be independent normally distributed random errors with mean 0 and variances σ_b^2 and σ^2 respectively. Since only one observation is available in each cell of the arrangement, additivity between fish and treatment is assumed, allowing the mean square partition for interaction between fish and treatment to be used as an estimate of experimental error (Kuehl 1994).

3.2.4. Estimating precision error in the toothfish fishery

A total of 110 fish were sampled on board the B/F *Cisne Verde* between 4-12 April 1997 while fishing on the shelf slope to the north and north-west of South Georgia. Bottom depth ranged from 1240-1840m. The randomised sampling design outlined in Chapter 2 was used to sample fish from each day's catch: primary sampling units consisted of successive series of 10 coils of line, from which one series was randomly selected, and all fish caught on the series sampled for otoliths. A second sample of otoliths from 100 fish was collected haphazardly from longlines set in international waters near the Falkland Islands in the South Atlantic between 27-28 February 1997. Full biometric data were taken for both samples; otoliths were wiped clean and placed in paper envelopes to dry. The otoliths were prepared using the grinding technique outlined above.

For the first sample, two readers read the otoliths twice each. Each reading was completed in a single day, the order of otoliths was randomised between readings, and repeat readings occurred 7-10 days after the first reading. For the second sample, the sections were read twice at Old Dominion University and sent to the National Institute of Water and Atmospheric Research (NIWA) in New Zealand where they were read once using NIWA criteria. One otolith was crushed during preparation and another considered unreadable. All readings were made without auxiliary information or reference to any previous set of readings. As all fish in our sample were all taken in either February or April, the outer annulus was not counted.

Data from both samples were analysed using the first model in Section 3.2.3. As two readings were taken by each reader, the South Georgia sample was also analysed using the second model. All analyses were carried out using SAS Version 4.0.

3.3. Results

3.3.1. South Georgia sample

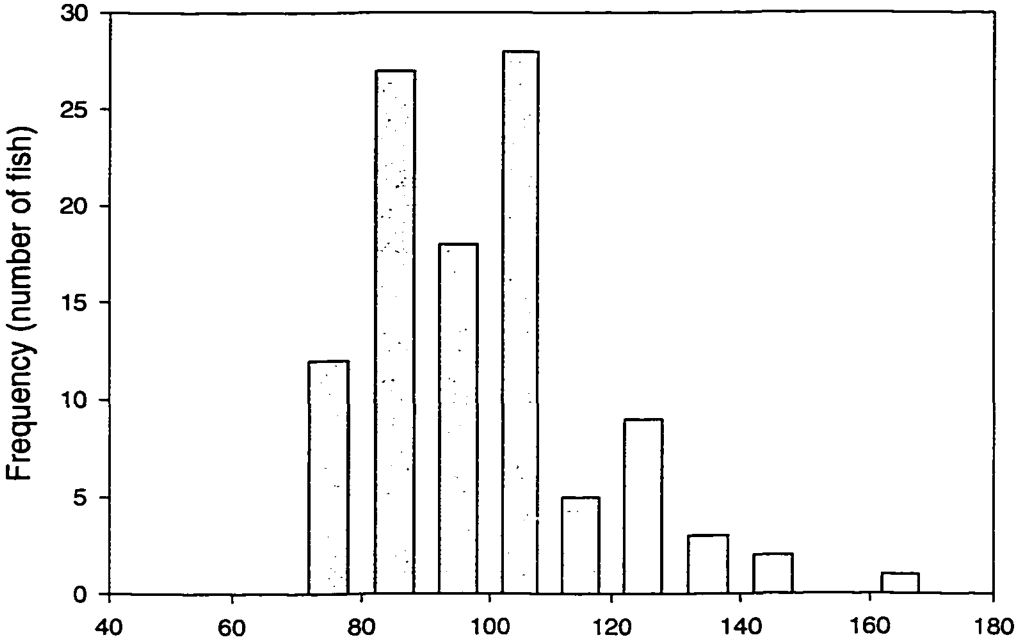
The distribution of TL in the sample is shown in Figure 3.2. Scatter plots of ages estimated within reader and between reader are shown in Figure 3.3. The analysis of variance is shown in Table 3.2. The residuals showed no trends with age or order of reading, but outliers occurred for readings from two fish: in one case, where a single value from a single treatment differed widely from the other three, the reading was winsorized, i.e. replaced with the nearest value (Sokal and Rohlf 1995), and a single degree of freedom subtracted from the calculation of mean square error. In the other case, all four values diverged: it was concluded that the otolith was unreadable, and the fish was removed from the sample. Remaining residuals were distributed normally and fulfilled the assumption of homogeneity of variances between treatments, tested using $F_{max} = s_{largest}^2 / s_{smallest}^2$ with $\alpha = 0.05$. Using the test for nonadditivity, the sum of squares for error was partitioned into a 1 degree of freedom sum of squares for nonadditivity and a residual sum of squares. The null hypothesis for no nonadditivity was tested with the statistic:

$$F_0 = \frac{MS(Nonadditivity)}{MS(residual)} = \frac{11.59}{3.57} = 3.25$$

The null hypothesis was not rejected with $F_0 < F_{.05,1,308} = 3.92$. The effects were therefore considered to fulfil the assumption of additivity (that is, of no interaction effect between individual fish and reading).

The one-way ANOVA indicated that the full linear statistical model was better than the reduced model for describing the data, and therefore that there were significant differences between age readings. Furthermore, the interaction effect between readings for each reader was examined using the two-way ANOVA and found significant, confirming that a single observer showed bias between readings. The HSD and SNK procedures were used for pairwise comparison of all treatment means: the readings taken

a) South Georgia



b) International waters

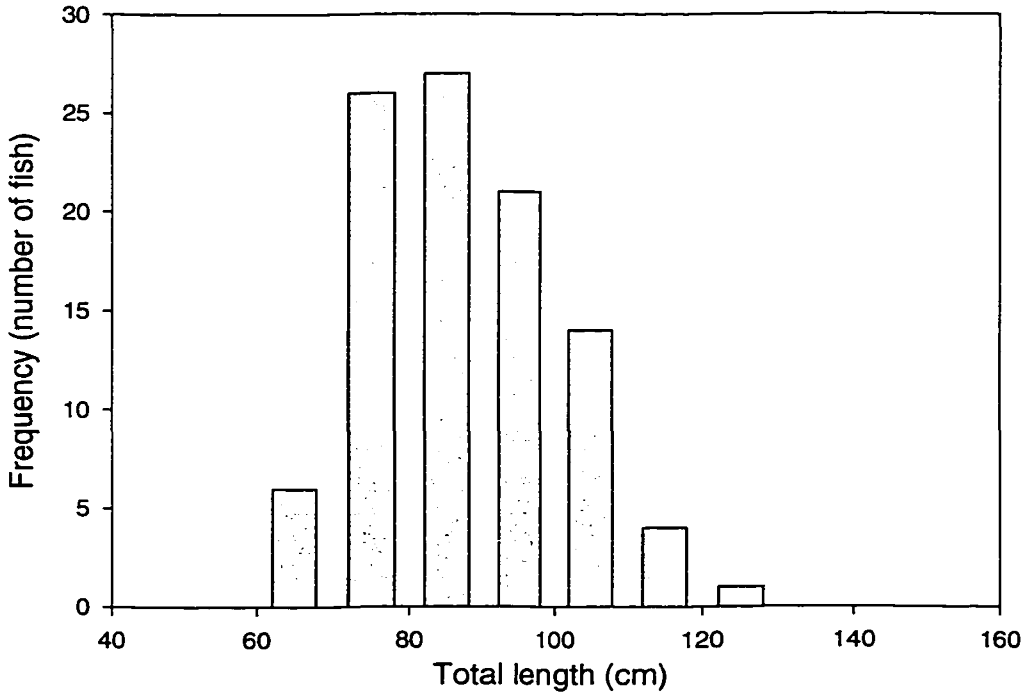


Figure 3.2: Total length of *D. eleginoides* sampled from the catch taken a) off South Georgia, and b) in international waters near the Falkland Islands.

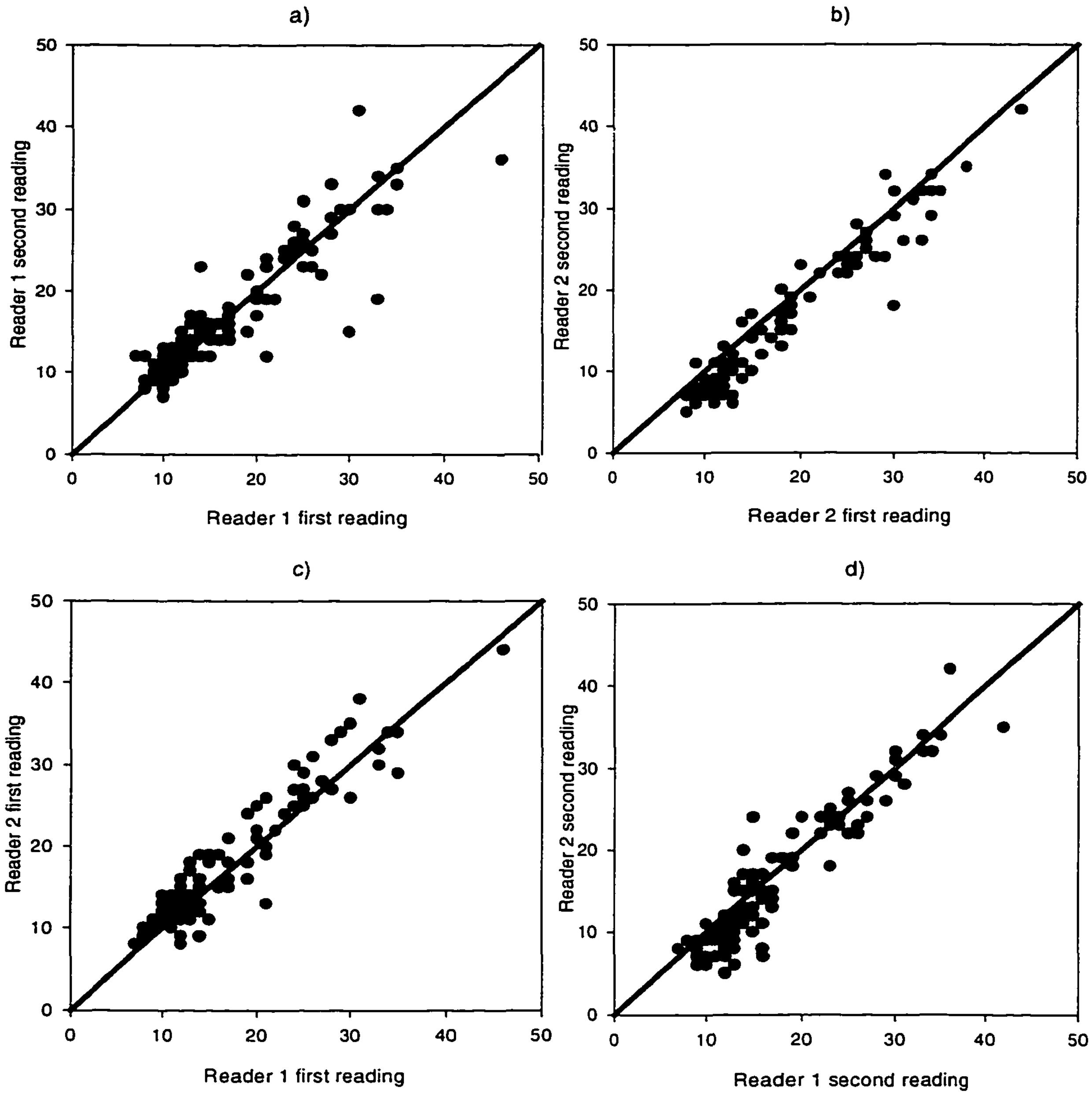


Figure 3.3: Age frequency plots summarizing pairwise comparisons of age estimates for South Georgia sample between a) two readings by Reader 1, b) two readings by Reader 2, c) first reading by Reader 1 and Reader 2, d) second reading by Reader 1 and Reader 2.

Table 3.2: Results from randomised block ANOVA, used to detect bias and estimate precision variability for repeated age estimates by two readers, for Patagonian toothfish sampled off South Georgia.

Source	*df*	*sum of squares*	*Mean square*	*F*	*Pr>F*
Total	415	26340.9			
fish	103	25033.6	243.0	67.6	0.0001
treatment	3	196.2	65.4	18.2	0.0001
error	309	1111.1	3.6		

by Reader 2 were found to be significantly different from each other ($\alpha = 0.05\%$), and from both readings taken by Reader 1. The bias shown by Reader 2 in his first reading was estimated as $\mu\text{-}\mu_{21} = 0.89$, and for his second reading as $\mu\text{-}\mu_{22} = -1.03$. The variation in precision of age estimation remaining after the bias had been accounted for was estimated by the MSE, $\sigma^2 = 3.60$. The power of the analysis was estimated in an *a posteriori* manner using:

$$\phi = \left[a - 1(\frac{MS_{among} - MS_{within}}{a \times MS_{within}} \right] = 3.59$$

Which, at v1 = 3 and v2 = 412 degrees of freedom, gives a power of c100%.

The South Georgia sample was also examined for sex differences in precision. Data were divided by sex, and each subset of data analysed separately with ANOVA (Table 3.3). Male and female populations were assumed independent. The residuals for both populations showed no trends with age, and the same outliers were eliminated as for the main analysis. In both cases, the remaining residuals were distributed normally, and differences between variances were not significant. For females, the second reading by Reader 2 was significantly different (HSD and SNK tests, $\alpha = 0.05$), indicating bias in age estimation. For males, both readings by Reader 2 were significantly different (SNK test, $\alpha = 0.05$), indicating bias in age estimation. I also examined if variances in estimating age were the same for males and females ($H_0: \sigma_m^2 = \sigma_f^2$) using the variance test:

$$F = \frac{\sigma_f^2}{\sigma_m^2} = \frac{3.84}{3.23} = 1.19$$

The null hypothesis was not rejected with $F_0 < F_{.05,189,117} = 1.47$. Taken together, these results indicated that Reader 2 changed his criteria between readings, particularly for males, but once this bias was accounted for, the precision of age estimation remained similar between sexes.

Table 3.3: Results from randomised block ANOVAs, used to test for differences in age precision between sexes. Data are repeated age estimates by two readers, for Patagonian toothfish sampled off South Georgia.

Females:

Source	*df*	*sum of squares*	*Mean square*	*F*	*Pr>F*
Total	255	20251.9			
fish	63	19409.9	308.1	80.1	0.0001
treatment	3	115.3	38.4	10.0	0.0001
error	189	726.7	3.8		

Males:

Source	*df*	*sum of squares*	*Mean square*	*F*	*Pr>F*
Total	159	5948.3			
fish	39	5483.1	140.6	43.5	0.0001
treatment	3	86.8	28.9	8.9	0.0001
error	117	378.4	3.2		

3.3.2. Sample taken from international waters.

Scatter plots of ages estimated by Reader 1 versus Reader 2 are shown in Fig. 3.4. The analysis of variance is shown in Table 3.4. As the NIWA reader only read the otoliths once, each reading was considered as a separate treatment. The residuals showed no trends with age or order of reading, but all values for one fish diverged: the otolith was concluded to be unreadable, and the fish removed from the sample. Remaining residuals were not distributed normally (Shapiro-Wilks test, $\alpha = 0.05$), and the data was square-root, log, and $\log_{10}$ transformed: the distribution of the square-root transformed data was closest to normal (Shapiro-Wilks test, $\alpha = 0.05$, $P<W = 0.0108$), but showed evidence of platykurtosis. Testing for equality of variances using $F_{max} = s_{largest}^2/s_{smallest}^2$, there was no significant difference ($\alpha = 0.05$) between treatments. As a consequence of the Central Limit Theorem, means follow the normal distribution more closely than variates so, with a large sample size (n>50), the ANOVA design is robust to departures from normality (Sokal and Rohlf 1995). Thus, the square-root transformed data was considered sufficently normal for the analysis to proceed. The null hypothesis for no nonadditivity was tested: F_0= 4.3282, which lay between the critical areas $F_{.05,1,189} = 3.92$ and $F_{.025,1,189}$. The effects were concluded to be marginally nonadditive, and there was some interaction between fish and reading.

The treatment effect was found to be significant. The HSD and SNK procedures for pairwise comparison of treatment means indicated that the second reading by the ODU reader and the reading by the NIWA reader were significantly different from each other ($\alpha = 0.5\%$), but the first reading by the ODU reader was not significantly different from the other two. The bias shown by the NIWA reader was estimated as $\mu - \mu_{11} = -0.028$, and for the second reading by the ODU reader as $\mu - \mu_{21} = 0.038$. The variation in precision of age estimation remaining after the bias had been accounted for was estimated by $\sigma^2 = 0.033$. Calculating power for the analysis in an a posteriori manner, $\phi = 7.04$, which at v1 = 2 and v2 = 294 degrees of freedom, gives a power of ca. 100%. These estimates are based on the transformed data, and the probabilities of statistical inference

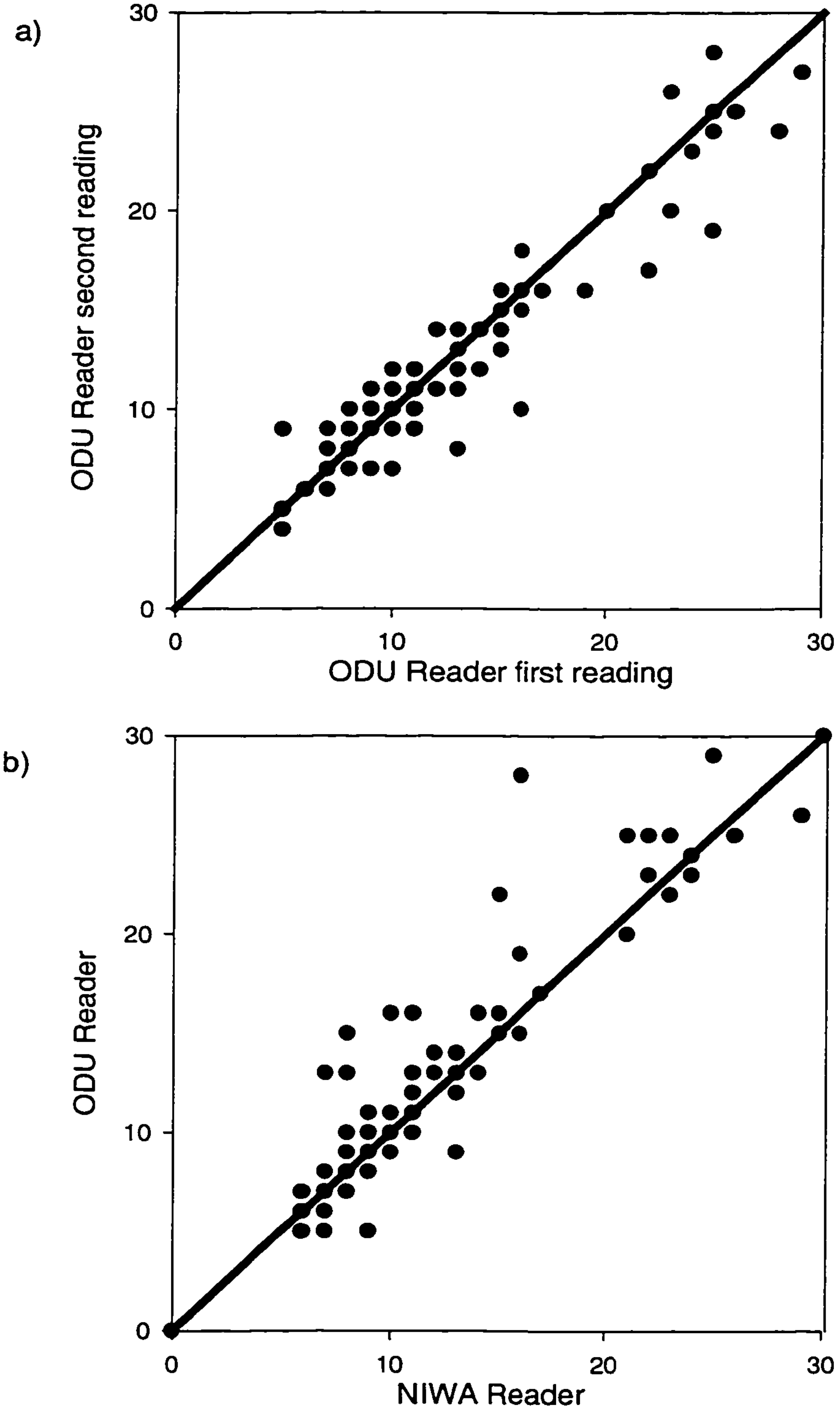


Figure 3.4: Age comparison plots summarizing pairwise comparisons of age estimates for sample from international waters, between a) two readings by the Reader from Old Dominion University, and b) one reading by the ODU Reader and the reader from the NIWA.

Table 3.4: Results from randomised block ANOVA, used to detect bias and estimate precision variability for repeated age estimates by two readers, for Patagonian toothfish sampled in international waters. Dependent variable is square-root transformed age.

Source	*df*	*sum of squares*	*Mean square*	*F*	*Pr>F*
Total	287	142.25			
fish	95	135.82	1.430	43.8	0.0001
treatment	2	0.23	0.114	3.5	0.0326
error	190	6.20	0.033		

apply only to the new scale of measurement, so cannot be directly compared to the sample from South Georgia.

3.4. Discussion

3.4.1. Preparation methodology and criteria

The grinding technique avoided the problem of locating the nucleus using features on the surface of the otolith. Otoliths could be oriented precisely, and the operator could monitor progress by frequent checks, so that the grinding plane could be accurately placed through the otolith nucleus and between crenellations. It allowed more flexibility in choosing a region of the otolith with clearer microstructure which, combined with prior baking, improved results considerably over the sectioning technique using light and SEM, and the break and burn technique previously tried. Preparation by grinding was slower than the other techniques, but the resulting slide was easier to handle and store than broken otolith halves, saving time in the later stages of processing. The better images may improve precision as well as reduce the number of unreadable images, so a given standard of precision can be achieved with fewer samples.

The otolith structure was highly crenellated and complex. It was frequently difficult to discriminate between checks, yearly annuli and structures corresponding to non-targetted time-scales. Within-otolith variability was large, and annuli frequently became compressed or obscured approaching crenellations. Although the criteria provided a framework for interpreting otolith structure, criteria were not easy to define objectively, resulting in subjective interpretation contributing to precision error. Considerable experience was needed before the criteria could be used with ease. On this evidence alone, care is needed to assess the error in age data for Patagonian toothfish and maintain quality standards.

3.4.2. The ANOVA model

The analytic model allowed bias and variability to be assessed within an ANOVA framework, avoiding the problems encountered by Kimura et al. (1979). The randomized block design with blocking on individual fish allowed error due to reading treatment to be isolated and integrated over a representative sample population. The model allowed the variation due to bias and precision to be accounted for properly, unlike the methods of percentage agreement, APE and CV. Precision bias could be examined without treating readings by one observer as replicates, avoiding Kimura's (1979) assumption of no within-reader bias. So long as the residuals conformed to the assumptions of ANOVA, the known properties of the model could be applied without improvisation, allowing unbiased estimation of reader bias and residual variability in a form that could be easily incorporated in further analyses of the age data. Blocking by fish also compensated for differences in age distribution: so long as the residuals show no trend with age, the reader bias and precision variability could be compared between samples of different ages. However, care is necessary in evaluating σ^2: the estimate includes error from all treatment groups, so within and between reader precision are confounded in the present design. To estimate within-reader variation, data should be sub-divided by reader and analysed separately for each reader. To estimate between-reader variation, separate pairwise analyses should be performed.

The performance of the model depends on how well data fulfill ANOVA assumptions. However, even when assumptions are not fulfilled, the model represents a useful alternative to age-bias plots (Campana et al 1995) and assymetry tests (Hoenig et al. 1995) for detecting non-linear bias. Conventional analysis of residuals using the model presented would detect these trends, with the advantage of allowing use of the numerous quantitative tests and techniques available in ANOVA for treating residual trends. Once identified, changes in bias can be examined by blocking. Data can be transformed if it departs from the assumptions of normality and homogeneity of variances, using power transformations or transformations to convert data with known distributions (Kuehl 1994). In the last resort, rank transformation may be useful. As long as fish are selected randomly from a defined population, the results from this approach can be used directly to provide population estimates of bias and residual variability.

3.4.3. The model applied to toothfish populations

For the sample from South Georgia, the analysis was straightforward: the ANOVA assumptions were fulfilled, including the absence of interaction between the effects due to individual fish and to reading. The hierarchical randomized design used for sampling covered between-line and between-SSU scales at which significant variation in TL occurred (see Chapter II), and inferences were made for a defined population, the catch taken by the B/F *Cisne Verde* between 4-12 April 1997. Treatments were significantly different, showing Reader 1 to be consistent, whereas Reader 2 changed criteria between readings, estimating age both higher and lower than Reader 1. Considering reading as a factor allowed treatments to be partitioned into factors of reader and reading, so that the bias between readings by a single reader was clearly demonstrated by the interaction term. Due to the positive interaction effect, only the simple effects between individual treatments were considered. These results demonstrate that for toothfish, repeat readings by one reader cannot be considered true replicates: in an analysis like that by Kimura et al. (1979), error would include biases between readings. In any analysis of variance, precision should be assessed between readings by each reader as well as between readers.

In contrast, for the sample from international waters, the results were less reliable: the sampling design was not randomized, undermining the fundamental assumption of independence, and there was no sampling frame, so the limits of inference for the analysis were not well defined. Although a probability-based sampling design would remedy both problems, residuals did not fulfill the assumption of normality, and the factors were marginally nonadditive. This indicated that distributions of residuals differed for the two samples. If distributions are stable for the two areas, this may indicate a difference in the precision of otoliths, due possibly to toothfish biology, but also means that estimates of bias and error in international waters near the Falkland Islands would frequently be based on transformed data. Even with the same treatment regimes for both samples therefore, comparisons between precision estimates for samples taken from catches in these international waters and off South Georgia may be difficult.

On the other hand, the differences in distribution may be methodological. Age estimates recorded for 85% of the fish by the NIWA reader and during the first reading by the ODU reader, fell within two years or less of each other, with the ODU reader consistently giving higher readings. However, differences were much higher for the remaining 15% of fish. Microstructure for these fish was difficult to discern, or occurred at different scales. In six cases, the NIWA reader interpreted several narrow opaque and translucent zones as multi-banded annual increments, whereas the CQFE reader interpreted each narrow paired opaque and translucent zone as an annual increment. The reverse was true for two other otoliths but, in all these cases, the distribution of repeated age estimates would be bimodal rather than normal, accounting for the platykurtotic departure from normality found in the data. As the contentious paired and opaque zones occurred in the region of narrow regular annuli, discrepancies in interpretation occured in older fish but not younger, accounting for the marginal interaction between fish and reading. Fine-tuning criteria so that readers count the contentious zones similarly may therefore result in assumptions being fulfilled for further precision analyses of fish from these international waters, allowing comparisons with South Georgia data.

In contrast, data from samples from the New Zealand sector prepared at NIWA by baking and sectioning and read once by the same NIWA and ODU readers (Horn 1999), indicated a similar residual distribution to the data from South Georgia. The residuals from the New Zealand samples fulfilled all assumptions, although readings were significantly different (α =0.05)(Table 3.5), with treatment means deviating from the overall mean by -0.54. Precision variability σ^2 was estimated to be 4.096, and using the variance test:

$$F = \frac{\sigma_{NZ}^{2}}{\sigma_{SG}^{2}} = \frac{4.10}{3.60} = 1.14$$

Table 3.5: Results from randomised block ANOVAs, used to estimate precision variability for otoliths prepared at NIWA. Data are repeated age estimates by two readers.

Source	*df*	*sum of squares*	*Mean square*	*F*	*Pr>F*
Total	111	2332.5			
fish	55	2074.0	37.7	9.21	0.0001
treatment	1	33.2	33.2	8.11	0.0062
error	55	225.3	4.1		

The null hypothesis of no difference in residual precision between the two studies was not rejected with $F_0 < F_{.05,111,415} = 1.32$, indicating precision to be similar between areas once bias had been accounted for. Note, however, that the comparison was not symmetrical: σ^2_{NZ} included only two between-reader treatments, whereas σ^2_{SG} included four treatments within and between readers.

As for the sample from international waters, ODU Reader 2 and the reader from NIWA used similar criteria when reading the sample prepared at NIWA, but Reader 2 consistently gave slightly higher ages than the NIWA reader, and may have been identifying structures near the nucleus differently. Kalish and Timmiss (1998) noted that there may be differences in the 'readability' of otoliths from different localities. However, the evidence presented here indicated that, once the effect due to bias was removed, the precision of readings from the South Georgia sample and the New Zealand sample were similar, even using different preparation techniques. More work is needed to separate out the effects of preparation techniques. Similarly, localities can be compared by preparing samples from both areas using one technique, and randomly mixing them in a single set for both readers to read.

3.4.4. Construction of reference collection

Data for the ANOVA consisted of four repeat age readings for each fish, from which the overall mean can be used as an estimate of fish age, and standard deviation and variance used as measures of readability for each otolith. Otoliths can be used in training new readers and re-familiarizing experienced readers. By randomly inserting otoliths in subsequent samples, sample readings for the inserted otoliths can be compared to their previously estimated means. The ANOVA model or a simple paired t-test can then be used to detect bias; and precision variability can be estimated to ensure the sample reading remains within defined bounds. Where bias is found to be significant, the reading can be adjusted by the discrepancy between the standard mean age and sample mean age of the inserted otoliths. Age criteria are difficult to define objectively; instead, the sample set can be used as a reference collection with standardized ages.

Some of the readings were significantly biased in both the sample from South Georgia and international waters. For the South Georgia sample, the overall mean could still reasonably be used as an estimate of age for a reference collection because the biases were small, the mean was not skewed by a large bias from one set of readings, and fell between the means of the two consistent readings by Reader 1. However, the estimates of dispersion incorporated the bias for the two readings by Reader 2. To remove the effect due to bias, Readings 1 and 2 by Reader 2 for each fish were adjusted by:

$$\tau = \bar{y}_{..} - \bar{y}_{2.}$$

The standard deviations and variances were then re-calculated (Table 3.6). A similar procedure was followed for the sample from international waters.

Repeated use of reference otoliths in this way will generate data that will allow the distribution of age estimates to be examined for individual fish. If normal, confidence intervals can be calculated for the age of each fish in the reference collection. Platykurtotic departures from normality, detectable by a simple Kolmogorov-Smirnov test, will indicate the need for fine-tuning criteria.

Recognition artefacts are a problem for collections as small as these. Repeat readings of more samples can be made with the original reference otoliths randomly inserted; once adjusted for bias, the new sample can be incorporated within the reference collection. Incorporating new otoliths will renew the reference collection regularly, reducing reader recognition due to discolourations or signs of wear.

3.4.5. Conclusion

During the course of this study, in response to debate within WG-FSA over the use of scales in age determination for Patagonian toothfish, I undertook further work (Ashford et al. 2000) comparing ages estimated using otoliths and scales. The results

Table 3.6: Means, standard deviations (SD) and variances (Var) for ages estimated twice by two readers for sample set taken off South Georgia.

Fish	*Mean age*	*SD*	*Var*		*Fish*	*Mean age*	*SD*	*Var*
1	19.5	1.25	1.57		26	13.0	1.37	1.87
2	11.0	0.83	0.69		27	12.3	1.22	1.49
3	14.0	1.86	3.45		28	9.5	2.34	5.45
4	11.3	2.96	8.77		29	8.0	2.69	7.23
5	42.0	4.25	18.10		30	12.0	0.81	0.65
6	18.5	3.69	13.64		31	9.0	0.05	0.00
7	26.3	1.54	2.36		32	12.3	2.96	8.77
8	14.3	0.98	0.95		33	9.0	1.44	2.08
9	23.5	1.77	3.12		34	11.0	2.01	4.02
10	11.0	0.81	0.65		35	11.0	1.95	3.79
11	12.5	0.55	0.31		36	21.2	1.71	2.93
12	10.3	0.48	0.23		37	20.8	1.73	2.98
13	17.0	1.20	1.43		38	24.5	0.98	0.96
14	33.3	3.45	11.89		39	9.5	1.02	1.05
15	31.5	2.39	5.73		40	9.3	1.87	3.49
16	27.5	1.24	1.55		41	23.8	0.94	0.88
17	32.2	2.16	4.69		42	16.0	0.86	0.74
18	27.8	2.66	7.05		43	10.0	1.86	3.45
19	25.3	2.39	5.71		44	9.5	0.55	0.31
20	26.0	2.69	7.24		45	11.5	4.08	16.68
21	34.0	1.13	1.28		46	9.8	1.25	1.57
22	36.5	4.51	20.36		47	13.0	1.84	3.38
23	21.8	2.18	4.75		48	31.5	1.28	1.64
24	31.8	2.08	4.32		49	10.5	1.88	3.53
25	27.5	2.97	8.84		50	8.0	0.81	0.65

51	27.0	2.21	4.89		78	9.8	0.93	0.87
52	28.5	2.43	5.91		79	16.3	1.67	2.78
53	12.8	0.93	0.87		80	11.8	1.68	2.83
54	25.8	1.72	2.98		81	11.3	1.54	2.36
55	10.3	1.71	2.94		82	15.3	2.09	4.36
56	11.5	1.27	1.60		83	14.5	4.33	18.79
57	11.5	0.60	0.36		84	24.8	1.27	1.61
58	9.0	0.81	0.65		85	9.0	1.43	2.04
59	22.3	2.66	7.10		86	9.3	1.51	2.27
60	15.5	1.90	3.60		87	15.0	1.44	2.08
61	11.3	0.96	0.91		88	10.5	2.66	7.05
62	19.3	1.52	2.32		89	10.8	0.52	0.28
63	12.0	0.77	0.59		90	31.3	2.10	4.41
64	11.0	1.13	1.28		91	11.3	0.92	0.84
65	16.8	1.27	1.63		92	13.3	2.88	8.30
66	12.3	0.99	0.97		93	12.0	0.81	0.65
67	13.5	1.88	3.53		94	25.0	1.45	2.10
68	14.0	1.99	3.98		95	24.5	1.27	1.60
69	16.3	1.87	3.49		96	23.5	0.98	0.96
70	11.0	0.81	0.65		97	15.5	1.28	1.64
71	11.3	1.46	2.14		98	31.8	2.11	4.44
72	16.8	3.11	9.68		99	16.3	2.22	4.94
73	8.8	0.98	0.96		100	16.0	1.19	1.43
74	18.8	1.54	2.37		101	16.3	1.31	1.71
75	16.8	0.51	0.26		102	11.5	0.95	0.90
76	11.8	0.45	0.20		103	15.0	1.39	1.93
77	10.0	0.86	0.74		104	21.8	1.92	3.70

demonstrated that fish were estimated consistently younger using scales compared to using otoliths, especially for older fish. As a result, at the CCAMLR meeting in 2000, the WG-FSA agreed that otoliths should be used in the future for estimating age in toothfish.

Although the otolith annuli laid down by older fish are less compressed than for scales and the age of older fish is therefore likely to be more accurately estimated, the results of the current study demonstrate that otolith structure is nevertheless complex. Care must be taken to interpret structures consistently, even for experienced otolith agers, and variability in readings must be quantified in a way that can be incorporated in decisions on age-based analysis and modelling (Kimura and Lyons 1991). The approach can be applied to error distributions that depart from normal using techniques commonly used in ANOVA, so that it is likely to be be applicable to other species.

Even given the complexity of the otolith, the results indicate substantial agreement over criteria between the age laboratories at Old Dominion University and NIWA. Horn (1999) has used mean CV and bias plots to demonstrate similarities in readings between NIWA and other laboratories estimating age in toothfish. The standardised age incorporated in the reference collection developed in this study is therefore likely to be reasonably consistent with other laboratories, and represent a CCAMLR standard that can proceed to be tested. Further work is needed to define this standard and improve consistency in age readings between laboratories. Further work is also needed to establish that different preparation techniques do not lead to differences in age estimation.

Extra precautions must be used when ageing catches of fish whose otoliths are difficult to read. Reading procedures can be experimentally manipulated to assess their influence on precision and accuracy (eg. Kimura et al. 1992). The sampling regime can be refined to achieve a target precision at minimum cost, or the lowest variance for a target cost (Lai 1987). The effect of remaining ageing error on management models can be assessed by sensitivity analyses (eg. Bradford 1991), or incorporated into ASA models

(Fournier and Archibald 1982), so that management decisions are taken which correctly reflect the level of uncertainty in the data.

In the previous chapters, in response to needs identified by the WG-FSA of CCAMLR, I have fulfilled two of the three requirements for an age estimation programme (Morison et al. 1998). I developed a probability-based sampling program, a preparation methodology that produces clear and consistent images for reading, and a quality assurance methodology to assess error in a statistically rigorous manner using a reference collection based on a conceptual model of the relationship between otolith structure and time. In the next chapter, I will turn to the third requirement and test the conceptual model using a validation test. In Chapter V, I shall use reference otoliths to control for data quality.

CHAPTER IV

AGE VALIDATION

4.1. Introduction

4.1.1. The problem

Accurate age estimation allows a powerful suite of models and other tools to be deployed to understand fish populations. But inaccuracies can result in biased estimates of age toward younger or older individuals. Despite warnings by earlier workers (eg. Van Oosten 1923, Hile 1936), Beamish and McFarlane (1983) found, in a survey of 500 studies published between 1907-1980, that 35% did not consider the possibility that ages may be incorrect, and less than 3% validated their age estimation technique for all year classes. In an additional sample of 75 studies published more recently between 1965-1980, only 40% mentioned validation, and none validated the full age range.

Incorrect ages can profoundly affect stock management. Beamish and McFarlane (1987) described a case for sablefish off the west coast of North America where, prior to 1981, fish caught in the commercial fishery were estimated to be 3-8 years using scales. However, an age method using otoliths was validated, and it demonstrated that fish ranged between 4-40 years, indicating slower growth and a much less productive fishery. Sustainable exploitation rates should have been 20-30% of those based on incorrect scale ages: management strategies using otolith ages resulted in a stable fishery (McFarlane et al. 1985).

Similarly, otolith age estimates of Pacific Ocean perch indicated fish were considerably older than the scale estimates used previously (Beamish and McFarlane 1983). Estimates of natural mortality using otoliths were considerably lower leading to a more conservative management strategy. The loss of wholesale value to Canadian industry due to over-exploitation before the correction was estimated to be $4 million in 1981 Canadian dollars. For white suckers, Stewart (1926) validated the yearly growth of scales in young fish; and based upon this limited validation, subsequent studies using scales indicated white suckers grew quickly to ages 4-7 years, and that few survived beyond age 9 or 10. Annual mortality was considered high after maturity even though active growth continued. Beamish and Harvey (1969) used a method based on fin-rays that had been validated for the full age range; they demonstrated that the oldest age was 23 years, and large numbers of fish in unexploited populations survived after growth slowed or ceased. Yearly annuli on scales were difficult to identify after five years, but were distinct on fin ray sections: validation of the full age range was necessary to prevent errors in age estimation and evaluate the importance of older fish as a component of the population (Beamish and McFarlane 1983).

Under-estimation of age using scales is usually due to compression of circuli in older fish, and Beamish and McFarlane (1987) listed examples of studies identifying species for which scales should not be used for this reason. Although less pronounced in otoliths, changes in incremental pattern with age result in compression in the annuli of older sablefish, which can obscure annuli, leading to under-estimation of age; older fish may also not form annuli every year (Beamish et al. 1983). Ages assigned beyond the oldest validated age should not therefore be assumed to be correct. Age estimation can also be inaccurate in younger ages: using otoliths, readers over-estimated the age of known-age juvenile sablefish in 50% of cases when they had no previous knowledge of age. Once knowing the age, 15% were still over-aged (Beamish et al. 1983).

4.1.2. Validation methods

Wilson et al. (1983) defined validation as ‘the temporal confirmation of an increment’, used to determine the accuracy of an age determination. Beamish and McFarlane (1983) gave a wider meaning as ‘proving a technique (of age estimation) is accurate’, where accuracy can be proven. Francis (1995) refined this meaning, proposing that a validation is the process of estimating quantitatively how accurate an age determination method is.

The four methods commonly employed as validations are marginal increment analysis (MIA), laboratory rearing, mark and recapture, and radiometric dating. For marginal increment analysis of annuli, samples are taken monthly for a year and the width of the outer annulus examined (Bagenal and Tesch 1978). The outer annulus is for each age group, and the mean monthly width is plotted by age group against month. If the annuli are yearly, a characteristic dip should be seen corresponding to where the new annulus begins to grow. To validate the full age range, the sample should include fish from all year classes, and be subdivided by year (e.g. Barbieri et al. 1994). For the second method, fish are reared in a laboratory or mesocosm so that ages are known. Known ages are then compared with ages estimated independently from hard parts. Thirdly, wild fish can be tagged, injected with tetracycline or another chemical to mark growth structures, and released. The period between tagging and recapture can be compared to the increase in age independently estimated from the deposition of the mark to the edge of the growth structure (eg. Beamish et al. 1983). Fourthly, radiometric dating techniques can be used, where age estimates are calculated from the ratio of radioisotopes in otoliths and are compared to independent estimates from the incremental structure (eg. Kastelle et al. 1993, Campana and Jones 1997).

The conceptual model of the relationship between otolith structure and time is used to predict annuli that are laid yearly (see Chapter III). MIA tests whether the annuli correspond to a year, conforming to the definition of a validation by Wilson et al. (1983). But the accuracy of the age data is inferred, not directly tested, and the test is graphical and subjective, with no statistical measure of accuracy. The other methods all address the accuracy of age data directly, and allow the use of probability-based statistical techniques to test and measure accuracy.

However, these direct methods depend on certain assumptions. Rearing studies assume that growth of reared fish will be representative of wild fish, allowing extrapolation of results. Yet growth structures and biological parameters are modified by laboratory conditions, and it is dubious to argue that the laboratory animals can be considered a random sample of the wild population. Mark-and-recapture allows the use of a randomized sampling design, but the marking process can affect growth and biological parameters like survival (eg. Beamish et al. 1983), so tagged individuals that have been recaptured may show bias with respect to the population. The technique also assumes chemical marks are laid immediately or at a known time after injection. Radiometric techniques avoid these assumptions but assume a constant specific activity of radioisotopes incorporated into the otoliths, a known initial activity ratio, and no external loss or gain to the otolith of any radioisotopes in the decay chain (Kastelle et al 1993). Departure from any of these assumptions introduces error into the 'known' ages estimated: known ages tend to be quite variable, and the technique has not been used to estimate the age of individual fish.

Other validation methods are sometimes used. Length-frequency modal analysis can be used where length modes are well separated and serve as proxies for age classes, particularly where dominant year classes occur. Modes should be followed between years to ensure they are yearly, especially when fish are exposed to cycles in the environment that alter their growth. However, accuracy cannot be directly estimated (Francis 1995), and modes usually become indistinct with age due to mortality and slowing growth, restricting validations to younger year classes. Similarly, counts of daily micro-increments can be used to validate annuli as yearly (e.g. Radtke and Hourigan 1990), but do not allow measurement of the accuracy of the age estimated by the technique, and depend on validation of the micro-increments: micro-increments must be deposited consistently every day throughout the year over the entire life. For hard parts analysis, the counts of increments are compared between calcified structures, typically with otoliths assumed to be accurate : the technique does not test directly for accuracy or timing, and cannot be considered a validation technique unless one of the hard parts has been independently validated. ^{14}C from nuclear explosions which was subsequently incorporated in otoliths has also been used to provide a reference date (Campana and Jones 1997); however, the method is expensive, fish samples must be available from fish living prior to the build-up of ^{14}C, and the reference date is subject to error due to rates of oceanographic dispersal and biological uptake that are frequently unknown. In black drum (Campana and Jones 1997) there appears to be a one-year delay between atmospheric ^{14}C and subsequent ^{14}C increases in coastal migratory fish.

4.1.3. Antarctic fish

Age estimation has been validated for only a few species of Antarctic fish, despite wide discrepancies in age estimates from different readers under different conditions (eg. Anon 1982; Coggan et al. 1990, Kock 1990). North (1988) examined the marginal increment of otoliths and scales from a suite of fish from the South Atlantic, and found evidence to support the hypothesis that adjacent opaque and hyaline zones in otoliths, and widely and closely-spaced sclerites in scales, represent one year of life. Burchett (1983) examined the marginal increment in otoliths of juvenile *Notothenia rossii*, and found that the opaque and hyaline zones were laid on an annual basis. Ashford and White (1993) found similarly for otoliths from juvenile *Notothenia coriiceps*. Daniels (1983) found the hyaline zone in otoliths of *Harpagifer bispinis* was readily visible in samples taken in autumn and winter, and an opaque border became visible in most otoliths collected in September. Micro-increments have been shown to be daily in *Trematomus newnesi* (Radtke et al. 1989) and *Nototheniops nudifrons* (Radtke and Hourigan 1990), and have been used to age adult fish.

In Patagonian toothfish, no validation has yet been published, but Kalish and Timmiss (1998) presented a preliminary report at the 1998 CCAMLR meeting, in which they examined the level of bomb carbon (^{14}C) present in the nuclei of otoliths from large fish captured recently. Using ages estimated independently by readers, they indicated that the temporal pattern of carbon deposition in the nuclei conformed to the temporal pattern of release from nuclear explosions, indicating that age estimates were accurate.

4.1.4. This study

Of the four common methods used, mark and recapture was not considered feasible: the logistics were too demanding, costs too high, and the required time-frame for results too long. The poor condition of toothfish captured by longline and the rate of illegal fishing risked a high mortality of tagged fish and a poor rate of return. Similarly, laboratory rearing was impracticable because of the logistics and time needed to cover the full age range.

Radiochemical techniques give good results over long periods of time but tend to be less accurate distinguishing age of young fish for which age estimation has proved most difficult. They are also costly and demanding on laboratory infrastructure. MIA is effective for earlier year classes where otolith growth is fast, but can be less so where annuli growth is reduced with age and mortality reduces the number of older fish available. MIA provides a more limited test of the conceptual model and demands a greater sampling effort. However, MIA was selected because it was cheaper and the requisite equipment was available; samples could be obtained through the legal toothfish fishery, and were likely to give good results for the younger age classes contributing most heavily to the fishery.

In this chapter, I present results of a MIA used to test the conceptual model for toothfish developed in Chapter III. I put forward a methodology with a view to putting MIA on a more quantitative footing, and discuss the implications of the results. I further discuss some conceptual problems with MIA, and suggest ways in which known-age methods can be used to test and estimate the accuracy of the conceptual model within an ANOVA framework.

4.2. Methods and Materials

Samples were obtained opportunistically from collections made by fishing authorities within the CCAMLR jurisdiction and off southern Chile. All collections were made purposively, mostly using longlining but some by trawling. To ensure coverage of the full age range, total lengths were divided into 10cm intervals from 10-140cm and 20cm intervals from 141-180cm and over 180cm; otoliths from 10 fish from each interval were selected within each monthly period. Despite this, coverage of ages was not consistent between all months: even where only longlining was used, the size range captured varied between areas, and the age range was distributed differently between areas and months, even within age blocks.

Otoliths were processed using the grinding method described in Chapter III. All sections were randomly sorted and examined using a Leica MZ8 microscope. Age was estimated using the conceptual model described in Chapter III, and the outer two annuli measured from the inner edge of the hyaline zone, using the Optimas image analysis package. My scientific question was: are annuli in the otoliths of toothfish laid yearly? My hypothesis was that each combination of opaque and hyaline zones was yearly, created by faster growth of the opaque zone relative to the hyaline zone; therefore the mean width of the outer annulus for each year class should be larger for consecutive months until the annulus is fully deposited, after which it should be considerably smaller as a new annulus starts to develop. If annuli correspond to more or less than yearly intervals, this trend should not occur and the hypothesis should be rejected.

Data were grouped by 5-year age classes (1-5, 6-10, 11-15, 16-20, 21+ years). To compensate for variation in the age and size ranges of samples between area and month (and thus the effect of individual fish rather than month on the width of the outer annulus), the ratio of outer/inner annulus width for the two outer annuli was calculated for each fish. The means of the ratios pooled for all blocks were estimated for each month with standard errors, and the estimates were plotted by month. To examine effects due to age, the means of the ratios and their standard errors were estimated by block for each month, and similarly plotted by month.

As inferences from the graphical analysis were qualitative and decisions on the hypothesis tests largely subjective in nature, further tests of differences between months in the width of the outer annulus were attempted using ANOVA. The statistical question addressed was: are there significant differences in the width of the outer annulus between months, consistent with annual deposition of the outer annulus? Each month was considered a fixed treatment effect in a comparative observational study. Because data were missing from cells between months, a fixed blocking factor consisting of each 5-yr age class was not used; the discrepancies were haphazard and did not allow a balanced incomplete block design instead. To compensate for the effect due to individual fish, the ratio of the widths of the outer two annuli was used as the response variable. Proceeding with a simple one-way analysis for unequal replications, let y_{ij} be the ratio for a fish taken in the *i*th month. Under the assumptions of the fixed effects model:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ijk} \qquad (56)$$

where μ = the general mean

τ_i = the effect of the ith level of the factor month

ε_{ijk} = the experimental error.

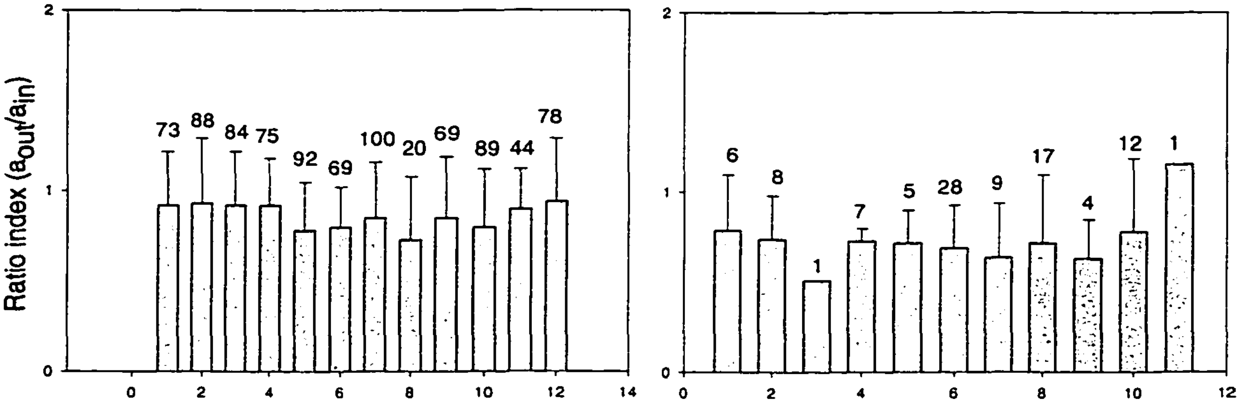
The assumption of normality of residuals was tested using the Kolmogorov-Smirnov test, and the assumption of homogeneity of variances was tested graphically. Individual treatment comparisons were made between months by pairwise HSD and SNK tests.

4.3. Results

The provenance and date of each sample are shown in Table 4.1. Most samples were taken by longline and younger ages were not taken, whereas the sample for May from Kerguelen was taken by trawling and older fish were not sampled. In July, only 19

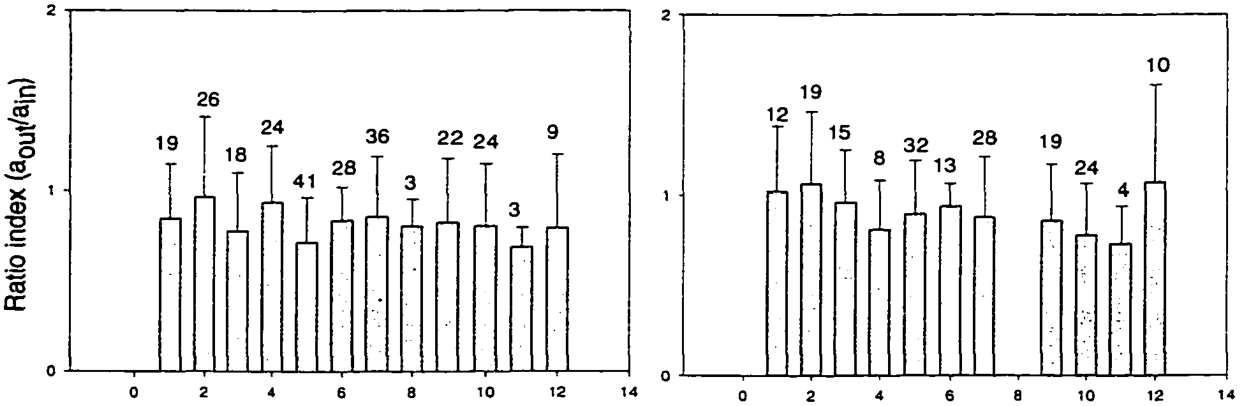
Table 4.1: Provenance and date of monthly samples used to test for yearly annuli by marginal increment analysis of the otoliths of Patagonian toothfish.

Month	*Year*	*Fishing Area*	*N*
December	1995	Southern Chile	73
January	1996	Southern Chile	88
February	1996	Southern Chile	84
March	1994	South Georgia	75
April	1997	Falkland Islands	92
May	1995	Kerguelen	69
June	1997	S. Atlantic international waters	100
July	1982/84	Kerguelen	20
August	1997	Prince Edward Islands, S. Ocean	69
September	1997	Prince Edward Islands, S.Ocean	89
October	1997	Prince Edward Islands, S. Ocean	44
November	1997	Prince Edward Islands, S. Ocean	78



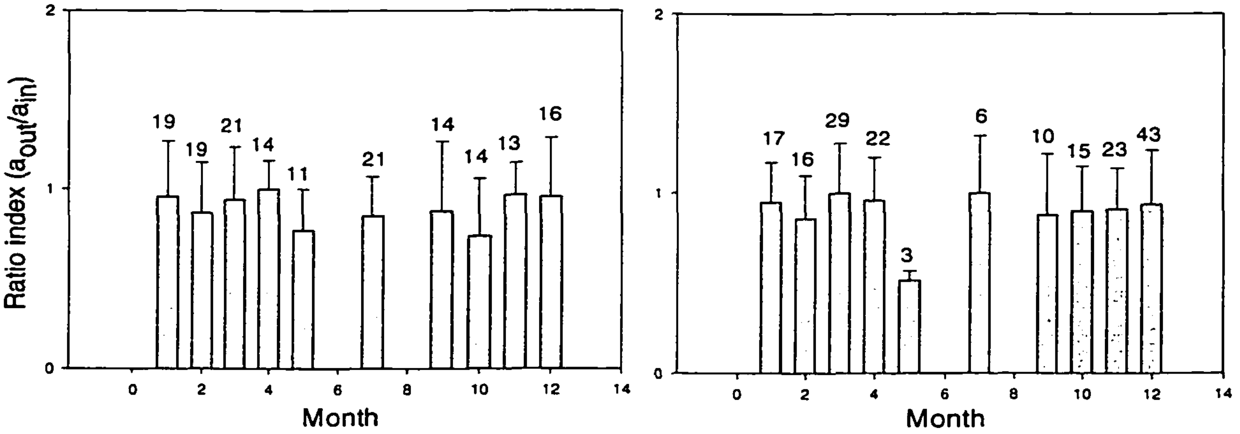
a) All fish pooled

b) Ages 1-5 yrs



b) Ages 6-10 yrs

c) Ages 11-15 yrs



e) Ages 16-20 yrs

f) Ages over 20 yrs

Figure 4.1: Plots of mean of increment width ratio index versus month for a) all ages pooled, and b-f) 5-year age classes.
Month 1 = December.
Ratio index = (width outer anulus)/(width inner annulus).
Columns represent means, bars represent the standard error of the mean, *n* is shown above the bar.

fish were sampled, by a sampling trawl during a non-commercial survey at Kerguelen, and samples were missing for all year class groupings except years 1-5.

For the graphical analysis, results for the pooled data are shown in Figure 4.1a. Some evidence of a decrease in the mean of the ratio of the outer/inner annuli was found between April and September, which may be interpreted as consistent with the trend predicted. However, results by age-class grouping (Figure 4.1b-f) did not show a similar effect. For ages 1-5, the mean ratio of outer/inner annuli decreased from December to August, and increased from September. For ages 6-10, there was little evidence of a trend except generally slightly downwards. For ages 11-15, there was a trend downwards, with some evidence of an increase in November. For ages 16-20, there was no evidence of a trend. For ages 20+, there may have been a decrease in April. In none of the year-class groupings did the data show a pattern in the ratio of outer/inner annulus width consistent with that hypothesized for a yearly annulus.

For the ANOVA model, residuals fulfilled the assumption of homogeneous variances and were normally distributed after removal of a single outlier ($\alpha = 0.05$). Significant differences were found between months ($\alpha = 0.05$) (Table 4.2). The HSD and SNK comparison tests showed a grouping for all means except July, and another grouping for means for April to September (SNK) or October (HSD).

4.4. Discussion

The results from the pooled data analysed graphically and by ANOVA demonstrate some slowing of growth as predicted from my hypothesis, indicating that annuli may be laid down yearly. However, the differences between means in the ANOVA is much less than predicted and the results are further weakened by the lack of expected trends in the graphical analysis by age-class grouping. Overall, the evidence to validate annuli as yearly is inconclusive.

Table 4.2: Results from one-way ANOVA, used to test for differences in marginal increment width between months, for the otoliths of Patagonian toothfish. Dependent variable is the increment width ratio index $=(a_{outer}/a_{inner})$.

Source	*df*	*Sum of squares*	*Mean square*	*F*	*Pr>F*
Total	880	83.5			
month	11	3.4	0.31	3.34	0.0002
error	869	80.1	0.09		

Sampling throughout the year in the Antarctic is challenging and this collection of toothfish otoliths is the first with year-round coverage. Despite this, the limits on the sampling possible meant that the inferential limits of the study were unclear: the frame could be viewed as the whole species range with poorly defined sampling coverage; or it could be viewed as only those fish sampled, because other fish have zero probability of being sampled. Sampling from different years also made temporal scope unclear.

There were several sources of variation that were ultimately not accounted for in the design. Although a ratio was used as the *y*-variable to address effects due to individual fish allowing cohort data to be pooled (both overall and within the 5-year age groupings), the design assumed a linear relationship in the slowing of growth rates between succeeding years over the age range sampled, and demanded that, outside this relationship, individual fish changed little in the timing and rate of growth between years. It also assumed that the population remained constant during the year, with no unattributed effects due to migration, fish availability, or behavioural changes in bait uptake. Where these assumptions did not hold, distribution of age between monthly samples, rate of growth, variation within otoliths, and geographic area may all have acted to influence MSE, bias treatment means or confound the effect due to treatment. Most importantly, interpretion of the edge of the otolith was elusive due to light and edge artefacts (eg. North 1988), contributing reader error that may be biased or random but was not quantifiable.

4.4.1. the ANOVA model

Although likely error due to several sources meant that the validation test was not conclusive, the ANOVA block design provided a quantitative approach, allowing a probability-based test of monthly effects that improved on the subjectivity of the graphical approach. For a more conclusive validation, the ordering of treatments in the multiple comparisons tests should demonstrate significant differences between months either side of the transition between one annulus and the next, with a sliding set of relationships between consecutive months as the new annulus grows.

The experimental design for the ANOVA could be improved in several ways. Selecting fish by age from a previously aged sample population would avoid empty cells and achieve a balanced design, but implies a prior large-scale programme of year-round representative sampling and age estimation. For a comparative observational study using ANOVA, units should be randomly selected from within each treatment population (Kuehl 1994). A consistent sampling design between months with defined frame and randomized selection of sampling units would clarify the inferential limits, guard against spatial error, and avoid confounding the treatment effect. Sampling within a single fishing area would restrict the inferential limits to that area, whereas stratified sampling across areas would increase the scope. Sampling in consecutive months would limit the scope to the year of sampling: the inferential limits could be extended temporally by repeat sampling each year or in years selected randomly from within a frame of possible years. This would be especially appropriate if conditions change (e.g. through fishing mortality or low prey availability, leading to changes in growth and more stress events that may affect annuli appearance and deposition).

4.4.2. Conceptual considerations for MIA in relation to other validation methods

Francis (1995) has pointed out some confusion from which validation studies in general have suffered. The main methods used in validation studies address different scientific hypotheses: MIA tests the hypothesis that predicted annuli are yearly, whereas known-age methods directly test the hypothesis that age estimations are accurate. Secondly, the methods are often used for somewhat different epistemological purposes: hypothesis-testing as against measurement of the accuracy of age data.

Implicit in Francis's (1995) critique is a confusion over the continuing attempt to verify or prove that the ages estimated for a fish population using a particular aging methodology are correct (eg. Beamish and McFarlane 1983). This is embodied in the label of 'validation' for the class of tests, but the approach is inconsistent with current ideas of hypothesis-testing (see Chapter I). The confusion may be historical in origin: the question of validating age data arose in an epistemological context that sought to verify hypotheses, frequently with non-quantitative tests. MIA was among the earliest methods to be used to test the accuracy of age estimation and conformed nicely to this context. The known age approach is essentially a process of falsification: under the ANOVA approach, the scientific hypothesis is that age is accurate, consistent with the statistical null hypothesis that there is no difference between the two populations of known age and estimated age. An attempt is made to reject this using a probability-based test of significance. However, although hypothesis-testing is certainly of interest, if the purpose is to quantify accuracy, this is measured by the distance between the means of the two age sets, and is essentially a question of estimation. Maximum likelihood estimates have the most desirable properties (Edwards 1992, Hilborn and Mangel 1997): if the age data is distributed normally, they are the same as means estimated by least squares, but if the distribution is other than normal, maximum likelihood estimates are better for measuring accuracy.

The graphical test for MIA can easily be converted over to a falsification approach. The distinction is largely one of semantics, but plots may be interpreted differently if the purpose is to reject rather than support a hypothesis of yearly deposition of annuli. However, in the quantitative approach to MIA outlined, the logical structure is not straightforward: the null hypothesis (of no difference between months) must be rejected for the scientific hypothesis (that age is accurate) to survive. This is less elegant than the known-age approach; more importantly, the linkage between the scientific and null hypotheses is weaker, and it risks rejecting the scientific hypothesis illegitimately through lack of power or the influence of nuisance variation.

MIA also suffers from confusion over the periodicity it tests. Periodicity of annuli can be sub-annual or supra-annual, especially if the population is influenced by environmental cycles with a period greater than a year. A single transition between annuli will not discriminate between annual and supra-annual periodicity; whereas no transition will not discriminate between lack of periodicity and supra-annual periodicity sampled between transitions. Annual periodicity can only be discriminated by sampling two consecutive transition events. On the other hand, if annulus growth occurs as a short period of fast growth (eg. Ashford and White 1993, 1995), monthly samples may not fully capture the change, and the hypothesis of yearly periodicity will be illegitimately rejected.

A final source of confusion is when MIA is presented as an observation of the progression of the marginal increment through the year, rather than as a test of a predicted annulus from a prior model. It is difficult to see how a study can be conducted *tabula rasa*, and a prior definition of an annulus is in fact usually being tested but with a lack of logical and statistical rigour: frequently, samples are read non-randomly with the month of capture known, and the results are confounded with the reader's conscious and sub-conscious biases.

4.4.3. Conclusion

The ANOVA model presented may provide a more quantitative footing for MIA, but reader interpretation of the edge effects observed in toothfish and the difficulties of collecting samples in the Antarctic within a well-defined sampling frame, mean that MIA will likely remain problematic for toothfish. In any case, MIA remains a rather cumbersome test of a limited hypothesis. Furthermore, it has no direct quantitative measure of variation in age estimates which can be integrated into a framework for assessing error in population age data.

For validation of estimated ages in toothfish, further work would be better concentrated on a known-age method which allows repeat readings by different observers, and estimates of bias and precision related to known age instead of a standard. The ANOVA model developed in Chapter III can be used, with known age as a control treatment in addition to the reading treatments, allowing error due to missing yearly annuli or sub-annual checks to be accounted for, and all sources of error integrated into a single set of estimates.

Of the known age methods, laboratory rearing of toothfish is likely to remain impracticable, not least for the time scale needed to validate the full age range. Risk of mortality during such a long period would be high. Although fish might be kept in tanks at island stations in the Southern Ocean and marked at a known date, older fish will be needed, and capture, transport to the facility and aquarium management are likely to be prohibitively expensive. In contrast, mark and recapture can be performed from research vessels at sea and sampling can be by randomized design so that, unlike laboratory rearing, the results of the validation test can be inferred for the population sampled. Mark and recapture has the further advantage over radiochemical techniques that its assumptions can be tested more easily. Tests of dosage schedules can be used to minimize mortality associated with the capture and marking process and estimate the period between injection and deposition of the mark in the otoliths. Logistically, the condition of captured toothfish could be improved by marking from research vessels using trawls which now operate regularly in most management areas. The substantial increase in fishing recently in several management areas means that the age range in the fishery is likely to be less, reducing the time scale needed for the project; the continued fishing means that mortality is higher than in the population prior to the increase in exploitation, resulting in higher yearly tag returns. The decline in biomass and improved enforcement of regulations has meant that illegal fishing has declined within the CCAMLR area and Falkland Islands conservation zones; combined with improved monitoring in both areas through observer programs, recapture rates are likely to be reasonable. Evidence of a population boundary between southern South America and the CCAMLR area (see Chapter V) means that recapure rates within the latter are likely to be better than off the Falkland Islands where the tagged fish may migrate into waters where enforcement and monitoring programs are not as effective. The question remains however, whether enough fish could be marked to ensure the returns necessary for statistical analyses: although the 30% of the population necessary for good biomass estimates (Seber 1982) is probably not possible under present logistical and cost restraints, it may be possible for the purposes of estimation of age precision and validation, identifying dispersal range, and calculating individual growth rates.

Ideally, if a sufficiently large proportion of the toothfish population were marked each year in an ongoing tagging programme, subsequent samples would include marked fish, allowing age error to be estimated directly for the target population. However, a tagging program of this size would be expensive; smaller programs instead would mean fewer known-age fish and few opportunities to repeat validations over time. However, reference collections for quality control programs can be built cheaply using toothfish without known ages, if standardised ages of recaptured fish can be estimated and tested against their known ages, allowing reference collections to be corrected and expanded. Thus, otoliths from toothfish of standardized and known age can be randomly mixed in a single set to be read twice each by two readers. The data for standard age and known age fish can then be treated separately: using the ANOVA model from Chapter 3, standard age otoliths can be tested for bias between the standard and estimated ages, and known age otoliths for bias between known age and estimated age. Residual variance for both analyses can also be compared and tested to see if they are significantly different under the same conditions, and whether there are significant sources of random variability not accounted for in precision analyses using standard ages.

CHAPTER V

STOCK STRUCTURE INDICATED BY GROWTH: DOES GROWTH VARY BETWEEN FISHING AREAS?

5.1. Introduction

Most models available in stock assessment assume discrete stocks which can be managed as separate units. Gulland (1983) highlighted the importance of selecting stocks at appropriate spatio-temporal scales to allow treatment as if homogeneous and independent without unacceptable error. The scale must not be too large to neglect important within-stock differences: if several sub-groups behave differently and movement is restricted, pooling data may seriously misrepresent the dynamics of one or more sub-group, especially if fishing effort is variable spatially. But the scale must be sufficiently large to reduce the importance of interactions with fish outside: if too small so that interactions with fish in other unit stocks are large, management may be compromised by contrary management choices taken for other stocks.

The concept of a unit stock, however, has had an erratic history, marked by considerable differences over definition and the weight of factors to consider in delineating a stock. Booke (1981) has pointed out that the word has been used to describe groupings at species, race, population and sub-population levels; and has shifted in meaning from figuratively the trunk of a tree (as opposed to the roots or branches), used historically to signify the source of a line of descent. In the twentieth century, Marr (1957) defined a stock as a group of fish ‘characterized by similarities which are not heritable but are induced by the environment’, and which may contain members of several different sub-populations. More recently, the meaning came to include genetic similarity and was explicitly incorporated in a population biology framework. Thus, Gulland (1971) defined the unit stock as ‘a self-contained and self-perpetuating group, with no mixture from the outside, and within which the biological characteristics and impact of fishing are uniform. Such a unit stock would also be a genetic unit.' However, the importance of management considerations was emphasized by Larkin (1972) in his definition of a stock as 'a population of organisms which, sharing a common gene pool, is sufficently discrete to warrant consideration as a self-perpetuating system which can be managed'.

At the Stock Concept International Symposium, Booke (1981) emphasized the life history cycle as a stock property and incorporated a spatial dimension in his general definition of a stock 'as a species group, or population, of fish that maintains and sustains itself over time in a definable area'. He further gave a more precise genetic definition : 'a population of fish maintaining and sustaining Hardy-Weinberg equilibrium'. At the same meeting, Ihssen et al. (1981) added a temporal dimension, defining a stock as 'an intra-specific group of randomly mating individuals with temporal and spatial integrity,' while Maclean and Evans (1981) argued for a genetic perspective in management decisions, recognizing that fish species are sub-divided into local populations with genetic differences that are adaptive. They further pointed out geographical, ethological and ecological barriers that restrict dispersal and gene-flow, and argued that managers needed to better understand the processes governing population structure and isolation, in making the compromises necessary between the biological and socio-economic considerations in a fishery.

More recently, evidence has increased that environmental, as well as biological, factors may influence the structure of populations. Large- and meso-scale dynamic processes, in conjunction with behaviour, may influence patterns of larval and adult distribution, and geographic boundaries like fronts may isolate populations (e.g. Loeb et al. 1993, Kingsford 1993). Front dynamics provide possible mechanisms for return transport of larvae: for example, southward flowing shelf water off the east coast of the United States is entrained by the northward flowing Gulf Stream, forming a narrow band of northward flowing shelf water (Lillibridge et al. 1990). On the other hand, larval mixing can occur across oceanic fronts through the formation of warm and cold core rings from front undulations, or intrusions of slope water along the pycnocline (Cowen et al. 1991, Myers and Drinkwater 1989).

In the Southern Ocean, fronts in the ACC cross the distribution range of Patagonian Toothfish between the Falkland Islands and South Georgia, and across the Kerguelen plateau. These may act to isolate populations within environments that differentially influence population vital rates. Meso-scale processes also play an essential role in the distribution of water masses in the Southern Ocean. Rings have been observed associated with the Polar Front (Gordon 1988), and meander-like displacements of warm and cold surface waters lead to intensification in the flow of water from the Drake Passage to the Bransfield Strait (Stein 1989). Cross-frontal injections of newly formed Antarctic Intermediate Water have been recorded in the Crozet Basin (Park and Gamberoni 1994), while eddy activity along the frontal zone north of the Kerguelen Plateau likely enhances cross-frontal exchange of water masses (Park et al. 1993). These processes may provide a mechanism for transport and mixing. Hofmann et al. (1998), using oceanographic models of the South Atlantic, demonstrated that krill off South Georgia may recruit from spawning areas off the Antarctic Peninsula, using fronts within the Antarctic Circumpolar Current for transport along the southern edge of the Scotia Sea. Strong ACC current flow eastward may influence movement of different toothfish life stages (SC-CAMLR-XIV, Annex 5): mixing may occur between areas, and be directional rather than random, with recruitment from upstream sites.

Several examples of how environmental processes influence the structure of fish populations are summarized by Cushing (1981). Arctic cod larvae hatch in northern Norway and ride the West Spitzbergen and North Cape currents until they reach the Svalbard shelf and the banks of the south-eastern Barents Sea, where they settle. Adolescent fish move into deeper water to join the adults in their spawning migration, returning to their spawning site in northern Norway. The spawning migration may be directly against the currents or fish may ride in a counter-current on the shelf edge. The population is contained geographically by the currents; larvae arise on a single spawning ground and continuity, or coherence, is maintained between generations. Similarly, European eels spawn in the Sargasso sea and elvers drift across the Atlantic within the North Atlantic gyre to metamorphose and move up rivers, returning finally to sea as adults for their spawning migration. American eels spawn nearby to the west and use the gyre for transport, but grow quicker than the European eel and metamorphose after one year instead of 2.5 years, reflecting the shorter time needed to reach the North American continent. North Sea herring occur in four spawning groups with differences in meristic characters and in vital rates such as recruitment, growth and mortality. However, adults from three of these groups appear to overlap spatially in feeding grounds in the central North Sea. Cushing (1981) shows that, if herring migrate in tidal streams using selective tidal transport, the differential spawning grounds would nevertheless segregate the populations throughout the life cycle.

The environment often drives discrepancies in vital rates between segregated populations. Mortality rates through predation and dispersal may vary in time and space and with life stage, differentially influencing the life history pattern. Growth rates are influenced by resource availability and temperature (e.g. Clarke and North 1991), which also vary in space and time. Temperature can influence movement: for example, Arctic cod may be limited in their movement into the Barents Sea by the 2°C isotherm, allowing a coridoor which would vary in size with season and year (Cushing 1981). If isolation persists sufficently over time, divergent selection pressures will lead to differential genetic adaptation: life history strategies and vital rates shift towards those sectors of each population most adapted to take advantage of their specific environmental context.

On the other hand, linkages between populations through movement and mixing may homogenize genetic differences, or modify the vital rates and persistence of linked populations (e.g. Levins 1969, Pulliam 1988, Polachek 1990). Gauldie (1991) has argued that for fisheries, stocks linked by migration and with heterogeneous growth rates may be more realistic than the idea of isolated stocks with homogeneous growth rates. Detailed knowledge of structure and movement within and between populations is therefore critical in delimiting the domain and understanding the limitations of a model when applied to a particular stock.

5.2. How to examine population structure?

A number of methods have been developed to elucidate stock structure in fish populations (Ihssen et al. 1981, Maclean and Evans 1981). Meristics and morphometrics can be used to distinguish between groups in different areas. Discrepancies in population or physiological parameters between populations, or differences in otolith shape and increment patterns, may be used to infer stock boundaries. But the methods may confound genetic and environmental factors: differences may be due to ecological separation during part of the life history rather than reproductive isolation.

Instead, stocks can be distinguished by genetic differences. However, mixing rates of 10% are sufficient to mask distinctions between populations, yet low enough to necessitate managing the populations separately, especially when large environmental differences separate mortality and life history characteristics. Genetic differences accrue slowly through evolution, especially if selection forces on two separated populations are similar: the most sensitive techniques for genetic discrimination, using micro-satellite DNA, cannot detect events occurring more recently than several thousand years ago (Anon 1994). Yet environmental changes, including long-term cycles governing the extent of pack-ice from the Antarctic continent (Murphy et al. 1995, White and Peterson 1996), may lead to variations in population structure and mixing on a time-scale that would not be detectable by genetic differences, but should be accounted for in management.

Mark and recapture methods give information on movement and mixing in space and time, and between populations, but assume random mixing of tagged and untagged fish, good reporting, no loss of tags, and no changes in fitness of tagged fish (Ricker 1975). Although reporting can be improved using rewards, 30% of a population needs to be tagged to achieve the required level of returns to estimate population abundance (Seber 1982). Long-lived species of low fecundity are characterized by low mortality rates, so that returns occur over a long time period, limiting the usefulness of the method.

Studies have been largely restricted to adult populations because of logistical problems in marking sufficient numbers of young with tags to achieve good returns (Cushing 1981), and the legal prohibition in the United States of chemical markers because of the possible exposure to human consumers. As a result, in studies of population structure using mark-recapture, it is hard to show if adults return to spawning grounds where they themselves were spawned (Cushing 1981), and spawning site fidelity has frequently to be assumed instead. Parasite markers specific to a population can be used, but they must not affect mortality differentially and must occur in a high proportion of individuals in one population relative to the other, which limits the technique's usefulness.

These methods provide different types of information and can be used in combination to make inferences about stock structure and movement. Only genetic techniques directly address the issue of population isolation; present tagging methods can demonstrate isolation during the later life history but not during the early stages, and provide only very general quantitative information on proportions of mixing between stocks. Alternatively, naturally-occuring chemical markers are laid down in the otoliths during growth (Mulligan et al 1987, Edmonds et al.1992) which reflect the water through which the fish moves. These can be linked to time and fish age using otolith annuli and micro-increments, and the chemical markers used to examine spawning separation and early life-history movement retroactively (Jones 1992, Kalish 1990, Edmonds et al. 1991, Campana et al. 1994, Thorrold et al.1999). However, movement must be on a time-scale and between water bodies with sufficiently distinctive chemical characteristics to be detectable. Chemical signals that are effective in spatial discrimination appear to be highly specific, and little work has been attempted in fish that are exclusively marine.

5.3. Growth

Knowledge of growth rates through the life cycle are fundamental in assessing the biomass productivity of a fishery, and how it is affected by different management choices (e.g. Beverton and Holt 1957, Ricker 1975). Fish generally exhibit a period of accelerating increases in absolute length during the early period of life, followed by a deceleration. The age when the changeover occurs may vary: if prolonged, the size-at-age relationship may be approximately linear; more commonly, size is asymptotic in older fish (Ricker 1975).

Growth rates can be estimated using length-at-age data, by sampling across the population at a single time. However, Hilborn and Walters (1992) warn that size information can be grossly misleading and that investigators should examine carefully how data are taken. Inaccurate estimation of growth rates occur when mortality is stronger for larger fish than smaller fish within a cohort so that smaller fish tend to survive (or vice versa). Sampling may have a similar effect: due to gear if it is selective for size within cohorts, or does not cover all age classes equally; due to differences in fish behaviour and distribution with size; or due to an inadequate sampling design (Ricker 1975, Hilborn and Walters 1992). Further error will be contributed if individual variation in growth rate is large; if sex differences are not accounted for; or if age-specific growth rates show trends between succeeding year-classes (Ricker 1975), due to environmental cycles or long-term changes in fishing pressure. If these effects occur, the population growth rate estimated will be different from the true growth rate (Ricker 1975).

Alternatively, somatic growth rates can be estimated for individual fish over their life-span by measuring the increments laid down in scales and otoliths and back-calculating using the relationship between somatic and incremental growth (Ricker 1969, 1975). But at a population level, back-calculated size estimates become biased in fish of younger ages under the same conditions of size-selective mortality that adversely affects size-at-age data (Lee 1912). The method depends on the relationship between somatic and otolith or scale growth, and error will be introduced if the relationship is weak. Furthermore, measurements within a fish are not independent, so that error within a fish is auto-correlated relative to other fish (Jones 2000).

Mark-recapture can be used to examine growth rates by measuring the difference in size between first capture and recapture. However, tagging is subject to the problems discussed in Section V.2. above, and growth rate may be affected by tagging. Biases will result if this varies with size (for example, due to greater trauma suffered by larger fish during handling out of the water, or during retrieval from deeper water) (Gulland 1983).

Modes in the distribution of fish size frequency can identify age groups; growth rates can be estimated from the distance between the modes, or directly by following their progression over time. However, the technique is particularly vulnerable to the problems discussed above in using size data, and depends on good discrimination between year classes, which is usually only true for the youngest ages and species which have a short spawning season. Environmental cycles can also confound the method by generating pulses of recruitment, and size-selective biases will occur as age groups recruit to the fishery, larger individuals recruiting earlier, sometimes by as much as several years (Gulland 1983).

The most common model used for describing growth (eg. Beverton and Holt 1957, Barbieri et al. 1993, Piner 1998) in fish length was developed by von Bertalanffy (1938):

$$l_t = L_\infty(1 - e^{-K(t-t_0)}) \qquad (6)$$

where L_∞ is the mean maximum length, t_0 is the theoretical age at which length=0, and K is the Brody growth coefficient, governing the rate at which L_∞ is approached.

Alternative asymptotic models to describe growth are the Richards equation, the Gompertz equation, and the logistic equation. Schnute (1981) demonstrated that these, along with the von Bertalanffy, are special cases of a single general model. Sainsbury (1980) has found that individual variation can lead to under-estimates of K when length-at-age and growth increment data are used, which can lead to serious over-estimation of mean length-at-age in the case of growth increment data. Francis (1988) has pointed out that von Bertalanffy growth parameters can differ in meaning for estimates calculated from tagging and from age-length data.

5.4. This study

Mark and recapture was considered impracticable to examine population structure in Patagonian toothfish, for the reasons given in Section IV.1.4. Although directly addressing the issue of isolation, genetic determination was considered insufficiently sensitive to likely levels of mixing, and ignored the possibility of recent changes in population structure relevant to current fishery management. Chemical markers in otoliths were considered more promising but possibly insensitive to the levels of variability found between Antarctic water bodies. Addressing variation in growth rates would allow growth to be estimated for several management areas, and be less costly than the other methods. Population growth rate estimated from age-length data may diverge from the true growth rate but avoids the problems over recruitment in frequency modal analysis, and the relationship between otolith and somatic growth rates in back-calculation. Growth inferences based on age-length data were therefore selected for a first examination of stock structure.

Differences in growth rates between toothfish management areas are predicted to arise if the stocks are segregated in time or space, and divergent selection pressures operate over enough time for adaptation to occur. But similar differences will occur if fish are segregated only after spawning, with different environmental exposure that modifies growth rates differentially while segregated. The fish will be genetically identical, mixing randomly during reproduction, but will still need to be managed as separate stocks because of divergent vital rates affecting dynamics and biomass productivity. Of course, differences will be maintained only if there is little mixing between the stocks. Therefore, differences in growth rates between management areas can be used to infer the presence of discrete stocks.

On the other hand, similar growth rates between areas may be due to a single stock with random mixing, but would also be true for segregated stocks exposed to similar environmental and selection pressures. The latter case is unlikely, demanding that the large suite of variables affecting a population be similar or compensatory. Even so, the possibility remains that vital rates could be similar but the stocks different, so that management decisions for one would not affect the other. Alternatively, populations may be genetically distinct with different vital rates, but mix over their post-spawning life-history. In these cases, growth rates based on age-length data will not discriminate from the single-stock scenario.

Beverton and Holt's (1957) dynamic pool model uses parameters derived from the von Bertalanffy function. Growth models using the function have been developed for Patagonian toothfish, and suggest stock boundaries between the region encompassing southern South America and the Scotia Arc, and the southern Indian Ocean. Additionally, some information on parasite loadings suggest greater differences between the southern Patagonian shelf and South Georgia, than for fish from southern Chile to the southern Patagonian shelf (SC-CAMLR-XIV, Annex 5). The growth data used are compromised by inconsistencies in age estimation methods, sampling and sample sizes, and by the use of Ford-Walford plots to derive estimates, rather than non-linear regressions. However, the information indicates two hypotheses which can be tested: i) that there is a stock boundary between the Falkland Islands and South Georgia, and ii) that there is another boundary between South Georgia and Kerguelen.

In this chapter, I estimated Von Bertalanffy parameters for each location, and constructed models describing growth rates predicted from competing hypotheses of stock separation and mixing between locations. Assuming that the conceptual model developed in Chapter III is applicable to all areas, size-at-age data will allow relative comparison between management areas. I use likelihood methods to select between hypotheses in a confrontation with the data, and discuss the implications of these results to understanding toothfish population dynamics, stock management, and their potential relationship to the oceanographic structure of the Southern Ocean. The accuracy of the growth parameter estimates when used in models of toothfish population dynamics is also considered.

5.5. Methods and materials

I ask the scientific question: do differences in growth between locations indicate that Patagonian toothfish in the Southern Ocean represent one stock or more than one stock? Specifically, do fish found off the Falkland Islands, South Georgia, Kerguelen and Heard Island demonstrate similar Von Bertalanffy growth parameters or significantly different growth parameters? Using normal likelihood procedures, models can be set up predicting the situations under each hypothesis: 1) with separate growth curves between locations (H_{Ω}); 2) with one growth parameter equal ($H_{w1\text{-}3}$); and 3) with no differences in growth curves between locations (H_{w4}) :

$$H_{\Omega} : l_{ij} = L_{\infty i}(1 - e^{-K_i(t_{ij} - t_{0i})}) \tag{57}$$

$$H_{w1} : l_{ij} = L_{\infty}(1 - e^{-K_i(t_{ij} - t_{0i})}) \tag{58}$$

$$H_{w2} : l_{ij} = L_{\infty i}(1 - e^{-K_i(t_{ij} - t_{0i})}) \tag{59}$$

$$H_{w3} : l_{ij} = L_{\infty i}(1 - e^{-K_i(t_{ij} - t_{0})}) \tag{60}$$

$$H_{w4} : l_{ij} = L_{\infty}(1 - e^{-K(t_{ij} - t_{0})}) \tag{61}$$

Selection between models is based on the relative likelihood calculated from the data using the normal likelihood function (Kimura 1980, Edwards 1992).

Otoliths and length data were taken from the full age range of toothfish sampled by longliner between 11 February and 7 March 1997 off the Falkland Islands, 21 April and 21 May 1997 off South Georgia, and 25 February and 10 March 1997 off the Kerguelen Islands. Total lengths were divided into blocks as for the validation study in Chapter IV, and 20 fish sampled as far as possible within each block. Samples by longliner for South Georgia and Kerguelen were obtained using the randomised design outlined in Chapter II: samples for each length block were taken every fishing day until the block allocation was fulfilled, so that sampling was over a longer period for blocks with fewer fish in the catch. Samples for the Falkland Islands were taken haphazardly from the catch by longliner, and assumed to be representative (ie. that captured fish came aboard in random order). Samples were also taken by commercial trawl at Kerguelen between 8 March and 8 April 1997: sampling was again haphazard, taking fish of the desired length range that could be obtained from each net, and assumed randomized mixing within net and no conscious or unconscious selective tendencies by the observer. Samples from Heard Island were randomly selected from a sample taken haphazardly from the catch by commercial trawl between 5 September and 24 September 1993 (and included six samples from similar sampling taken between 11-21 June 1990), making the same assumptions for the original catch sampling as for Kerguelen. Some larger fish had been taken from the catch sample for other projects: these were assumed to have been taken randomly and did not exceed TL 83.7 cm. For the Kerguelen trawl collections, samples were taken over consecutive fishing days until the allocation for each block was fulfilled. Numbers and dates of collection are shown in Table 5.1.

Data for total length and sex were recorded, except for Falkland Islands where standard length was recorded. To standardize length measurements, total length and standard length were measured for fish taken while sampling off South Georgia. The relationship by sex was closely described using linear regression (Ricker 1975): for females, the relationship was $TL = 1.116SL+1.48$; for males, it was $TL = 1.16SL-2.04$ ($R^2_{female} = 99.5\%$; $R^2_{male} = 99.7\%$). Using these relationships, SL data from the Falkland Islands were weighted by sex to convert to TL. Age was estimated using the methodology developed in Chapter III. All otolith sections from the samples and 59 otoliths from the reference collection were pooled and randomly ordered. To control for data quality, age estimated from reference otoliths was compared to their standardized ages, and tested for significant differences using the ANOVA model developed in Chapter III (Table 5.2). Power was calculated in an a posteriori manner: $\phi = 3.32$ with

Table 5.1: Sample sets of *Dissostichus eleginoides* used for comparison of growth rates.

Location	*gear*	*no. females*	*no. males*	*dates of collection*
Falkland Islands	longline	93	67	11 Feb – 07 Mar 1997
South Georgia	longline	176	80	21 Apr – 21 May 1997
Kerguelen	longline	91	68	25 Feb – 10 Mar 1997
Kerguelen	trawl	72	78	08 Mar – 08 Apr 1997
Heard Island	trawl	63	33	05 Sep – 24 Sep 1993
				(11 Jun – 21 Jun 1990)

Table 5.2: Comparison between age estimated and standardized age for 58 reference otoliths, using the randomised block ANOVA model developed in Chapter III (standardized age = mean of four age estimations undertaken in Chapter III). Dependent variable: age; treatment factor: two levels (reading, standardized age); blocking factor: 58 levels (individual fish); α=0.05.

Source	*df*	*sum of squares*	*Mean square*	*F*	*Pr>F*
Total	115	8523.8			
fish	57	8387.3	147.1	64.5	0.0001
treatment	1	6.4	6.4	2.8	0.0992
error	57	130.1			

$v_1 = 1$ and $v_2 = 116$ degrees of freedom, giving a power of ca. 100%. No significant difference was found between the age data from the growth study and standardised ages ($\alpha = 0.05$), thus assuring consistency in aging.

To check that data from both gears could be pooled, the data were first examined for differences in growth prameters between the catches taken by trawl and longline at Kerguelen. The Von Bertalanffy model was fitted by least squares for pooled data and separately by sex: because of the form of the equation describing the normal likelihood function, least squares estimates are the same as maximum likelihood estimates when the assumptions of normality and constant variance are met. Residuals were examined for normality using the Shapiro-Wilks test ($\alpha = 0.05$) and error trends. The estimates from the pooled fit were used to parameterize the constrained parameters in the competing growth models outlined above, and estimates from the male and female fits for the θ_i parameters. Likelihood ratio (LR) tests were undertaken between the models to examine differences between capture. The unconstrained model (separate curves for each gear) could be expected to have the highest support, but some criterion is needed to test whether this is sufficient to justify its selection over the simpler models. The LR (Λ) can be used to give the test statistic:

$$-2\log(\Lambda) = -N\log\left(\frac{\sigma_0^2}{\sigma_{ai}^2}\right)$$

which has a chi-squared distribution, and this was used to test for the significance of differences between the unconstrained and other models, and by implication between growth parameters (Kimura 1980).

To check that data from both sexes could be pooled, a similar procedure was followed to examine differences in growth estimators due to sex for each location. Based on the results of these preliminary tests, comparisons between locations were made by sex and only between sampling sets using the same capture method. Thus, LR tests were performed for each sex between trawl samples from Kerguelen and Heard Island. To allow comparisons between each location sampled by longline, simultaneous LR tests using an adjusted α-level were performed pairwise for each sex between longline samples from Kerguelen and South Georgia, Kerguelen and Falkland Islands, and Falkland Islands and South Georgia. All analyses were undertaken using SAS PROC NLIN.

5.6. Results

Plots of total length against age are shown in Figure 5.1 for females and males in each sample set. These indicated differences in distribution between sexes for the Kerguelen and South Georgia samples, and differences between capture methods: trawls appeared to catch younger fish overall, although older females were caught off Kerguelen; coverage of the first few year classes was considerably better by trawl than by longline. The plots also indicated differences between management areas: old and large females were caught off Kerguelen by trawl, but did not appear in trawls off Heard Island; variation appeared to be distributed differently with age for the Falkland Islands than for South Georgia and Kerguelen.

Estimates of growth parameters by sample set for each sex are given in Table 5.3, with estimated standard deviations. Data were found to be normal for male and female data except those from South Georgia ($p<0.0001$) and females taken by longline off Kerguelen ($p<0.0072$). Residuals showed little expansion with age in males, especially after the first 5 years, but showed some expansion in females taken off Kerguelen by longliner and trawl, and more strongly in females taken off South Georgia. The analysis for Kerguelen females was unlikely to be greatly affected by the level of heteroscedasticity; and the method should be robust to limited violation of normality. However, the violation was greater for both South Georgia sexes, and was little improved by square root or log-normal transformations. Parameter estimates by least squares are likely to depart from maximum likelihood estimates.

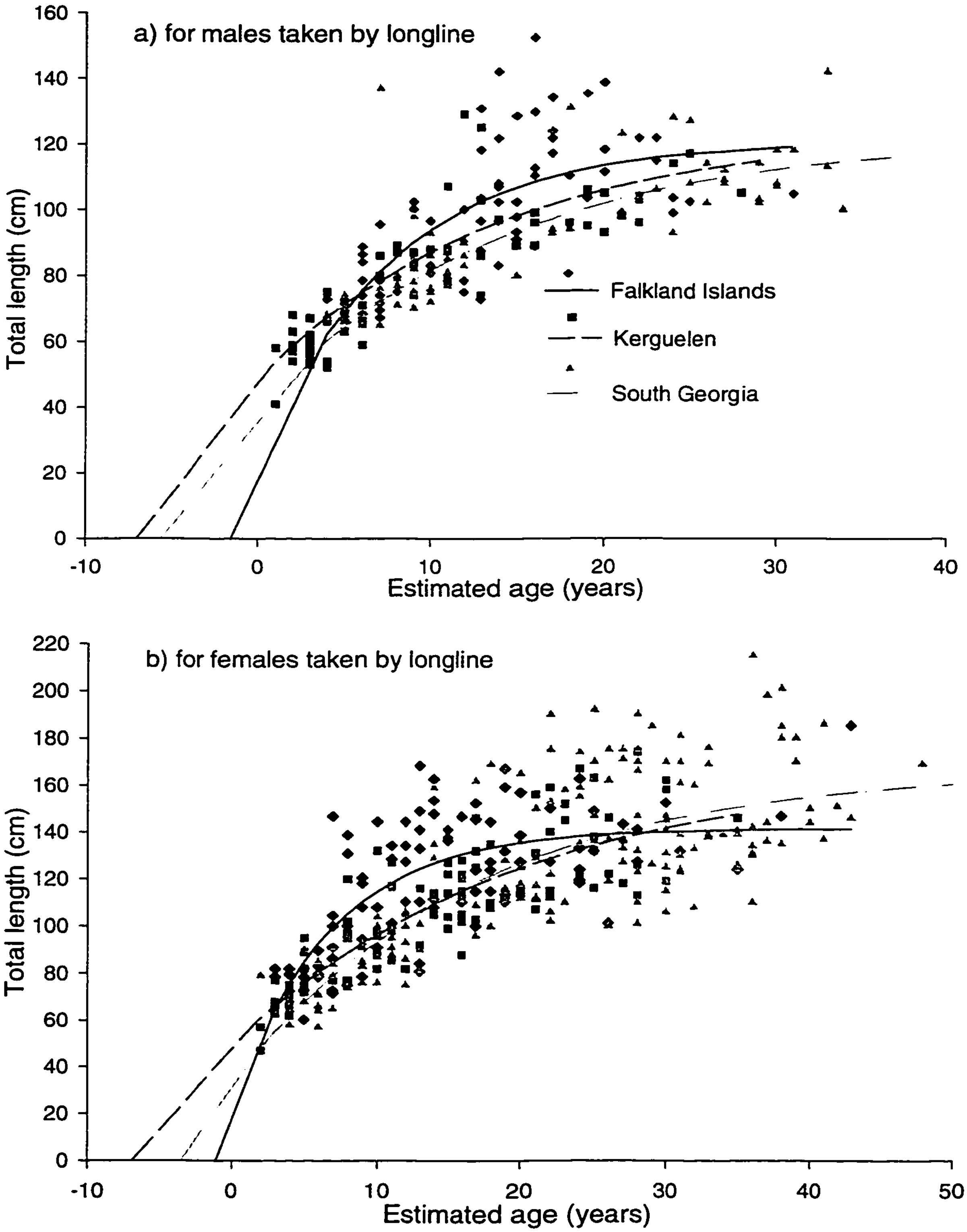


Figure 5.1: Scatter plots of total length versus estimated age, fitted with the von Bertalanffy function: a) for males taken by longline, b) for females taken by longline, c) for males taken by trawl, and d) for females taken by trawl.

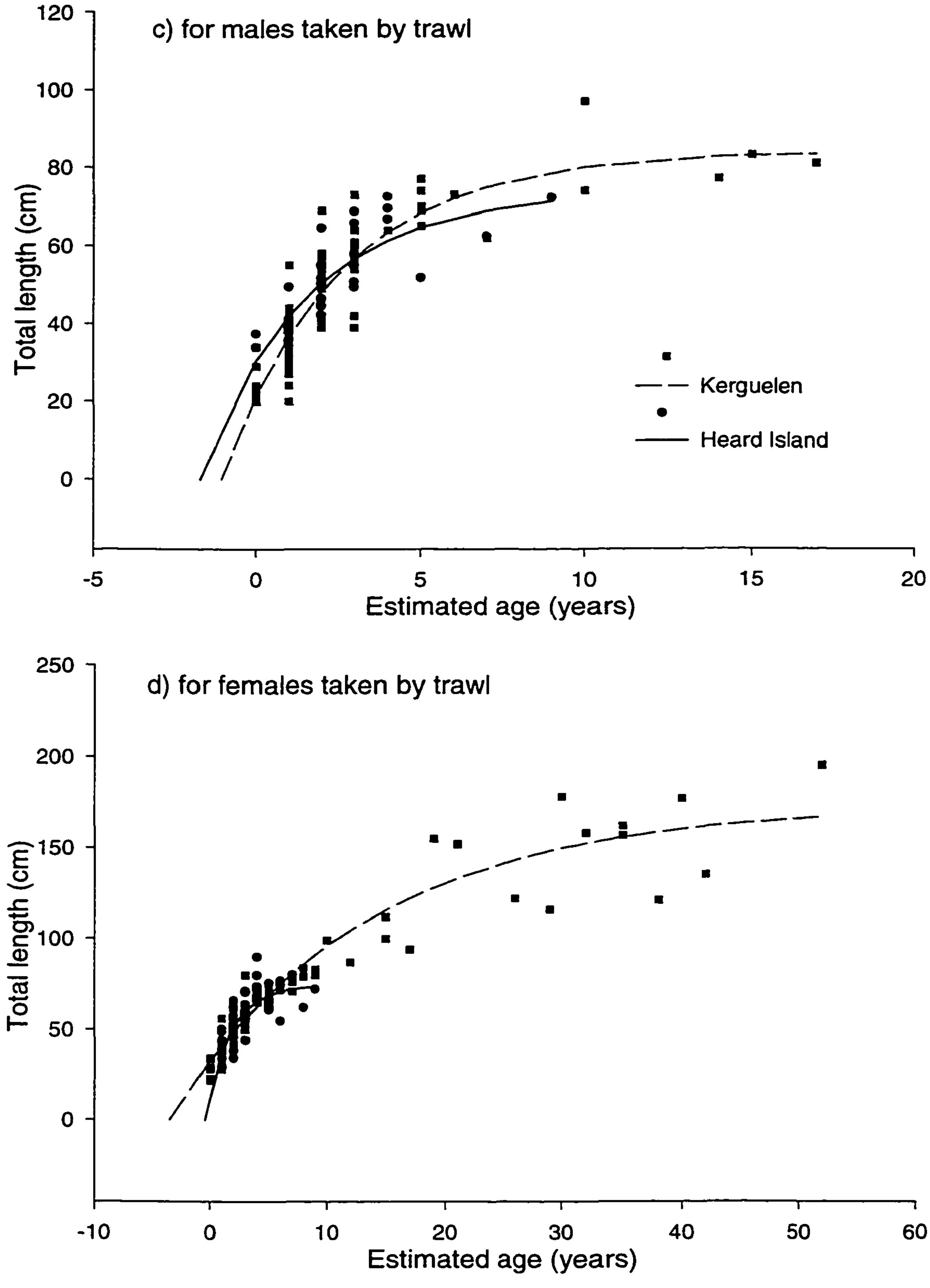


Figure 5.1: continued.

Table 5.3: Estimates of Von Bertalanffy growth parameters for males and females, for sample sets shown in Fig 5.2. Standard deviations in brackets

Males:

	n	l_∞	*K*	t_0
a) Falkland Islands (longline)	67	120.7(7.9)	0.13(0.05)	-1.55(2.72)
b) South Georgia (longline)	80	122.1(7.2)	0.07(0.02)	-5.71(2.78)
c) Kerguelen (longline)	68	124.6(9.3)	0.07(0.02)	-7.11(1.73)
d) Kerguelen (trawl)	78	83.8(4.0)	0.28(0.04)	-1.05(0.20)
e) Heard Island (trawl)	33	73.9(7.8)	0.31(0.12)	-1.71(0.69)

Females:

	n	l_∞	*K*	t_0
a) Falkland Islands (longline)	93	141.4(5.7)	0.15(0.04)	-1.10(1.61)
b) South Georgia (longline)	176	167.0(9.0)	0.06(0.01)	-3.64(1.80)
c) Kerguelen (longline)	91	168.1(20.78)	0.05(0.02)	-7.00(2.55)
d) Kerguelen (trawl)	72	172.9(9.2)	0.06(0.01)	-3.48(0.64)
e) Heard Island (trawl)	63	74.4(4.4)	0.48(0.13)	-0.46(0.42)

5.6.1. Differences between capture methods

The results from the LR tests are shown in Table 5.4. For males, the H_Ω model (all von Bertalanffy parameters differed by gear-type) better fit the data for both males and females, indicating that trawling and longlining took different fish. A better fit was obtained with separate VBGF parameters, indicating all parameters were significantly different and the H_Ω model was clearly the most appropriate model for the data. This was not true for females: the H_Ω model was not better than the partially constrained models and some parameters could be pooled. These data, taken together, provide some evidence for differences due to gear, sufficient to justify treating separately samples taken by different capture methods.

5.6.2. Differences between sexes

The results from the LR tests are shown in Table 5.5. The H_Ω model fit the data better than the H_{w4} model in all samples except Heard Island, indicating significant differences in growth between sexes ($\alpha = 0.05$). At Heard Island, the difference was significant at the $\alpha = 0.10$ level, and the H_Ω model (all VBGF parameters differed) had significantly more support than the H_{w3} model (L_∞ and K differ), indicating significant differences for t_0 and weak evidence for differences in growth between sexes. At Kerguelen, the H_Ω model fit the data significantly better than all the constrained models for the sample taken by trawl, indicating strong evidence for differences in growth by sex, reflecting the disparity in distribution due to the lack of large males. This was not true for the sample taken by longliner: H_Ω better fit the data than H_{w4}, but was not better than the partially constrained models, so that some parameters could be pooled. Taken together, these data provide some evidence for differences in growth between sexes at Kerguelen, over and above the differences due to gear type. Similar results were obtained for South Georgia and the Falkland Islands: the partially constrained models (H_{w1}, H_{w2}, H_{w3}) did not have significantly different support from the unconstrained model and could equally well be selected, indicating weak evidence of differences due to sex. Differences

Table 5.4: Likelihood ratio tests comparing Von Bertalanffy parameter estimates for male and female toothfish taken by longline and trawl off Kerguelen ($\alpha = 0.05$).

Hypothesis	*constraints*	*sum of squares*	*$-N \log(\sigma_\Omega^2/\sigma_\omega^2)$*	*df*	*p*
a) Males (n = 146)					
H_Ω	none	9494.0			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	11333.8	25.90	1	0.001
$H_{\omega 2}$	$K_1=K_2$	11679.0	30.24	1	0.001
$H_{\omega 3}$	$t_0 1=t_0 2$	12018.7	34.40	1	0.001
$H_{\omega 4}$	all	12881.1	44.50	3	0.001
b) Females (n = 163)					
H_Ω	none	31763.8			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	31772.2	0.04	1	ns
$H_{\omega 2}$	$K_1=K_2$	31786.2	0.11	1	ns
$H_{\omega 3}$	$t_0 1=t_0 2$	32323.3	2.85	1	ns
$H_{\omega 4}$	all	34205.6	12.07	3	0.01

Table 5.5: Likelihood ratio tests comparing Von Bertalanffy parameter estimates for male and female toothfish for each location ($\alpha = 0.05$).

Hypothesis	*constraints*	*sum of squares*	$-N \log(\sigma_\Omega^2/\sigma_\omega^2)$	*df*	*p*
a) Falkland Islands (longline) n = 160					
H_Ω	none	50513.7			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	51318.4	2.53	1	ns
$H_{\omega 2}$	$K_1=K_2$	50530.0	0.05	1	ns
$H_{\omega 3}$	$t_01=t_02$	50521.6	0.02	1	ns
$H_{\omega 4}$	all	65172.9	40.76	3	0.001
b) South Georgia (longline) n = 256					
H_Ω	none	92752.7			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	93496.8	2.05	1	ns
$H_{\omega 2}$	$K_1=K_2$	92785.6	0.09	1	ns
$H_{\omega 3}$	$t_01=t_02$	92839.5	0.24	1	ns
$H_{\omega 4}$	all	117888.1	61.39	3	0.001
c) Kerguelen (longline) n = 159					
H_Ω	none	25086.3			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	25458.3	2.34	1	ns
$H_{\omega 2}$	$K_1=K_2$	25129.7	0.27	1	ns
$H_{\omega 3}$	$t_01=t_02$	25086.5	0.00	1	ns
$H_{\omega 4}$	all	35362.1	54.58	3	0.001

Table 5.5: continued.

d) Kerguelen (trawl) n = 150

H_{Ω}	none	16127			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	20094	32.98	1	0.001
$H_{\omega 2}$	$K_1=K_2$	19472	28.27	1	0.001
$H_{\omega 3}$	$t_0 1=t_0 2$	17936	15.95	1	0.001
$H_{\omega 4}$	all	20839	38.45	3	0.001

c) Heard Island (trawl) n = 96

H_{Ω}	none	5150.2			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	5150.3	0.00	1	ns
$H_{\omega 2}$	$K_1=K_2$	5219.4	1.28	1	ns
$H_{\omega 3}$	$t_0 1=t_0 2$	5373.4	4.07	1	0.05
$H_{\omega 4}$	all	5511.3	6.50	3	ns

between sexes for all sample sets were therefore sufficient to justify treating them separately in subsequent analyses.

5.6.3. Differences between locations.

There were significant differences in growth between female fish caught by trawl at Kerguelen and Heard Island (Table 5.6), reflecting the absence of larger fish in the Heard Island sample. All parameters were significantly different and the H_{Ω} model was the most appropriate model for the data. For males, however, no significant differences were found, reflecting the absence of larger males taken by trawl at both locations.

For collections made by longline, the pairwise significance level was adjusted to $\alpha = 0.0167$ to allow simultaneous comparisons between three locations with an experiment-wise significance of $\alpha = 0.05$. There were no significant differences in growth between South Georgia and Kerguelen fish (Table 5.7): differences in residuals between the competing hypotheses were extremely low for both sexes, indicating that growth rates between locations were very similar. There were differences between the Falkland Islands and South Georgia (Table 5.8a), and Falkland Islands and Kerguelen (Table 5.8b): the H_{Ω} model was significantly better supported by the data than the $H_{\omega 4}$ model in all cases. However, for males, the partially constrained models also fit the data well, indicating individual parameters were not significantly different, and providing weak evidence for growth differences between locations. For females, the partially constrained models were not significantly different from the H_{Ω} model, but residuals were larger and p-levels were marginal in several cases, indicating stronger evidence for growth differences between locations.

Table 5.6: Likelihood ratio tests comparing Von Bertalanffy parameter estimates for male and female toothfish taken by trawl off Kerguelen and Heard Island ($\alpha = 0.05$).

Hypothesis	*constraints*	*sum of squares*	$-N \log(\sigma_\Omega^2/\sigma_\omega^2)$	*df*	*p*
a) Females (n = 135)					
H_Ω	none	15071.8			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	16431.9	11.70	1	0.001
$H_{\omega 2}$	$K_1=K_2$	16391.6	11.30	1	0.001
$H_{\omega 3}$	$t_0 1=t_0 2$	16040.9	8.40	1	0.005
$H_{\omega 4}$	all	16599.2	13.03	3	0.005
b) Males (n = 111)					
H_Ω	none	6206.3			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	6255.5	0.88	1	ns
$H_{\omega 2}$	$K_1=K_2$	6208.5	0.04	1	ns
$H_{\omega 3}$	$t_0 1=t_0 2$	6279.3	1.30	1	ns
$H_{\omega 4}$	all	6654.8	7.74	3	0.1

Table 5.7: Likelihood ratio tests comparing Von Bertalanffy parameter estimates for male and female toothfish taken by longline off South Georgia and Kerguelen ($\alpha = 0.05$).

a) Females (n = 267)

Hypothesis	*constraints*	*sum of squares*	*-N* $log(\sigma_\Omega^2/\sigma_\omega^2)$	*df*	*p*
H_Ω	none	103998.8			
$H_{\omega 1}$	$L_\infty 1=L_\infty 2$	103999.3	0.00	1	ns
$H_{\omega 2}$	$K_1=K_2$	104068.3	0.18	1	ns
$H_{\omega 3}$	$t_0 1=t_0 2$	104409.0	1.05	1	ns
$H_{\omega 4}$	all	105078.3	2.76	3	ns

b) Males (n = 148)

Hypothesis	*constraints*	*sum of squares*	*-N* $log(\sigma_\Omega^2/\sigma_\omega^2)$	*df*	*p*
H_Ω	none	13840.2			
$H_{\omega 1}$	$L_\infty 1=L_\infty 2$	13844.5	0.05	1	ns
$H_{\omega 2}$	$K_1=K_2$	13847.9	0.08	1	ns
$H_{\omega 3}$	$t_0 1=t_0 2$	13859.8	0.21	1	ns
$H_{\omega 4}$	all	13899.5	0.63	3	ns

Table 5.8: Likelihood ratio tests comparing Von Bertalanffy parameter estimates for male and female toothfish taken by longline off a) Falkland Islands and South Georgia; and b) off Falkland Islands and Kerguelen ($\alpha = 0.05$).

a) Falkland Islands and South Georgia

Hypothesis	*constraints*	*sum of squares*	$-N \log(\sigma_{\Omega}^2/\sigma_{\omega}^2)$	*df*	$<p$
i) Females (n = 269)					
H_{Ω}	none	119375.6			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	121683.2	5.15	1	0.025
$H_{\omega 2}$	$K_1=K_2$	121495.6	4.73	1	0.05
$H_{\omega 3}$	$t_0 1=t_0 2$	119832.3	1.03	1	ns
$H_{\omega 4}$	all	128419.8	19.64	3	0.001
ii) Males (n = 147)					
H_{Ω}	none	23890.9			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	23894.7	0.00	1	ns
$H_{\omega 2}$	$K_1=K_2$	24221.3	2.01	1	ns
$H_{\omega 3}$	$t_0 1=t_0 2$	24136.7	1.50	1	ns
$H_{\omega 4}$	all	27154.0	18.80	3	0.001

Table 5.8: continued.

b) Falkland Islands and Kerguelen

Hypothesis *constraints* *sum of squares* $-N \log(\sigma_\Omega^2/\sigma_\omega^2)$ *df* *<p*

i) Females (n = 184)

H_Ω	none	56242.5			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	56987.8	2.42	1	ns
$H_{\omega 2}$	$K_1=K_2$	58116.5	6.03	1	0.025
$H_{\omega 3}$	$t_0 1=t_0 2$	57467.4	3.96	1	0.05
$H_{\omega 4}$	all	62328.5	18.90	3	0.001

ii) Males (n = 135)

H_Ω	none	19357.6			
$H_{\omega 1}$	$L_{\infty 1}=L_{\infty 2}$	19371.0	0.09	1	ns
$H_{\omega 2}$	$K_1=K_2$	19752.1	2.72	1	ns
$H_{\omega 3}$	$t_0 1=t_0 2$	19829.4	3.25	1	ns
$H_{\omega 4}$	all	21878.0	16.52	3	0.001

5.7. Discussion

5.7.1. Population structure

The results support the hypothesis of stock separation between South Georgia and Falkland Islands. This was seen for longline samples for both sexes and for comparisons between Falkland Islands with both South Georgia and Kerguelen; given the available data, their consistent agreement shows strong evidence for separate stocks. Although there were growth differences between Kerguelen and Heard Island, the data are better explained by female migration than separation of stocks. Females caught by trawl off Heard Island were mostly young, which indicates they may migrate away when older. Meanwhile, older females were caught by trawl off Kerguelen. Heard Island may therefore represent a nursery area linked to Kerguelen; a possibility that is enhanced by the proximity of the two areas on the same continental shelf system, facilitating movement. Older males were also present at Kerguelen, as shown by the longline data, but were not caught by trawl; a similar migration may therefore occur for males, but the evidence is weaker.

The results do not support the hypothesis of separate stocks at South Georgia and Kerguelen. Fish taken at the two locations may therefore be from one stock. Alternatively, there may be separate populations but with similar vital rates, or separation only at spawning. The latter case is not likely given the considerable distance between the islands. But populations may have similar vital rates due to similar environments; however, this is also unlikely as much higher fishing levels occurred at South Georgia than Kerguelen between 1993-1996. Similar fishing levels at Kerguelen occurred too recently to be reflected in these data. If stocks are separate, the data should reflect reduced competition from larger and older fish at South Georgia, leading to faster early growth, and therefore a higher K, than at Kerguelen.

In the experimental design, by necessity sampling was blocked on length: this meant coverage was not equal between ages because, as fish growth slows with age, the blocks of larger lengths covered more age classes. The design may also have resulted in selectivity for younger fish at large sizes. Again by necessity, Heard Island was sampled during a different year and season from the other sample sets. All analyses assumed that size-at-age of the captured populations was representative of size-at-age of the natural populations, yet differences due to gear indicate this to be unlikely. Size selective mortality rates may have compounded any biases in the calculation of growth parameters.

Little evidence exists on how these assumptions may affect conclusions. To address some of these difficulties, data were examined by sex and capture gear but the differences in sampling design between the Falkland Islands and both South Georgia and Kerguelen remain, confounding the test for growth differences. Similarly, temporal differences between the Heard Island (1993) and Kerguelen (1997) trawl sample sets confound the test for growth differences between them. However, the difference between years is likely to be smoothed by the number of cohorts at both locations. Fishing pressure overall is unlikely to explain the differences: the increase in fishing pressure at Kerguelen is more likely to reduce differences by reducing the number of large females. In any case, the Kerguelen samples were taken before changes in growth due to increasing fishing pressure would have reasonably been expected.

Overall, the samples are the best possible given the operational difficulties of coordinating sampling across four management areas located in remote areas of the Southern Ocean: those from all but Heard Island were collected within a three month period; and all were taken by trained observers on board licensed fishing vessels. Previous data on growth differences between management areas are likely to have suffered from similar problems and, unlike this study, did not standardize methodology and readers to estimate ages between management areas, or randomize the reading of samples to avoid conscious or unconscious bias by age readers. For previous studies, these effects were therefore confounded with those for management area. Assuming that otolith structure can be interpreted similarly for all management areas, the data in this study represent the best available for testing hypotheses on population structure using growth differences.

Sample sets were divided by sex in this analysis, and sampling was stratified on length so that some age classes were represented by few data points or none. Larger samples taken using random selection by age from a previously aged representative sample population would correct this imbalance, reduce stochasticity in mean lengths-at-age, and strengthen the assumption of normality as the sample increased. Departure from constant variance can be corrected by minimizing the sum of squares:

$$\Sigma(n_i / s_i^2)(\bar{l}_i - \mu(t_i))^2$$

where s_i^2 is the sample variance of the lengths of the n_i individuals aged t_i. Larger, balanced year classes would allow weights to be used and lead to convergence of the ML estimators on the true value for the population parameter.

5.7.2. Growth

Care is needed when the growth parameters estimated are incorporated in management models. Even though the collection methods used were the best available given the exigencies of working in the Southern Ocean, sampling was nevertheless only from the commercial catch population. The linkage between the catch population and the natural population is not well understood. These data show some selection of large fish by longline compared with trawl, indicating bias in the gear-specific estimates. For an active method like trawling, size selectivity by mesh and trawl mouth size may be important; for a passive method like longlining, selectivity may involve many more variables than the distribution of the population in time and space. The size of hooks may influence the number and size of fish captured.

Even so, the sample design allows conclusions on the relationships between different sampled populations, and these results can be used to select the best growth estimates available for use in modelling toothfish populations, and to determine how better estimates could be obtained. The results from the Kerguelen data indicate that trawling may not sample older males compared to longlining, but will sample the first few year classes better for both sexes. Once available and vulnerable to the gear, selection of younger fish is likely to be less biased by trawl; at larger sizes, longlining is more effective in obtaining numbers of large fish of both sexes. Differences in size-at-age occur between sexes for both capture methods, and occur for all sample sets. The best estimates of growth should therefore be by sex in these populations, and should incorporate younger year classes taken by trawl, and older year classes taken by longline (especially for males), with numbers balanced for each age category. With larger numbers within each age group, mean size-at-age weighted by ($\sqrt{n_i}$) will also prevent violation of the assumptions of normality and constant variance, and may best describe growth over the life-span (Kimura 1980). To guard against hook selectivity, several hook sizes should be used when sampling to ensure that the larger fish in the population are represented.

Previous estimates of Von Bertalanffy growth parameters for *Dissostichus eleginoides* are given in Table 5.9. Estimates for the South Atlantic region by Zakharov and Frolkina (1976), Shust et al. (1990) and Aguayo (1992) were based on length-at-age data. Methodologies used to age fish were not clear, and neither were estimation methods in the two earlier studies. Aguayo (1992) used the Walford method which is less reliable than the non-linear method. No comparisons in aging methodology were made between investigators and sampling was exclusively from catches by longliner. Despite the agreement between studies, the uncertainties in methodologies mean that these results should be treated with caution.

More recently, Horn (1999) compared results from age readings between laboratories working on toothfish, showing substantial agreement between readers. Cassia (1998), one of the readers from Horn's (1999) study, sampled catches from longliners off South Georgia, and found no difference between scales and otoliths for estimating age, and used measurements of the distance between nucleus and annulus to measure growth. Although her results agreed with those for the earlier studies, the method she used assumed a linear relationship between the measure and fish length, and also assumed no size-selective mortality. Neither assumption was tested; nor were

Table 5.9: Values for Von Bertalanffy parameters estimated for *Dissostichus eleginoides* in previous studies.

Area	*sex*	L_∞	*K*	t_0	*Source*
Patagonian Shelf	all	204.3	0.0563	-0.545	Zakharov and Frolkina (1976)
South Georgia	all	174.8	0.0712	-0.005	Shust et al (1990)
	all	210.8	0.0644	0.783	Aguayo (1992)
	all	170.8	0.0916	-0.031	Aguayo (1992)
	all	207.0	0.0748	-0.289	Cassia (1998)
Southern Chile	all	216.1	0.062	-0.877	Aguayo (1992)
Macquarie Island	male	138.4	0.072	-1.37	Kalish and Timmiss (1998)
	female	205.3	0.045	-1.54	Kalish and Timmiss (1998)
NZ EEZ - Sub-area 88.1	male	134.3	0.118	0.08	Horn (1999)
	female	158.7	0.085	-0.35	Horn (1999)

differences due to sex although all data were pooled. Horn (1999) observed differences for growth curves fit by sex for Patagonian toothfish captured by trawl and longline in the southern Pacific, with females growing faster and larger than males. Similarly, Kalish and Timmiss (1998) were involved in the comparison of age readings between laboratories, and found differences between sexes for fish caught by trawl off Macquarie Island, with L_∞ for female fish substantially larger. Although differences between these studies may indicate differences in growth between regions, Horn (1999) attributed the differences between his estimates and those of Kalish and Timmiss (1998) to differences in sampling gear rather than region, and the differences of these two studies to that of Cassia's (1998) results may also be attributable to differences in sampling gear. It may also be due to the treatment of data by sex; when pooled for sex, estimates of growth parameters in Kalish and Timmiss's (1998) study were similar to those for Cassia (1998). Additionally, even though comparisons in age readings between these investigators indicated substantial agreement, age interpretation may change with time (Campana et al. 1995), and this was not controlled for in comparisons between the studies.

The estimates for growth in this study improve on the previous estimates by treating by sex and gear; using standardized methodology including randomized catch sampling of longline catches off South Georgia and Kerguelen; and a quality control programme for age data with random presentation of otoliths for reading and comparison with reference otoliths. Comparisons between my estimates are therefore more reliable than comparisons with other studies.

Even so, previous work (Ashford and Horn 1999, Horn 1999) indicated that my age readings are broadly similar to those for other laboratories, and some comparisons may indicate further differences between regions. Thus, comparing samples taken by trawl in this study and that of Kalish and Timmiss (1998), L_∞ is lower for both sexes at Kerguelen than Macquarie and K is substantially higher for males; however, the discrepancy between sexes is less at Macquarie. Data for trawl and longline for

Kerguelen in this study can be pooled to allow comparisons with Horn's (1999) results. Thus, for females

$$l_{ij} = 162.9(1 - e^{-0.065(t_{ij} + 3.66)})$$

and for males:

$$l_{ij} = 105.9(1 - e^{-0.15(t_{ij} + 2.02)})$$

indicating the growth curve for females may be similar between the two regions, but L_∞ may be substantially higher for males in the southern Pacific. Although I tended to give higher estimates of age possibly due to differences identifying the first annulus (Ashford and Horn 1999), leading to lower estimates for L_∞ and K, this is unlikely to explain that degree of discrepancy. It also does not explain regional differences in the discrepancy of L_∞ between sexes. Further work is needed to see if growth at Macquarie Island and in the southern Pacific Ocean is different from that found in this study for the southern Atlantic and Indian Oceans.

5.8. Conclusion

The results presented in this chapter were presented to the 2000 meeting of the WG-FSA (Ashford et al. 2000). At the same time, a molecular study by Smith and Gaffney (2000) indicated similar conclusions for population structure: mitochondrial DNA showed a distinct break between samples from the South American shelf and samples from the Southern Ocean. Two distinct groups were further revealed within the Southern Ocean: one group included the Ross Dependency and Macquarie Island, the other included Heard Island, Kerguelen and South Georgia. The Working Group concluded that the two studies provided strong evidence of a separation between stocks within the CCAMLR area and those around the South American shelf. It also noted the evidence presented here of differences in growth parameters estimated from samples taken by longline and trawl at Kerguelen, and between sexes off South Georgia, Kerguelen and the Falkland Islands.

CHAPTER VI

CONCLUSION

The results from this study show that there is spatial variability at regional scales in growth parameters derived from size-at-age distributions, from which population boundaries and linkages may be inferred. Thus, relative to rival hypotheses, these data support the hypotheses i) that the Falkland Islands is a separate stock from South Georgia and Kerguelen, and ii) that South Georgia and Kerguelen represent one stock. The hypotheses predict that movement is low enough between the Falkland Islands and South Georgia/Kerguelen for growth differences to be preserved, but movement between South Georgia and Kerguelen is sufficient to smooth out any differences in individual growth rate. This implies different migration rates and linkages between the three areas. These linkages are likely to be largely determined by spatial distribution and current flows, primarily eastward in the ACC.

Traditional population dynamics models assume panmictic populations and no persistent spatial differences in vital rates (Hanski and Simberloff 1997). The approach emphasises birth and mortality rates, often assuming no immigration and emigration. In contrast, a spatial approach is based on the premise that species occur in assemblages of spatially structured local populations inter-linked by migration. These local populations make up a larger discrete metapopulation in which the population dynamics are more or less affected by the finer scale linkages. In this view, the critical processes in modelling dynamics are immigration and emigration: the linkages between local populations, or connectivity, must be identified and then quantified.

Early spatial models by MacArthur and Wilson (1967) were at the community level and dealt only with species richness, examining changes in species number with time and predicting the equilibrium number of species. Levins (1969) addressed population questions more directly using terrestrial organisms, but developed his model using presence/absence data that emphasized local extinction. Spatial effects are most likely to be seen in the dynamics of local populations: the source/sink model of Pulliam (1988) incorporated stage-based population data, and parameters for survival and recruitment, with an emphasis on movement between spatially separated local populations.

The results for toothfish underscore the importance of developing spatial models that also incorporate age-based population dynamics. Given the spatial distribution of toothfish and the potential linkages between populations through oceanographic features, spatial modelling of population dynamics may have considerable potential. Beverton and Holt (1957) developed a dynamic pool model that incorporated a spatial approach using age-based data, with a diffusion term to describe movement between areas. Similarly, Caddy (1975) took issue with fisheries models that assumed no spatial distribution of biomass or effort over a fishing ground, linking spatial variation in fishing mortality to local availability of biomass. The diffusion term is unlikely to describe toothfish movement, but fishing mortality is spatially variable over the range of toothfish, incorporating international waters with little regulation, and national and CCAMLR zones with regulatory regimes.

The work of Beverton and Holt (1957) anticipated the recent development of refugia as a management tool (Pauly 1993). A rationale for a spatial approach to marine refugia was outlined by Polachek (1990). Caddy (1993) used a simple cohort model to explore the possibility of a rotating harvest scheme for a sedentary red coral, creating refuges in time, while Holland and Brazee (1996) developed a matrix model to describe the dynamics of marine reserves, showing that red snapper populations protected from fishing can act as a sink for populations in fished areas, when fishing mortality is moderate to heavy (and connectivity is moderate). But refugia demand enforcement to be maintained, and for toothfish the costs are prohibitive due either to remoteness, or often lack of resources where the fishery is within national EEZs.

However, spatial distribution is relatively well known, consisting of a continuous population on the South American continent and a series of islands and banks of differing area with diverse regulatory regimes. Although refugia are unlikely to be a useful tool for management, the present study shows there are likely to be linkages between areas, and spatially-based population models will greatly enhance our understanding of their effect on stock dynamics and production. To develop these, information is needed on age structure, recruitment and connectivity between areas. The methodologies developed in this study provide the basis for obtaining this information representatively with rigorous quality control of age data. Estimates of growth for Kerguelen, South Georgia and Falkland Islands can be taken from this study, but further work is needed to elucidate age structure and examine the rates and scales of movement-at-age.

Movement-at-age is likely to be heavily influenced by environmental variables. Alexander and Roughgarden (1996) and Botsford et al. (1993) have attempted to incorporate terms describing marine physical features in models of population dynamics. Hofmann et al. (1998) have developed a coupled physical-biological model for krill recruitment to South Georgia along ACC core currents. If the ACC and its fronts provide a corridoor from South Georgia to Kerguelen, mixing of the ACC with the Weddell Gyre and Agulhas Retroflection may increase mortality of life-stages, and shifts in latitude may direct the larvae towards or away from the Kerguelen Plateau, potentially affecting recruitment.

Groundfish surveys at South Georgia have indicated large inter-annual variation in toothfish age structure and recruitment (Everson 1991). This stochasticity could be demographically based (cf. May 1975), or due to discrepancies in sampling scale (Underwood 1996). Alternatively, variability may be environmentally driven. The extreme seasonality and inter-annual variation found in the Southern Ocean is complicated by cycles or trends: anomalies in the maximum extent of sea-ice precess around the Antarctic continent with a period of 7-9 years (Murphy et al. 1995), and inter-annual variation in the Antarctic may be caused by ENSO signals propagating to higher latitudes through the atmosphere (White and Peterson 1996). Hence, this is a very difficult environment to sample for population dynamics.Variability in the physical environment at different temporal scales may profoundly influence dynamics, and determine life-history strategy. For instance, several years of zero recruitment to settlement plates by marine vertebrates were followed by a single year of massive recruitment (Dayton 1989); recruitment variability at McMurdo Sound was associated with changes in the incidence of anchor ice, which may in turn have been associated with changes in the upwelling of cold deep water (Barry and Dayton 1988). Priddle et al. (1988) showed that models of recruitment failure or mortality did not explain all the observed variability in krill populations in the Scotia Sea, and pointed to large-scale ocean-atmosphere processes as a likely source. Trathan et al. (1993) found strong correlations in meso-scale krill distribution with hydrographic variability, with active ontogenetic migration potentially also playing a role.

Toothfish population dynamics are also unlikely to be stationary, and further work is needed to understand the role of variability at different time scales. For this, randomized sampling is needed, especially at higher temporal and spatial scales to cover the periods of oscillations in the physical environment. Age structure can be compared between populations in time and space based on the ageing methodology developed in this study, assuming that interpretation of criteria remains similar for all populations. However, further age validation studies are necessary to measure the accuracy of age data and check this assumption. With these data, mortality and recruitment can be assesed. However, the estimation of migration parameters is more difficult, yet the results of the stock structure study presented here indicate that they may be critical in understanding toothfish population dynamics. Further work is needed to develop the methodology to estimate and assess movement-at-age in toothfish.

REFERENCES

Aguayo, M. 1992. Preliminary analysis of the growth of *Dissostichus eleginoides* from the austral zone of Chile and South Georgia. Document *WG-FSA-92/30.* CCAMLR, Hobart, Australia.

Alexander, S.E., and J. Roughgarden. 1996. Larval transport and population dynamics of intertidal barnacles: a coupled benthic/oceanic model. *Ecol.Monogr.*, 66: 259-275.

Anonymous. 1982. Working Party on Antarctic Fish Biology. Report of the Second Workshop on the Aging of Antarctic Fishes, Orono, Maine, USA, 7-10 June 1982. *BIOMASS Rep. Ser.*, 17: 1-30.

Anonymous. 1989. *Ocean circulation.* Oceanography course team, Open University. Open University, Milton Keynes and Butterworth-Heinemann, Oxford: 238pp.

Anonymous. 1994. *An assessment of Atlantic bluefin tuna.* Committee to review Atlantic Bluefin Tuna, National Research Council. National Academy Press, Washington, D.C: 148 pp.

Anonymous. 1996. *Scientific Observer's Manual, System of International Scientific Observation.* CCAMLR, Hobart, Australia.

Anonymous. 2000. *Improving the collection, management, and use of marine fisheries data.* Ocean Studies Board, Commission on Geosciences, Environment and Resources, National Research Council. National Academy Press, Washington, D.C..

Ashford, J.R. 1993. *An exploratory fishing expedition for Dissostichus eleginoides to the the South Sandwich Islands - Feb/Mar 1993.* Scientific Observer's Report, Foreign and Commonwealth Office, London.

Ashford, J.R. 1994. *Fishing operations for Dissostichus eleginoides by the Chilean longliner B/F Friosur V at South Georgia, April-May 1994.* Scientific Observer's Report, Foreign and Commonwealth Office, London.

Ashford, J.R. 1999. Paradigms lost. Letters to *Science,* 284: 591.

Ashford, J.R., and M.G. White. 1993. Cross-sectional structure and validation of the timing of annulus formation in otoliths of the Antarctic fish *Notothenia coriiceps* Richardson (*Nototheniidae*). *Cybium,* 17(2): 153-163.

Ashford, J.R., and M.G. White. 1995. An annual profile of growth in the otoliths of immature *Notothenia coriiceps* Richardson in relation to the environment at Signy Island, Antarctica. *Ant. Sci.,* 7(4): 345- 349.

Ashford, J.R. and G. Duhamel. 1996. A randomised sampling design used in the longlining fishery for *Dissostichus eleginoides* in the Southern Indian Ocean. Document *WG-FSA-96/56*. CCAMLR, Hobart, Australia.

Ashford, J.R. and G. Duhamel. 1997. Protocol: Randomised sampling designs for use in sampling *Dissostichus eleginoides* in a longline fishery. Document *WG-FSA-97/4*. CCAMLR, Hobart, Australia.

Ashford, J.R., and I. Everson. 1997. An assessment of longlining operations for *Dissostichus eleginoides* on board the Chilean-registered longliner B/F *Cisne Verde* during March-May 1997 around South Georgia (Sub-area 48.3). Document *WG-FSA-97/41*. CCAMLR, Hobart, Australia.

Ashford, J.R., and J.P. Croxall. 1998. An assessment of CCAMLR measures employed to mitigate seabird mortality in longlining operations for *Dissostichus eleginoides* around South Georgia. *CCAMLR Sci.*, 5: 217-230.

Ashford, J.R., and S. Wischniowski. 1998. Criteria for aging the otoliths of *Dissostichus eleginoides* from South Georgia (Sub-area 48.3) and an analysis of aging precision. Document *WG-FSA-98/52*. CCAMLR, Hobart, Australia.

Ashford, J.R. and P. Horn. 1999. An inter-laboratory comparison of ages estimated for *Dissostichus eleginoides* using otoliths. Document *WG-FSA-99/56*. CCAMLR, Hobart, Australia.

Ashford, J.R., K. Robinson, and M.G. White. 1993. A method for preparing large numbers of otolith sections for viewing by scanning electron microscope. *ICES J. Mar. Sci.*, 50: 227-229.

Ashford, J.R., J. Croxall, P. Rubilar, and C.A. Moreno. 1994. Seabird interactions with longlining operations for *Dissostichus eleginoides* at the South Sandwich Islands and South Georgia. *CCAMLR Sci.*, 1: 143- 153.

Ashford, J.R., J. Croxall, P. Rubilar, and C.A. Moreno. 1995. Seabird interactions with longlining operations for *Dissostichus eleginoides* around South Georgia, April to May 1994. *CCAMLR Sci.*, 2: 111-121.

Ashford, J.R., P.S. Rubilar, and A.M. Martin. 1996a. Interactions between cetaceans and long-line fishery operations around South Georgia. *Mar. Mammal Sci.*, 12(3): 452-457.

Ashford, J.R., P.S. Rubilar, C.A. Moreno, and C.M. Jones. 1996b. A shore-based randomised design used in the artesanal longline fishery for *Dissostichus eleginoides* off northern Chile. Document *WG-FSA-96/54*. CCAMLR, Hobart, Australia.

Ashford, J.R., G. Duhamel, and M. Purves. 1998. A protocol for randomised sampling of longlines in the Southern Ocean fishery for *Dissostichus eleginoides*: System of International Scientific Observation. Document *WG-FSA-98/60*. CCAMLR, Hobart, Australia.

Ashford, J.R., C. Jones, and I. Everson. 2000. Stock structure and growth in Patagonian toothfish (*Dissostichus eleginoides*) in the Southern Ocean. Document *WG-FSA-00/44*. CCAMLR, Hobart, Australia.

Ashford, J.R., C. Jones, S. Wischniowski, S. Bobko, and I. Everson. 2000. A comparison between otoliths and scales for use in estimating the age of *Dissostichus eleginoides* from South Georgia. Document *WG-FSA-00/28*. CCAMLR, Hobart, Australia.

Bagenal, T.B., and F.W. Tesch. 1978. Age and growth. In: Bagenal, T.B (Ed.). *Methods for assessment of fish production in fresh waters 3rd ed.* Blackwell Scientific Publications, Oxford: 101-136.

Baranov, F.I. 1918. On the question of the biological basis of fisheries. *Nauchn. Issled. Ikhtiologicheskii Inst. Izv.*, 1: 81-128.

Barbieri, L.R., M.E. Chittenden, C.M. Jones. 1993. Age, growth, and mortality of Atlantic croaker, *Micropogonias undulatus*, in the Chesapeake Bay region, with a discussion of apparent geographic changes in population dynamics. *Fish. Bull.*, 92: 1-12.

Barry, J.P., and P.K. Dayton. 1988. Current patterns in McMurdo Sound, Antarctica, and their relationship to local biotic communities. *Polar Biol.*, 8: 367-376.

Beamish, R.J., and H.H. Harvey. 1969. Age determination in the white sucker. *J. Fish. Res. Board Can.*, 26: 633-638.

Beamish, R.J., and D.A. Fournier. 1981. A method for comparing the precision of a set of age determinations. *Can. J. Fish. Aquat. Sci.*, 38: 982-983.

Beamish, R.J., and G.A. McFarlane. 1983. Validation of age determination estimates: the forgotten requirement. *Trans. Am. Fish. Soc.*, 112: 735-743.

Beamish, R.J., and G.A. McFarlane. 1987. Current trends in age determination methodology. In: Summerfelt, R.C. and G.E. Hall (eds). *Age and Growth of Fish*. Iowa State University Press, Ames.

Beamish, R.J., G.A. McFarlane, and D.E. Chilton. 1983. Use of oxytetracycline and other methods to validate a method of age determination for sablefish. In: *Proceedings of the Second Lowell Wakefield Fisheries Symposium, Anchorage, Alaska*. Alaska Sea Grant Report 83-8: 95-116.

Bedford, B.C. 1983. A method for preparing sections of large numbers of otoliths embedded in black polyester resin. *J. Cons. Perm. Int. Exp. Mer,* 41: 4-12.

Begon, M., and M. Mortimer. 1986. *Population ecology: a unified study of animals and plants.* Blackwell Scientific Publications, Oxford.

von Bertalanffy, L. 1938. A quantitative theory of organic growth. *Hum. Biol.*, 10: 181-213.

Beverton, R.J.H., and S.J. Holt. 1957. *On the dynamics of exploited fish populations. Fisheries Investment Series 2, Vol. 19.* UK Ministry of Agriculture and Fisheries, London.

Bjordal, A., and S. Lokkeborg. 1996. *Longlining.* Fishing News Books, Blackwell Science Ltd., Oxford.

Boehlert, G.W., and M.M. Yoklavich. 1984. Variability in age estimates in *Sebastes* as a function of methodology, different readers, and different laboratories. *California Fish and Game*, 70 (4): 210-224.

Booke, H.E. 1981. The conundrum of the stock concept - are nature and nurture definable in fishery science? *Can. J. Fish. Aquat. Sci.,* 38: 1479-1480.

Botsford, L.W., C.L. Molony, A. Hastings, J.L. Largier, T.M. Powell, K. Higgens, and J.F. Quinn. 1993. The influence of spatially and temporally varying oceanographic conditions on meroplanktonic metapopulations. *Deep Sea Res.*, 41: 107-145.

Bowker, A.H. 1948. A test for symmetry in contingency tables. *J. Amer. Stat. Assoc.*, 43: 572-574.

Bradford, M.J. 1991. Effects of ageing errors on recruitment time series estimated from sequential population analysis. *Can. J. Fish. Aquat. Sci.*, 48: 555-558.

Burchett, M.S. 1983. Age and growth of the Antarctic fish *Notothenia rossii* from South Georgia. *Bull. Br. Antarct. Surv.,* 60: 45-61.

Caddy, J.F. 1975. Spatial model for an exploited shellfish population and its application to the Georges Bank scallop fishery. *J. Fish. Res. Board. Can.*, 32: 2330-2344.

Caddy, J.F. 1993. Background concepts for a rotating harvest strategy with particular reference to the Mediterranean red coral, *Corallium rubrum*. *Mar. Fish. Rev.*, 55: 10-18.

Campana, S.E., and J.D. Neilsen. 1985. Microstructure of fish otoliths. *Can. J. Fish. Aq. Sci.*, 42: 1014-1032.

Campana, S.E., and C.M. Jones. 1997. Radiocarbon from nuclear testing applied to age validation of black drum, *Pogonias cromis*. *Fish. Bull.*, 96: 185-192.

Campana, S.E., A.J. Fowler and C.M. Jones. 1994. Otolith elemental finger-printing for stock identification of Atlantic cod (*Gadus morhua*) using laser ablation ICPMS. *Can. J. Fish. Aquat. Sci.*, 51(9): 1942-1950.

Campana, S.E., M.C. Annand, and J.I. McMillan. 1995. Graphical and statistical methods for determining the consistency of age determinations. *Trans. Am. Fish. Soc.*, 124: 131-138.

Cassia, M.C. 1998. Comparison of age readings from scales and otoliths of the Patagonian toothfish (*Dissostichus eleginoides*) from South Georgia. CCAMLR *Sci.*, 5: 191-203.

Chang, W.Y.B. 1982. A statistical method for evaluating the reproducibility of age determination. *Can. J. Fish. Aq. Sci.*, 39: 1208-1210.

Chester, A.J. 1984. *Sampling statistics in the Atlantic menhaden fishery*. NOAA Technical Report NMFS 9. Dept. of Commerce, United States.

Chikov, V.N., and Yu.S. Melnikov. 1990. On the question of fecundity of the Patagonian Toothfish, *Dissostichus eleginoides*, in the region of the Kerguelen Islands. *Voprosy Ikhtiologii*, 30(5): 864-866.

Chilton, D.E., and R.J. Beamish. 1982. *Age determination methods for fishes studied by the groundfish program at the Pacific Biological Station*. Dept. of Fisheries and Oceans, Ottawa: 102 pp.

Chittenden, M. 1989. Sources of variation in Chesapeake Bay pound-net and haul-seine catch compositions. *N. Amer. J. Fish. Man.*, 9: 367-371.

Christensen, J.M. 1964. Burning of otoliths, a technique for age determination of soles and other fish. *J. Cons. Int. Explor. Mer.*, 29: 73-81.

Clarke, A., and A.W. North. 1991. Is the growth of polar fish limited by temperature? In: di Prisco, G., B.Maresca, and B. Tota (Eds.). *Biology of Antarctic Fish*. Springer-Verlag, Berlin: 54-69.

Coggan, R.A., K. Skora, A. Murray, and M.G. White. 1990. A comparison between age determinations of the Antarctic fish *Notothenia gibberifrons* Lonnberg using scales and otoliths. *Cybium,* 14: 43-55.

Cowen, R.K., J.A. Hare and M.P. Fahay. 1991. Beyond hydrography: can physical processes explain larval fish assemblages within the Middle Atlantic Bight? *Bull. Mar. Sci.,* 53(2): 567-587.

Cox, P.A. 2000. Will tribal knowledge survive the millenium? *Science,* 287: 44-45.

Cushing, D.H. 1981. *Fisheries Biology.* 2nd.ed.Univ.Wisc.Press, Madison, Wisconsin. 200 pp.

Dayton, P.K. 1989. Interdecadal variation in an Antarctic sponge and its predators from oceanographic climate shifts. *Science,* 245: 1484-1486.

Daniels, R.A. 1983. Demographic characteristics of the Antarctic punderfish, *Harpagifer bispinis* antarcticus. *Mar. Ecol. Prog. Ser.,* 13: 181-187.

Delury, D.B. 1947. On the estimation of biological populations. *Biometrics,* 3: 145-167.

Dorval, E. 1998. *Effect of sampling errors on estimates of recruitment and fishing mortality from separable virtual population analysis.* M.Sc. Thesis, Old Dominion University, Norfolk, United States.

Edmonds, J.S., N. Caputi and M. Morita. 1991. Stock discrimination by trace-element analysis of otoliths of orange roughy (*Hoplostethus atlanticus*), a deep-water marine teleost. *Aust. J. Mar. Freshwater Res.,* 42: 383- 389.

Edmonds, J.S., R.C.J. Lenanton, N. Caputi and M. Morita. 1992. Trace elements in the otoliths of yellow-eye mullet (*Aldricheta forsteri*) as an aid to stock identification. *Fish. Res.,* 13: 39-51.

Edwards, A.W.F. 1992. *Likelihood, 2nd. edition.* Johns Hopkins University Press, Baltimore.

Efron, B., and R.J. Tibshirani. 1993. *An introduction to the bootstrap.* Chapman and Hall, New York.

Everson, I. 1980. *Antarctic fish age determination methods.* BIOMASS Handbook 8: 24 pp.

Everson, I. 1991. Stock assessment of the Patagonian toothfish (*Dissostichus eleginoides*) at South Georgia. Document *WG-FSA-91/20.* CCAMLR, Hobart, Australia.

Fisher, R.A. 1921. On the ‘probable error’ of a coefficient of correlation deduced from a small sample. *Metron,* 1(4): 3-32.

Fisher, R.A. 1922. On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. A.*, 222: 309-368.

Fournier, D., and C.P. Archibald. 1982. A general theory for analysing catch at age data. *Can. J. Fish. Aquat. Sci.,* 39: 1195-1207.

Francis, R.I.C.C. 1988. Are growth parameters estimated from tagging and age-length data comparable? *Can. J. Fish. Aquat. Sci.*, 45: 936-942.

Francis, R.I.C.C. 1995. The analysis of otolith data - a mathematician’s perspective (what, precisely, is your model?). In: Secor, D.H., J.M. Dean, and S.E. Campana (Eds). *Recent developments in fish otolith research.* University of South Carolina Press, Columbia, United States: 81-96.

Gauldie, R.W. 1991. Taking stock of genetic concepts in fisheries management. *Can. J. Fish. Aquat. Sci.,* 48: 722-731.

Gordon, A.L. 1988. Spatial and temporal variability within the Southern Ocean. In: Sarhage, D. (Ed). *Antarctic Ocean and resources variability.* Springer-Verlag, Berlin, Heidelberg.

Gould, S.J. 1977. *Ever since Darwin: reflections in natural history.* W.W. Norton and Co., New York.

Gould, S.J. 2000. Deconstructing the “Science Wars” by reconstructing an old mould. *Science,* 287: 253-261.

Gregory, R.L. 1977. *Eye and Brain: the psychology of seeing.* Third edition.Weidenfeld and Nicholson, London.

Gulland, J.A. 1965. *Estimation of mortality rates.* Annex to Arctic Fisheries Working group Report (meeting in Hamburg, January 1965). ICES, C.M. 1965, Document 3 (mimeo), Copenhagen.

Gulland, J.A. 1969. Manual of methods for stock assessment. Part I. Fish Population analysis. *FAO Man. Fish. Sci.,* 4: 1-154.

Gulland, J.A. 1971. Ecological aspects of fishery research. *Adv. Ecol. Res.*, 7: 115-176.

Gulland, J.A. 1983. *Fish stock assessment: a manual of basic methods.* John Wiley and Sons, New York.

Hanski, I., and D. Simberloff. 1997. The metapopulation approach, its history, conceptual domain, and application to conservation. In: Hanski, I.A., and M.E. Gilpin (Eds.). *Metapopulation biology: ecology, genetics, and evolution.* Academic Press, San Diego, California: 5-26

Hanski, I., and C.D. Thomas. 1994. Metapopulation dynamics and conservation: a spatially explicit model applied to butterflies. *Biol. Conserv.*, 68: 167-180.

Hastings, A. 1997. *Population Biology: concepts and models.* Springer-Verlag, New York.

Hilborn, R., and C.J. Walters. 1992. *Quantitative fisheries stock assessment: choice, dynamics and uncertainty.* Chapman and Hall, New York.

Hilborn, R., and M. Mangel. 1997. *The ecological detective: confronting models with data.* Princeton University Press, New Jersey.

Hile, R. 1936. Age and growth of the cisco, *Leucicthys artedi* (Le Sueur), in the lakes of the North-eastern Highlands, Wisconsin. *U. S. Bur. Fish. Bull.*, 48: 211-317.

Hoffman, E.E., J.M. Klinck, R.A. Locarnini, B. Fach and E. Murphy.1998. Krill transport in the Scotia Sea and environs. *Ant. Sci.*, 10(4): 406-415.

Hoenig, J.M. 1983. Empirical use of longevity data to estimate mortality rates. *Fish. Bull.*, 82: 898-903.

Hoenig, J.M., M.J. Morgan and C.A. Brown. 1995. Analysing differences between two age determination methods by tests of symmetry. *Can. J. Fish. Aquat. Sci.*, 52: 364-368.

Holland, D.S., and R.J. Brazee. 1996. Marine reserves for fisheries management. *Mar. Res. Econ.*, 11: 157-171.

Horn, P.L.1999. Age and growth of Patagonian toothfish (*Dissostichus eleginoides*) and Antarctic toothfish (*D. mawsoni*) in waters from the New Zealand Exclusive Economic Zone to CCAMLR Subarea 88.1. Document *WG-FSA-00/43.* CCAMLR, Hobart, Australia.

Hureau, J.C., and C. Ozouf-Costaz. 1980. Age determination and growth of *Dissostichus eleginoides* Smitt, 1898, from Kerguelen and Crozet Islands. *Cybium* 3e Serie, 1980(8): 23-32.

Ihssen, P.H., H.E. Booke, J.M. Casselman, J.M. McGlade, N.R. Payne, F.M. Utter. 1981. Stock identification: materials and methods. *Can. J. Fish. Aquat. Sci.,* 38: 1838-1855.

Jones, C.M. 1986. Determining the age of larval fish with the otolith increment technique. *Fish. Bull. U.S.* 87: 265-272.

Jones, C.M. 1992. Development and application of the otolith increment technique. In: Stevenson, D.K., and S.E. Campana (Ed.). *Otolith microstructure examination and analysis.* Can. Spec. Publ. Fish. Aquat. Sci., 117: 1-11.

Jones, C.M. 2000. Fitting growth curves to retrospective size-at-age data. *Fish.Res.,* 46: 123-129.

Kalish, J.M. 1990. Use of otolith microchemistry to distinguish the progeny of sympatric anadromous and non-anadromous salmonids. *Fish. Bull.,* 88(4): 657-666.

Kalish, J.M., and T.A. Timmiss. 1998. Determination of Patagonian toothfish *Dissostichus eleginoides* age, growth and population characteristics based on otoliths. Document *WG-FSA-98/40.* CCAMLR, Hobart, Australia.

Kastelle, C.R., D.K. Kimura, A.E. Nevissi, D.R. Gunderson. 1994. Using Pb-210/Ra-226 disequilibria for sablefish, *Anoplopoma fimbria,* age validation. *Fish. Bull.,* 92: 292-301.

Kimura, D.K. 1980. Likelihood methods for the von Bertalanffy curve. *Fish. Bull.,* 77(4):765-776.

Kimura, D.K., and J.J. Lyons. 1991. Between-reader bias and variability in the age-determination process. *U. S. Fish. Bull.,* 89: 53-60.

Kimura, D.K., R.R. Mandapat, and S.L. Oxford. 1979. Method, validity, and variability in the age determination of yellowtail rockfish (*Sebastes flavidus*), using otoliths. *J. Fish. Res. Board. Can.,* 36: 377-383.

Kimura, D.K., J.J. Lyons, S.E. MacLellan, and B.J. Goetz. 1992. Effects of year-class strength on age determination. *Aus. J. Mar. Freshwater Res.,* 43: 1221-8.

Kingsford, M.J. 1993. Biotic and abiotic structure in the pelagic environment: importance to small fishes. *Bull. Mar. Sci.,* 53(2): 393-415.

Kingsland, S.E. 1995. *Modeling nature: episodes in the history of population ecology.* University of Chicago Press, Chicago.

Kock, K.-H. 1990. Results of the CCAMLR fish otoliths/scales/bones exchange system. In: *Selected Scientific Papers* 1989 (*SC-CAMLR-SSP/).* CCAMLR, Hobart, Australia: 51-68.

Kock, K-H. And J. Selling. 1996. Report on scientific observations on board the Chilean longliner, 'Puerto Ballena', in statistical subarea 48.3 from March to May 1996. Document *WG-FSA-96/22*. CCAMLR, Hobart, Australia.

Kock, K.-H., G. Duhamel, and J.-C. Hureau. 1985. *Biology and status of exploited Antarctic fish stocks: a review.* BIOMASS Scientific Series no.6, SCAR: 143 pp.

Konforkin, I.N., and A.N. Kozlov. 1992. Pre-spawning and spawning biology of the Patagonian toothfish, *Dissostichus eleginoides*, around South Georgia (Subarea 48.3). Document *WG-FSA-92/13*. CCAMLR, Hobart, Australia.

Korovina, V.A., A.V. Neyelov, and Ye.P. Bondarenko.1988. The anatomy and histology of the intestine of the Patagonian Toothfish (*Dissostichus eleginoides* Smitt). *Trudy Zoologicheskogo instituta AN SRR*, 181: 56-61.

Kuehl, 1995. *Statistical principles of research design and analysis.* Wadsworth Publ. Co., California.

Kuhn, T.S. 1962. *The structure of scientific revolutions.* University of Chicago Press, Chicago.

Lai, H.L. 1987. Optimum allocation for estimating age composition using age-length key. *Fish. Bull.*, 85(2): 179-183.

Lai, H.L., and D.R. Gunderson. 1987. Effects of ageing errors on estimates of growth, mortality and Yield per Recruit for walleye pollock (*Theragra chalcogramma*). *Fish. Res.*, 5: 287-302.

Larkin, P.A. 1972. The stock concept and management of Pacific salmon. In : Simon, R.C., and P. Larkin (Ed.). *The stock concept in Pacific salmon.* H.R. MacMillan Lectures in Fisheries. University of British Columbia , Vancouver, Canada: 11-15.

Lee, R.M. 1912. *An investigation into the methods of growth determination in fishes.* Cons. Explor. Mer, Publ. de Circonstance 63:35p.

Leslie, P.H., and D.H.S. Davis. 1939. An attempt to determine the absolute number of rats on a given area. *J. Anim. Ecol.*, 8: 94-113.

Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Amer.*, 15: 237-240.

Lillibridge, J.L., G. Hitchcock, T. Rossby, E. Lessard, M. Mork, and L. Golmen. 1990. Entrainment and mixing of shelf/slope waters near the surface of the Gulf Stream. *J. Geophys. Res.*, 95: 13065-13087.

Loeb, V.J., A.K. Kellermann, P. Koubbi, A.W. North, and M.G. White. 1993. Antarctic larval fish assembalges: a review. *Bull. Mar. Sci.*, 53(2): 416-449.

Lokkeborg, S. 1990. *Rate of release of potential feeding attractants from natural and artificial bait.* M.Sc. Thesis, University of Bergen, Norway: 118 pp.

Lokkeborg, S., and A. Bjordal. 1992. Species and size selectivity in longlining fishing - a review. *Fish. Res.*, 13(3): 311-322.

MacArthur, R.H., and E.O. Wilson. 1967. *The theory of island biogeography.* Princeton University Press, Princeton, New Jersey.

MacLean, J.A., and D.O. Evans. 1981. The stock concept, discreteness of fish stocks and fisheries management. *Can. J. Fish. Aquat. Sci.*, 38: 1889-1898.

MacLellan, S.E. 1997. How to age rockfish (*Sebastes*) using *S. Alutus* as an example – the burnt otolith section technique. *Canadian Technical Report of Fisheries and Aquatic Sciences* 2146: 39 pp.

Magee, B. 1973. *Popper.* William Collins Sons and Co. Ltd, Glasgow.

Magee, B. 1997. *Confessions of a philosopher.* Weidenfeld and Nicolson, London.

Martin, J.H., R.M.Gordon, and S.E.Fitzwater. 1990. Iron in Antarctic waters. *Nature*, 345: 156-158.

Marr, J.C. 1957. The subpopulation problem in the Pacific sardine (*Sardinops caerulea*). *U. S. Fish Wildl. Serv. Spec. Sci. Rep. Fish.*, 208: 108-125.

Matsuoka, T., J. Kasu and H. Nagaleta. 1992. Capture process in vertical longline fishing. *Nippon Suisan Gakkaishi - Bulletin of the Japanese Society of Scientific Fisheries*, 58: 213-222.

Matsuoka, T., J. Kasu and H. Nagaleta. 1995. Catch and fish-loss model for vertical longline fishing with particular emphasis on injury of unharvested fish. *Fish. Sci.*, 61(5): 760-765.

May, R.M. 1975. Biological populations obeying difference equations: Stable points, stable cycles and chaos. *J. Theor. Biol.*, 49: 511-524.

McFarlane, G.A., W. Shaw, and A.V. Tyler. 1985. Sablefish: coastwide stock assessments. In: Tyler, A.V., and G.A. McFarlane (Eds.). *Groundfish stock* *assessments off the west coast of Canada in 1984 and recommended yield options for 1985.* Canadian Manuscript Report of Fisheries and Aquatic Sciences 1813.

Megrey, B.A. 1989. Review and comparison of age-structured stock assessment models from theoretical and applied points of view. *Amer. Fish. Soc. Symp.*, 6: 8-48.

Miller, D. 1999. Being an absolute skeptic. *Science,* 284: 1625-1626.

Morison, A.K., S.G. Robertson, and D.C. Smith. 1998. An integrated system for production fish aging: image analysis and quality assurance. *N. Amer. J. Fish. Management*, 18: 587-598.

Mulligan, T.J., F.D. Martin, R.A. Smucker and D.A. Wright. 1987. A method of stock identification based on the elemental composition of striped bass *Morone saxatilis* (Walbaum) otoliths. *J. Exp. Mar. Biol. Ecol.*, 114: 241-248.

Murphy, E.J., A. Clarke, C. Symon, and J. Priddle. 1995. Temporal variation in Antarctic sea-ice: analysis of a long-term fast-ice record from the South Orkney Islands. *Deep-Sea Res. I,* 42(7): 1045-1062.

Myers, R.A., and K. Drinkwater. 1989. The influence of Gulf Stream warm core rings on recruitment of fish in the northwest Atlantic. *J. Mar. Research,* 47: 635-656.

Neilson, J.D. 1992. Sources of error in otolith microstructure examination. In: Stevenson, D.K., and S.E. Campana (Eds). *Otolith microstructure examination and analysis.* Can. Spec. Publ. Fish. Aquat. Sci. 117: 115-125.

North, A.W. 1988. Age of Antarctic fish: validation of the timing of annuli formation in otoliths and scales. *Cybium,* 12(2): 107-114.

Olsen, S., and T. Laevestu. 1983. *Fish attraction to baits and effects of currents on the distribution of smell from baits.* Proc. Rep. Northwest and Alaska Fisheries Center, Seattle: 50 pp.

Orsi, A.H., T. Whitworth, and W.D. Nowlin. 1995. On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep-Sea Res. I* , 42(5): 641-673.

Park,Y-H., and L. Gamberoni. 1994. Cross-frontal exchanges of Antarctic Intermediate Water and Antarctic Bottom Water in the Crozet Basin. *Deep-Sea Res. Special Issue (Antares 1).*

Park, Y-H., L. Gamberoni, and E. Charriaud. 1993. Frontal structure, water masses, and circulation in the Crozet Basin. *J. Geophysical Res.*, 98(c7): 12361-12385.

Pauly, D. 1993. Foreword to Beverton, R.J.H., and S.J. Holt. 1957: *On the dynamics of exploited fish populations.* Fisheries Investment Series 2, Vol 19. UK Ministry of Agriculture and Fisheries, London.

Pearson, K. 1900. *The Grammar of science. Second edition.* Adam and Charles Black, London.

Piner, K.R. 1999. *Growth and yield-per-recruit modeling of spot (Leiostomus xanthurus) in the Chesapeake Bay, and a comparison of biological reference points.* Ph.D. dissertation, Old Dominion University, Norfolk, United States.

Polachek, T. 1990. Year round closed areas as a management tool. *Nat. Res. Model.,* 4: 327-354.

Pope, J.G., and J.G. Shepherd. 1982. A simple method for the consistent interpretation of catch-at-age data. *J. Cons. Int. Explor. Mer,* 40: 176-184.

Popper, K.R. 1959. *The logic of scientific discovery.* Hutchinson, London. Orig. Pub.(1934) as: Logik der Forschung. Springer-Verlag, Vienna.

Priddle, J., J.P. Croxall, I. Everson, R.B. Heywood, E.J. Murphy, P.A. Prince, and C.B. Sear 1988. Large-scale fluctuations in distribution and abundance of krill - a discussion of possible causes. In: Sarhage, D. (Ed.). *Antarctic Ocean and resources variability.* Springer-Verlag, Berlin, Heidelberg.

Pulliam, H.R. 1988. Sources, sinks and population dynamics. *Amer. Naturalist,* 132: 652-661.

Radtke, R.L., and T.F. Hourigan. 1990. Age and growth of the Antarctic fish *Nototheniops nudifrons. Fish. Bull.,* 88: 557-571.

Radtke, R.L., T.E. Targett, A. Kellermann, J.L. Bell, and K.T. Hill. 1989. Antarctic fish growth: profile of *Trematomus newnesi. Mar. Ecol. Prog. Ser.,* 57: 103-117.

Ricker, W.E. 1958. *Handbook of computations for biological statistics of fish populations.* Bull. Fish. Res. Board Can. 119: 300 pp.

Ricker, W.E. 1969. *Effects of size-selective mortality and sampling bias on estimates of growth, mortality, production, and yield.* J. Fish. Res. Board Can. MS Rep. 961: 38 pp.

Ricker, W.E. 1975. *Computation and interpretation of biological statistics of fish populations.* Bulletin 191, Dept. Environment, Fisheries and Marine Service, Ottawa, Canada.

Royce, W.F. 1972. *Introduction to the fishery sciences.* Academic Press, New York, N.Y: 351 pp.

Rubilar, P., C.A. Moreno, J.R. Ashford, and I. Everson. 1993. An exploratory fishing expedition for *Dissostichus eleginoides* around the South Sandwich Islands, Antarctica. Document *SC-CAMLR-XII/BG/4.* CCAMLR, Hobart, Australia.

Rubilar, P., C.A. Moreno, J.R. Ashford, and I. Everson. 1994. Depletion experiment of *Dissostichus eleginoides* stock in the south of South Georgia Island (Antarctica). Document *WG-FSA-94/31.* CCAMLR, Hobart, Australia.

Russell, F.S. 1931. Some theoretical considerations on the 'overfishing' problem. *J. Cons. Explor. Mer,* 6: 3-27.

Sainsbury, K.J. 1980. Effect of individual variability on the von Bertalanffy equation. *Can. J. Fish. Aquat. Sci.,* 37: 241-247.

Schaefer, M.B. 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Bull. Inter-Am. Trop. Tuna. Comm.,* 1(2): 27-56.

SC-CAMLR. 1992. *Report of the eleventh meeting of the Scientific Committee for the Conservation of Antarctic Marine Living Resources.* CCAMLR, Hobart, Australia.

SC-CAMLR. 1994. *Report of the thirteenth meeting of the Scientific Committee for the Conservation of Antarctic Marine Living Resources.* CCAMLR, Hobart, Australia.

SC-CAMLR. 1995. Report of the workshop on methods for the assessment of *Dissostichus eleginoides.* In : *Report of the fourteenth meeting of the Scientific Committee for the Conservation of Antarctic Marine Living Resources, Annex 5.* CCAMLR, Hobart, Australia.

SC-CAMLR. 1996. Report of the Working Group on Fish Stock Assessment. In : *Report of the fifteenth meeting of the Scientific Committee for the Conservation of Antarctic Marine Living Resources.* CCAMLR, Hobart, Australia.

SC-CAMLR. 1997. Report of the Working Group on Fish Stock Assessment. In : *Report of the fifteenth meeting of the Scientific Committee for the Conservation of Antarctic Marine Living Resources.* CCAMLR, Hobart, Australia.*Report of the working group on fish stock assessment.* CCAMLR, Hobart, Australia.

Schnute, J.T. 1981. A versatile growth model with statistically stable parameters. *Can. J. Fish. Aquat. Sci.,* 38: 1128-1140.

Seber, G.A.F. 1982. *The estimation of animal abundance and related parameters.* Oxford University Press, New York.

Shust, K.V., P.S. Gasiukov, R.S. Dorovkikh, and B.A. Kenzhin. 1990. The state of *Dissostichus eleginoides* stock and TAC for 1990/91 in Subarea 48.3 (South Georgia). Document *WG-FSA-90/34.* CCAMLR, Hobart, Australia.

Skora, K.E., and J.Sosinski. 1983. Observations on the ichthyofauna distribution in the regions of the Scotia Sea and Antarctic Peninsula. *Polish Polar Res.*, 4(1-4): 49-55.

Smith, P., and P. Gaffney. 2000. Population genetics of Patagonian toothfish *Dissostichus eleginoides* and fillet identification of Patagonian toothfish and Antarctic toothfish *D. mawsoni.* Document *WG-FSA-00/53.* CCAMLR, Hobart, Australia.

Sokal, R.R., and F.J.Rohlf. 1995. *Biometry.* 3rd.ed. W.H.Freeman and Co., New York: 887 pp.

Somerton, D.A., and B.S. Kikkawa. 1995. A stock survey technique using the time to capture individual fish on longlines. *Can. J. Fish. Aquat. Sci.*, 52(2): 260-267.

Stein, M. 1989. Seasonal variation of water masses in Bransfield Strait and adjacent waters. *Arch. Fisch Wiss.*, 39(Beih.1): 15-38.

Stewart, N.A. 1926. Development, growth and food habits of the white sucker, *Catostomus commersoni* Le Sueur. *U. S. Bur. Fish. Bull.*, 42: 147-184.

Thompson, S.K.1992. *Sampling.* John Wiley and Sons, New York: 343 pp.

Thompson, W.F., and F.H. Bell. 1934. *Biological statistics of the Pacific halibut fishery. 2. Effect of changes in intensity upon total yield and yield per unit of gear.* Rep. Int. Fish. (Pacific Halibut) Comm. 8: 49 pp.

Thorrold, S.R., C.M. Jones, P.K. Swart, and T.E. Targett. 1998. Accurate classification of juvenile weakfish (*Cynoscion regalis*) to estuarine nursery areas based on chemical signatures in otoliths. *Mar. Ecol. Prog. Ser.*, 173: 253.

Trathan, P.N., J. Priddle, J.L. Watkins, D.G.M. Miller, and A.W.A. Murray. 1993. Spatial variability of Antarctic krill in relation to mesoscale hydrography. *Mar. Ecol. Prog. Ser.*, 98: 61-71.

Tukey, J.W. 1949. One degree of freedom for non-additivity. *Biometrics,* 5: 232-242.

Underwood, A.J. 1996. Spatial patterns of variance in density of intertidal populations. In: Floyd, R.B., A.W. Sheppard, and P.J. de Barro (Eds). *Frontiers of Population Ecology.* CSIRO Publishing, Melbourne: 369-389.

Underwood, A.J. 1997. *Experiments in ecology.* Cambridge University Press, Cambridge.

Van Oosten, J. 1923. The whitefishes (*Coregonus clupeaformis*). A study of the scales of whitefishes of known ages. *Zoologica* (New York), 2: 381-412.

White, M.G. 1991. Age determination in Antarctic fish. In: Di Prisco, G., B. Maresca, and B. Tota (Eds.). *Biology of Antarctic Fish.* Springer-Verlag, Berlin: 87-100.

White, W.B., and R.G. Peterson. 1996. An Antarctic circumpolar wave in surface pressure, wind, temperature and sea-ice extent. *Nature,* 380: 699-702.

Wilson, C.A., E.B. Brothers, J.M. Casselman, C.L. Smith, and A. Wild. 1983. Glossary. In: E. Prince and L. Pulos (Eds.). *Proceedings of the International Workshop on Age Determination of Oceanic Pelagic Fishes: Tunas, Billfishes, Sharks.* NOAA Technical Report (NMFS) 8: 207-208.

Yefremenko, V.N. 1979. Larvae of six species of fishes of the family *Nototheniidae* from the Scotia Sea. *Voprosy Ikhtiologii,* 19(6): 1068-1076.

Young, Z., R. Gili, and L. Cid. 1995. Prospeccion de bacalao de profundidad entre las latitudes 43°S y 47°S. *Informe Tecnico IFOP-SUBPESCA*, Chile: 46 pp.

Yukhov, V.L. 1972. The range of fish of the genus *Dissostichus* (family *Nototheniidae*) in Antarctic waters of the Indian Ocean. *Voprosy ikhtiologii,* 12(2): 384-385.

Zakharov, G., and Z. Frolkina. 1976. Some data on the distribution and biology of the Patagonian toothfish (*Dissostichus eleginoides*) occurring in the southwest Atlantic. *Trudy AtlantNIRO*, 65: 143-150.

Zhivov, V.V., and V.M. Krivoruchko. 1990. On the biology of the Patagonian Toothfish, *Dissostichus eleginoides*, of the Antarctic part of the Atlantic. *Voprosy ikhtiologii,* 30(5): 861-864.

VITA

Julian R. Ashford
Biological Sciences Department,
110 Mills Godwin Jr. Building,
Old Dominion University,
Norfolk, VA 23529.

Education

B.Sc. (Honours) Zoology. 1983. University of Durham, United Kingdom.
M.Sc. by thesis. 1993. University of Durham, United Kingdom,.
Ph.D. Ecological Sciences. 2001 Old Dominion University, Virginia.

Previous Employment

1997,1994,1993. Scientific Observer for the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR).
1991-1992. Contract Marine Biologist, British Antarctic Survey.
1986-89. Assistant Marine Biologist, British Antarctic Survey. Signy Island Research Station, South Orkneys, Antarctica.

Publications

'An assessment of CCAMLR measures employed to mitigate seabird mortality in longlining operations for *Dissostichus eleginoides* around South Georgia.' - J.R. Ashford and J.P. Croxall. *CCAMLR Science* (1998) 5: 217-230.

'Interactions between cetaceans and long-line fishery operations around South Georgia.' - J.R. Ashford, P.S. Rubilar and A. Martin. *Marine Mammal Science* (1996): 12(3): 452-457.

'Seabird interactions with longlining operations for *Dissostichus eleginoides* around South Georgia, April-May 1994.' - J.R. Ashford, J.P. Croxall, P.S. Rubilar and C.A. Moreno. *CCAMLR Science* (1995), 2: 111-121.

'An annual profile of growth in the otoliths of immature *Notothenia coriiceps* Richardson in relation to the environment at Signy Island, Antarctica.' - J.R. Ashford and M.G. White. *Antarctic Science* (1995) 7(3): 265-277.

'Cross-sectional structure and validation of the timing of annulus formation in the otoliths of *Notothenia coriiceps* Richardson.' J.R. Ashford and M.G. White. *Cybium* (1993), 17(2): 153-163.

'A method for preparing large numbers of otolith sections for viewing by scanning electron microscope.' - J.R. Ashford, K. Robinson, and M.G. White. *ICES J. Mar. Sci.* (1993), 50: 227-229.