A novel method for obtaining age compositions from ancillary biological data and its potential for cost reductions in stock assessments

Simon Robertson and Vladimir Troynikov

State Government Victoria
Department of Primary Industries

Australian Government
Fisheries Research and Development Corporation

Project No. 2005/023

A novel method for obtaining age compositions from ancillary biological data and its potential for cost reductions in stock assessments

Simon Robertson and Vladimir Troynikov

Fisheries Victoria, Fisheries Research Branch,
Department of Primary Industries, Queenscliff, Victoria

April 2010

Project No. 2005/023

ISBN 978-1-74264-135-5 (print)

Preferred way to cite:
Robertson S. and Troynikov V. (2010) A novel method for obtaining age compositions from ancillary biological data and its potential for cost reductions in stock assessments. Final report to Fisheries Research and Development Corporation Project No. 2005/023. Primary Industries Research Victoria, Queenscliff.

Published by Fisheries Victoria, Department of Primary Industries, Queenscliff, Victoria, 3225.

Formatted/designed by Fisheries Victoria Queenscliff
Printed by Fisheries Victoria Queenscliff, Victoria

The Fisheries Research and Development Corporation plans, invests in and manages fisheries research and development throughout Australia. It is a statutory authority within the portfolio of the federal Minister for Agriculture, Fisheries and Forestry, jointly funded by the Australian Government and the fishing industry.

Table of Contents

List of Tables

List of Figures

NON-TECHNICAL SUMMARY

2005/023	Title: A novel method for obtaining age compositions from ancillary biological data and its potential for cost reductions in stock assessments

Principal Investigator: **Simon Robertson**
Address: Fisheries Victoria
Fisheries Research Branch
PO Box 114, Queenscliff, 3225
Tel: (03) 5258 111 Fax: (03) 5258 0270
Email: Simon.Robertson@dpi.vic.gov.au

Co-Investigator: **Vladimir Troynikov**
Address: As above
Tel: (03) 5258 0111 Fax: (03) 5258 0270
Email: Vladimir.Troynikov@dpi.vic.gov.au

Objectives:

1. During first 12 months provide a 'Proof of Concept Study' on two species.
   One of the species with a short longevity and stable age composition over time and the second, with variable recruitment and higher longevity.
2. Using Central Ageing Facility data, determine which commercially important fish stocks and associated data provide sufficient information for estimation of the age composition from length frequencies.
3. Apply limited length-at-age data with extensive length-frequency data to improve time series of age compositions for fish stock assessment.
4. Determine the appropriate sample size for collection of age data.
5. Examine robustness of the numerical methods to length-specific selectivity.
6. Compare the new techniques with existing numerical methods.
7. Develop user friendly software and data environment for numerical reconstruction of cohorts and age groups.
8. Disseminate method to a broad audience of end-users.

Non-technical Summary:

Fredholm first kind equation (FFK) method for estimating age composition provides the potential of using one year's ageing data, and the length frequency data from another year to reconstruct the age composition from the year where ageing samples were not taken. This works by using the distributions of length-at-age from the aged sample and decomposing the target length frequency into the underlying age-classes. The successful application of this method would allow the age composition to be calculated from the length frequency, without the prerequisite of samples aged from that year or period. This would have several major benefits over the current age-length key (ALK) system. The current ALK does not allow the age composition to be estimated from a length frequency where no ageing data were sub-sampled. To understand the FFK methods' utility to fisheries science, it is necessary to understand the current ALK method used to reconstruct the age composition from a length frequency distribution.

Age length Key

The ALK method has been used as a fisheries management tool since its development by Fridriksson (1934). Age composition for the South East Fishery (SEF) is obtained currently by the age-length key in a three-step process. These are:

- Randomly sample the catch and measure each fish.
  This provides the length frequency from which the age composition will be determined. The numbers of fish measured for this step would be relatively large (>5000).
- Randomly collect a sub-sample from the larger length frequency sample and remove the otoliths / scales or vertebrae and age these samples using traditional techniques.
  The numbers of samples used for ageing would normally be in the hundreds.
- Construct an age-at-length matrix (distribution of ages at each length-class).
  This matrix provides information on the proportion of each age-class at a given length-class. The proportion for each age-class in a given length-class is then multiplied by the frequency at each length-class from the larger sample to provide the age composition of the larger length frequency sample which has not been aged.

A fundamental requirement of the ALK method is that the sub-sample of material (usually otoliths) used for ageing is obtained from the samples which comprise the larger length frequency data set. This is because the information contained in larger length frequency information reduces the error when it is used to reconstruct the age composition (Kimura 1977).

When this fundamental requirement is met, the ALK method provides a robust and accurate estimate of the age composition in the length frequency sample. However, the ALK method is both time consuming and costly due to the requirement to sub-sample ageing data from the larger length frequency sample.

The age-estimation process is expensive costing in excess of $AUD15.00 per sample. With larger multi-species fisheries, and each species requiring 100s of fish to be aged, costs can become excessive. The time necessary to age the samples and determine the age composition is another factor, To generate the age structure of a population, hundreds of sampled fish must be aged. This process typically takes many months. The technical expertise required to age samples to the required precision is not a common skill. The nexus of these three issues (cost, time and technical expertise) drives the motivation to develop new and more efficient ways to gather the same age composition data. The FFK method has the potential to obviate these problems.

The FFK method

The FFK method uses the *a priori* distribution of length-at-age to decompose a mixed length-at-age distribution to the underlying age frequencies. For example, each age-class has a distribution of length. Known distributions of length-at-age can be used to determine what combinations and frequencies of these age-classes constitute the unknown length frequency distribution. The FFK method is a series of equations that does this iteratively.

The FFK method theoretically provides an alternative to the ALK method for determining age composition. The method is outlined in the following two steps:

- Use *a priori* ageing data (samples that have been aged in the past, or data from the year of collection) to construct a length-at-age matrix. The information in the matrix contains the distribution of length-classes at each age-class. Past data can only be used if the length-at-age distributions have remained constant through time.
- Use a series of equations (Fredholm First kind equations) iteratively to determine which combination of length-at-age distributions would be necessary to construct the observed length distribution in the larger sample (Troynikov and Robertson 2005).

We have demonstrated that this method converges to the true underlying distributions of age (Troynikov and Robertson 2005).

The major advantage of the FFK method is that it provides a mechanism where age composition data can be derived using length frequency data and age composition data from different years and sampling events. The requirement that age composition data are derived from the same year and sampling events as the length composition data to which they are to be applied is a major restriction imposed on the ALK technique but is not a prerequisite for the FFK method.

All methods have restrictions. The FFK method is based on the assumption that the *a priori* distributions of length-at-age are similar to the distributions of length-at-age in the larger length frequency sample for which the age composition is sought.

The FFK method was robust and generated accurate age compositions within a simulation framework for both generated and fisheries data sets.

When the method was applied to fisheries data obtained from the South East Fishery, the method failed. The reasons for this failure are unclear. Failure may be attributed to non-stable length-at-age or biases introduced though sampling protocols. Biases that are introduced through the sampling process also affect the ability to obtain accurate age composition data using the ALK method. The FFK method was successfully applied to real-world fisheries data obtained from France.

The project was terminated after the second milestone for failure to reconstruct age compositions from the South East Fishery.

OUTCOMES ACHIEVED

The results from this project demonstrated that the FFK method has the potential to be used in fisheries management. It could not demonstrated that the FFK method was applicable to data currently available for the species chosen from the South East Fishery. As a result it was decided to terminate the project after the first year.

The failure of the method to reconstruct age compositions for two species from the SEF using length-at-age distributions from one year and length frequency data from another year was perplexing. The reasons for the failure of the method using real fisheries data for these species suggest that the sampling protocols used did not suit the application of this method or alterations of the length-at-age distributions between years obviated the technique. Whatever the reasons for the failure, i.e. either non-stable length distribution at age or sampling bias, the finding has implications for the current stock assessment process and the current data collection methods should be reviewed.

Where the data were sampled in a regime that was considered appropriate, the FFK method produced results that would provide age compositions that were of sufficient accuracy (close to those obtained using traditional ALK techniques).

In all simulations using both fisheries data and generated data used to test the technique, the results demonstrated that the technique does work at a level that would provide accurate age composition data. Further, we demonstrated the technique works on some real-world fisheries data sets.

Keywords: School Whiting, Blue Grenadier, cod, age composition, FFK, Fredholm first kind equations

Acknowledgments

We gratefully acknowledge funding from the FRDC. We would like to thank Terence Walker for his ideas and comments on the manuscript. We would also like to thank Leanne Gunthorpe who also provided useful comments on the manuscript. Matthew Koopman for providing Integrated Scientific Monitoring Program (ISMP) length frequency data for School Whiting and grenadier is also acknowledged and thanked. To the agers, Therese Stokie, Corey Green and Kyne Krusic-Golub, who aged the School Whiting and grenadier – thank you.

We would like to thank Ronan Fablet and Robert Bellail (IFREMER, France) for providing age and length frequency data for additional trials of the technique.

Finally, we would like to thank the anonymous reviewer whose input made this report more concise, readable and a better final report.

FINAL REPORT

05/023	**A novel method for obtaining age compositions from ancillary biological data and its potential for cost reductions in stock assessments**

Background

This project trialled the implementation of a new method for determining the age composition using samples of length frequency. The successful application of this technique would allow a significant reduction in the need for annual ageing of samples and makes collected ageing data more valuable.

Catch age-composition data provide one of the most important pieces of information needed to manage a fish stock. These data provide key information on the condition of a stock and provide a basis for effective management strategies. Age-composition data are important because the effects of fishing often manifest in the age composition of a stock long before they are evident from analysis of the length frequency or catch-per-unit-effort data. Age-composition data are therefore fundamental for stock assessments, which are in turn used for setting quotas. The quotas can then be adjusted either up or down to ensure that the fishery is managed responsibly within the framework of sustainability. It has been estimated that approximately eight million fish are aged on a worldwide scale to this end (Campana 2001). In the South East Fishery (SEF) age data are collected from approximately 10,000 fish annually,

Direct ageing techniques using otoliths,scales, vertebrae or spines of fish are an expensive and time-consuming process. The application of the approach involves extensive preparation, and the quality of the resulting age estimates relies on experienced and skilled technical staff. Given the large number of fish species which are aged, high demand for age-structured time-series data and limited resources, ageing often provides insufficient resolution across all species and all years. Many time series of ageing data are of short duration or have missing years which results in insufficient information for reliable quantitative fish stock assessments.

The problem with the age-structured approach to stock assessment is that ageing is an expensive process. To gain an accurate insight into the population dynamics of a fish stock, hundreds or thousands of fish need to be aged depending on longevity and stock structure. The older the fish, generally the more samples need to be aged, this is demonstrated in the SEF stock assessment process through sample size modelling. . With complex stock structuring, samples need to be aged from each group. To detect changes in age composition over time requires this process to be repeated each year.

These factors were recognised when the age-length key (ALK) was developed in 1934 (Fridriksson 1934). The ALK method requires a smaller sample from the catch to be aged, usually numbering in the hundreds. The ALK method requires a large sample of the catch (which is representative of the stock, subject to gear selectivity and catchability) to be measured for length. This is required to ensure the structure of the catch is adequately represented, Typically thousands of fish are measured. It is then a straightforward mathematical exercise to multiply the proportions of ages in each length class from the aged sample by the number of fish in each length class in the larger length frequency sample. The age frequency for the catch is then calculated.

The ALK method provides a robust measure of the age composition of the catch. There is however a severe limitation with this technique. The sub-sample used for the initial ageing must be sampled from the larger sample. The proportions of age at each length class are affected by factors such as mortality, variable recruitment and migration, which may not be stable between sampling events. The net effect is that ageing must be carried out on an annual basis (or the time-step from which the age composition is required) to accurately assess the composition of the stock.

Over the past five years, cutting-edge developments in mathematical techniques at the Department of Primary Industries (DPI) Queenscliff theoretically allow this requirement to be removed by using the Fredholm First Kind (FFK) equation method. Using 10 years of School Whiting data from the Central Ageing Facility, the FFK method was tested using sub-sampled age-at-length data to represent data in a typical sample collection. The age composition of the larger sample was then reconstructed using the FFK method and compared to the age composition from the traditional age-length key. The age composition was then altered by removing age classes (to simulate changes in age frequency over time) and reconstructions of age composition were again compared. The results were very promising and were presented at the 3rd World Otolith Symposium in Townsville, Australia, 2004, where the results were well received. The results of this pilot study were published in a peer-reviewed journal (Troynikov and Robertson 2005). The value of the new approach was recognised by several world-leading fish ageing scientists as an important scientific breakthrough. This study extends the application of the FFK method to temporally disparate data.

Need

Age-composition data provide the key information necessary to effectively manage fisheries. The study provides a mechanism where reliable estimates of age composition data can be gained using length frequency data and age composition data from different years and sampling events, which has previously been impossible. The benefits would be a reduced need for production ageing, more timely age-composition data and the ability to construct age-composition data from historical length frequency data where no samples were collected for ageing.

Currently the Age-Length Key (ALK) is the most widely used numerical method for assessing the age composition in a large sample of length-frequency data. However, the application of ageing data in this approach is restricted to the original sample of length distribution (ageing data from the same year the length-frequency sample is taken). Due to this severe limitation, the ageing information must be regenerated for each new data sample. Using the Fredholm First Kind equations, previous years ageing data can be used to generate the underlying age composition from the current length-frequency data. Furthermore, the ageing data may be added to include many years, improving the robustness of the statistic which can then be used to decompose the underlying age distribution from the given length frequency.

As noted by a number of referees, the major problem with the current ALK methods is variable recruitment. We have demonstrated that the FFK technique is tolerant to the most extreme changes in age frequency (Troynikov and Robertson 2005). These extreme changes in age frequency, tested in that study, were greater than any changes that could occur naturally through recruitment variability or failure.

The cost of collecting ageing data is high, with approximately $178,000 spent each year on ageing samples from commercially important species within the South East Fishery. Due to the cost, not all the commercially important species can be aged, so species that are aged are prioritised on a scientific and social-political basis. The cost-benefit of applying the FFK approach is intuitively a large reduction in cost of ageing to industry and more timely information on the age structure of the population. A cost-benefit analyses will need to be conducted on a species by species basis. This is a function of different cost structures for ageing different species and different numbers of samples that need to be aged for each species. These different numbers of estimates that need to be made for each species are primarily due to the longevity of individuals of those species and stock structuring.

The successful application of the FFK method to determine age composition would benefit the South East Trawl Fishery, the Great Australian Bight Trawl Fishery and the Gillnet, Hook and Trap Fishery. Further, the success of this technique would benefit any stock assessment programs that rely on age-structured data.

An important benefit of the project is the ability to reconstruct age composition using *a priori* length frequency data and current age composition data. In this scenario the age composition could be reconstructed for species where samples were not aged but length-frequency data were collected. This would enable age-structured population analysis where, previously, the lack of ageing data prevented these stock assessment techniques from being applied. The net effect of this approach is to greatly improve the knowledge base from which commercial species are managed. One of key advantages of this approach is that it is complementary with current methods and can provide temporal and spatial coverage of age-composition information, which is currently cost prohibitive.

Objectives

1. During first 12 months provide a 'Proof of Concept Study' on two species. One of the species with a short longevity and stable age composition over time and the second, with variable recruitment and higher longevity.
2. Using Central Ageing Facility data, determine which commercially important fish stocks and associated data provide sufficient information for estimation of the age composition from length frequencies.
3. Apply limited length-at-age data with extensive length-frequency data to improve time series of age compositions for fish stock assessment.
4. Determine the appropriate sample size for collection of age data.
5. Examine robustness of the numerical methods to length-specific selectivity.
6. Compare the new techniques with existing numerical methods.
7. Develop user friendly software and data environment for numerical reconstruction of cohorts and age groups.
8. Disseminate method to a broad audience of end-users.

Outcomes

The proposal was developed in two parts, the first component is a 'Proof of Concept Study' where the use of the Fredholm First Kind Equations to provide age compositions from length frequency data were further examined. If this was not successful, the project would be terminated at the end of the second milestone. The second and third year would examine the application of the FFK method to a broad range of species.

The results with simulations using both real data (from School Whiting) and generated distributions were outstanding. However, the FFK method was not proven to be accurate when decomposing length-frequency data for Blue Grenadier and School Whiting to the respective age classes between years. Several other species were also examined. The patterns of inconsistencies with respect to age composition reconstruction were similar. Although the method works very effectively, the data from the SEF were inappropriate for use with this method. The project ceased after the second milestone.

Successful application of the FFK method was achieved for a dataset supplied by the French government research agency (IFREMER). The dataset was for Atlantic cod (*Gadus morhua*). The dataset spanned five years of catch and ageing data sampled at monthly intervals. The results from these data successfully demonstrated that the technique worked. The results were accurate. The FFK method should be considered to be a useful tool for fisheries management.

Chapter 1. Mathematical descriptions

Introduction

This chapter provides a technical overview of the FFK method used in this project. Essentially the FFK method decomposes a length-frequency sample into the underlying age classes. Each age class has an associated length distribution. The combined length frequency is constituted by an unknown group of age classes. Using *a priori* length distributions for each age class, the FFK method determines which combinations of age-classes constitute the length frequency sample. The technique is theoretically and demonstrably robust managing both missing age classes and strong modal progressions.

An underlying requirement for the FFK method is that the length distributions at each of the age-classes remains stable between the *a priori* sample and the distributions of length at each age-class within the length frequency sample for which the age composition is unknown.

To determine the age composition of a length frequency sample, the sub-sample used for ageing must be drawn from the length frequency sample otherwise there is no guarantee on the accuracy of the resultant age composition. The FFK method has a similar condition for formulating the initial matrix. The length-at-age matrix is Bayes corrected by the length frequency from which the aged sample was drawn. Once the matrix is determined, it can then be applied using the FFK method to unknown length frequency samples to determine the age composition. The ALK method requires a sub–sample of age-at-length data from each length frequency where age composition data are required.

The technical description of the mathematical principles that underpin this project is described in more detail below.

Technical description

Mathematical formulation of the problem

Several statistical parametric and non-parametric methods have been proposed within the mixture distribution approach. An overview of the statistical properties and some numerical results from these methods can be found in Clark (1981), Kimura and Chikuni (1987), Hoenig *et al* (1993), Gavaris and Eckhaute (1994) and Elmore *et al* (in press). In the framework of the mixture distribution approach, practical problems have different input information that require the use of different numerical methods.

A "weak quasi-solution method" based on a technique proposed in Troynikov (2004) was adopted. The technique was developed using the regularisation methods for "ill-posed problems" (Tichonov *et al.* 1995). The credibility of regularisation methods for problems with noisy data is well known in physics, engineering and many other fields of applied science (Bukuchansky and Goncharsky, 1994). Using this approach, both continuous and discrete mixture distribution problems can be addressed and either continuous or discrete age distributions can be estimated. This method ensures uniform convergence, which is a very important characteristic for a numerical method.

The mathematical formulation of the problem can be viewed as the Fredholm First Kind equation

$$\int_{t_1}^{t_2} P(x \mid t)g(t)dt = f(x), \quad x \in [x_1, x_2], \qquad (1)$$

where $g(t)$ is an unknown density function of age, $f(x)$ is a density function of the age-dependent variable, where x can be scalar (e.g. fish length), or multi-dimensional (e.g. fish length and otolith weight) age-dependent measurements. Kernel $P(x \mid t)$ of the operator of equation (1) is the conditional density function of x on age t. For instance, $P(x \mid t)$ can be the density function of fish length x at age t.

In practice, functions $P(x \mid t)$ and $f(x)$ are known as the approximations $P_h(x \mid t)$ and $f_\delta(x)$, where subscripts h and δ denote errors in the right hand side and the operator of the equation. Thus, in practice, equation (1) should be substituted by the equation

$$\int_{t_1}^{t_2} P_h(x \mid t)g(t)dt = f_\delta(x), \quad x \in [x_1, x_2], \qquad (2)$$

In equation (2) the value of the error in the left hand side is measured in terms of the operator norm and in the right hand side in the norm of L2[x1,x2] functional space. It is well known that equation (1) and (2) belong to the class of linear ill-posed problems, where a small change (error) in the data can lead to large alteration of the solutions. For each particular data set the solution of equation (2) may be obtained, but there is no guarantee that the solutions of equation (2) will converge to the solution of equation (1) when the errors in data converge to zero (Tichonov *et al.* 1995). To obtain the solutions of equation (2) that converge to the solution of (1) the information about values of errors in (2) and a priori information on the "class of smoothness" of the anticipated solution *g(t)* are required (Tichonov *et al*. 1995).

In the context of estimating the age distribution using empirical population data, the values of errors are usually not available and the information on the anticipated solution $g(t)$ is often subjective and unreliable. To overcome this situation, a weak quasi-solution method was considered in Troynikov (2004). Equations (1) and (2) have been converted to the Fredholm first kind equations with respect to the anti-derivative *q(t)* of the unknown function;

$$q(t) = \int_{t_1}^{t} g(\tau)d\tau.$$

A compact set of distribution functions *q(t)* (Tichonov, *et al*. 1995) is the natural set of well-posedness in this approach. Note this method addresses both discrete and continuous mixture distribution problems, therefore discrete and continuous age distributions can be estimated. The method guarantees uniqueness and uniform convergence (convergence in $C[t_1, t_2]$ functional space) of the weak quasi-solution when the data errors converge to zero. This type of convergence corresponds to the reduction of the maximum error in the approximate solution. Convergence in $C[t_1, t_2]$ implies convergence in $L2[t_1, t_2]$ (reduction of quadratic error), therefore uniform convergence is a most desirable characteristic of any numerical method. Using this method, knowledge of the value of the errors in data is not necessary. In contrast to the weak quasi-solution method, in all other existing methods the set of density functions *g(t)* is the set of anticipated solutions. Intuitively, the weak quasi-solution method provides the best convergence because, in the functional space $L2[t_1, t_2]$, the set of anticipated solutions (set of distribution functions) is a "small" subset of non-negative, square-integerable functions on $[t_1, t_2]$.

The age distribution function $q(t)$ is used to calculate the percentages $p_{i,k}$ of a population that belong to any given age interval $[t_i, t_k) \in [t_1, t_2]$;

$$p_{i,k} = 100 \times (q(t_k) - q(t_i))\%.$$

Denote $p_i = p_{i,i+1}$ as the percentages for one year interval. To evaluate the performance of the method, the quadratic average error

$$(\Sigma(a_i - p_i)^2 / n)^{1/2} \quad , \ i=1,\ldots, n,$$

was calculated, where a_i are the estimates obtained using ALK method.

Empirical approximation to P(x|t)

In general, ageing data that are used in the ALK method cannot be directly applied for empirical estimation of size-at-age distributions. Empirical size-at-age distributions are biased because usually samples of age data are not uniformly random, but rather size stratified to equalise representation of all size intervals. To correct for the bias, we use Bayesian decomposition of size-at-age distribution ;

$$P(x|t) = P(t|x)/ P(t), \ P(t) = \int P(t|x) f(x)dx, \qquad (3)$$

where $P(t|x)$ and $f(x)$ can be estimated directly from age and size data and the integral is approximated by the summing. The second equation in (3) is ALK estimate of age composition, therefore it is compulsory that the age data be a sub-sample of the size data.

Chapter 2. Simulations

Introduction

To test the applicability of the method from first principles, a series of simulations was undertaken. The simulations used generated data. The simulations were designed to stress the method in stages, with each simulation presenting a more difficult mixture distribution problem. The final simulation approximated the difficulty associated with overlapping length-at-age distributions found in fisheries data.

All of the test distributions generated in this component of the study were normally distributed. Normal distributions were chosen for convenience within the simulation framework. The FFK method requires no parameterisation of the underlying distributions. These simulations could have been generated using other distribution functions.

The first simulation was undertaken to test the FFK method on data with two age classes from a mixed distribution. The age classes were characterised by

- two separate modes
- a high degree of separation between the means (35 and 55 respectively)
- the same standard deviation (5).

This simulation is comparable to testing two age classes within a mixed distribution and approximates distributions of length for two separate age-classes, which have not yet reached the asymptote of a growth curve.

The second simulation tests the FFK method on data with two age classes that are characterised by

- one mode
- the same mean (35) and
- different standard deviations (5 and 10).

This simulation is comparable to testing the method on two age-classes, which have reached an asymptotic length on some growth function.

The third series of simulations combines the first two scenarios. In most ageing studies more than two age-classes would be present. This simulation tests the method on data with six age classes that are characterised by

- three means;
- each mean had two standard deviations.
- overlapping distributions of all six classes in some areas of the sample space.

This simulation is comparable to testing the method on the challenges represented with length frequency data sampled from the South East Fishery.

Method

Data generation for simulations

The techniques used to generate the data were common to all datasets used in this component of the study. All data were generated using the normal distribution data generator in Microsoft Excel™.

Given a mean and a standard deviation the MS Excel random number generator was used to generate normally distributed data for each class. The initial seed was set to a predetermined integer to enable generation of the same set of random numbers for further studies. Resultant numbers from the random number generation were binned to the lowest whole integer. For each age class, 400 random numbers were generated for the first two simulations. The third simulation contained 800 random numbers per age class. These normally distributed random numbers, parameterised in this simulation by user-supplied mean and standard deviation, provided the data that represented the population distribution of a given variable, in this instance an analogy for length-at-age. The means, standard deviations and initial seed values used to generate the datasets for the first two simulations are shown in Table 1.

The distributions of length-at-age for class A and class B, which represented the population distributions, are shown in Figures 1 and 2 for simulation 1 and simulation 2 respectively. Displayed in figures are the frequency distributions of length for each age-class. As these data are simulated, the Class represents an age-class, and the Category represents length.

Table 1. Mean, standard deviation and initial seed values used to generate data for simulation 1 and simulation 2.

Simulation	Class	Mean	Standard Deviation	Seed
Sim. 1	Class A	35	5	550
	Class B	55	5	650
Sim. 2	Class A	35	10	550
	Class B	35	5	650

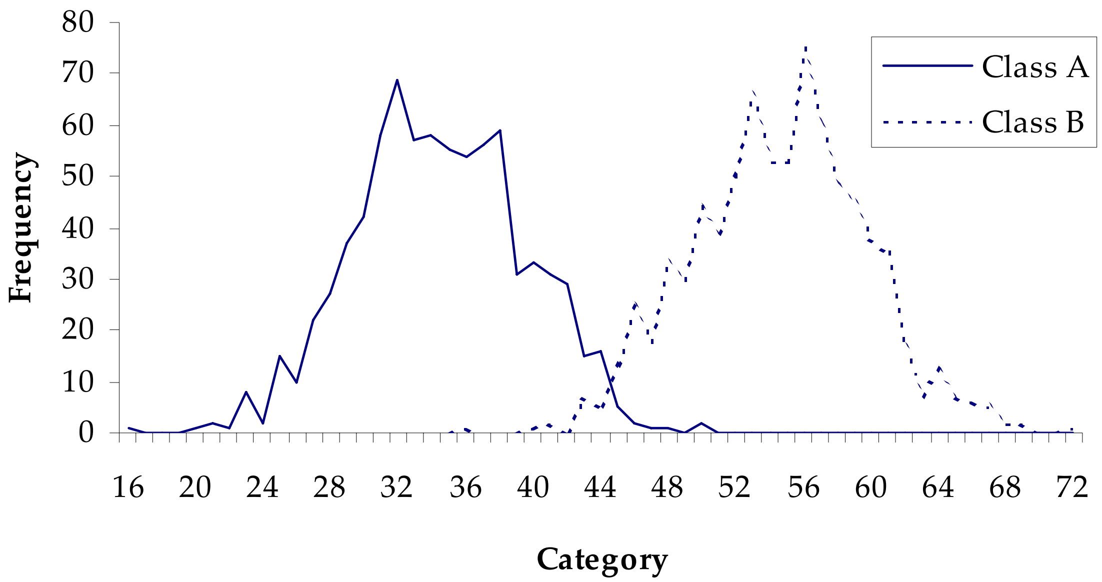


Figure 1. Original distributions of class A and class B generated using a random number generator used for simulation 1. N=400 for each class.

For the first and second set of simulations, five separate sets of relative abundance of each age class were explored. The proportions of each class used are shown in Table 2. The number of samples in each test were chosen to be equal, biased to one class, biased to the other class, all one class and all the other class. The percentages used for each class were decided within this framework. The same proportions of each class were used for both the first and second simulations.

The numerical analyses program that processed the mixed distributions and determined the underlying proportions for each class using the FFK method was written in MS FORTRAN. The data inputs for this process consisted of two matrices. The first matrix contained the *a priori* information on the distribution of lengths for each of the age classes. This matrix was dimensioned to the number of age classes by the number of length bins in each age class. The second matrix supplied to the FORTRAN program was the mixed length distribution that was comprised of an unknown proportion of each of the age classes. The second matrix was a one-dimensional vector with the known distribution (observed distribution) appended to the bottom of the distribution. This was used to calculate the error between the observed and the predicted proportions of each class within the mixed distribution that contained the unknown proportions of each class.

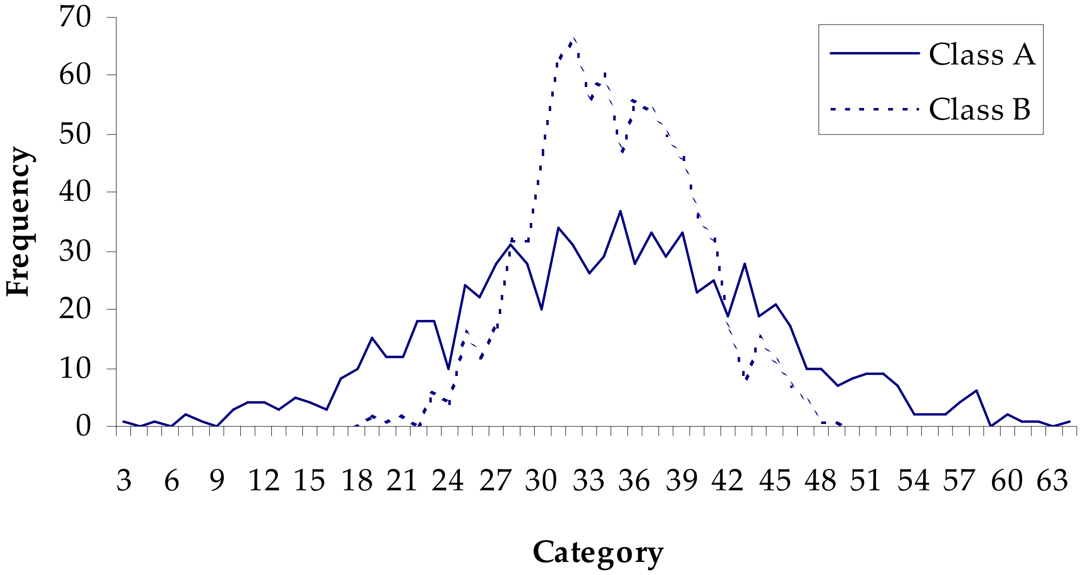


Figure 2. Original distributions of age class A and class B generated using a random number generator used for simulation 2. N=400 for each age class.

Table 2. Proportion of age classes A and B used in the first and second simulations.

	Class	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Percentage	Class A	50	27	64	100	0
	Class B	50	73	36	0	100
N	Class A	400	216	512	800	0
	Class B	400	584	288	0	800

To prepare the matrix that provided the information that was used in the decomposition, the frequency distribution of each known class was calculated and then summed to one, an empirical probability distribution. Each frequency distribution associated with each given class was then analogous to an empirical probability distribution of length for a given age-class. These data provided the information used to decompose the unknown distributions.

The matrix for a two-class problem consisted of a two-dimensional array, each of the dimensions consisting of the empirical probability distribution of length. The first two elements, $a[0,0]$ and $a[0,1]$ gave the x and y dimensions of the array used internally by the FORTRAN program. These values represent a[number of age-classes, number of length-classes]. In the first simulation where two different means with the same standard deviation were used, these values were $a[2,57]$. Where the same mean and different standard deviations were used, the size of the first matrix was *a*[2,63]. The distribution for each class was the frequency distribution from the generated random numbers expressed as an empirical probability.

The proportions of each class used in the first two series of simulation tests were the same for both datasets. The values ranged from 0% of class A and 100% of class B to the inverse of this (100% class A and 0% class B). Three other proportions of each class were used, giving a total of five simulation tests. The proportions chosen for each class tested in these simulations was arbitrary, chosen primarily to stress the method. These proportions are shown in Table 2. Classes were designated as class A and class B in this component of the study.

To select samples which comprised the unknown combined samples, each class was randomly re–sampled with replacement *N* times (Table 2) from the class A and class B population distributions to provide a one-dimensional unknown sample. The unknown sample was then a mixture of class A and class B at the predetermined quanta shown in Table 2. The combined distributions for each test were converted to an empirical probability distribution and the known proportions of each class appended to the array. These two matrices were then submitted to the FORTRAN program.

Data for the third simulation were generated using the same method in simulation 1 and simulation 2. Three groups of two means were selected. Each of the three groups had data generated using two standard deviations. Standard deviations were 5 units for groups A, C and E and 10 units of standard deviation for groups B, D and F. The seed value used for the generation of the datasets was also standardised so these data could be regenerated in further studies. These parameters for generating the data for each of the six normal distributions are given in Table 3.

Table 3. Mean, standard deviation and initial seed values used to generate data for simulation 3.

	Mean	Standard deviation	Seed
Class A	35	5	150
Class B	35	10	250
Class C	55	5	350
Class D	55	10	450
Class E	75	5	550
Class F	75	10	650

The third simulation consisted of a series of twelve trials using combined mixtures from these six classes. Each of the trials had altered relative abundance of each class (Table 4.). The length of the initial array used as the input to the FORTRAN program was *a*[6,100]. The larger array size was necessary to encompass the tails of the distributions at the extremes of the sample space. The initial distributions for each class are shown in Figure 3.

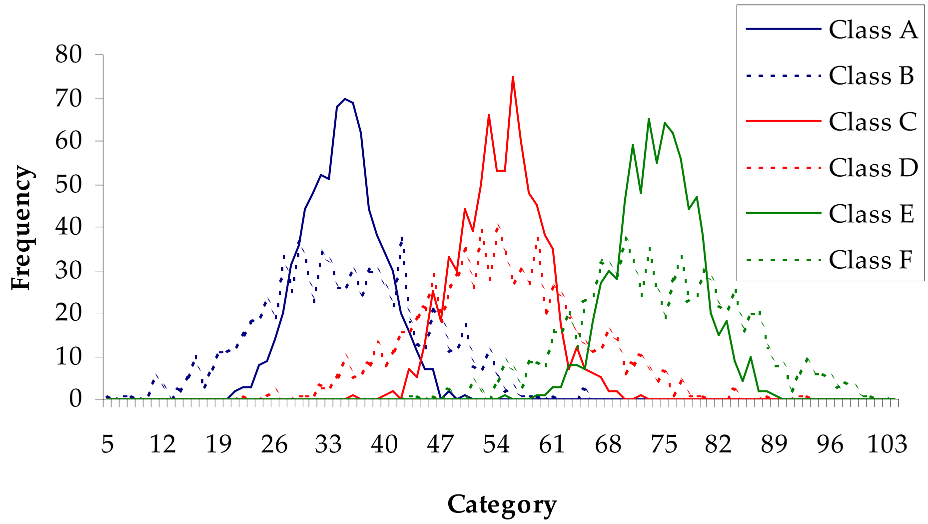


Figure 3. Original distributions of class A through F generated using a random number generator used for simulation 3. N=800 for each class.

Data for the mixed distributions were selected by randomly re-sampling with replacement *N* times (Table 4) from the dataset from the original distributions from each class (A-F) to provide a one-dimensional array which represented an unknown mixed distribution. The numbers chosen to comprise each class that are shown in Table 4 were arbitrarily chosen. The unknown sample distribution was then a mixture of classes A through F at predetermined quanta shown in Table 4. The combined distributions for each test were converted to an empirical probability distribution and the known proportions of each class appended to the array. These two matrices (the original distributions of each class) and the unknown one-dimensional array consisting of the re-sampled data from each class) were then submitted to the FORTRAN program.

Table 4. Data selected for the third simulation, which consisted of a total of 12 trials. Classes were designated as class A through to class F in this part of the study as, at this point, they do not represent real age data. Section (A) shows the percentage of each class that was randomly selected from the population distribution. Section (B) shows the N that comprised the distribution used in testing. Section (C) details the relative percentage of each class that comprised the 'population' under test.

		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10	Trial 11	Trial 12
A)	Class A	77	82	0	55	91	58	94	0	32	0	52	0
	Class B	79	0	98	41	86	16	39	100	0	0	75	0
	Class C	77	80	57	82	0	50	99	0	98	0	94	86
	Class D	100	17	14	0	31	40	67	68	0	87	0	0
	Class E	18	49	95	88	60	83	0	0	59	10	0	0
	Class F	35	49	92	67	51	0	14	61	0	37	0	0
B)	Class A	616	656	0	440	728	464	752	0	256	0	416	0
	Class B	632	0	784	328	688	128	312	800	0	0	600	0
	Class C	616	640	456	656	0	400	792	0	784	0	752	688
	Class D	800	136	112	0	248	320	536	544	0	696	0	0
	Class E	144	392	760	704	480	664	0	0	472	80	0	0
	Class F	280	392	736	536	408	0	112	488	0	296	0	0
	Sum	**3088**	**2216**	**2848**	**2664**	**2552**	**1976**	**2504**	**1832**	**1512**	**1072**	**1768**	**688**
C)	Class A	19.9	29.6	0	16.5	28.5	23.5	30	0	16.9	0	23.5	0
	Class B	20.5	0	27.5	12.3	27	6.5	12.5	43.7	0	0	33.9	0
	Class C	19.9	28.9	16	24.6	0	20.2	31.6	0	51.9	0	42.5	100
	Class D	25.9	6.1	3.9	0	9.7	16.2	21.4	29.7	0	64.9	0	0
	Class E	4.7	17.7	26.7	26.4	18.8	33.6	0	0	31.2	7.5	0	0
	Class F	9.1	17.7	25.8	20.1	16	0	4.5	26.6	0	27.6	0	0
	Sum	100	100	100	100	100	100	100	100	100	100	100	100

Results

Simulation distributions

The mixed distributions for simulation 1 (different means, same standard deviation) are shown in Figure 5A through Figure 9A. Mixed distributions from simulation 2 (same mean and different standard deviations are shown in Figure 10A through Figure 14A. Mixed distributions from simulation 3 (3 means by 2 standard deviations) are shown in Figure 15A through 26A.

The underlying frequencies of each class that comprise the mixed distribution are shown in Figure 5B through 26B for each of the tests in the simulations.

The original proportions of each class within the mixed distribution are shown in Figures 5C through Figure 26C. For direct comparison, the results from the mixture decomposition for each test are shown adjacent to the known frequencies.

Efficacy of FFK method

Simulation 1 – comparable to two age-classes which have not reached the asympote of a growth curve

The results for the reconstruction for simulation 1 show that the method decomposed the relative abundance of all classes for each of the five trials (Table 5, Figure 5C to Figure 9C). The maximum error for the difference between the observed and predicted composition of the two classes was 1% for trial 2.

The method also successfully decomposed the underlying relative abundance for trial 4 with 0% error (correct decomposition of the unknown distribution). Errors were approximately 0.5% or less for four out of the five trials from simulation 1 (Table 5).

Simulation2 – comparable to two age-classes which have reached the asympote of a growth curve
The decomposition of distributions into the underlying classes from simulation 2 was accurate (Table 6, Figure 5C through Figure 9C). The maximum error for this trial was 10% (trial 4) where the unknown distribution contained 100% of class A. The next highest error was 3.03% for a mixed distribution (64% class A and 36% Class B). All other assignment of relative abundance of classes from the mixed distributions was less than 1.5% error (difference between the known and calculated abundances expressed as a percentage of the combined distribution).

Simulation 3 – comparable to multiple age-classes collected from a real fishery
Decomposition of the combined distribution into the underlying distributions of each class (where six classes were included) again produced low errors. These are detailed in Table 7. The maximum error for these 12 trials was 6.3%. These errors were generated from classes within the same trial (trial 4). The distribution of errors from all trials (Figure 5) shows a modal error for all trials of close to 0%, which accounted for nearly 40% of all trials.

Table 5. Decomposition of the distributions for simulation 1 where the means of the classes was separated (35 and 55) and the standard deviation was the same (5).

Trial	Class	Observed	Predicted	Error
Trial 1	Class A	50%	50.56%	-0.56%
	Class B	50%	49.44%	0.56%
Trial 2	Class A	27%	28.00%	-1.00%
	Class B	73%	72.00%	1.00%
Trial 3	Class A	64%	64.12%	-0.11%
	Class B	36%	35.89%	0.11%
Trial 4	Class A	100%	100.00%	0.00%
	Class B	0%	0.00%	0.00%
Trial 5	Class A	0%	0.37%	-0.37%
	Class B	100%	99.63%	0.37%

Table 6. Decomposition of the distributions for simulation 2 where the means of the classes were the same (35) and the standard deviations were different (5, 10).

Trial	Class	Observed	Predicted	Error
Trial 1	Class A	50%	49.27%	0.73%
	Class B	50%	50.73%	-0.73%
Trial 2	Class A	27%	25.54%	1.46%
	Class B	73%	74.46%	-1.46%
Trial 3	Class A	64%	60.97%	3.03%
	Class B	36%	39.03%	-3.03%
Trial 4	Class A	100%	90.35%	9.65%
	Class B	0%	9.65%	-9.65%
Trial 5	Class A	0%	0.00%	0.00%
	Class B	100%	100.00%	0.00%

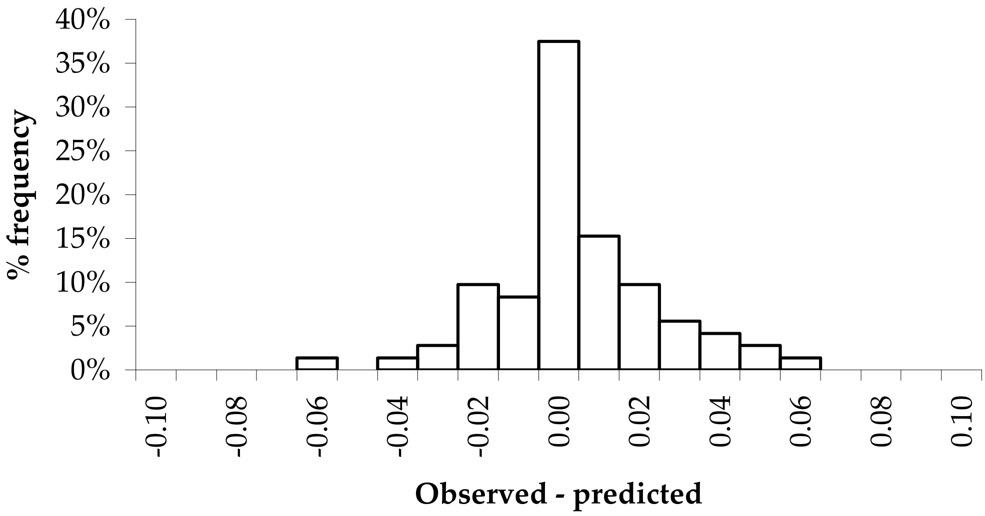


Figure 4. Frequency of observed minus predicted error distribution for simulation 3 (all trials), N=72.

Table 7. Decomposition of the distributions for simulation 3. The known abundance of each class, the estimated relative abundance for each class based on the *a priori* distributions of 'length' and the difference between the known and estimated abundances of each class expressed as a percentage.

Trial	Class	Observed	Predicted	Error	Trial	Class	Observed	Predicted	Error
Trial 1	Class A	19.95%	20.50%	-0.55%	Trial 7	Class A	30.03%	28.21%	1.82%
	Class B	20.47%	18.96%	1.50%		Class B	12.46%	14.70%	-2.24%
	Class C	19.95%	20.64%	-0.69%		Class C	31.63%	30.74%	0.89%
	Class D	25.91%	26.56%	-0.65%		Class D	21.41%	21.44%	-0.04%
	Class E	4.66%	4.31%	0.36%		Class E	0.00%	0.52%	-0.52%
	Class F	9.07%	9.03%	0.04%		Class F	4.47%	4.38%	0.09%
Trial 2	Class A	29.60%	27.14%	2.47%	Trial 8	Class A	0.00%	0.33%	-0.33%
	Class B	0.00%	2.50%	-2.50%		Class B	43.67%	43.72%	-0.05%
	Class C	28.88%	26.91%	1.97%		Class C	0.00%	3.29%	-3.29%
	Class D	6.14%	7.90%	-1.76%		Class D	29.69%	25.91%	3.78%
	Class E	17.69%	18.03%	-0.34%		Class E	0.00%	0.26%	-0.26%
	Class F	17.69%	17.52%	0.17%		Class F	26.64%	26.50%	0.14%
Trial 3	Class A	0.00%	0.77%	-0.77%	Trial 9	Class A	16.93%	16.39%	0.54%
	Class B	27.53%	22.32%	5.21%		Class B	0.00%	0.21%	-0.21%
	Class C	16.01%	22.30%	-6.29%		Class C	51.85%	52.61%	-0.76%
	Class D	3.93%	5.30%	-1.36%		Class D	0.00%	0.22%	-0.22%
	Class E	26.69%	24.40%	2.29%		Class E	31.22%	30.35%	0.86%
	Class F	25.84%	24.91%	0.94%		Class F	0.00%	0.22%	-0.22%
Trial 4	Class A	16.52%	14.25%	2.27%	Trial 10	Class A	0.00%	0.00%	0.00%
	Class B	12.31%	14.70%	-2.39%		Class B	0.00%	0.24%	-0.24%
	Class C	24.62%	23.63%	1.00%		Class C	0.00%	2.42%	-2.42%
	Class D	0.00%	1.99%	-1.99%		Class D	64.93%	62.60%	2.33%
	Class E	26.43%	29.64%	-3.21%		Class E	7.46%	10.29%	-2.83%
	Class F	20.12%	15.79%	4.33%		Class F	27.61%	24.46%	3.16%
Trial 5	Class A	28.53%	29.43%	-0.91%	Trial 11	Class A	23.53%	24.04%	-0.51%
	Class B	26.96%	25.15%	1.81%		Class B	33.94%	32.10%	1.84%
	Class C	0.00%	0.31%	-0.31%		Class C	42.53%	43.54%	-1.01%
	Class D	9.72%	12.08%	-2.36%		Class D	0.00%	0.11%	-0.11%
	Class E	18.81%	20.91%	-2.10%		Class E	0.00%	0.11%	-0.11%
	Class F	15.99%	12.12%	3.87%		Class F	0.00%	0.11%	-0.11%
Trial 6	Class A	23.48%	21.98%	1.51%	Trial 12	Class A	0.00%	0.00%	0.00%
	Class B	6.48%	7.62%	-1.15%		Class B	0.00%	0.01%	-0.01%
	Class C	20.24%	15.50%	4.74%		Class C	100.00%	98.92%	1.08%
	Class D	16.19%	20.98%	-4.79%		Class D	0.00%	0.01%	-0.01%
	Class E	33.60%	33.44%	0.16%		Class E	0.00%	0.01%	-0.01%
	Class F	0.00%	0.47%	-0.47%		Class F	0.00%	1.05%	-1.05%

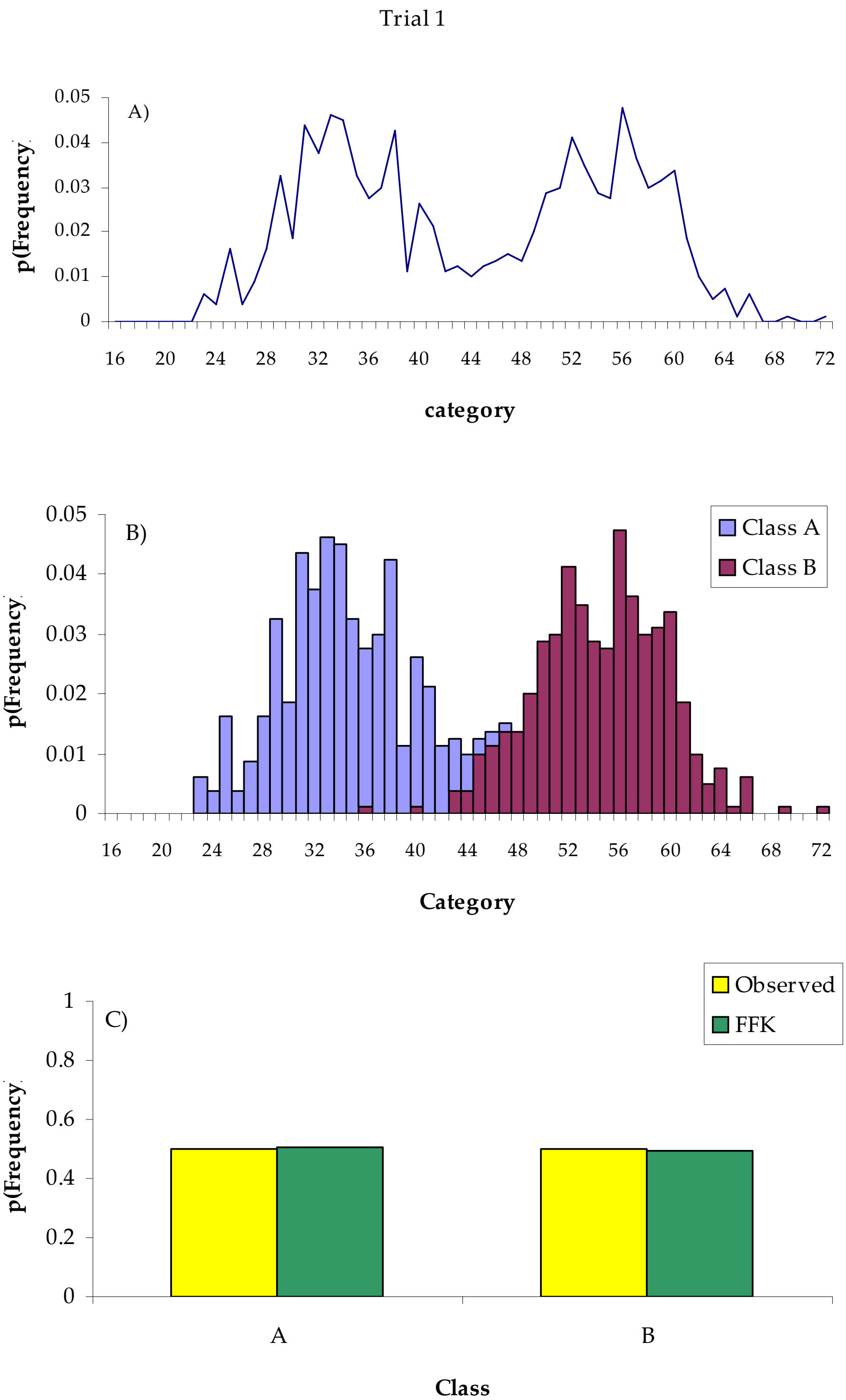


Figure 5. Decomposition of simulation 1 trial 1. Means were 35 and 55 with a standard deviation of 5. A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

Trial 2

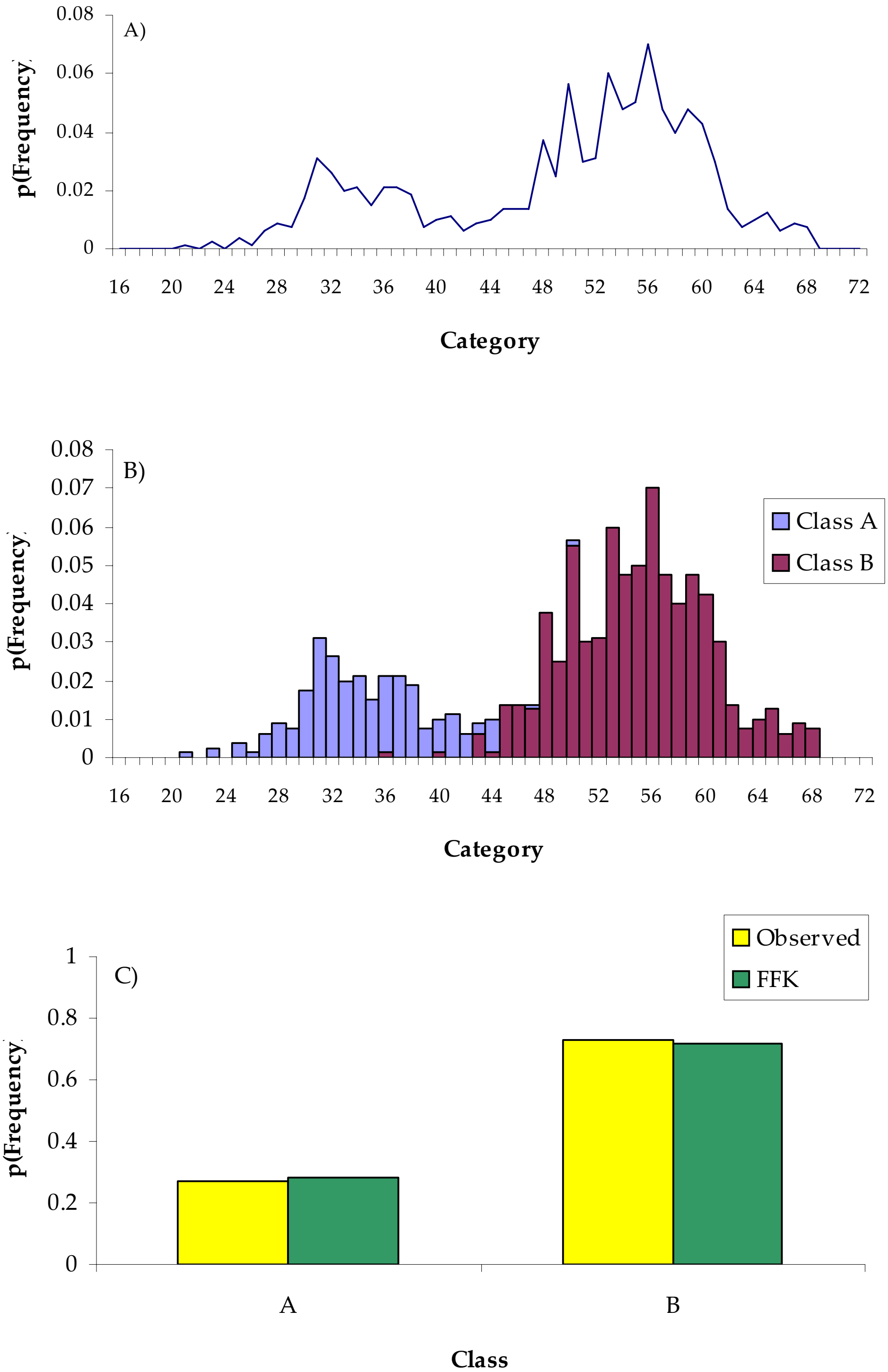


Figure 6. Decomposition of simulation 1 trial 2. Means were 35 and 55 with a standard deviation of 5. A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

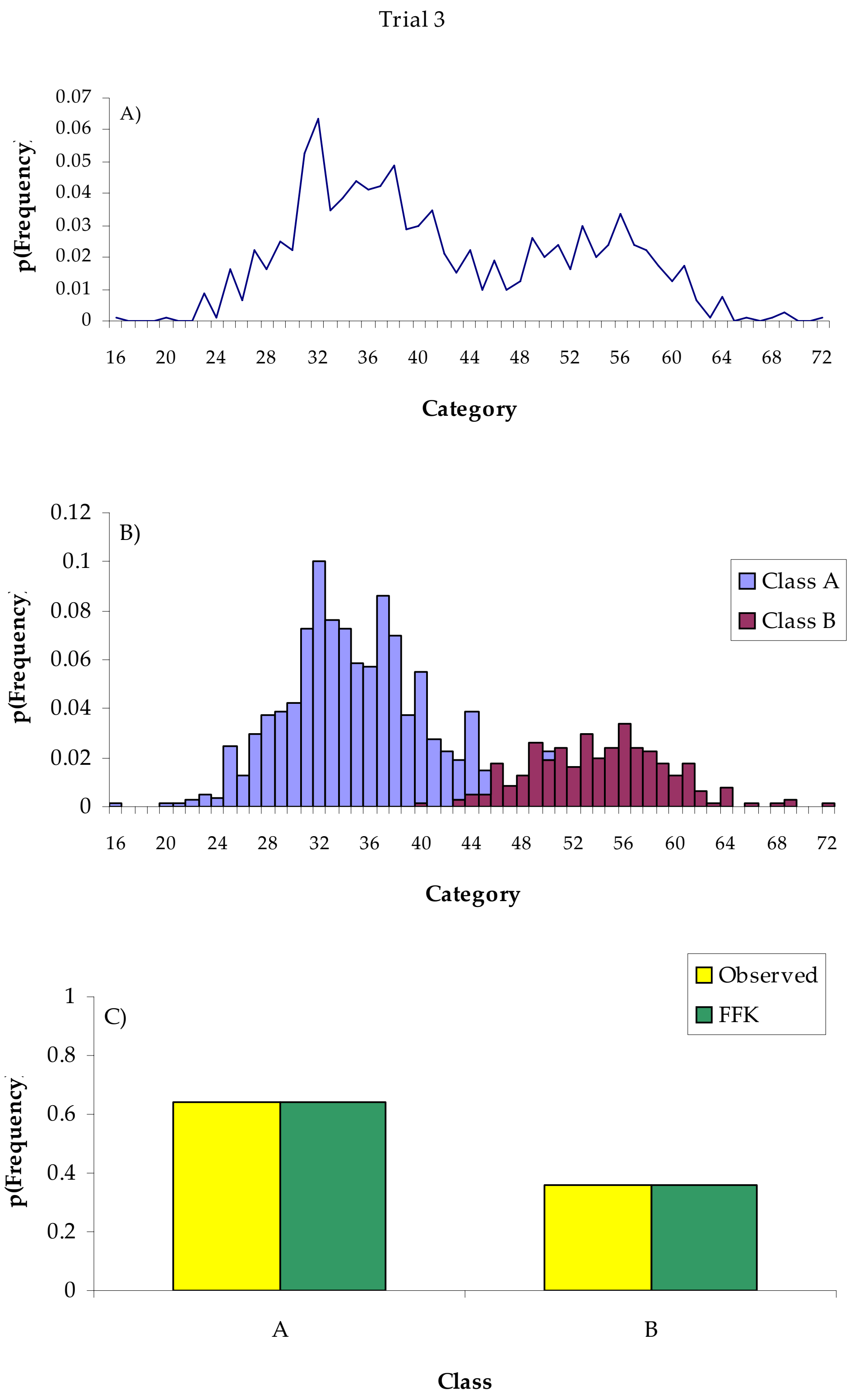


Figure 7.Decomposition of simulation 1 trial 3. Means were 35 and 55 with a standard deviation of 5. A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

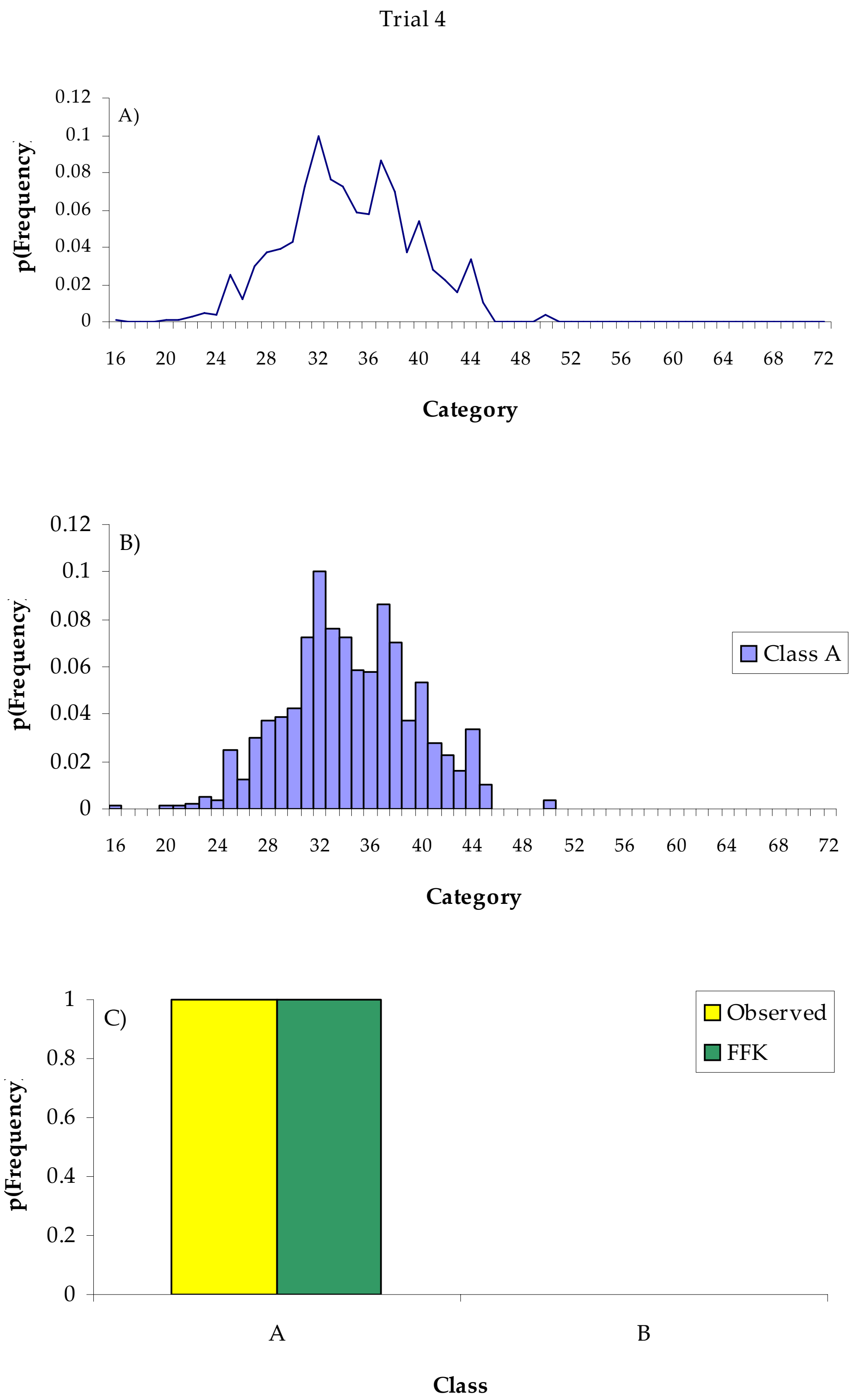


Figure 8. Decomposition of simulation 1 trial 4. Means were 35 and 55 with a standard deviation of 5. A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

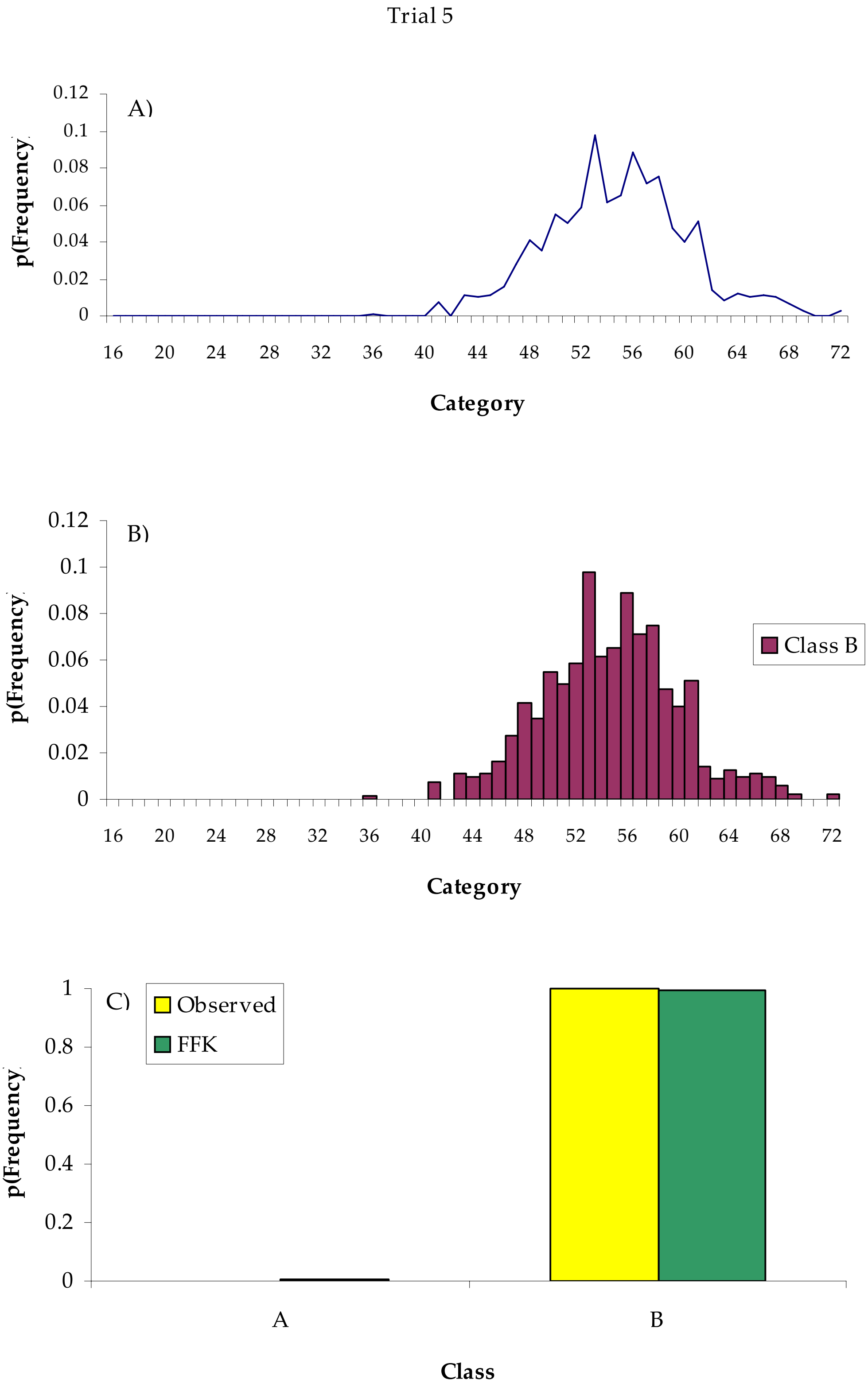


Figure 9. Decomposition of simulation 1 trial 5. Means were 35 and 55 with a standard deviation of 5. A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

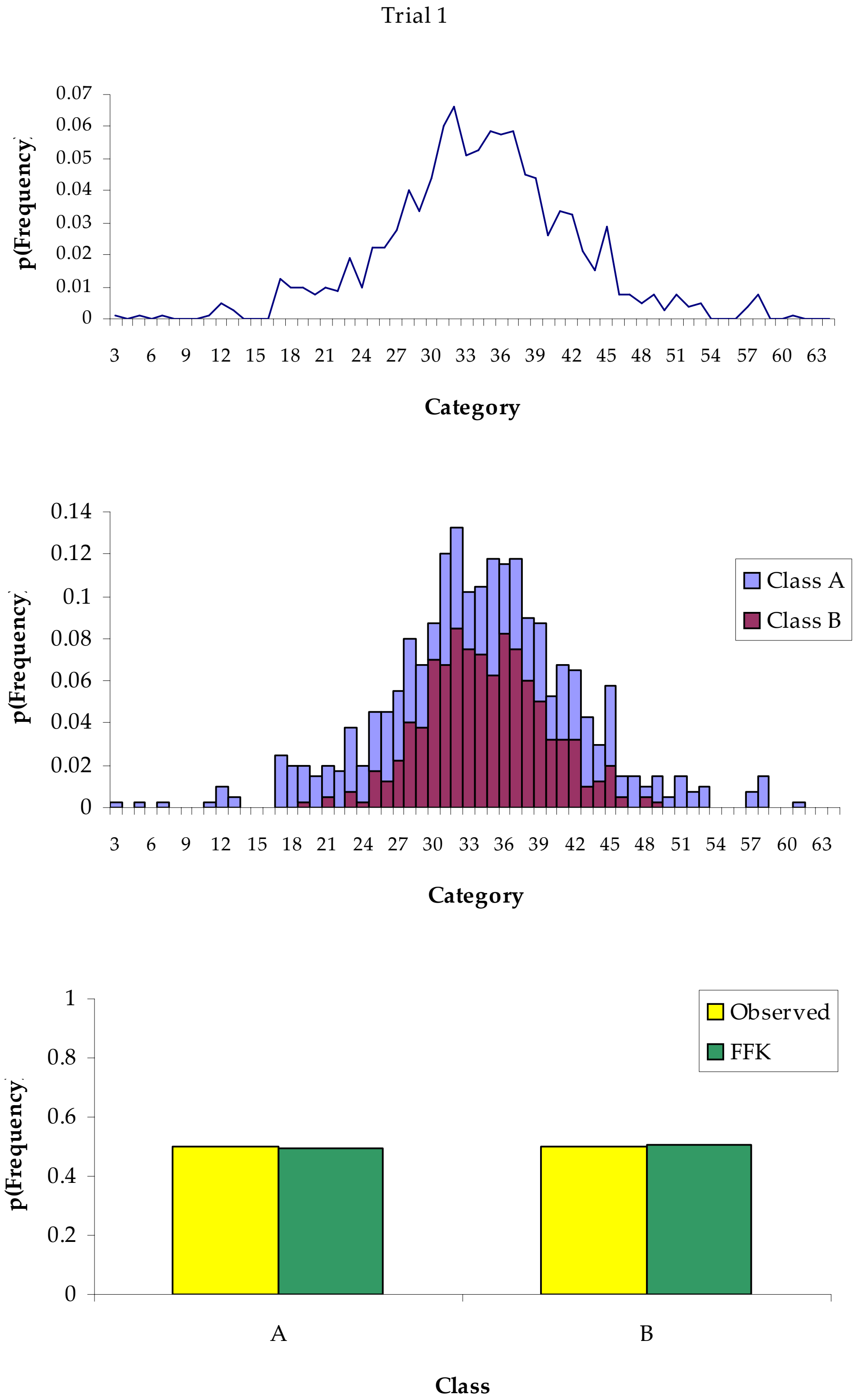


Figure 10. Decomposition of simulation 2 trial 1. Means were both 35 with a standard deviation of 5 and 10. A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

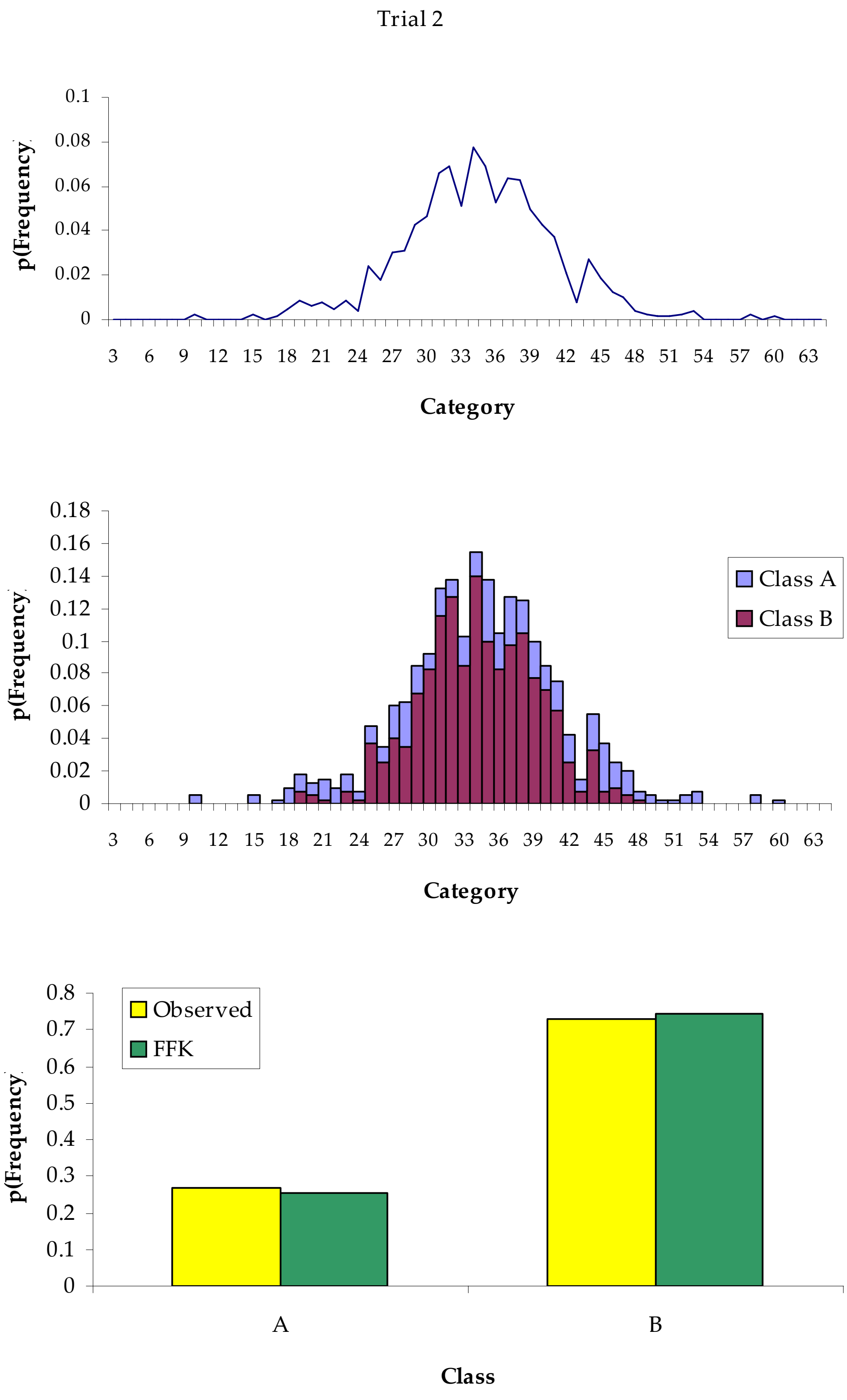


Figure 11. Decomposition of simulation 2 trial 2. Means were both 35 with a standard deviation of 5 and 10. A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

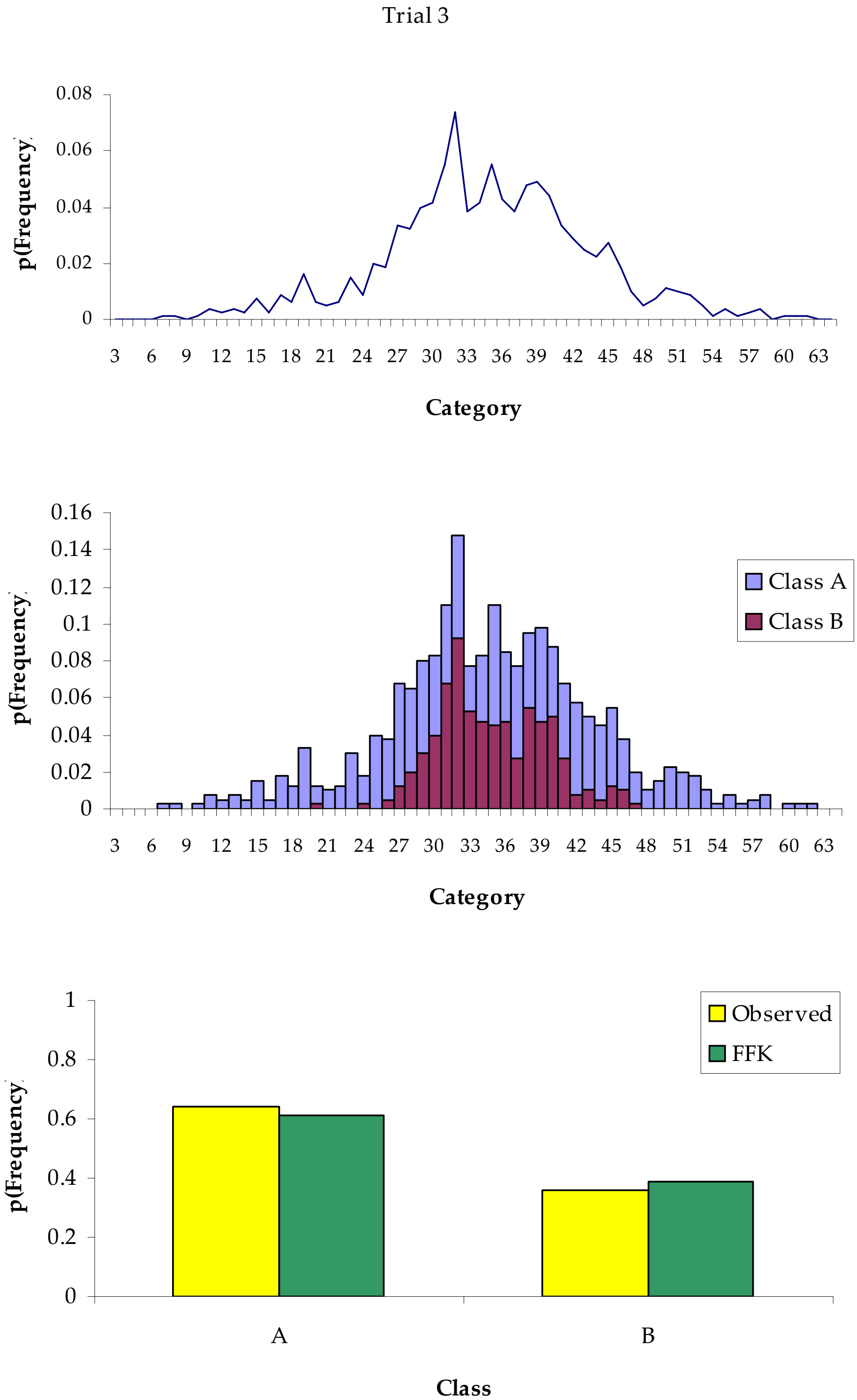


Figure 12. Decomposition of simulation 2 trial 3. Means were both 35 with a standard deviation of 5 and 10. A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

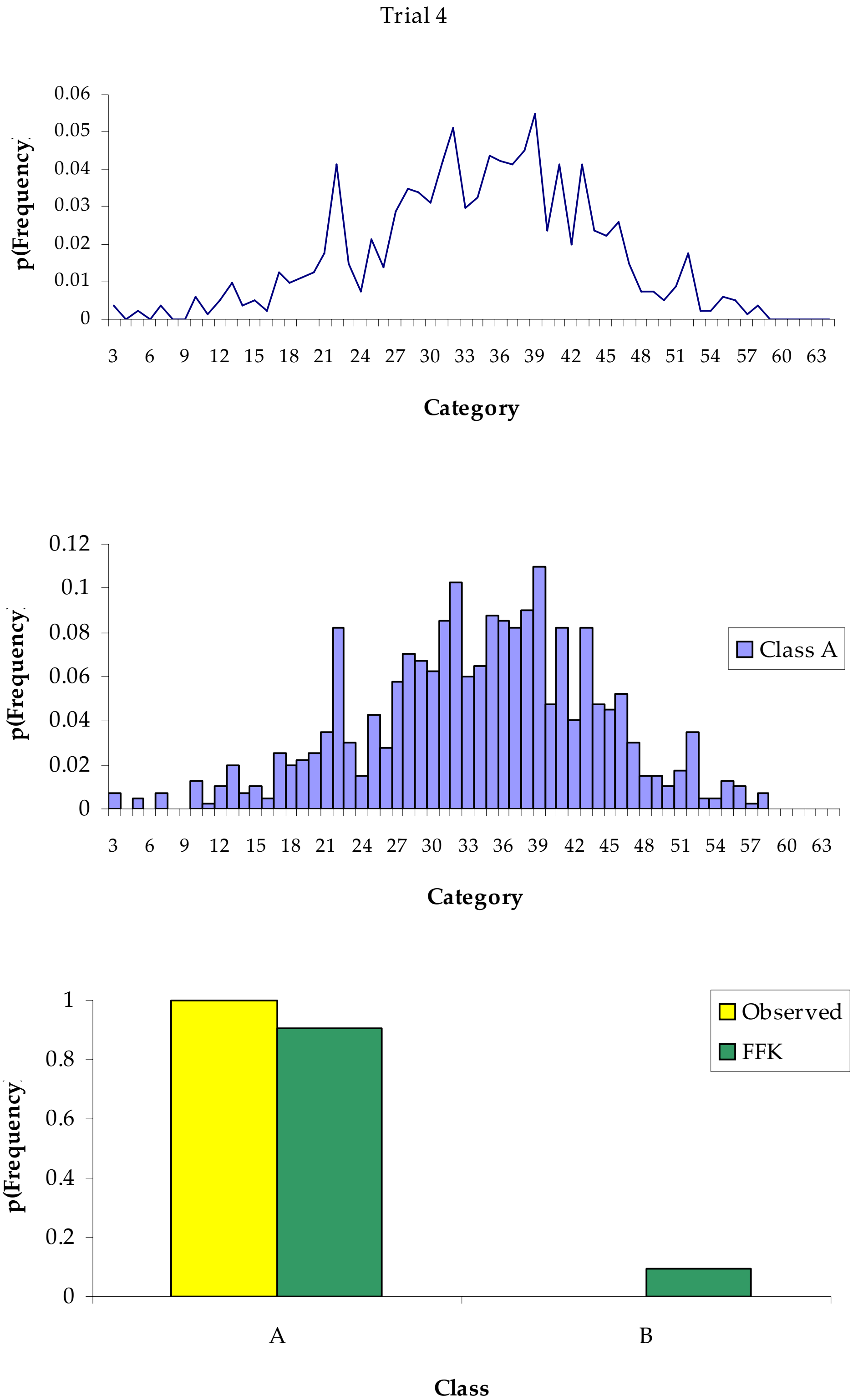
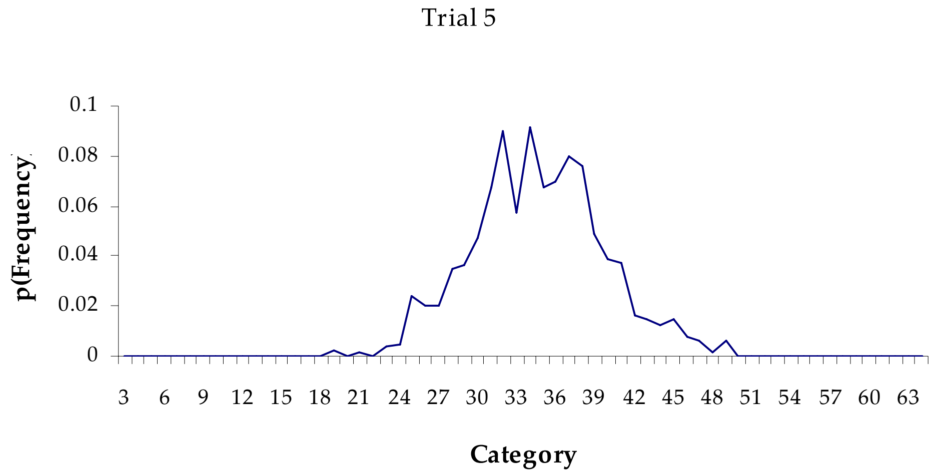


Figure 13. Decomposition of simulation 2 trial 4. Means were both 35 with a standard deviation of 5 and 10. A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.



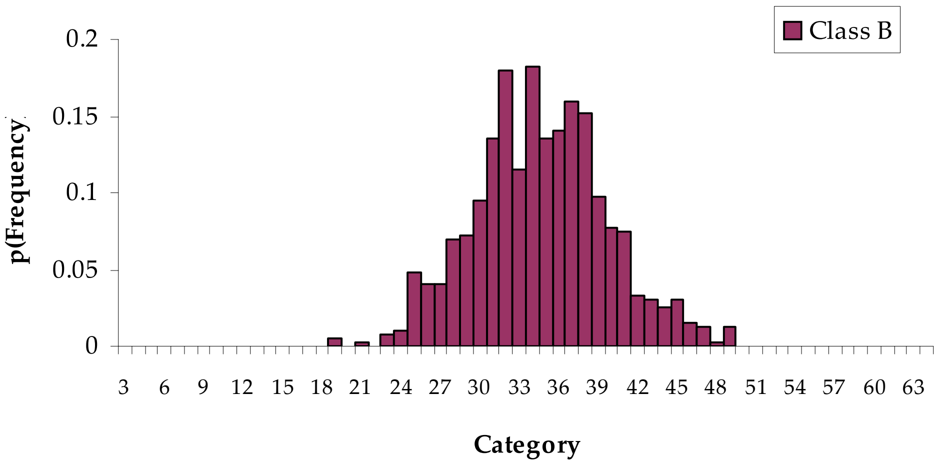


Figure 14. Decomposition of simulation 2 trial 5. Means were both 35 with a standard deviation of 5 and 10. A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

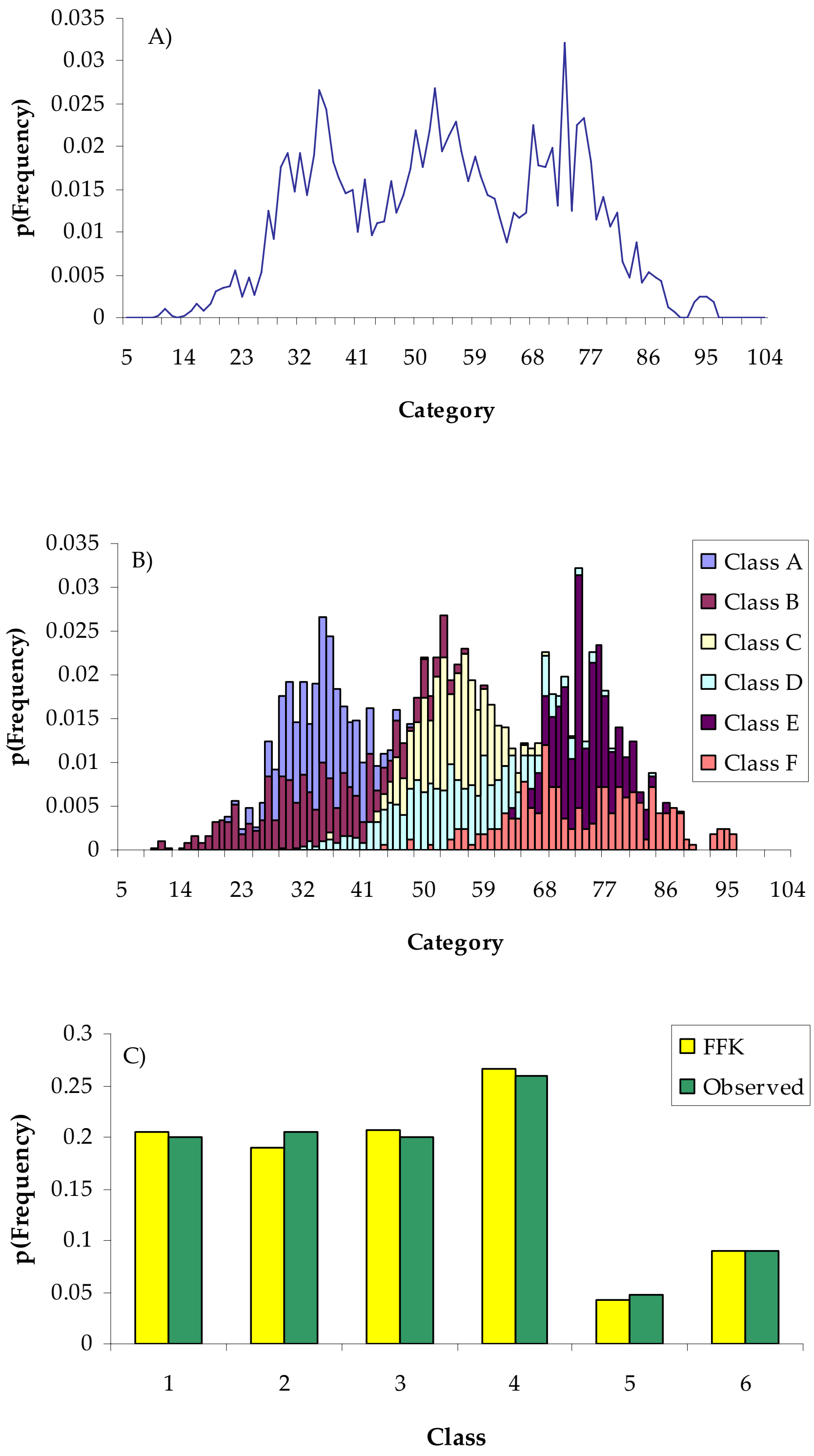


Figure 15. Decomposition of simulation 3 trial 1. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

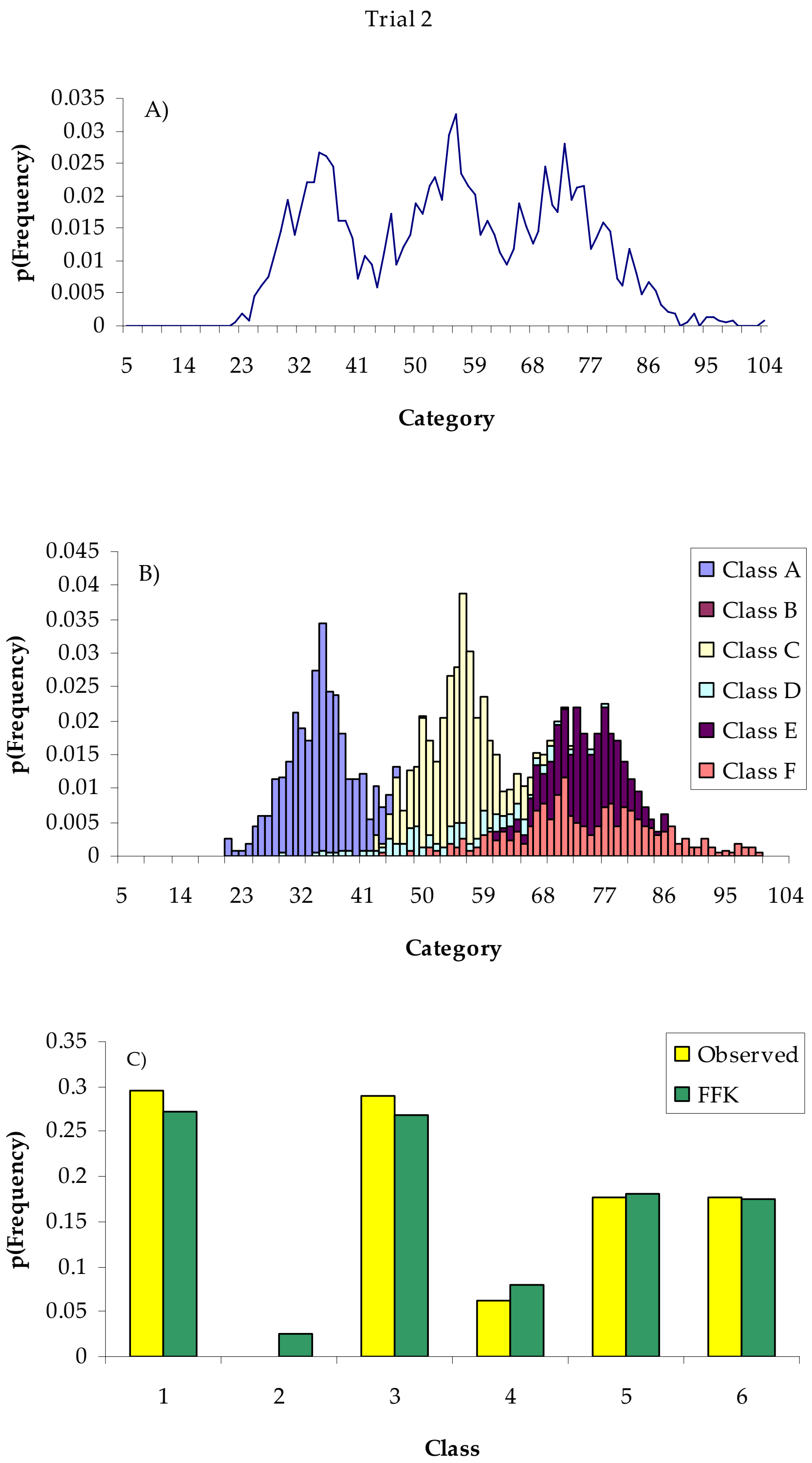


Figure 16. Decomposition of simulation 3 trial 2. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

Trial 3

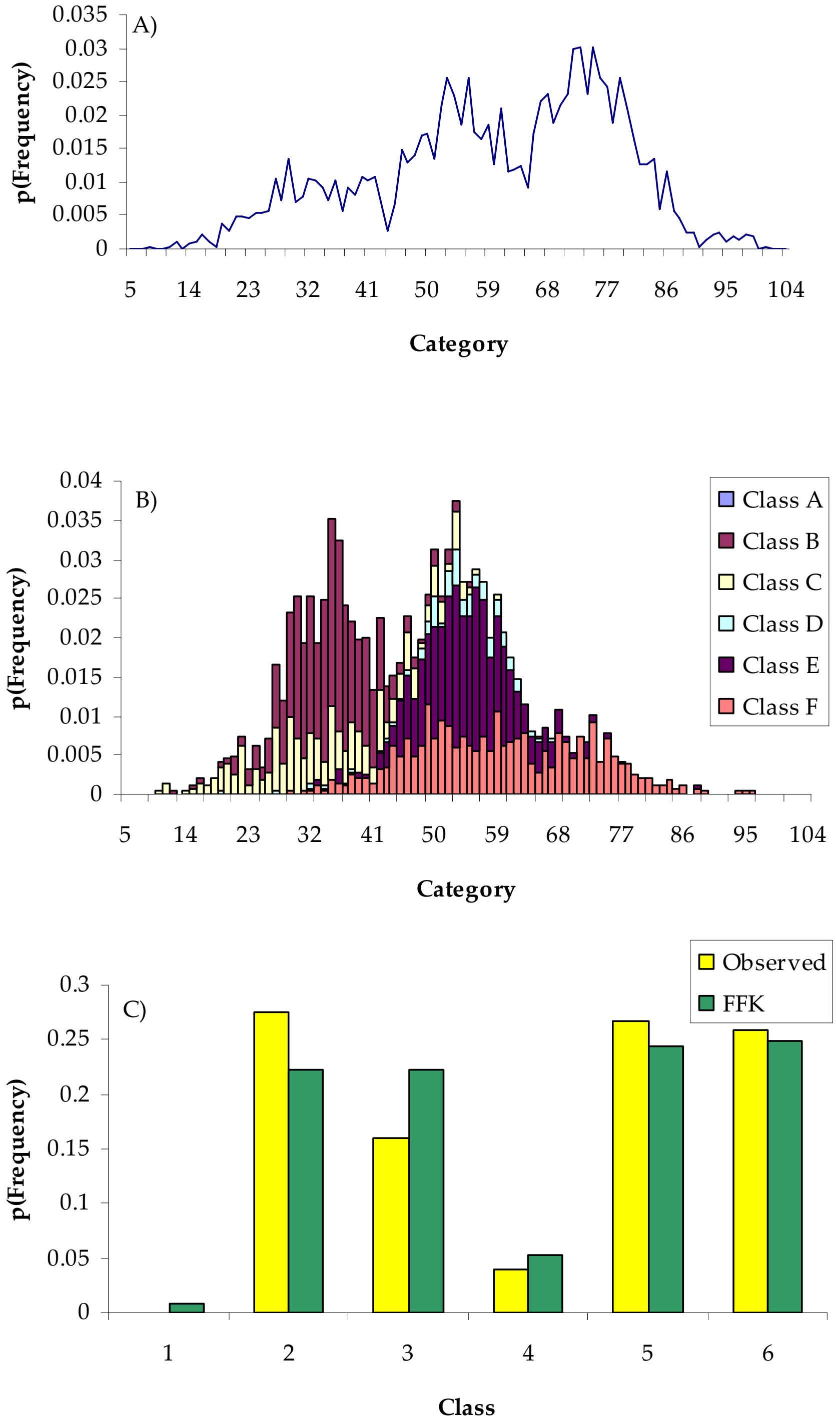


Figure 17. Decomposition of simulation 3 trial 3. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

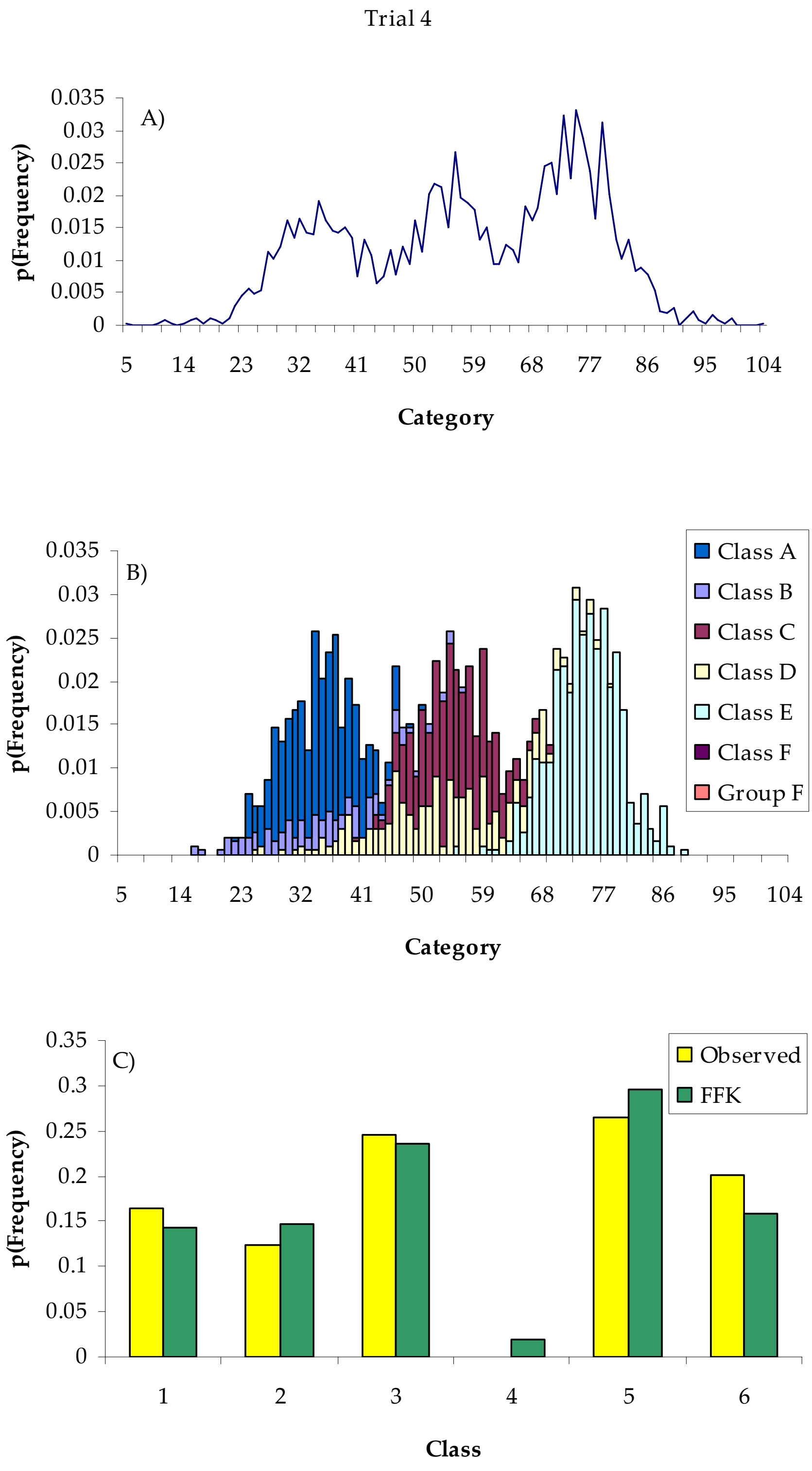


Figure 18. Decomposition of simulation 3 trial 4. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

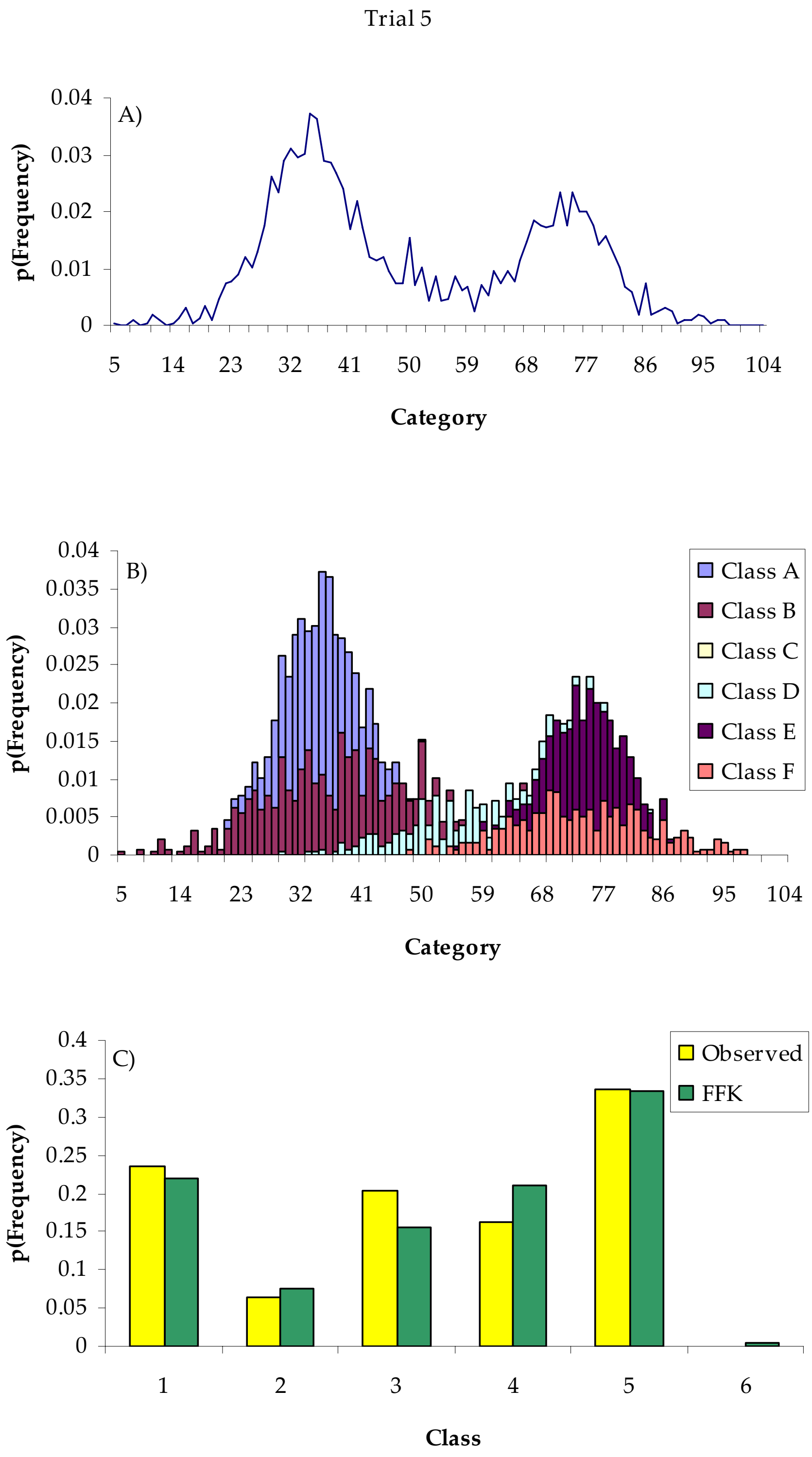


Figure 19. Decomposition of simulation 3 trial 5. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

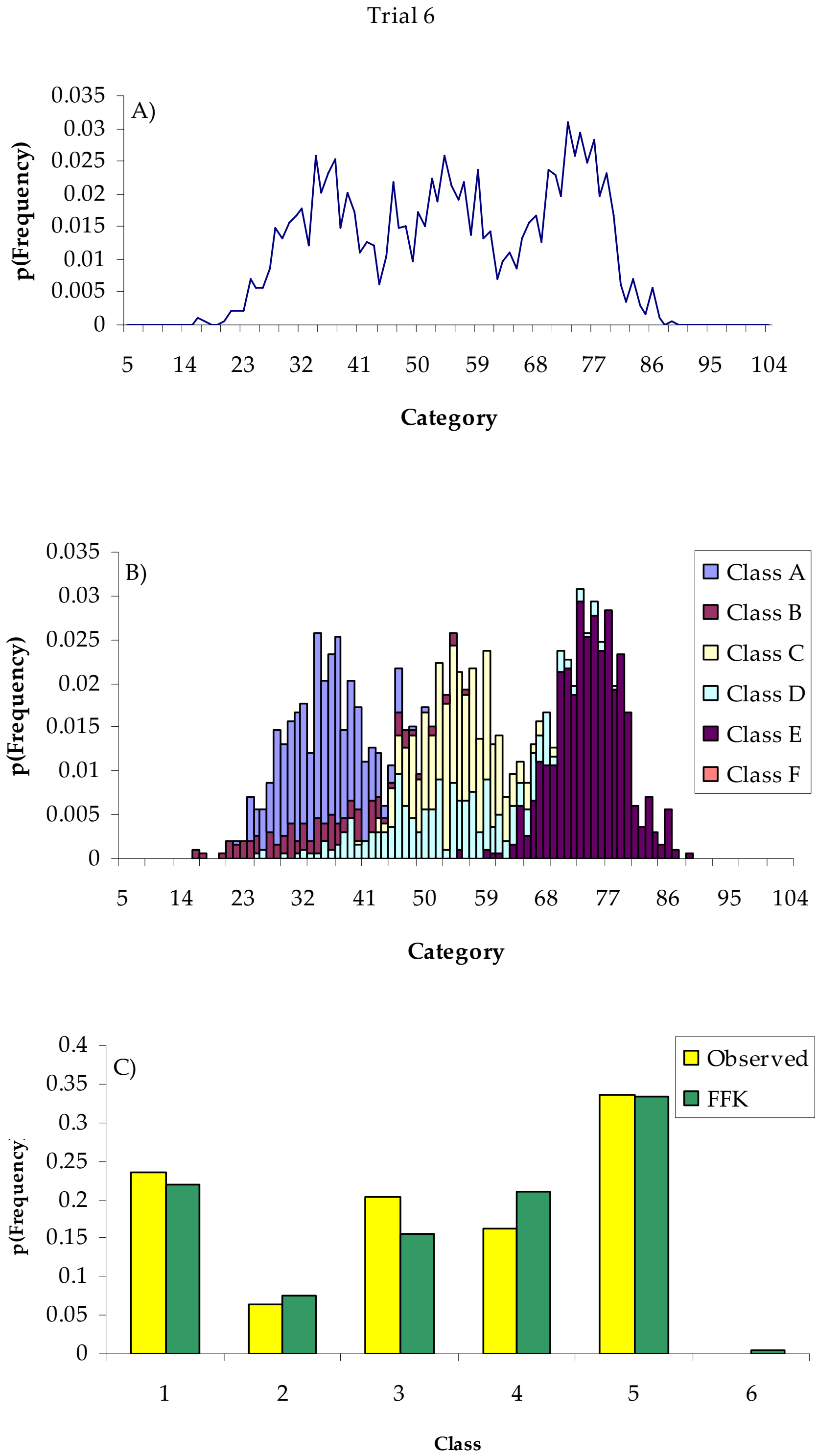


Figure 20. Decomposition of simulation 3 trial 6. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

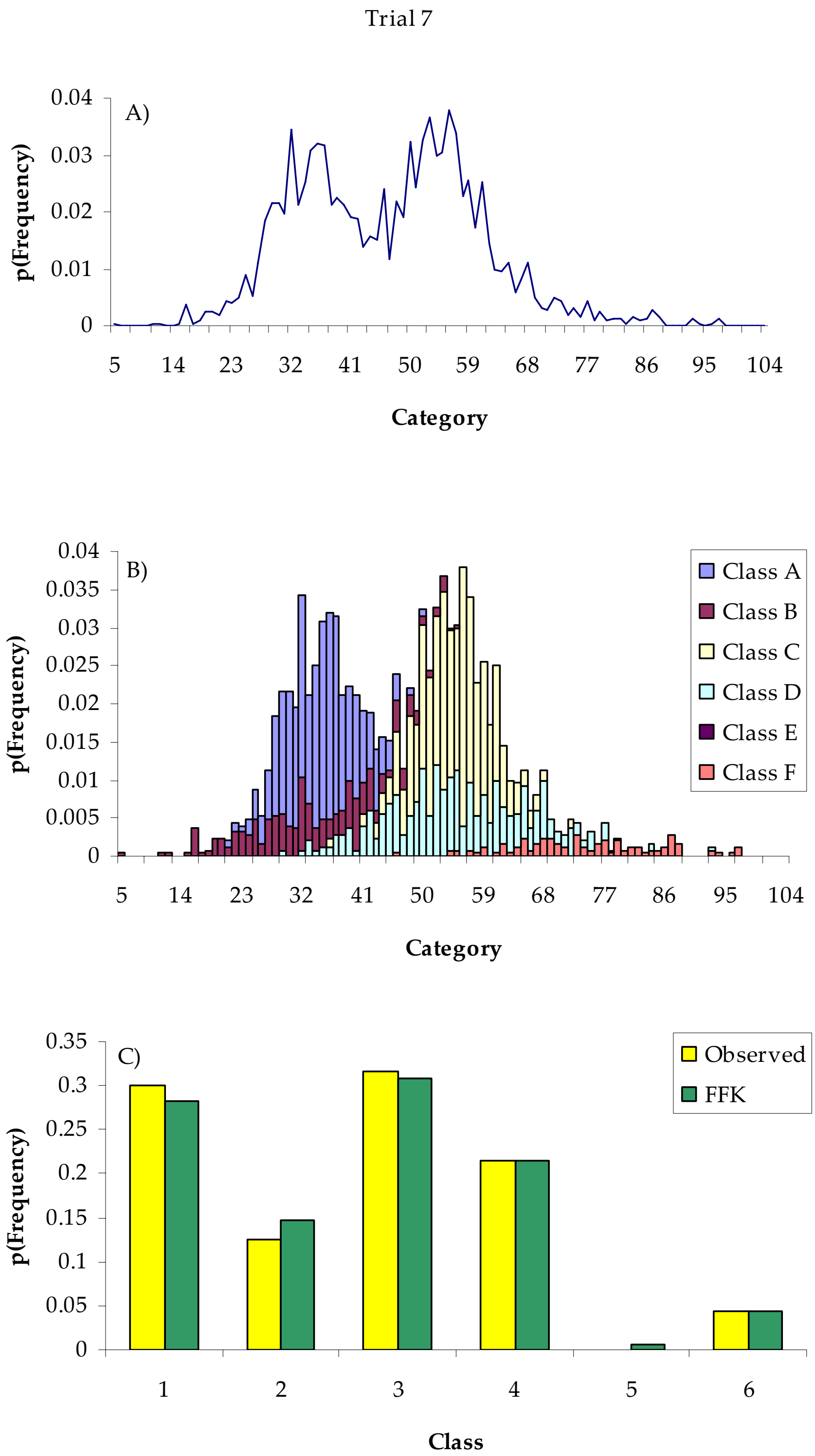


Figure 21. Decomposition of simulation 3 trial 7. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

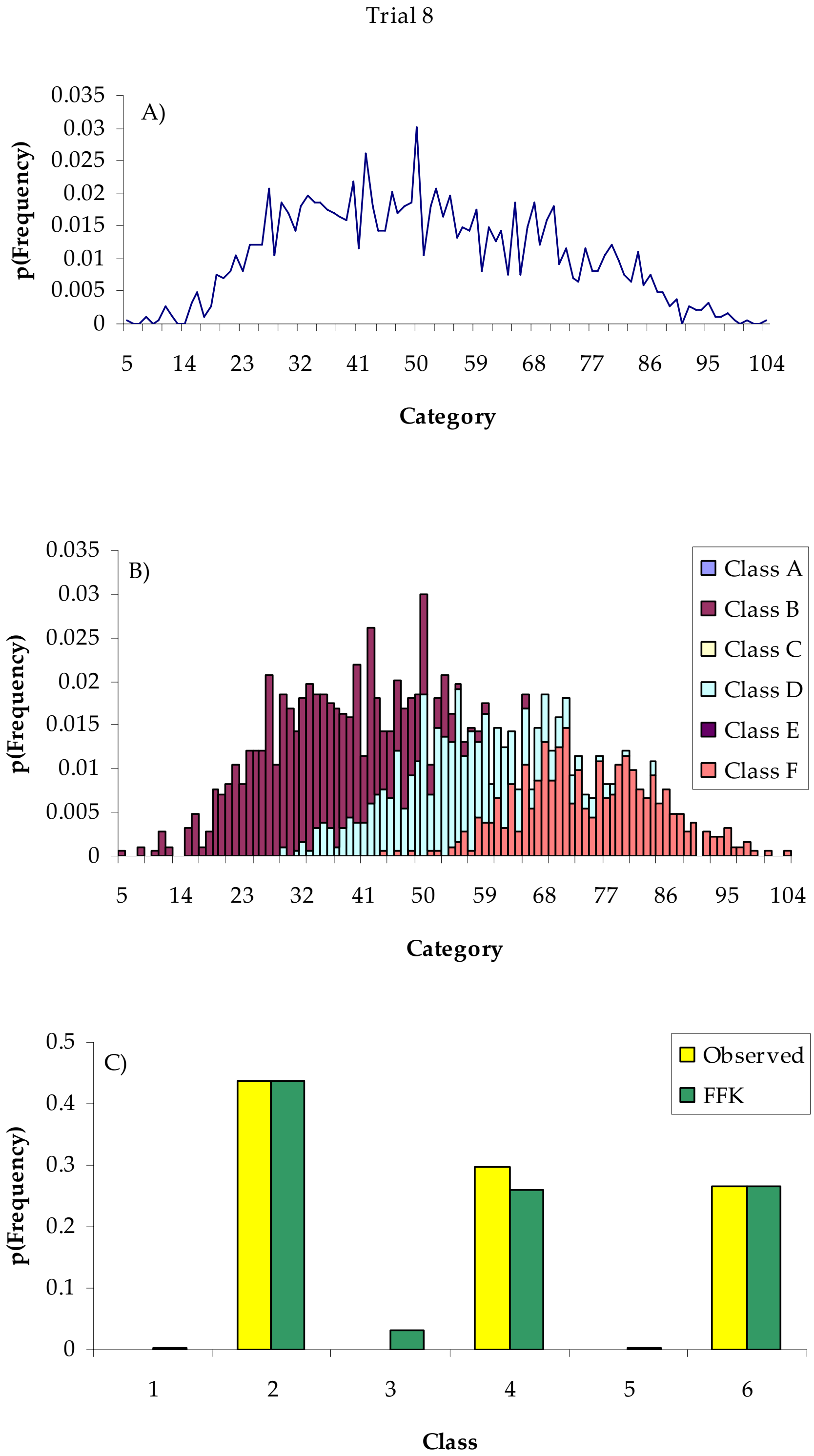
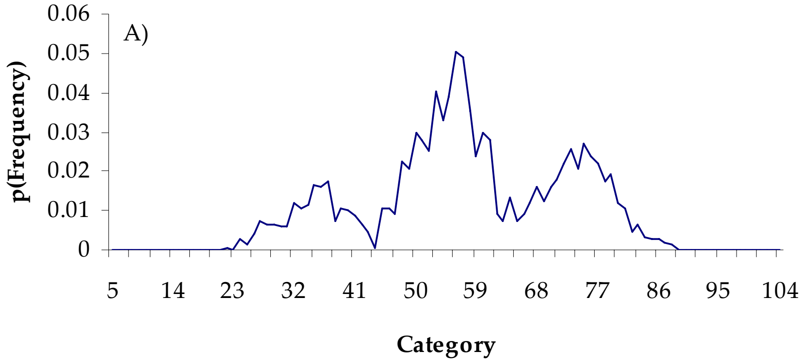
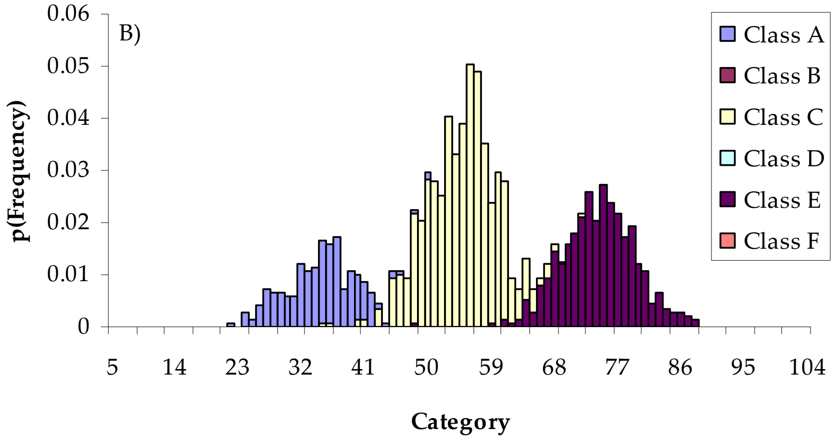


Figure 22. Decomposition of simulation 3 trial 8. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

Trial 9





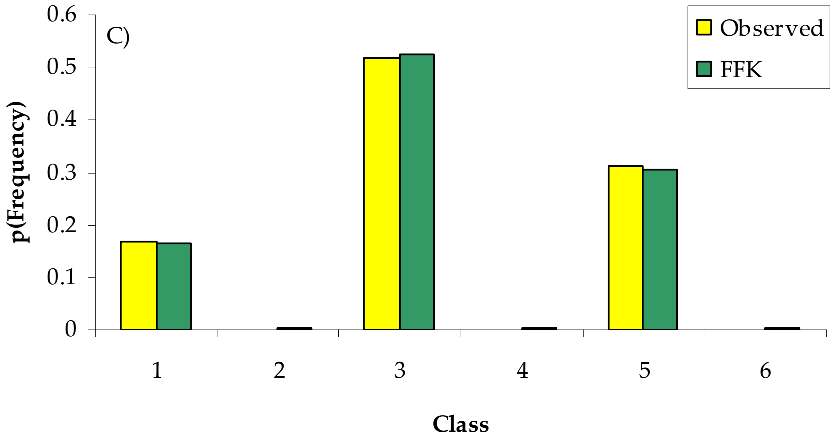


Figure 23. Decomposition of simulation 3 trial 9. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

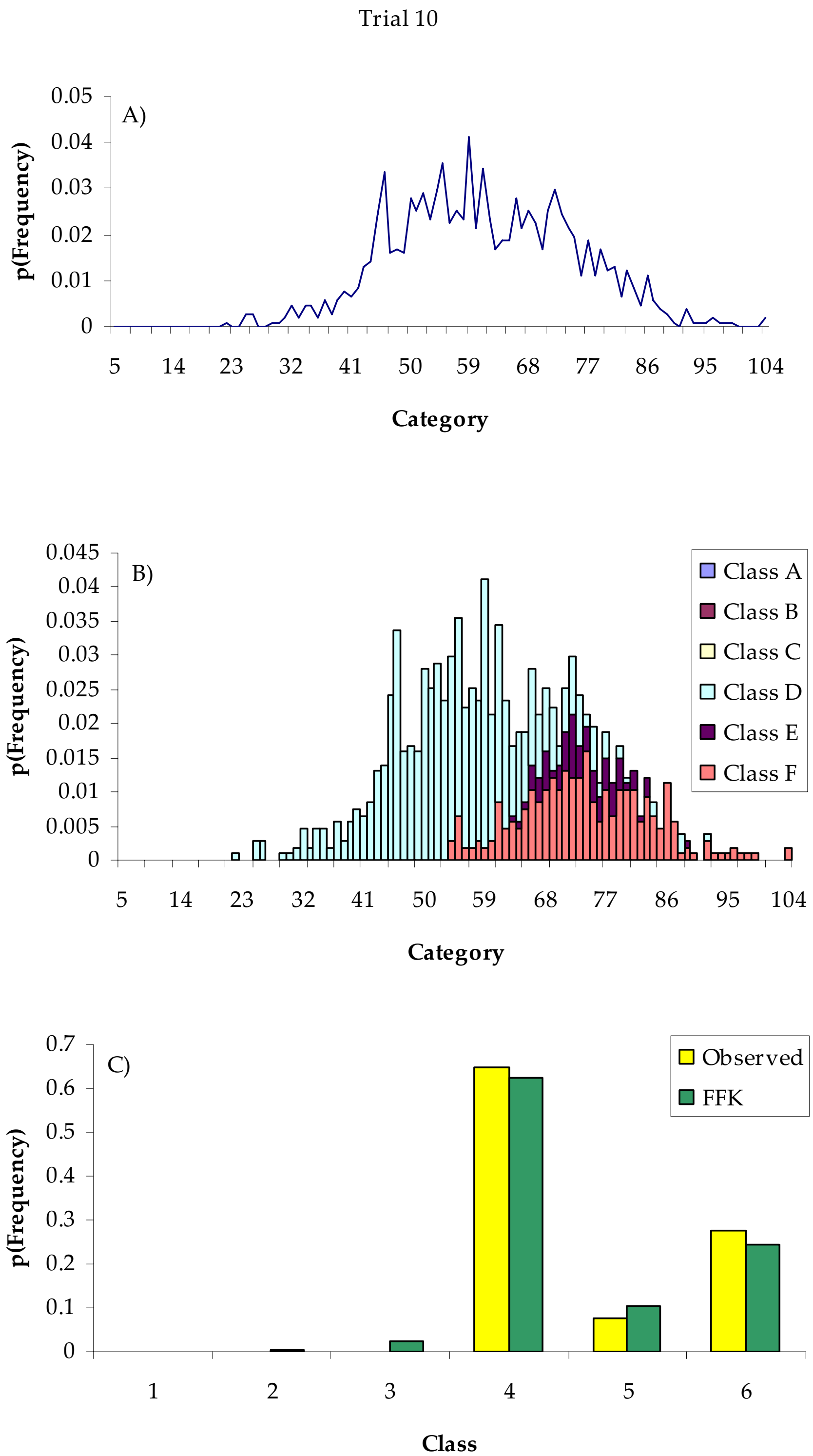


Figure 24. Decomposition of simulation 3 trial 10. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

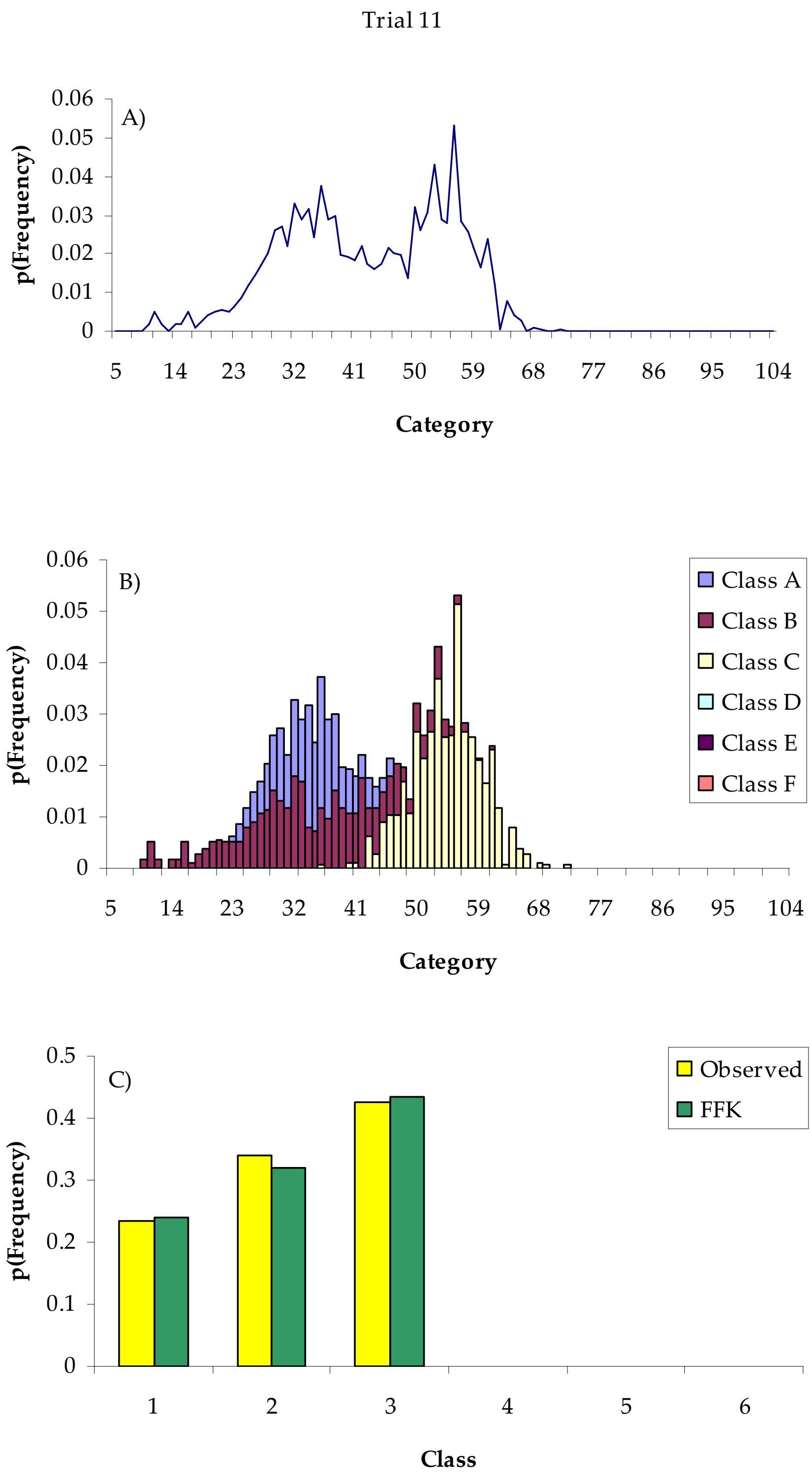
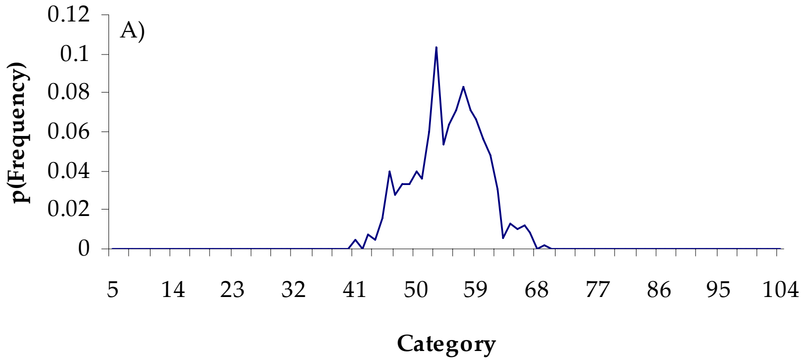
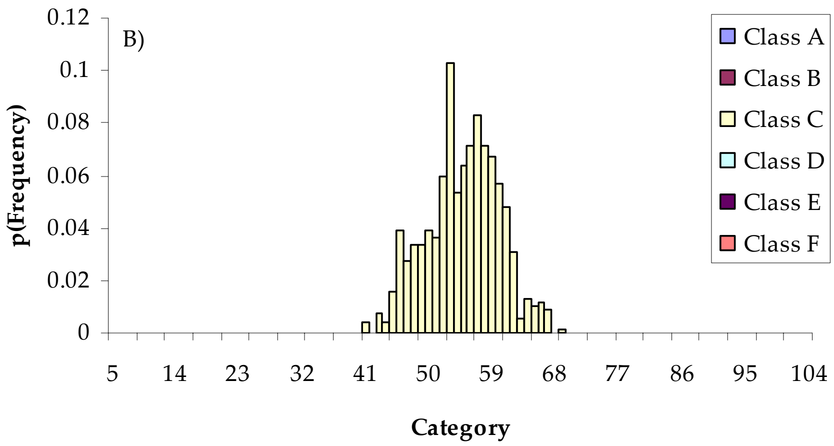


Figure 25. Decomposition of simulation 3 trial 11. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

Trial 12





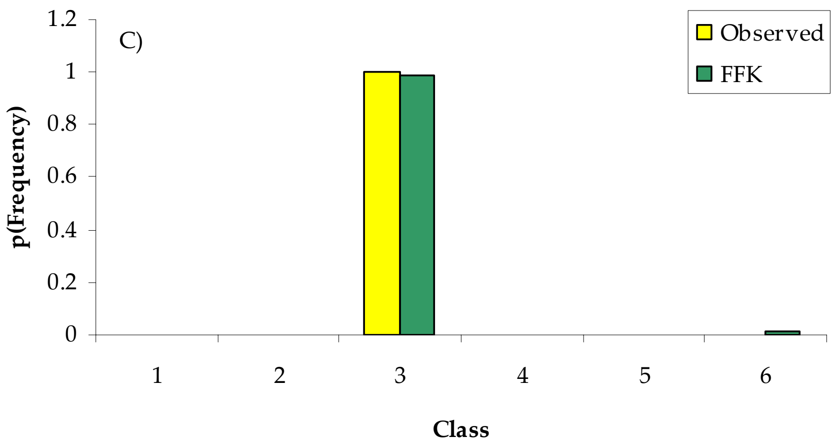


Figure 26. Decomposition of simulation 3 trial 12. Six classes (3 means (35, 55, 75) by 2 standard deviations (5,10)). A) The mixed distribution of the two classes, B) underlying distribution of each class and C) observed class frequency and frequency predicted by the FFK method.

Discussion

The FFK method successfully reconstructed the underlying classes from the first series of trials in simulation 1. The classes could be visually discriminated (Figures 5 to 9). It is conceivable that a number of techniques could have been employed to deliver similar results.

The FFK method was effective in decomposing length frequencies into the underlying age classes, where the classes had the same means (simulation 2, Figures 10-14) and where multiple classes had the same means and different standard deviations (simulation 3, Figures 15-26). Modal discrimination could not be used successfully to delineate the classes in these scenarios.

The FFK method was successful in decomposing length data into underlying age classes where three modes were present and where all six classes were included (see Table 4). Modal separation could not have been used successfully to classify these classes accurately, as examination of this mixed distribution would suggest only three classes. The FFK method required no parameterisation of the *a priori* distributions. Only the empirical *a priori* distributions and the mixed distribution constructed from an unknown number of classes were required as inputs.

The FFK method successfully separated six classes and predicted missing classes. Here, all of the initial distributions from which the mixed distributions were re-sampled were overlapping at the midpoint of the range (50). When the unknown distribution, which had a mean of 50 and the standard deviation was 5 (a single class), the method predicted the underlying class with a 98.9% accuracy. For all simulations, the FFK method effectively determined the relative abundances for each of the classes, reflecting the underlying unknown abundance of each class within the mixed distribution. This was achieved to a high level of precision.

The numbers chosen for the first two simulations (400) were relatively low. The third simulation, where initial distributions were re-sampled 800 times from each class, reflects sample sizes approximating real fisheries ageing data.

These simulated distributions, which were analogous to 'length-at -age', were constructed using the normal distribution function using predetermined means and standard deviations. The FFK method does not necessarily require the distribution of the age classes to conform to any predefined function as the empirical distribution is used.

The regime used to sample the distribution was chosen because it is similar to that undertaken for collecting fisheries data in Australia. This type of data collection would be analogous to randomly collecting length data from a sampling point (on-board or port based) and then sub-sampling the fish that were measured for ageing. The sample that was aged would then provide the known distributions of length–at–age (*a priori*), while a sample that was measured (with no material collected for ageing) would comprise the mixed unknown distribution.

Using a benchmark of 5% accuracy, a level which would be appropriate for fisheries assessments, the decomposition of the mixed distributions, using only *a priori* information on the empirical distribution of each class from the parent distribution, was successful.

In the next chapter the FFK method was trialled on School Whiting data and the effect of introducing a modal progression to the dataset.

Chapter 3. Simulations using School Whiting data

Introduction

The previous chapter demonstrated the utility of the FFK method to decompose mixed distributions with overlapping classes. The combined distributions ranged from two distributions with well-defined modes to multiple classes with the same mode and different standard deviations. The FFK method was successfully demonstrated on normal distributed data, where in some, the distributions overlapped to a high degree. In this chapter real fisheries data are used to test the efficacy of the FFK method.

The length-at-age data used in this chapter are observed length-at-age distributions. These data were collected from a variety of sources, including on-board observers and port measurers. All samples were aged at the Central Ageing Facility (CAF) using protocols described in Morison *et al* (1998). These age-composition data have been used as an integral part of School Whiting stock assessments since 1991.

The aim of this component of the study was to test the ability of the FFK method to detect modal progressions in fisheries length-at-age data. This is an extension of previous published work by Troynikov and Robertson (2005).

In this chapter, the total collection of aged data is treated as a population. The relative abundance of each age-class was increased by 30% from the number of samples within the original population, which simulated modal progression of a strong year-class over time.

Methods

School Whiting otolith collection

The dataset consisted of 7,349 aged School Whiting age estimates. Samples were primarily collected from Eastern Bass Strait (EBS - 95%) with 5% of the samples collected from East Australia (EastA). The majority of the School Whiting samples from EBS were collected from Lakes Entrance. The samples collected from EastA were from Greenwell Point, Kirrawa, Sydnam Inlet and Ulladulla. The fishing method for this species is Danish seine. Samples were collected at a frequency of approximately 600 per year. These were collected opportunistically by on-board observers from the Integrated Scientific Monitoring Program (ISMP) and port–based measurers at fish processors. All otolith samples were aged from whole otoliths under incident light. The numbers sampled at each month between 1991 and 2005 are shown in Tables 8 and 9.

Samples were collected at frequencies, which were representative of the fishing period. Numbers ranged from approximately 50 to 200 per month, with a target of 600 samples per year. Monthly totals for combined years ranged from 82 in August though to 1063 in March.

School Whiting otolith data

The length frequency of the School Whiting ranged from 9 cm to 26 cm with a single mode of 17cm (Figure 27). The number of classes within the combined length composition was 18. Data were not separated by sex as no significant differences in growth of the two sexes were observed for School Whiting (CAF data).

The ages ranged from one to eight years for the combined dataset. The modal age-class was three years. This dominant age-class accounted for approximately 40% of the data (Figure 28). All samples with an age-class greater than seven were binned to a 7+ group. The 7+ group still only accounted for 0.43% of the total sample.

Table 8. School Whiting samples collected by zone, area and year considered which were used to resample to construct a population.

	Zone	Area of Capture						
	East A			EAST A Total	EBS		EBS Total	Total
Year	Greenwell Pt.	Kirrawa Syd..I.	Ulladulla		Lakes Entrance	San Remo		
1991					97		97	97
1992					690		690	690
1993					1065		1065	1065
1994					501	249	750	750
1995					293	393	686	686
1996					603		603	603
1997					394		394	394
1998					690		690	690
2001			101	101	493		493	594
2002	103	54		157	403		403	560
2003	40			40	431		431	471
2004	40			40	609		609	649
2005					100		100	100
Total	183	54	101	338	6369	642	7011	**7349**

Table 9. Sample collection details by month and year (combined areas).

	Year													
Month	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**2001**	**2002**	**2003**	**2004**	**2005**	**Total**
1					98				101					199
2			104	135			99	127	22	254			100	841
3		282		269	195				120		197			1063
4						102	100	100						302
5			272		97	102	99	100	99					769
6			387			199	96					237		919
7		292	103	147		98								640
8									82					82
9		116			197				60			184		557
10	97		98	199							119	100		613
11			101		99	102			110	203	155	88		858
12								363		103		40		506
Total	97	690	1065	750	686	603	394	690	594	560	471	649	100	**7349**

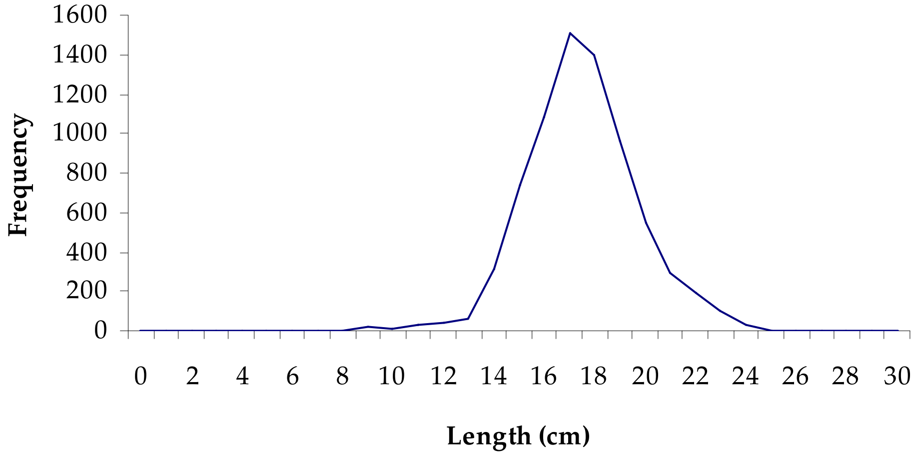


Figure 27. Length frequency of combined School Whiting samples used as the base population. N=7349

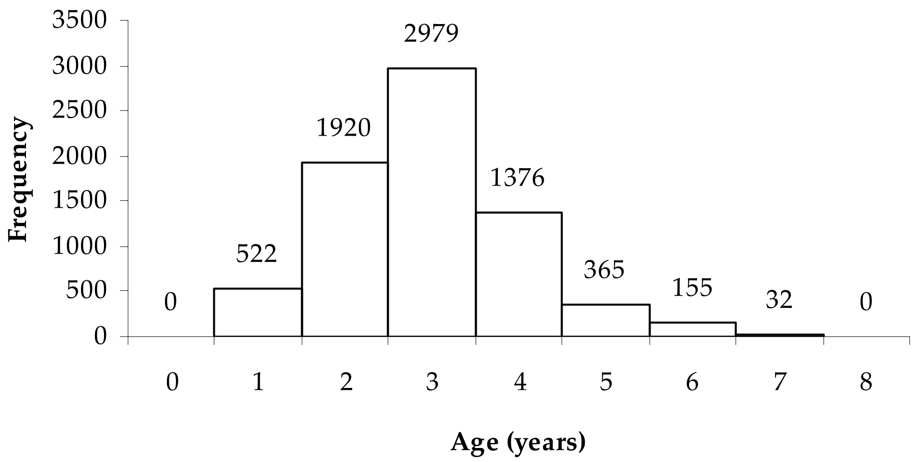


Figure 28. Age composition for combined School Whiting ageing data (1991-2005), Age-class 7 is the 7+ group. N=7349.

Sampling for modal populations

The distributions of length-at-age that constituted the matrix of *a priori* length-at-age information were calculated from the original population. The frequency distribution of each age-class was converted to an empirical distribution. These distributions of length for each age-class, which constitute the length-at-age matrix and provided the *a priori* information which was used for all tests in this chapter, are shown in Figure 29. The size of matrix used as input for the FFK method was *a*[7,18].

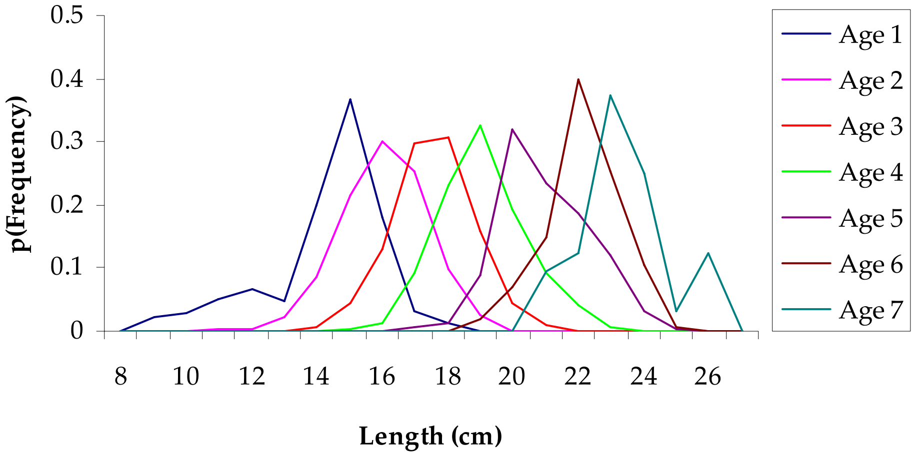


Figure 29. Empirical distributions of length-at-age for School Whiting selected from the orginal dataset.

To generate the population, which was used in the simulations, the original dataset was randomly re-sampled 7,349 times with replacement to provide the population on which all subsequent population modifications would be based. The re-sampled population was termed the base population. The modifications were made to increase the abundance of individual age-classes while the remaining age-classes remained at the base population abundance. Each successive age-class was increased in a separate simulation. The final simulation was to remove the modal age-class that was present in the base population, and removal of the original modal age-class over seven trials.

To determine the numbers for the increased abundance, 30% of each age-class was estimated. The result was rounded up to the next highest integer. For each of the six populations, the relative abundance of one age-class was increased by randomly re-sampling the base population from the target age-class to increase the abundance of the target age-class by the desired value. We termed the new populations with altered abundances of individual age-classes 'trial populations'. Each of the seven trial populations had the same age-class abundance configuration as the base population with the abundance of one age-class increased by 30%. The 7+ group accounted for low numbers and was not increased in abundance, but was included for all further analyses.

One additional population was trialled (trial 0) where the dominant age-class (3-year–old) was removed. This is a similar test used in Troynikov and Robertson (2005). The number of samples in each trial population is shown in Table 10.

The known age frequencies from each trial population were converted to an empirical probability by summing the distribution to one. This distribution was appended to the unknown distribution of length-at-age in the data input to the FFK. This was included so error between the known age composition and the estimated age composition derived using the FFK method could be calculated in the Fortran program. The error calculated in the program was a quadratic error between the observed and predicted age compositions.

For this component of the study, an adjustment of the length-at-age matrix was undertaken to remove biases. Potential bias in length-at-age data can be introduced from sampling and sub-sampling methods. The methods used to collect samples for ageing data traditionally used in fisheries management (the ALK method) are not necessarily random. For example data can be length stratified and not representative of the biological length-at-age distribution. The result is a matrix which represents age distribution at any given length class $P(a|l)$ that suit the requirements for the ALK method. The FFK method requires length distribution at a given age *P(l|a)*, which must be random and not a result of length-stratified sampling . Potential biases in length-at-age data are corrected by Bayesian inverting of conditional distributions (Equation 1).

$$P(l \mid a) = \frac{P(a \mid l).P(l)}{P(a)},$$

where

$$P(a) = \sum P(a \mid l).P(l)$$

Equation 1. Bayesian inversion of the conditional age-at-length to length-at-age distributions.

P(l) is the length distribution of larger sample. *P(a)* is the age composition which is estimated using the ALK method. *P(l|a)* is an unbiased estimate of length distribution which can then be used to reconstruct the age composition from independent length frequency data using the FFK method.

The length frequencies of the first three trials are presented in Figure 30. These frequencies have been converted to an empirical probability distribution. Length frequencies of trials 4-6 are presented in Figures 31 and 32.

Trials

After data selection, preparation and Bayes correction, the following simulations were conducted:

1. The age composition of the base population was first calculated by the FFK method using the length-at-age matrix derived from the original population and the length frequency distribution from the unknown (base population). The age composition of the base population was also calculated using the traditional ALK method. The ALK matrix was also constructed from the original population. Differences between the reconstruction results from both the FFK and ALK methods were expressed as a percentage of the known age composition for each age-class within the base population.
2. The simulation using the missing year-class three was tested next. Again the Bayes corrected matrix used by the FFK method was calculated from the original population and applied to the length frequency of the dataset with the missing age-class (trial 0). For this simulation the observed percentage for each age-class and the percentage of each age-class was calculated. The numbers of each age-class predicted by the FFK method and the known number at each age class within the base population were used to calculate the difference between these two values and the percentage error between the observed and predicted values were also assessed.
3. The same procedure used for simulation 2 was used in simulation 3. Each trial within simulation 3 had one age-class increased by 30% through random re-sampling. The FFK method used the Bayes corrected matrix from the original population to determine the underlying age-classes from the length frequency data of each trial containing the altered abundances.

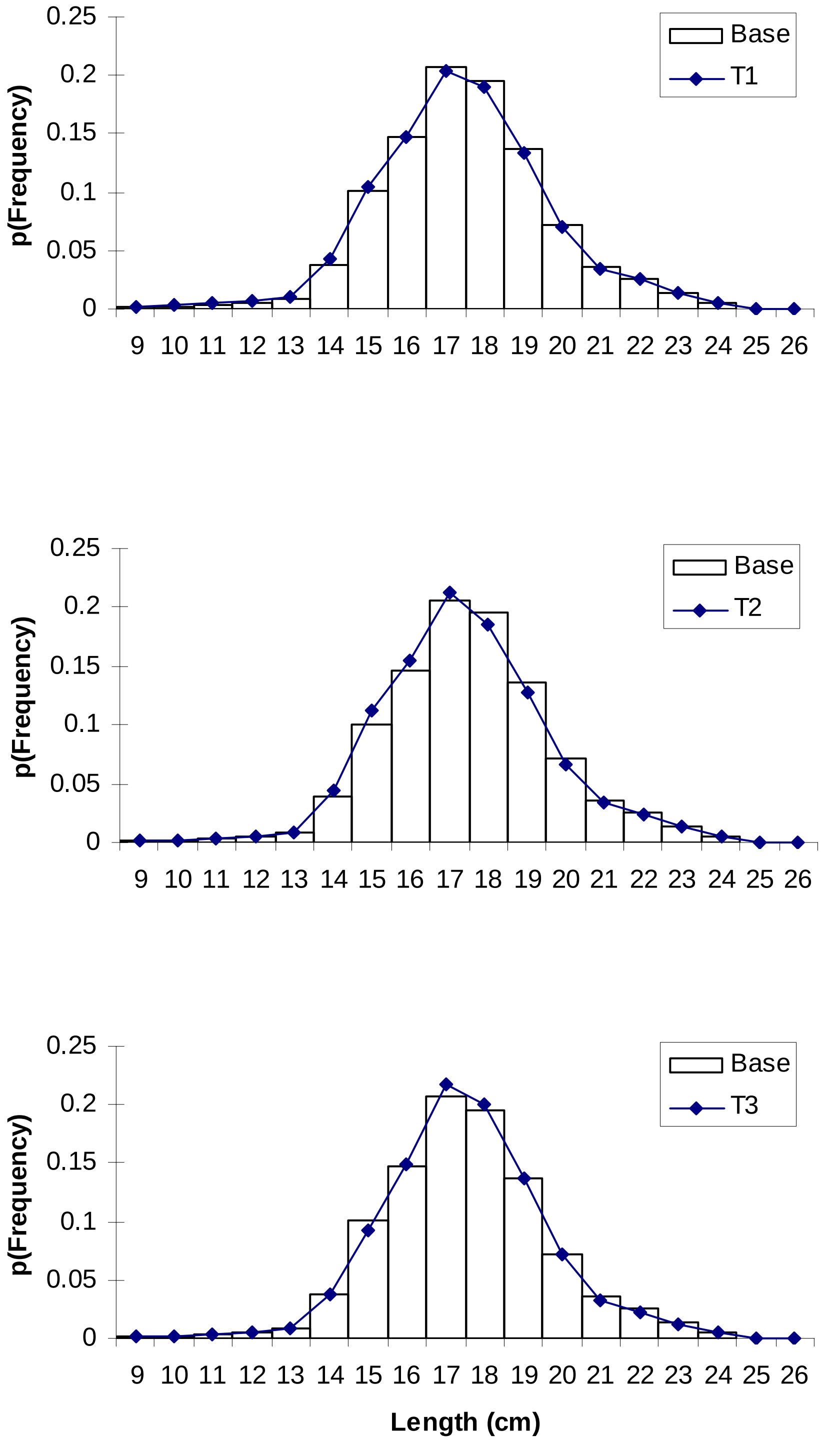


Figure 30. Base population length frequency and length frequency of trial populations with a) 30% increase in age-class 1 (T1), b) 30% increase in age-class 2 (T2), and c) 30% increase in age-class 3 (T3).

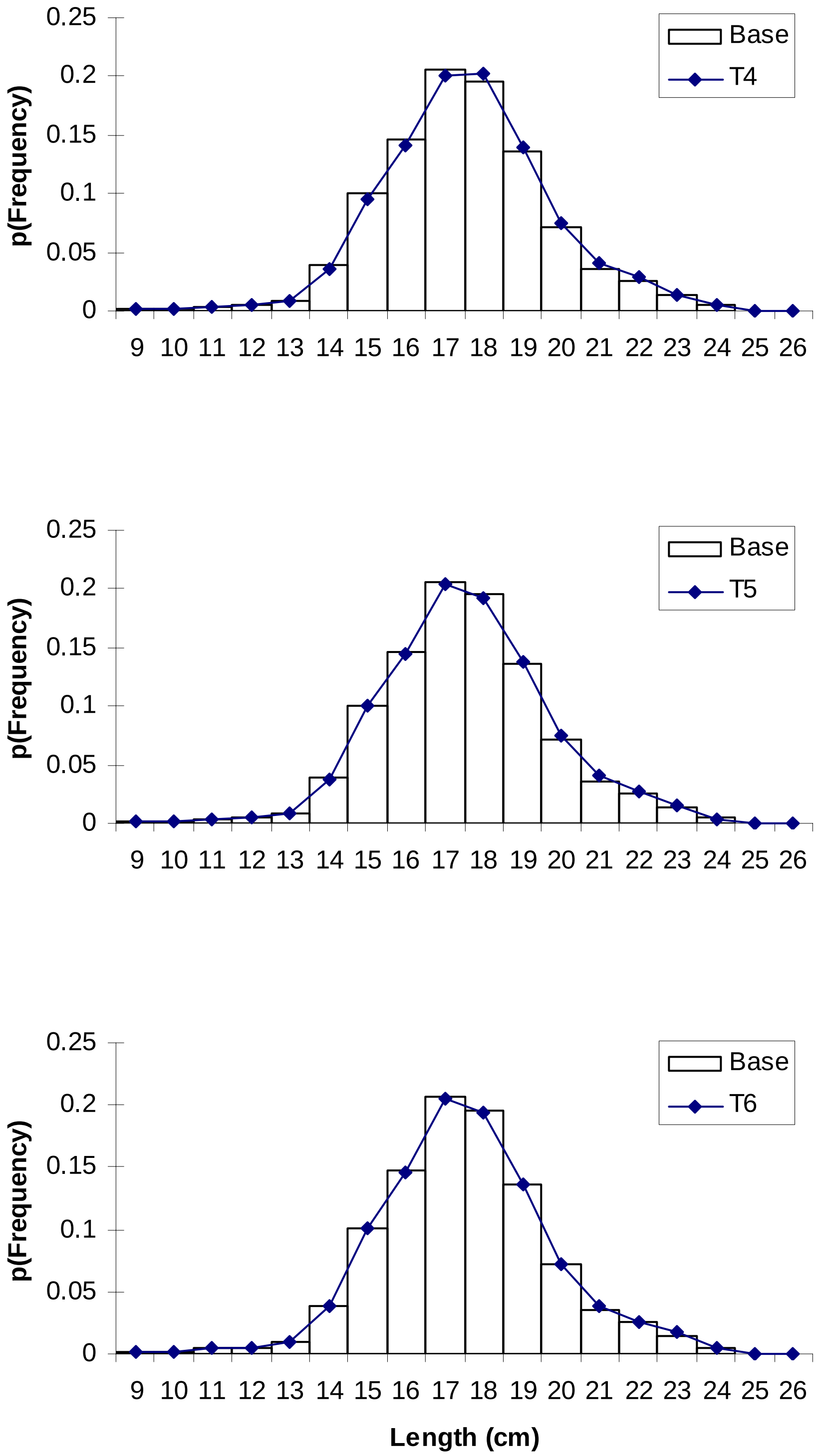


Figure 31. Base population length frequency and length frequency of trial populations with a) 30% increase in age-class 4 (T4), b) 30% increase in age-class 5 (T5), and c) 30% increase in age-class 6 (T6).

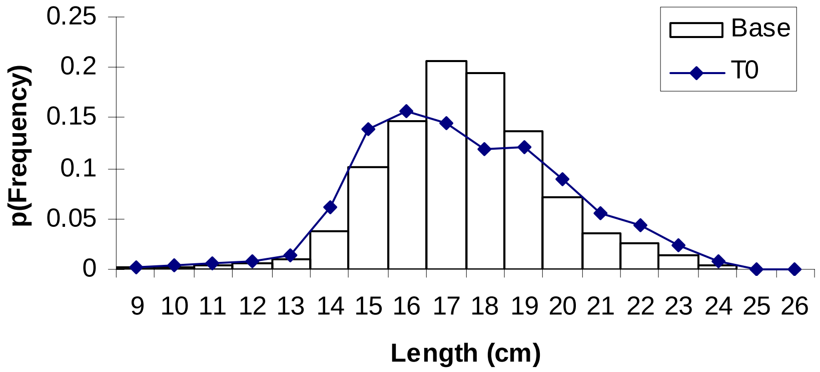


Figure 32. Base population and length frequency of trial 0 where the modal age-class of 3 was removed.

Table 10. Numbers of School Whiting used in each trial. The bold numbers indicate an increase of 30% over the base population age–class abundance for the first seven trials and the absence of age–class 3 in trial 0.

Age–class	Base pop.	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 0
0	0	0	0	0	0	0	0	0
1	522	**679**	522	522	522	522	522	522
2	1920	1920	**2496**	1920	1920	1920	1920	1920
3	2979	2979	2979	**3873**	2979	2979	2979	**0**
4	1376	1376	1376	1376	**1789**	1376	1376	1376
5	365	365	365	365	365	**475**	365	365
6	155	155	155	155	155	155	**202**	155
7	32	32	32	32	32	32	32	32
8	0	0	0	0	0	0	0	0
Total	7349	7506	7925	8243	7762	7459	7396	4370

Results

Simulation 1

The first simulation reconstructed the base population age composition using the length-at-age information from the original population. The decomposition into the respective age-classes from the base population length frequency from this trial was compared with the results obtained using the traditional age-length key method. These results provide a comparative measure of the efficacy of the FFK method. The ALK method produced slightly closer estimates of the unknown age composition than the FFK method. The results from this test are shown in Table 11 and Figure 33.

Table 11. The known relative abundance of each age-class from the base population and the predicted age composition from the ALK method and the FFK method.

Age–class	Base	A (ALK)	B (FFK)
1	7.10%	6.84%	6.07%
2	26.13%	25.98%	27.26%
3	40.54%	40.98%	38.78%
4	18.72%	18.79%	22.55%
5	4.97%	4.86%	1.54%
6	2.11%	2.11%	2.93%
7	0.44%	0.45%	0.87%

Figure 33. The known relative abundance of each age-class from the base population and the predicted age composition from the ALK method and the FFK method.

The results show that the ALK methods produced a result that was closer to the unknown values of relative abundance for each age-class within the base population than the FFK method. The largest error seen for this trial was the age-class 5, where the abundance was under-estimated by the FFK method. The fifth age-class only constituted about 4.9% of the total population. Results from other age-classes show relatively close agreement with the age composition to the base population age-class abundance for both methods.

Simulation 2
Where the modal age-class (3) was removed, the FFK method accurately decomposed the length frequency to the underlying age-classes. The maximum error for this was 1.28% for the first age-class. Where age-class three was removed, the error from the method was 0.69%. In a situation where an age–class was totally removed, the error in the estimated abundance could only be positive (Table 12).

The FFK method accurately determined the relative abundance of each of the other age-classes for trial 0. The maximum error for this test was 1.28% of the total population for age-class 1. The relatively low number of samples contributes to a higher proportion of mis-assignment of the frequency. The proportion of error between observed and predicted for age-class 3 was 0.69%. The error between the observed and predicted frequencies at the age-class with the next lowest abundance (age-class 7+), was 0.41%. Where sample size was large, error was low, for example 0.9% for age-class 2 (original n=1920, mis-assigned n=39) and 0.41% for age-class 4 (original n=1376, mis-assigned n=18).

Table 12. Trial (0) results where the modal age class was removed. Relative proportions of each age–class expressed as a percentage. Number predicted (N FFK) and known number (N Observed) are shown. The percentage error for each age–class with respect to the total population (n=7349) is also given. (% of pop)

Age-class	Observed %	FFK %	N FFK	N Observed	Diff. (N)	% of pop
1	12%	11%	466	522	-56	-1.28
2	44%	45%	1959	1920	39	0.90
3	0%	1%	30	0	30	0.69
4	31%	32%	1394	1376	18	0.41
5	8%	7%	324	365	-41	-0.93
6	4%	3%	146	155	-9	-0.20
7	1%	1%	50	32	18	0.41

Simulation 3
For all of the simulations a similar trend was evident. The FFK method generally under-estimated the known abundance when the abundance of the age-class had been artificially increased. The frequency predicted for each age-class was closer to the known age frequency than the age frequency of the base population. As all numbers for each age-class were the same in each trial except for the age-class with an altered abundance, a composite of all simulations is presented first which provides an overall summary (Figure 34, Table 13). Details of each trial are shown in Table 14 through Table 20. Only the first 6 age–classes are included in Table 13 and Figure 34 as age-class 7 was not altered.

The maximum error between the observed age composition and the age composition determined using the FFK method for each age-class (as a percentage of the total known population) was less than 5% for every trial.

Figure 34. Original base population, the number of the altered age-class

Table 13. Composite trial results with proportion of each age-class predicted from the length frequency (FFK), the true proportion of each age–class (Observed), the proportion of each age–class in the base population (Base). %A is the proportion between the observed and base percentage for each age-class and %B is the proportion between the FFK and base percentage for each age-class.

Age-class	Base	Observed %	FFK %	% A	% B
1	7.10%	9.05%	7.98%	78.52	88.21
2	26.13%	31.50%	31.13%	82.95	98.84
3	40.54%	46.99%	43.66%	86.27	92.92
4	18.72%	23.05%	23.22%	81.24	100.76
5	4.97%	6.37%	3.44%	77.99	53.97
6	2.11%	2.73%	3.11%	77.22	114.04

Table 14. Base age composition estimated using original population length-at-age data (n=7349).

Age-class	Observed %	FFK %	N FFK	N Observed	Diff. (N)	% of pop
1	6%	7%	446	522	76	1.04
2	27%	26%	2003	1920	-83	-1.14
3	39%	41%	2850	2979	129	1.75
4	23%	19%	1657	1376	-281	-3.82
5	2%	5%	113	365	252	3.43
6	3%	2%	216	155	-61	-0.82
7	1%	0%	64	32	-32	-0.44

Table 15. Trial 1 age composition estimated using original population length-at-age data. Age-class 1 increased by 30% (n=7506).

Age-class	Observed %	FFK %	N FFK	N Observed	Diff. (N)	% of pop
1	8%	9%	599	679	80	1.07
2	26%	26%	1978	1920	-58	-0.77
3	39%	40%	2890	2979	89	1.18
4	22%	18%	1628	1376	-252	-3.36
5	2%	5%	134	365	231	3.07
6	3%	2%	210	155	-55	-0.74
7	1%	0%	66	32	-34	-0.45

Table 16. Trial 2, age composition estimated using original population length-at-age data. Age-class 2 increased by 30% (n=7925).

Age-class	Observed %	FFK %	N FFK	N Observed	Diff. (N)	% of pop
1	6%	7%	507	522	15	0.20
2	31%	31%	2467	2496	29	0.36
3	38%	38%	3007	2979	-28	-0.36
4	19%	17%	1496	1376	-120	-1.51
5	3%	5%	203	365	162	2.04
6	2%	2%	174	155	-19	-0.24
7	1%	0%	70	32	-38	-0.49

Table 17. Trial 3, age composition estimated using original population length-at-age data. Age-class 3 increased by 30% (8243).

Age-class	Observed %	FFK %	N FFK	N Observed	Diff. (N)	% of pop
1	4%	6%	344	522	178	2.16
2	27%	23%	2207	1920	-287	-3.49
3	44%	47%	3599	3873	274	3.33
4	20%	17%	1661	1376	-285	-3.46
5	2%	4%	192	365	173	2.10
6	2%	2%	159	155	-4	-0.05
7	1%	0%	81	32	-49	-0.59

Table 18.Trial 4, age composition estimated using original population length-at-age data. Age-class 4 increased by 30% (n=7762).

Age-class	Observed %	FFK %	N FFK	N Observed	Diff. (N)	% of pop
1	7%	7%	524	522	-2	-0.03
2	23%	25%	1797	1920	123	1.58
3	41%	38%	3167	2979	-188	-2.43
4	23%	23%	1803	1789	-14	-0.18
5	2%	5%	159	365	206	2.66
6	3%	2%	262	155	-107	-1.38
7	1%	0%	50	32	-18	-0.24

Table 19. Trial 5, age composition estimated using original population length-at-age data. Age-class 5 increased by 30% (n=7459).

Age-class	Observed %	FFK %	N FFK	N Observed	Diff. (N)	% of pop
1	6%	7%	445	522	77	1.03
2	27%	26%	2005	1920	-85	-1.14
3	38%	40%	2847	2979	132	1.77
4	22%	18%	1669	1376	-293	-3.93
5	3%	6%	256	475	219	2.93
6	2%	2%	179	155	-24	-0.32
7	1%	0%	58	32	-26	-0.34

Table 20. Trial 6, age composition estimated using original population length-at-age data. Age-class 6 increased by 30% (n=7396).

Age-class	Observed %	FFK %	N FFK	N Observed	Diff. (N)	% of pop
1	6%	7%	450	522	72	0.97
2	27%	26%	1990	1920	-70	-0.94
3	39%	40%	2871	2979	108	1.45
4	22%	19%	1629	1376	-253	-3.43
5	2%	5%	154	365	211	2.86
6	3%	3%	230	202	-28	-0.38
7	1%	0%	71	32	-39	-0.53

Discussion

The robustness of the FFK method to large perturbations in the age-class strength was demonstrated in trial 0 (this study), reinforcing the initial evaluation of the method (Troynikov and Robertson, 2005). These demonstrate that the FFK method can accurately determine the age-composition from a length frequency from another (randomly re-sampled) population with the modal age-class removed.

The presentation of the length frequency for the base population and trials one to six demonstrate how close the altered length frequencies were to those of the base population. A relatively large change in abundance for an individual age-class has relatively little effect on the length frequency. This problem was examined in this component of the study. The FFK method closely approximated the underlying age composition given *a priori* information on the distribution of length-at-age.

The plots of the length frequency of the base and altered populations illustrate the difficulty of determining the underlying age composition from length frequency data.

The value of 30% increase in year-class strength was intended to simulate a realistic recruitment pulse. The removal of the modal age-class was included as a maximum stress test for the method, because while an age class has been removed, there is little change in the length distributions. The decrease from the modal value of the length frequency under this condition is approximately 40% from the initial value. Increases or decreases of 100% in a particular age-class would be unlikely in a fish stock. The FFK method can still accurately decompose the length frequency to the underlying age-composition. The FFK method reproduced (with high accuracy) the underlying age-classes.

As a generalisation, the FFK method tended to under-estimate the number of the age-class that was increased. The exception was trial 4 where the abundance of age-class 4 was slightly over-estimated. Where the predicted age-class abundance was less than the observed abundance, the age-classes at age+1 and age-1 were over-estimated. This can be seen in the age composition of trial 3 where the predicted abundance was under-estimated and the age-class 2 abundance was over-estimated by 287 individuals and age-class 4 was over-estimated by 285 individuals. Age-class 3 was under-estimated by a similar number (n=274). This level of accuracy is considered acceptable for application to fisheries data.

For age–class 5, the estimation of abundance was under-estimated in all trials. The reason for this is unknown.

Chapter 4. School Whiting age compositions

Introduction

In the last chapter, the effect of changing (increasing) the abundance of an age-class then using the distribution of length-at-age from the original population to decompose an unknown length frequency into the underlying age composition was tested. The results indicated the FFK method was successful in predicting the age composition in this scenario.

This chapter tests the FFK method using length-at-age distributions from one, or a combination of years, and a length frequency from another year, to determine the unknown age composition from the year from which the length frequency was obtained. This is an extension of the previous simulations. The FFK method is tested against conditions that would be used to assess age compositions with fisheries data. This application of the FFK method would be used in the stock assessment process.

The age compositions which were calculated using the FFK method were compared with those obtained from the age-length key. The age-length key is the method that is currently used in fisheries management to determine the age composition of a stock. The use of age compositions determined by the ALK method provided a benchmark against which the FFK method can be quantitatively assessed. The age compositions used in this chapter were catch adjusted to make the results comparable to those typically used in stock assessments.

Age composition is determined currently using ageing samples and length frequencies collected from the same year. The FFK method and the ALK method were trialled in this data environment where ageing samples and length frequency data were collected in the same year. This provided an initial benchmark for the FFK method.

The FFK method was also trialled to determine age compositions in years for which no age data exist but where length frequencies had been collected. In this scenario, the assumption was made that no age data exist in a given year but that length frequency data were collected. The FFK method was applied to calculate the age composition using the previous year(s) length-at-age matrix and the length frequency collected from a year with no ageing data.

Practical application of this approach would be to age a commercially important species every second year and take length frequency information in alternate years. The age composition could be calculated for the years when no biological samples were taken to age the fish. For these types of applications of the FFK method, ageing data could be provided by a cost effective and timely method. Using the CAF and Integrated Scientific Monitoring Program (ISMP) datasets from School Whiting, this scenario (1) will test this application of the FFK method and quantitatively compare the results with those obtained (known age compositions) using the ALK method.

The FFK method was also trialled in a scenario (2) for reconstructing age compositions from historical length frequencies using ageing data collected at a later time. Again, results were compared with age compositions calculated using the traditional ALK techniques. In both of these types of scenarios, the temporal efficacy of the FFK method was tested.

Multiple years of length-at-age data were also combined and the FFK method applied to decompose length frequencies into the underlying age compositions. This scenario (3) would be similar to adding to the sample size of ageing data to increase the resolution power of the FFK method using several different years of ageing data. The combined matrix of *a priori* length-at-age information was used to calculate the age composition from the following years' length frequency. Using multiple years of ageing data has the potential to value add the historical ageing data which currently can only be used to determine the age composition in the year from which they were collected.

Methods

Data

The data used in this Chapter was obtained from the Danish seine fishery of south-eastern Australia, which is the most common capture method for School Whiting in this fishery. Samples were collected for ageing and for the length frequency data.

The dataset used for catch adjusting the age compositions was supplied by the Integrated Scientific Monitoring Program (ISMP). These data also included catch weights and sample weight of fish that were measured. Length frequencies and ageing samples were collected from port-based scientific observers. The length frequencies were catch weighted and summed by year. To weight length frequencies to the catch, the proportion of the sample weight to the weight of the total catch was used to multiply the numbers of samples at each length class. This is a standard method used to determine the numbers of individuals at each length-class in the total catch. No ISMP length frequency data were available for School Whiting from Danish seine for 1994, 1999, 2000 and 2005.

The length frequency of the aged sample was added to the summed catch weighted length frequency for each year. Sample numbers were large and the relative proportion of the length frequency of the aged sample was comparatively small. No ageing was undertaken for 1999 and 2000, only length frequency data were available for the samples which were supplied to the CAF. Numbers of fish aged in 2005 were considered too low for the application of the technique to reconstruct the age composition. These data are summarised in Table 21.

The aged sample was a subset of the dataset used in the previous chapter. Only samples that were collected from port–based sampling were used. All samples collected onboard were not used. The total number of aged samples between (and including) 1994 and 2005 was 5401 samples. This equates to the target of about 600 samples per year. Years which had no corresponding weighted length frequency were discarded from further analyses (1994, 1999, 2000 and 2005).

Otolith samples, which were collected on an opportunistic basis by port-based measurers employed by the ISMP program, were aged at the CAF. Samples had been previously aged using a stereo-microscope coupled to an image analysis system under incident light using standard protocols (Morison *et al.* 2000).

The length frequency for the aged sample is shown in Table 22. The length frequency of the aged samples ranged from 9 to 27 cm. The age composition of the samples was similar to that from the previous chapter with a modal age of three years.

Table 21. Catch-adjusted length frequencies supplied from the ISMP program for School Whiting. The aged sample component is added to each year. Years 1999, 2000 and 2005 are composed of the length frequency data only.

	Year											
Length	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
5	0	0	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	1	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	16	0	0
10	0	0	0	0	0	0	0	7	1	7	0	0
11	0	0	0	0	0	0	0	18	1	8	17	0
12	0	2	0	0	0	0	0	246	1	183	430	0
13	0	23	3715	5204	1613	0	2	1521	208	1206	1748	0
14	8	10262	9605	9323	6027	17	24	15091	537	5217	9876	2
15	76	27718	6319	15895	50439	59	81	35168	2419	23378	17573	9
16	114	48222	24999	68484	278205	58	127	72240	13973	38760	31951	17
17	173	83211	50932	135706	575769	158	150	118194	32279	18750	51707	23
18	155	99390	38601	139213	489762	231	148	100934	32472	5328	51773	23
19	108	58906	23097	102238	264660	145	90	50326	20231	2878	26336	26
20	46	31305	10693	52992	167618	94	72	25774	7591	4146	9765	17
21	27	9104	10515	30065	102459	73	38	13772	2021	1772	3576	7
22	17	2516	1664	12701	41890	26	27	6939	528	663	1224	1
23	16	1007	1572	5359	13734	10	9	4123	74	205	866	0
24	9	1	184	207	3773	5	3	1043	19	24	235	0
25	1	0	0	0	1	1	0	304	0	5	73	0
26	0	0	0	0	1	0	0	78	0	0	18	0
27	0	0	0	0	1	0	0	34	0	0	0	0
Total	750	371667	181895	577387	1995951	877	771	445814	112356	102546	207170	125

Table 22. Length frequency of aged samples for which corresponding catch-weighted length frequency data were available. N=5401

Length-class	1994	1995	1996	1997	1998	2001	2002	2003	2004
5						1			
6									
7						1			
8									
9								16	
10						7	1	7	
11						18	1	8	
12		2				38	1	1	1
13		5			3	21	3	13	4
14	8	31	7	1	37	27	33	61	51
15	76	32	18	9	99	85	98	87	70
16	114	54	95	27	97	95	87	74	100
17	173	127	182	73	83	84	71	118	136
18	155	168	141	98	103	72	69	63	138
19	108	146	85	77	80	59	65	18	115
20	46	86	38	46	65	34	84	3	23
21	27	25	25	37	42	27	35	1	7
22	17	4	10	18	49	17	9	1	3
23	16	6	2	8	22	8	1		1
24	9	1			7	2	2		
25	1				1				
26					1				
27					1				
Total	**750**	**687**	**603**	**394**	**690**	**596**	**560**	**472**	**649**

Common data preparation

For all of the tests in this chapter, preliminary data preparation was necessary. The methods are described below.

For each year an age-at-length matrix was constructed. An intermediate matrix was generated by multiplying the distribution of age-at-length with the catch–weighted length frequency data, which had been converted to a distribution which summed to one. To produce the final *a priori* length-at-age matrix, the age-at-length matrix was converted to a length-at-age matrix by dividing each of the elements of the intermediate matrix by the sum of the columns in the same matrix. The final matrix contained the distribution of length at each age-class (each column sums to one).

These matrices were produced for each year where matching ISMP length frequency data were available. The size of the matrix was *a*[7,23]. This was larger than the previous chapter (*a*[7,18]) as the matrix encompassed the larger length distribution from the ISMP data. All matrices for each year or combination of years were produced using the same technique and were Bayes corrected (see Chapter 3).

The age composition from the catch for each year was calculated using the traditional ALK method for each year. The catch weighted length frequency was calculated for each year using the ISMP data. The age composition was calculated by multiplying the ALK matrix by the catch-weighted length frequency data. The resultant age composition was used as the known age composition to which all FFK results were compared. All age compositions from both the FFK and the ALK methods are presented as probability density functions (i.e. all sum to one).

Testing the FFK method in the same data environment as the ALK method is used

As an initial test, the FFK technique was used to determine the age composition from length frequencies collected within the same year as the age data. Here, a sub-sample was aged from a larger sample of the catch which was measured (this comprised the length frequency sample). Using the age composition data from the sub-sample, the age composition is determined from the catch using the ALK. This was done by multiplying the proportion of the numbers of individuals at each length class in the ALK by the larger measured length frequency sample. The length frequencies were adjusted by sample and catch weight to determine the overall age composition of the catch.

Using the prepared matrices, the FFK method was applied to seven years of ageing and length frequency data (1995, 1996, 1997, 1998, 2001, 2002 and 2003). The error between the age compositions determined using the ALK method and the FFK method was calculated as a quadratic error.

Testing the FFK method using single year matrices on disparate years

The FFK method was tested using length-at-age data from different years where the length frequency data were taken. The years from which the length-at-age data and the length frequency data were tested to determine the age composition are shown in Table 23.. For this component, two years were chosen (1995 and 2001) and the age composition was calculated from the catch-weighted length frequency for the next sequential three years for the first test year (1995) and the next two years for the second test year. This gave five comparisons. The length-at-age matrix from both 1995 and 2001 were Bayes adjusted before the age composition was calculated.

We also reduced the dimensionality of the problem by increasing the length resolution to 2 cm. This effectively reduced the size of the length-at-age matrix by half in the length dimension. One of the same base years (1995) was used and the age composition was calculated for the next sequential three years (1996, 1997 and 1998).

Table 23. Tests using length-at-age data from year (A) and catch-weighted length frequency data from year (B) at length resolution 1 cm (LR1) and length resolution of 2 cm (LR2)

LR1		**LR2**	
(A)	**(B)**	**(A)**	**(B)**
1995	1996	1995	1996
1995	1997	1995	1997
1995	1998	1995	1998
2001	2002		
2001	2003		

To test multiple years, length-at-age data for two or more years were combined. This was done by using a weighted average of the values in the Bayes corrected matrix. The matrix was weighted by the number of samples aged at each length class in each year. Two tests were conducted using three years of ageing data and one test was conducted using four years of ageing data (Table 24). The multiple year tests used length frequencies and length-at-age matrices at a resolution of one centimetre.

Table 24. Tests using combined length-at-age matrices from more than one year to determine the age composition from a length-frequency of 'unknown' age composition.

Multiple year ageing data Length resolution = 1cm	Target year
1995, 1996->1997	1997
1995 & 1996->1998	1998
2001 & 2002->2003	2003
1995, 1996,1997 & 1998->2001	2001

Results

Testing the FFK method in the same data environment as the ALK is used

Where the FFK method was used in the same data environment as the ALK method, the results demonstrate that the FFK method delivers results which are very close to the known age composition (Figure 35 through Figure 41). The known age composition was calculated using the ALK method. The maximum error for the comparison of the results for the two methods was 0.113% in 1996. All other quadratic errors were less than 0.1% (Table 25).

Testing the FFK method using single year matrices on disparate years

Where the length-at-age matrix was used to determine the age composition from the following year using the length frequency, the results were less successful (Table 26). The maximum quadratic error for this series of length frequency decompositions was 5.07% for the age composition reconstructed for 1996 using the 1995 length-at-age matrix. The lowest error was 1.81% for the reconstruction of the 2002 age composition from the 2001 length-at-age matrix.

Where the error was highest, the majority of the length frequency was assigned to the modal age-class. When the reconstruction error was lower, the proportion of age-classes assigned to the known modal age-class was still high (Figure 42 through Figure 45).

Table 25. Observed and predicted age compositions a) from years 1995-1998 and b) from year 2001-2003.
Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. Total error is the quadratic error between observed and predicted age composition.

A

1995		1996		1997		1998	
Observed	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**
0.11	0.11	0.01	0.01	0.01	0.01	0.04	0.04
0.22	0.22	0.21	0.21	0.19	0.19	0.22	0.22
0.29	0.29	0.46	0.46	0.43	0.43	0.29	0.29
0.29	0.29	0.24	0.24	0.22	0.22	0.24	0.24
0.06	0.06	0.05	0.05	0.13	0.13	0.11	0.11
0.02	0.02	0.02	0.02	0.02	0.02	0.07	0.07
0.01	0.01	0.00	0.00	0.00	0.01	0.02	0.02
Q-Error(%)	0.023		0.113		0.050		0.006

B

2001		2002		2003	
Observed	**FFK**	**FFK**	**Observed**	**Observed**	**FFK**
0.19	0.18	0.16	0.16	0.13	0.12
0.39	0.40	0.18	0.18	0.47	0.47
0.35	0.35	0.50	0.50	0.38	0.38
0.06	0.06	0.13	0.13	0.01	0.01
0.01	0.01	0.02	0.02	0.01	0.00
0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00
Q-Error(%)	0.074		0.057		0.049

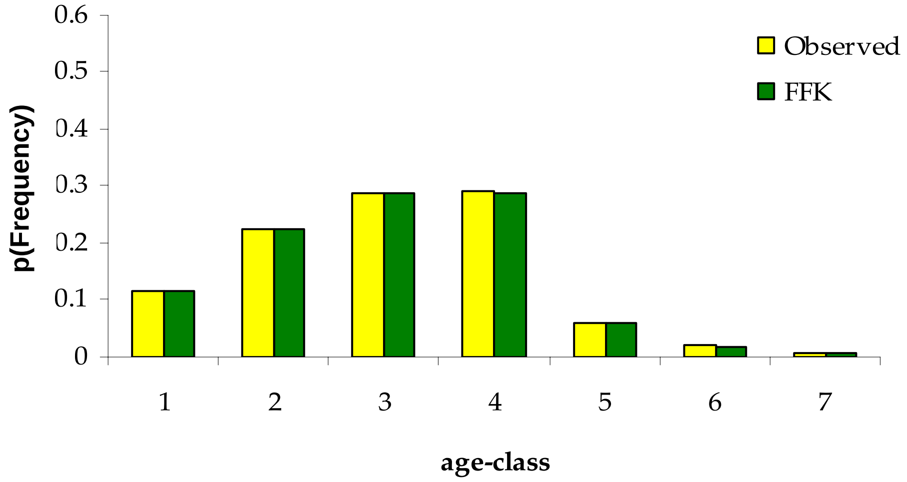


Figure 35. Age composition (1995) calculated using the FFK and the ALK method using the observed age composition length frequency data from the same year.

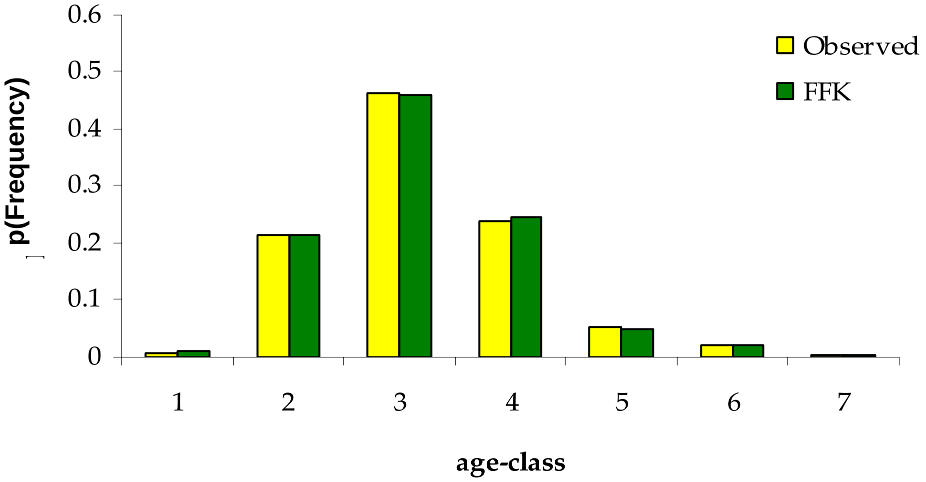


Figure 36. Age composition (1996) calculated using the FFK and the ALK method using the observed age composition length frequency data from the same year.

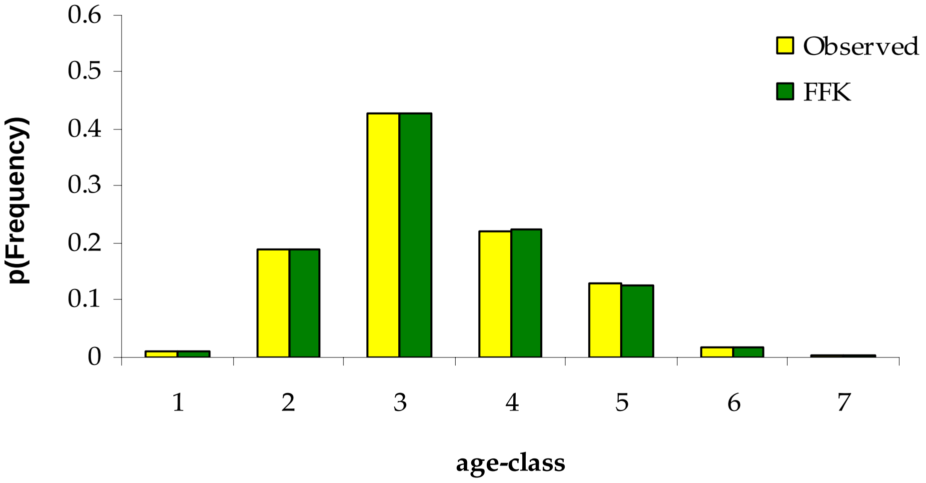


Figure 37. Age composition (1997) calculated using the FFK and the ALK method using the observed age composition length frequency data from the same year.

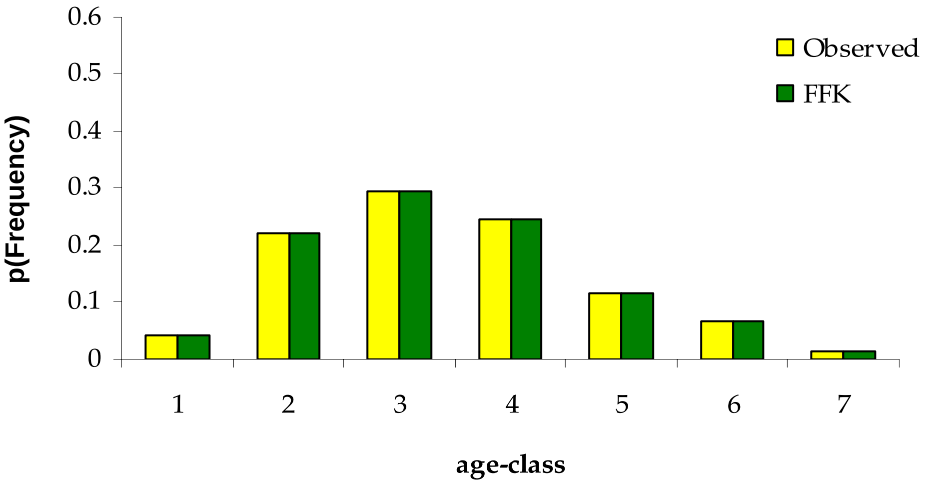


Figure 38. Age composition (1998) calculated using the FFK and the ALK method using the observed age composition length frequency data from the same year.

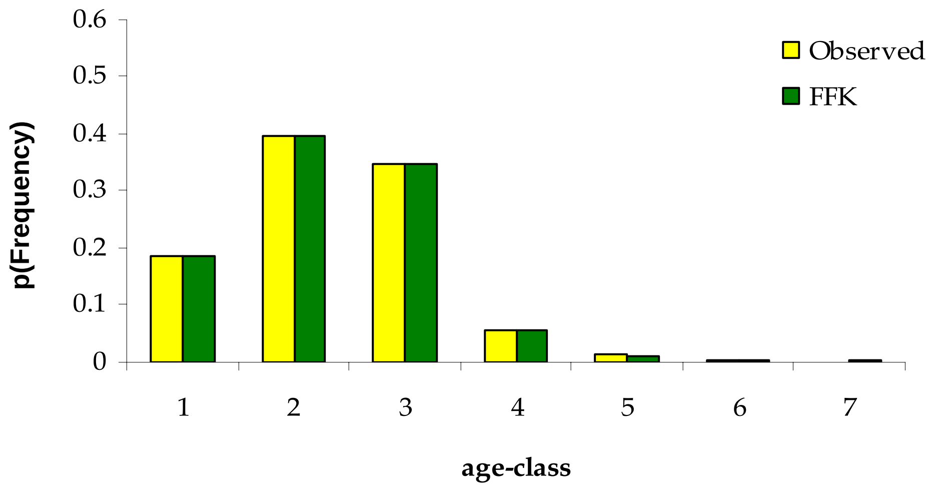


Figure 39. Age composition (2001) calculated using the FFK and the ALK method using the observed age composition length frequency data from the same year.

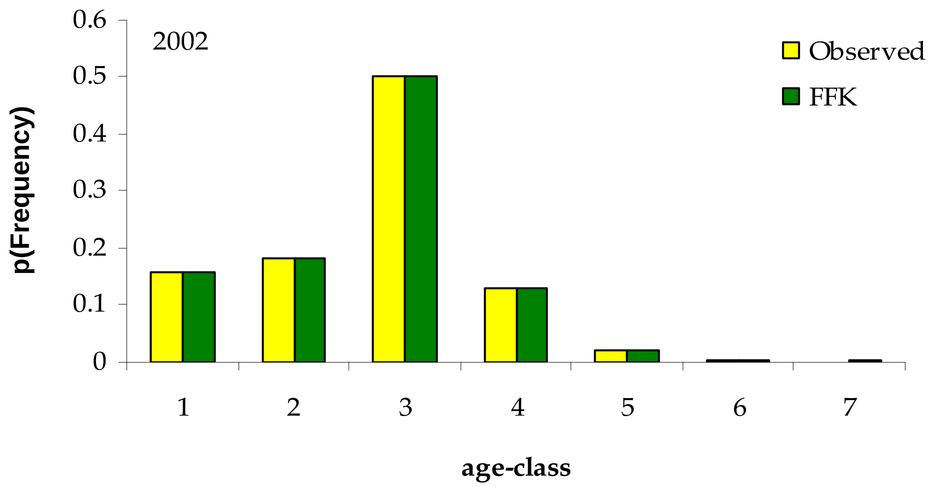


Figure 40. Age composition (2002) calculated using the FFK and the ALK method using the observed age composition length frequency data from the same year.

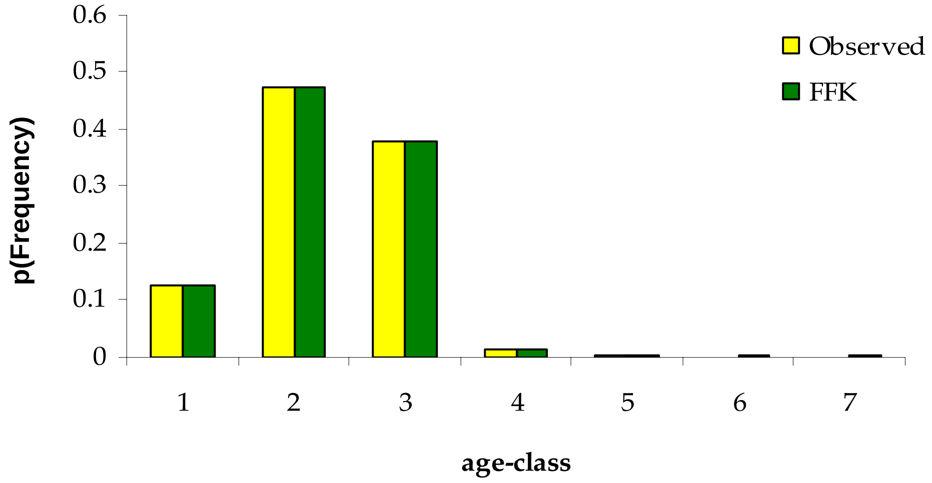


Figure 41. Age composition (2003) calculated using the FFK and the ALK method using the observed age composition length frequency data from the same year.

Table 26. Observed and predicted age compositions from years 1996-1998 using the 1995 length-at-age matrix and observed and predicted age compositions for years 2002 and 2003 using the 2001 length-at-age matrix.
Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. The measure of total error is the quadratic error between observed and predicted age composition.

Using 1995 length-at-age matrix						Using 2001 length-at-age matrix			
1996		1997		1998		2002		2003	
Observed	FFK	Observed	FFK	FFK	Observed	Observed	FFK	Observed	FFK
0.01	0.03	0.01	0.04	0.05	0.04	0.16	0.26	0.13	0.28
0.21	0.07	0.19	0.07	0.11	0.22	0.18	0.12	0.47	0.57
0.46	0.73	0.43	0.48	0.49	0.29	0.50	0.48	0.38	0.14
0.24	0.06	0.22	0.26	0.16	0.24	0.13	0.14	0.01	0.00
0.05	0.05	0.13	0.11	0.10	0.11	0.02	0.00	0.01	0.00
0.02	0.04	0.02	0.03	0.06	0.07	0.00	0.00	0.00	0.00
0.00	0.02	0.00	0.02	0.03	0.02	0.00	0.00	0.00	0.00
Q-error(%)	5.068		2.014		3.436		1.807		4.361

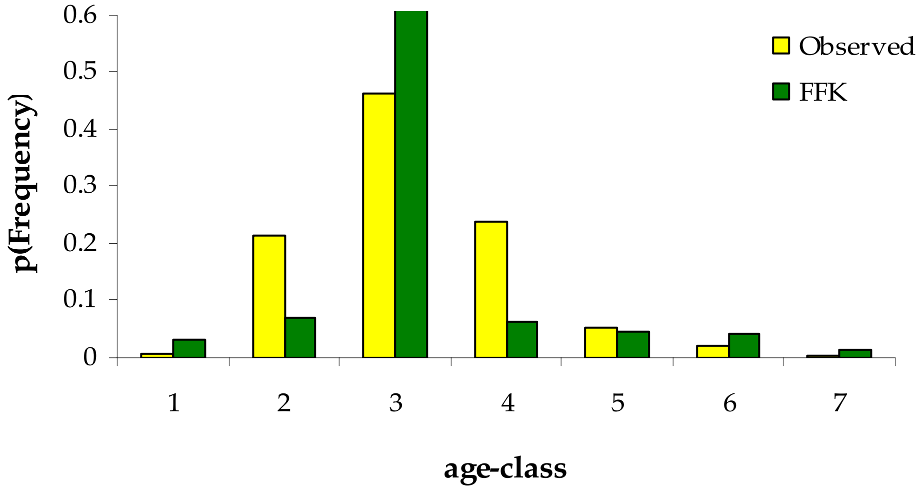


Figure 42. Age reconstruction for 1996 using the Bayes corrected 1995 length-at-age matrix.

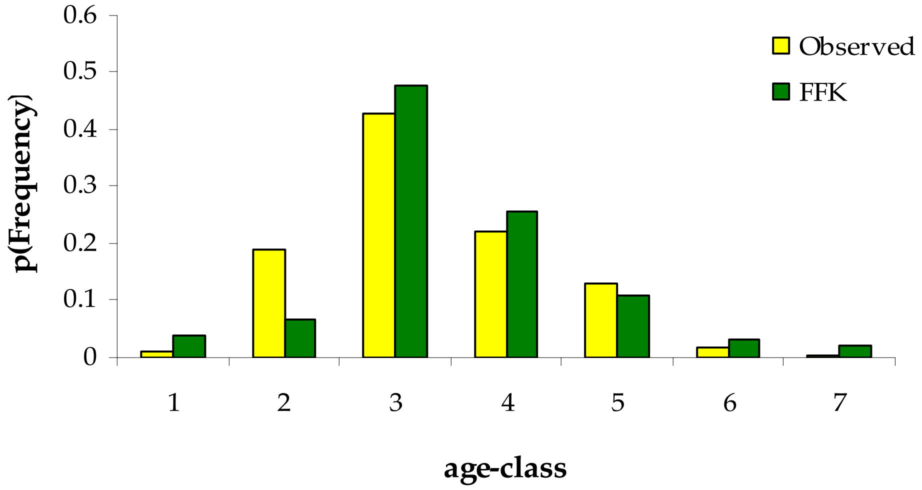


Figure 43. Age reconstruction for 1997 using the Bayes corrected 1995 length-at-age matrix.

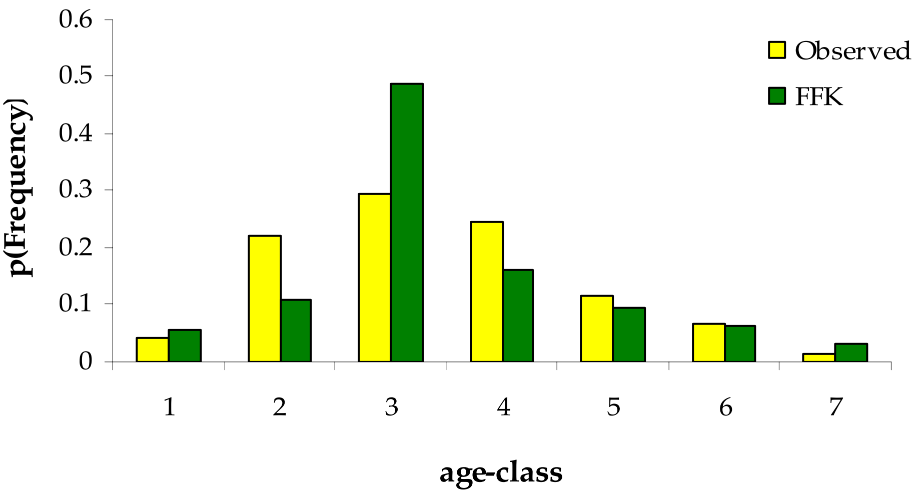


Figure 44. Age reconstruction for 2002 using the Bayes corrected 2001 length-at-age matrix.

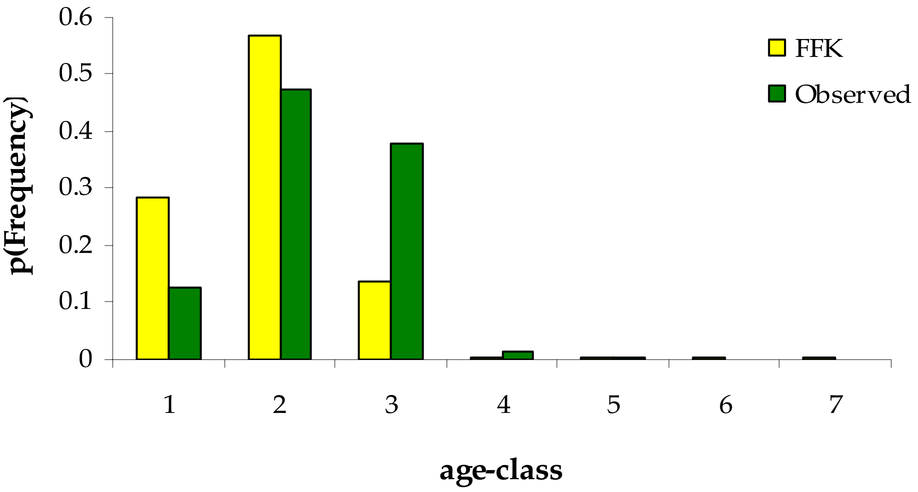


Figure 45. Age reconstruction for 2003 using the Bayes corrected 2001 length-at-age matrix.

Reduction of the length resolution of the matrices (increase from one to two centimetre length resolution) effectively reduced the dimensionality of the reconstruction problem. Using the length-at-age matrix from 1995 at a length resolution of two centimetres, the quadratic error was reduced in the three years from which an age composition was calculated (1996, 1997 and 1998). The results from these reconstructions are shown in Table 27.

Table 27. Observed and predicted age compositions from years 1996-1998 using the 1995 length-at-age matrix at a length resolution of two centimetres.

Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. Total error is the quadratic error between observed and predicted age composition.

Using 1995 length-at-age matrix at length resolution = 2cm

1996		1997		1998	
Observed	**FFK**	**Observed**	**FFK**	**FFK**	**Observed**
0.01	0.03	0.01	0.02	0.04	0.04
0.21	0.08	0.19	0.10	0.22	0.15
0.46	0.73	0.43	0.45	0.29	0.43
0.24	0.07	0.22	0.27	0.24	0.20
0.05	0.04	0.13	0.11	0.11	0.06
0.02	0.04	0.02	0.03	0.07	0.06
0.00	0.03	0.00	0.03	0.02	0.06
Q-error(%)	4.931		1.522		2.473

The same pattern of error in age-class assignment from the trials using a one centimetre length resolution was evident in the results obtained using the two centimetre resolution. The majority of samples were assigned to the modal age-class of the underlying age composition. The differences in the quadratic error between the one and two centimetre resolution trials were also consistent. The error reduced in all years by reducing the size of the length-at-age matrix. The maximum reduction in quadratic error was in the year with the lowest error in the one centimetre length resolution test.

The distributions of the age frequencies are shown in Figure 46 through Figure 48. Where the error was highest (age composition from the 1996 length frequency), the pattern of age-class assignments were similar (Figure 42 and Figure 46) for both length resolutions.

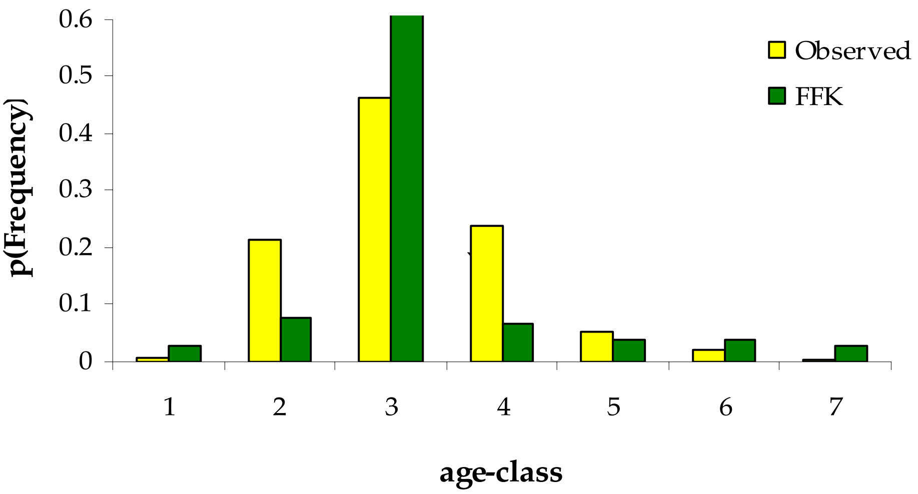


Figure 46. Age composition (1996) calculated using the FFK and the ALK method using the observed age composition length frequency data from the 1995 at a length resolution of two centimetres.

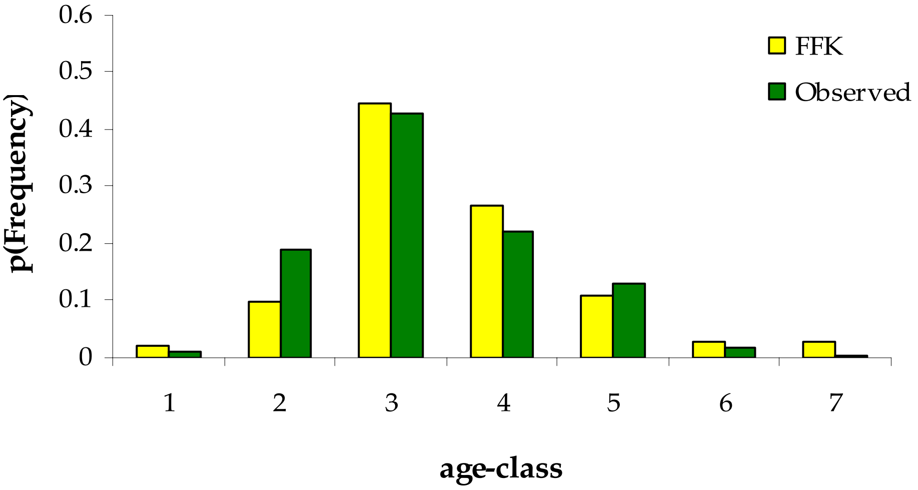


Figure 47. Age composition (1997) calculated using the FFK and the ALK method using the observed age composition length frequency data from the 1995 at a length resolution of two centimetres.

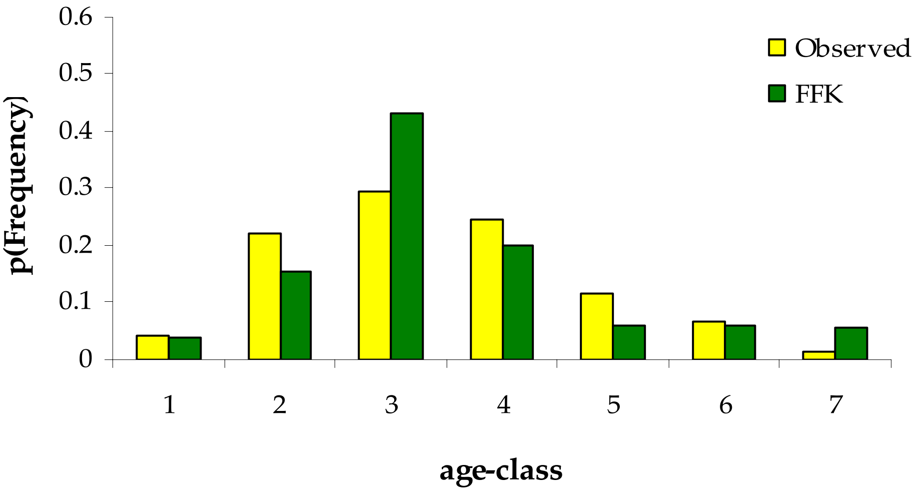


Figure 48. Age composition (1998) calculated using the FFK and the ALK method using the observed age composition length frequency data from the 1995 at a length resolution of two centimetres.

Testing the FFK method using multiple years

Where the a priori matrix was constructed from multiple years of ageing data to determine the age composition of a single year the results were closer to the known age distributions from the length frequencies under test. Three combinations of two years' ageing data were trialled and one combination of four years of ageing data were tested. The combinations of years used to make the length-at-age matrix and the length frequency under test are shown in Table 28. These trials were at a length resolution of one centimetre.

Table 28. Multiple years of ageing data used to determine the age composition from a length frequency.

Multiple year ageing data Length resolution = 1cm	Q-error(%)
1995, 1996->1997	1.79
1995 & 1996->1998	1.98
2001 & 2002->2003	3.56
1995, 1996,1997 & 1998->2001	3.74

These results were generally closer to the underlying age composition from the target length frequencies than the results of the previous trials. The lowest error was from using two years of ageing data to reconstruct the age composition from the following year (1995, 1996->1997). Where four years of ageing data were used in the length-at-age matrix the error was the highest (3.74%). The distributions of the age frequencies are shown in Figure 49 through Figure 52.

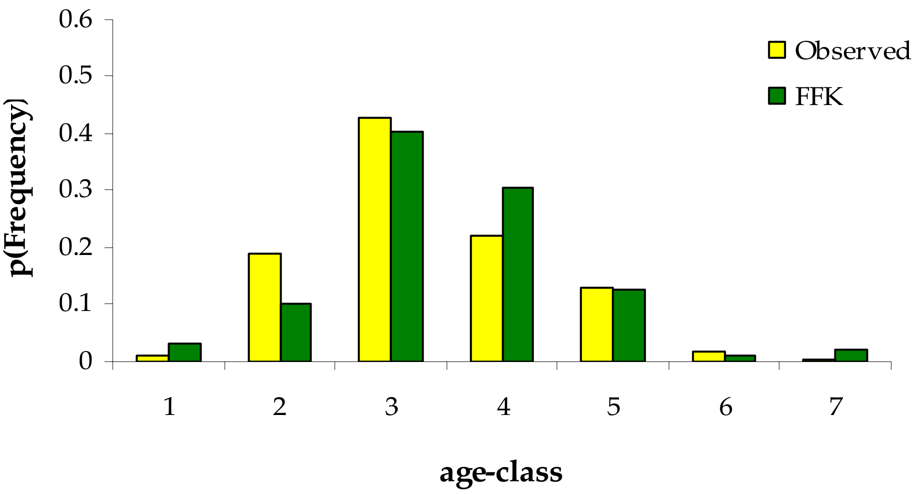


Figure 49. Age composition (1997) calculated using the FFK and the ALK method using the observed age composition length frequency data from the 1995 and 1996 length-at-age matrix.

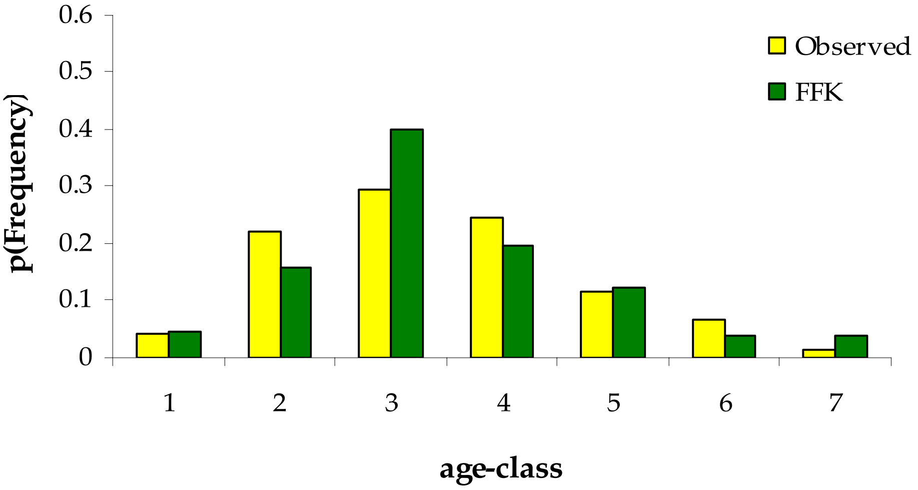


Figure 50. Age composition (1998) calculated using the FFK and the ALK method using the observed age composition length frequency data from the 1995 and 1996 length-at-age matrix.

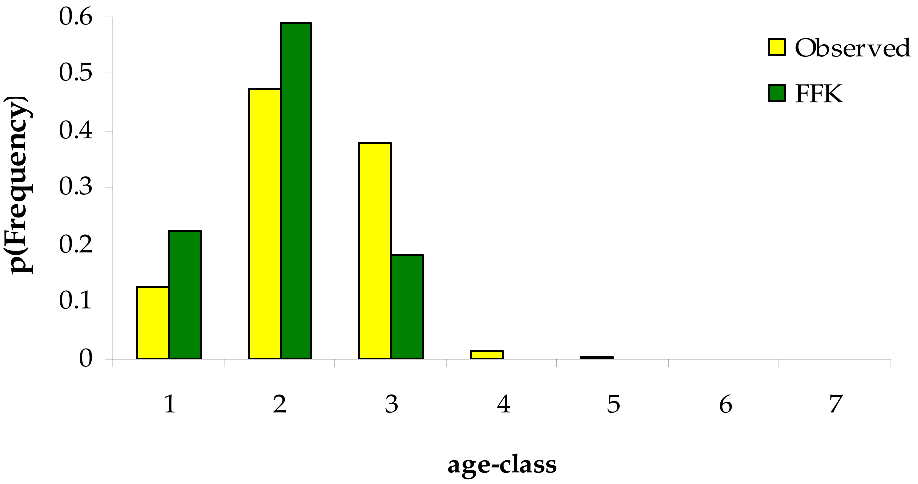


Figure 51. Age composition (2003) calculated using the FFK and the ALK method using the observed age composition length frequency data from the 2001 and 2002 length-at-age matrix.

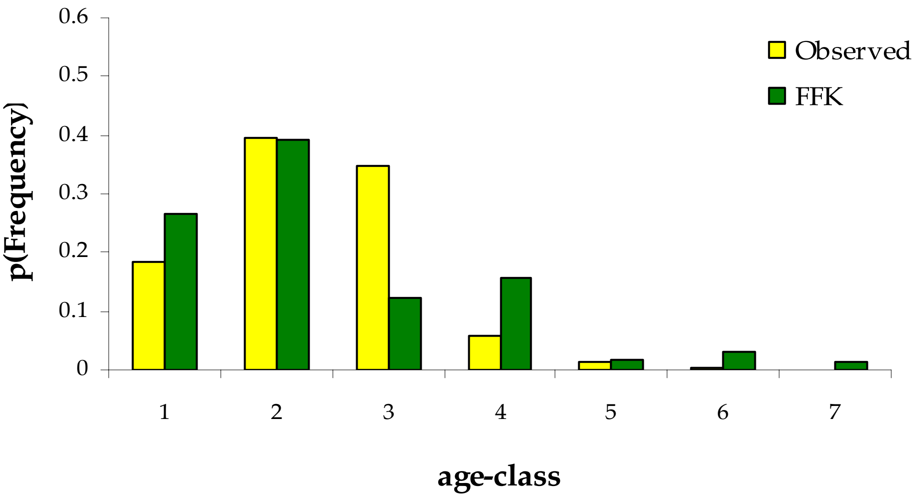


Figure 52. Age composition (2001) calculated using the FFK and the ALK method using the observed age composition length frequency data from the 1995, 1996, 1997 and 1998 length-at-age matrix.

Discussion

The application of the FFK method to real fisheries data in the framework that would be used by managers produced mixed results. Where the age data were sub-sampled from the larger length frequency set from the same year, the FFK method produced results that are equivalent to those obtained using the age-length key method. This is the same data environment as is currently used in fisheries management. This demonstrates that the results from the FFK method did converge to the underlying age composition.

Where the length-at-age matrix was used from a different year that the length frequency was taken, the results were less successful. Some length-at-age matrices produced results that were similar to those obtained using the ALK method (for example the 1998 age composition from the combined 1995 and 1996 length-at age matrix). The majority of the reconstructions of age compositions were of low accuracy. The 1996 age composition from 1995 length-at-age matrix determined that most of the samples were three-year-olds, with low abundances for all other age-classes.

The method could not be demonstrated to work effectively under any combination of year(s) that were trialled to produce results which would be considered appropriate for fisheries management for this species. The failure of the technique in these trials may be driven by the violation of the assumption that the length-at-age are similar to the length-at-age distributions within the length frequency with an unknown age composition. This can be caused through two separate factors. These are;

- biological changes in the length-at-age distribution in School Whiting between sampling periods;
- artificial changes in length-at-age as a function of sampling periods.

Biological changes in length-at-age may be caused by a variety of conditions (for example density dependent growth) that would render the technique invalid. Species with large variability in length-at-age between years will not be suitable to the application of this technique.

Artificial changes in length-at-age are a function of sampling. For example, where fish are sampled for age in the early part of the year, the length-at-age distributions will be different for fish sampled for length frequency taken later in the year. As fish growth is continuous until some asymptotic length is achieved, the period where sampling occurs for ageing, should approximate the time periods where length frequencies are taken. This would minimise the chance of introducing changes in length-at-age through sampling effects.

Another sampling related error may be introduced through the application of a Bayes adjustment. The data used for the Bayes adjustment may be biased, as a consequence of the sampling regime. The Bayesian adjustment is necessary, as the sub-sample used for ageing is not random from the larger length frequency. Collecting a truly random sample from the larger length frequency sample is nearly an impossible task. If the sample was random from the larger length frequency sample, the adjustment from the larger length frequency would not be necessary. For this adjustment to be valid, the ageing sample must be a sub-sample from the length frequency sample.

The exact reasons for failure of the FFK method were probably a combination of these factors.

Chapter 5. Blue Grenadier trials

Introduction

An objective of this project was to demonstrate the FFK method decomposing length frequencies to the age composition for two species harvested from Australia's South East Fishery. The previous chapter unsuccessfully demonstrated the FFK method using catch-adjusted length frequency data collected from the ISMP program for School Whiting.

Blue Grenadiers live to approximately 25 years of age. This species is fished in Australia's south-east fishery (SEF). The Blue Grenadier component of this study tests the FFK method on a species which is relatively long lived, and has episodic recruitment.

Within the SEF, the two largest Blue Grenadier fisheries are the Western Bass Strait fishery, and the Tasmanian west coast fishery. The Tasmanian west coast fishery targets spawning aggregations between July and September and is the largest of the Blue Grenadier fisheries in the SEF. The western Bass Strait fishery is predominately a summery fishery. Smaller numbers of grenadier are fished from other areas within the SEF, including East Bass Strait and Tasmania's east coast. Ageing and length data from the western Bass Strait and Tasmanian west coast fisheries are available at the CAF.

The trials in this chapter test the FFK method to reconstruct the age composition in the same way that the ALK method is currently used. Catch adjustments were not made to the Blue Grenadier data. Only length frequencies from samples which have been aged in the CAF were used. The FFK method was tested on data obtained from the winter and summer fishery, using length-at-age distributions from different years to the length frequency. The age compositions determined by the FFK method were compared with age frequency determined by ageing.

Methods

Data

The Blue Grenadier dataset available at the CAF is extensive. The data were first collected in 1984 and have been routinely collected since 1992. The CAF has aged otoliths from 27,434 fish. The majority of these samples (23,503, 85.67%) were obtained from Western Bass Strait (WBS) and West Tasmania (WTAS) with the balance (n=3,931 or 14.33%) sourced from the East Australia (EASTA), Eastern Bass Strait (EBS) and East Tasmania (ETAS) fisheries (Table 29).

Table 29. Samples of Blue Grenadier available at the CAF to test the FFK method.

Zone	All years	% All years	Grouped
EASTA	505	1.84	
EBS	1528	5.57	
ETAS	1898	6.92	14.33
WBS	8467	30.86	
WTAS	15036	54.81	85.67
Total	**27434**	**100**	**100**

The numbers of samples collected by year are shown in Table 30. As the sample sizes from East A, EBS and East Tasmania were lower than those of samples collected in Western Bass Strait and West Tasmania, it was decided to apply the FFK method to the Western Bass Strait and West Tasmania length frequencies only.

Table 30. Numbers of Blue Grenadier collected and aged which are held at the CAF.

Year	Zone EASTA	EBS	ETAS	WBS	WTAS
1984	0	0	0	0	511
1985	0	0	0	52	334
1986	0	0	0	0	174
1991	0	0	100	0	93
1992	0	0	219	487	481
1993	84	0	0	325	953
1994	0	284	0	196	938
1995	0	26	0	259	1251
1996	0	0	72	574	1186
1997	0	427	0	1186	280
1998	61	94	99	1155	779
1999	35	30	488	1295	1999
2000	27	92	434	792	1109
2001	31	567	97	196	1062
2002	12	0	240	358	1077
2003	31	0	49	404	1030
2004	23	0	0	412	1049
2005	201	8	100	776	730
Total	**505**	**1528**	**1898**	**8467**	**15036**

The numbers of samples were not collected consistently until 1998 for the Western Bass Strait sample (Table 30). Length frequency and length-at-age distribution information from the Western Bass Strait collected after 1997 were used in this study.

The progression of the stronger year classes also becomes obvious from 1998 in both the summer and the winter fishery (Table 31). Eight years of length frequency and length-at-age data, from 1998 to 2005, from both areas were used in this study.

Although the maximum age of Blue Grenadier is about 25 years, the majority of the samples are 15 years of age or below. For this reason the samples that were 15 years or over were included in a 15+ year group. Samples which were 0+ were only collected in 2005 from Western Tasmania. These samples were excluded from the analyses.

Age composition from Western Bass Strait and Western Tasmania are shown in Table 31 and Table 32 respectively. All samples were aged at the CAF using standard techniques using thin section otoliths under transmitted light and the ageing information stored in MS Excel (see Morison 1998 for a full description).

Each year, the ISMP aimed to collect 800 and1200 otolith samples for the Western Bass Strait summer fishery and the West Tasmanian winter fishery respectively. These sample numbers were determined for the stock assessment process.

Table 31. Age composition data for Western Bass Strait collected between 1998 and 2005, n=5388.

Age-class	Year 1998	1999	2000	2001	2002	2003	2004	2005
0	0	0	0	0	0	0	115	0
1	213	27	0	0	0	0	3	60
2	185	345	1	0	0	0	1	198
3	208	78	70	4	1	0	0	1
4	288	183	153	10	2	0	0	0
5	73	352	160	16	10	2	0	0
6	27	93	317	81	32	11	5	0
7	30	11	52	75	98	50	14	3
8	17	17	2	8	77	102	28	13
9	31	37	6	0	11	97	92	49
10	42	42	3	0	2	30	77	214
11	13	19	2	0	2	7	18	95
12	6	19	5	0	2	4	1	12
13	3	21	5	0	4	5	6	4
14	4	9	9	1	13	8	4	8
15	15	42	7	1	104	88	48	119
Total	**1155**	**1295**	**792**	**196**	**358**	**404**	**412**	**776**

Table 32. Age composition data for West Tasmania collected between 1998 and 2005, n=8835.

Age-class	Year 1998	1999	2000	2001	2002	2003	2004	2005
0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0
2	6	1	0	0	0	0	0	150
3	132	7	4	5	1	6	1	3
4	201	31	12	12	5	0	0	0
5	37	69	125	44	24	0	2	0
6	5	75	416	262	94	7	2	2
7	7	57	136	461	358	61	13	0
8	12	50	20	43	397	247	81	8
9	19	105	15	10	34	403	291	38
10	46	221	19	10	7	102	476	160
11	102	279	15	6	7	9	58	277
12	84	323	38	10	6	12	6	42
13	46	313	73	32	8	11	3	5
14	41	178	89	59	16	17	10	7
15	41	290	147	108	120	155	105	38
Total	**779**	**1999**	**1109**	**1062**	**1077**	**1030**	**1049**	**730**

The maximum length of the Blue Grenadier obtained from the Western Tasmania winter fishery was 121 cm (standard length). The largest fish taken in the Western Bass strait fishery was 114 cm. To reduce the number of length classes, all samples over 110 cm were assigned to the 110+ cm group. This accounted for 220 individuals (214 samples from West Tasmania fishery and six samples from the Western Bass Strait summer fishery). The length distributions for each fishery combined by years are shown in Figure 53 for Western Tasmania and Figure 54 for Western Bass Strait.

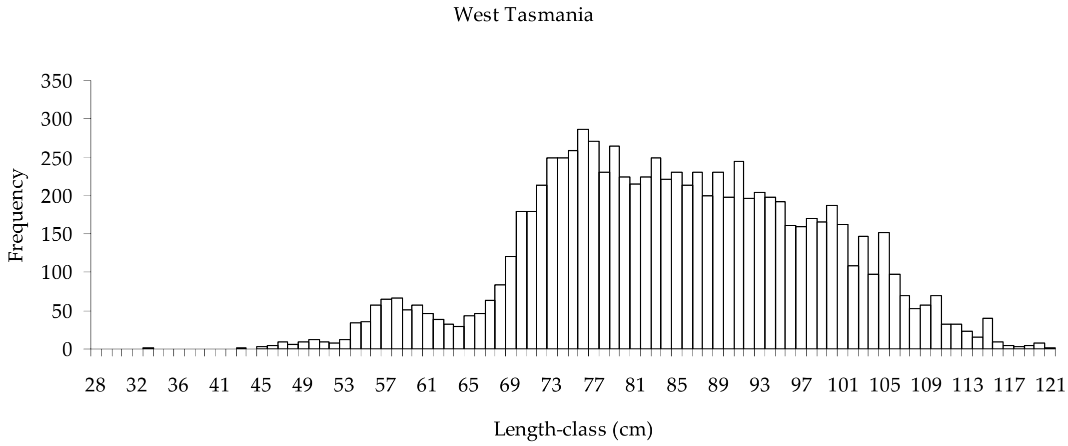


Figure 53. Length frequency for the combined Blue Grenadier data from 1998 to 2005 from West Tasmania, n=8,835.

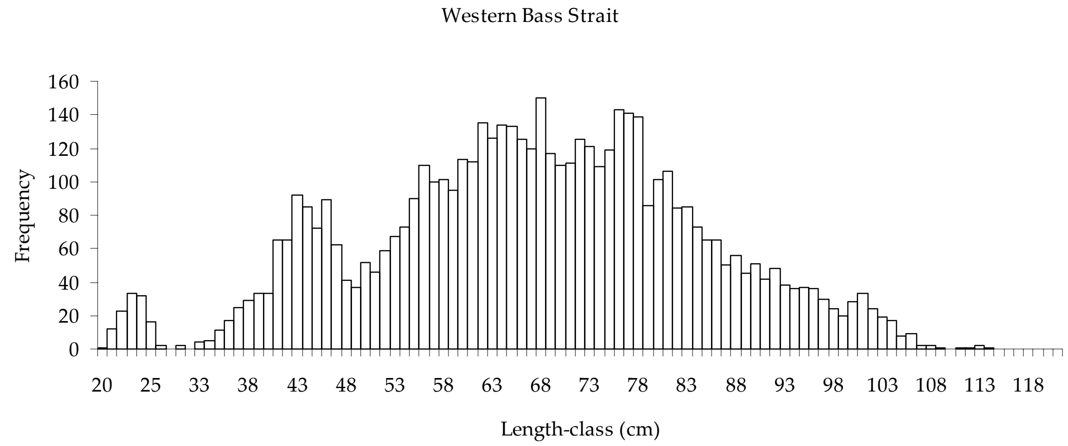


Figure 54. Length frequency for the combined Blue Grenadier data from 1998 to 2005 from Western Bass Strait, n=5,388.

Data preparation

The data preparation for both the West Tasmania and the Western Bass Strait datasets was the same. Many of the sample years from both areas contained missing year-classes. The length-at-age matrix must contain the length distribution from all age-classes. To obviate this problem, the length-at-age distribution from 1996 for each area was added to the length-at-age matrix for each year. These two matrices (one from each area) contain representation of all age-classes. Only one year contained 0+ individuals (Western Bass Strait, n=115, Table 31), the length-at-age distribution from this age-class was also added to each matrix. The age composition from the length-at-age data that was added to the base years to provide length-at-age distribution adapted for all age-classes in shown in Table 33. With the addition of these data, all matrices from 1998 to 2005 for both areas contained information on the distribution of length-at-age. The length-at-age distributions were Bayes-corrected by the combined length-frequency information. Length-at-age distributions were also adjusted by the weighted average of the sample size for each cell in the array.

Table 33. Age composition from 1996 length-at-age data added to the base year. * denotes 0+ samples added from the 2004 Western Bass Strait aged sample.

	Area	
Age-class	**West Tasmania**	**Western Bass Strait**
0	115*	115*
1	2	49
2	1	112
3	21	9
4	39	23
5	69	50
6	37	24
7	22	7
8	138	21
9	306	100
10	221	69
11	107	40
12	66	25
13	35	9
14	15	7
15	107	29
Total	1186	574

The size of the length-at-age matrix for each year for both areas was a[16,91]. This was the 16 age-class (0-15 years) and 91 length-classes (20-110 cm).

One additional length-at-age matrix was constructed using the length-at-age data from 1998 (Western Bass Strait data). This matrix was used to test the FFK method within the same data environment as the ALK would be used. The size of this length-at-age matrix was a[16,91], as 0+ samples were included in the analysis.

The age frequency from the aged samples for each year provided the observed age for each year. The use of the ALK was not necessary as the target age compositions were all aged samples, not a result of applying the ALK to a length frequency.

Trials

Trials within the same data environment were tested. These included an initial test using the 1998 length-at-age matrix to decompose the 1998 length frequency into the underlying age composition. A further sixteen trials were performed using the modified length-at-age matrices to determine the age composition.

Further trials were conducted but not reported as the FFK method failed to produce results which approximated the underlying age composition from any of these trials.

Table 34. Trials conducted using the FFK method for Blue Grenadier

Western Bass Strait		West Tasmania	
Matrix	Length frequency	Matrix	Length frequency
1998	1998		
1996, 1998	1998	1996, 1998	1998
1996, 1999	1999	1996, 1999	1999
1996, 2000	2000	1996, 2000	2000
1996, 2001	2001	1996, 2001	2001
1996, 2002	2002	1996, 2002	2002
1996, 2003	2003	1996, 2003	2003
1996, 2004	2004	1996, 2004	2004
998	1999		

Results

Figure 55 demonstrates the effectiveness of the FFK method used in the same data environment as the ALK is used. The results show close agreement between the observed abundance and the abundance calculated using the FFK method.

Figure 56 through to Figure 63 demonstrate the FFK method decomposing the length frequency from the same year for Western Bass Strait years 1998-2005 respectively.

Figure 64 through Figure 72 demonstrate the FFK method applied to length frequency data from Tasmania's west coast fishery.

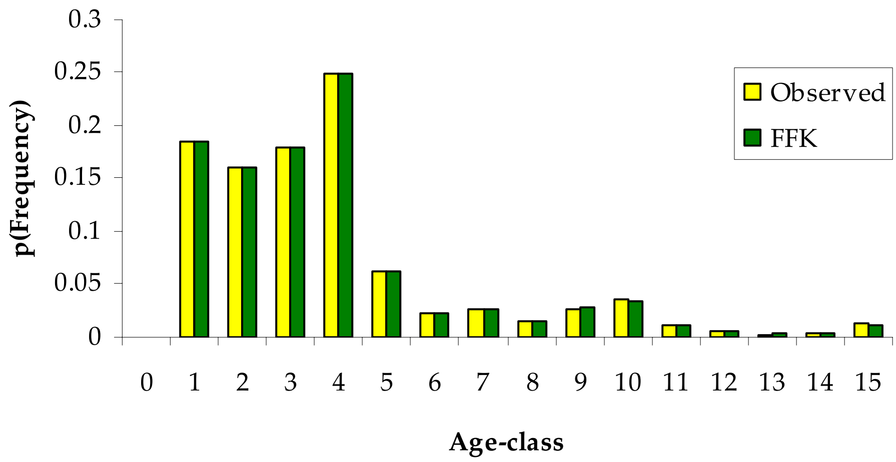


Figure 55. Age composition calculated from the 1998 Western Bass Strait length frequency using the length-at-age distributions from the same year.

Using the length-at-age distributions from each age-class, the FFK method reproduced the observed relative abundance of each age-class (Figure 55). The FFK method uses no information on the abundance of each age-class, only the distributions of length-at-age from the same year. This is akin to using the FFK method in the same data environment as the ALK would normally be used. The difference is that the ALK would normally be multiplied by a length frequency. The addition of the 0+ length-at-age distribution had no effect on the effectiveness of the FFK to determine the age composition from the length frequency.

With the addition of the 1996 length-at-age data, the results depart from the observed age composition.

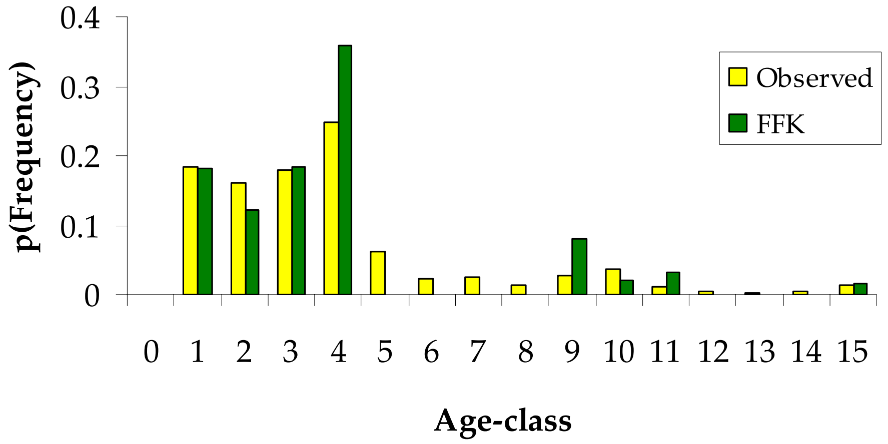


Figure 56. Reconstruction of the 1998 Western Bass Strait age composition using the 1998 and 1996 length-at-age matrix, with the addition of the length distribution from 0+ fish.

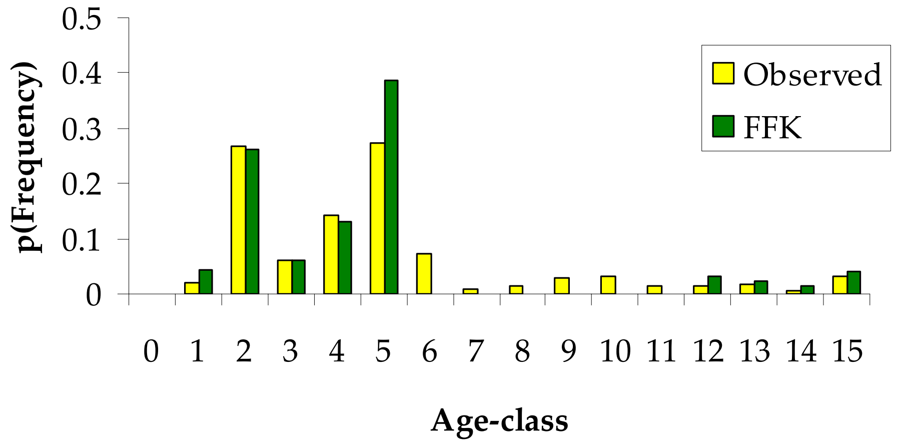


Figure 57. Reconstruction of the 1999 Western Bass Strait age composition using the 1999 and 1996 length-at-age matrix, with the addition of the length distribution from 0+ fish.

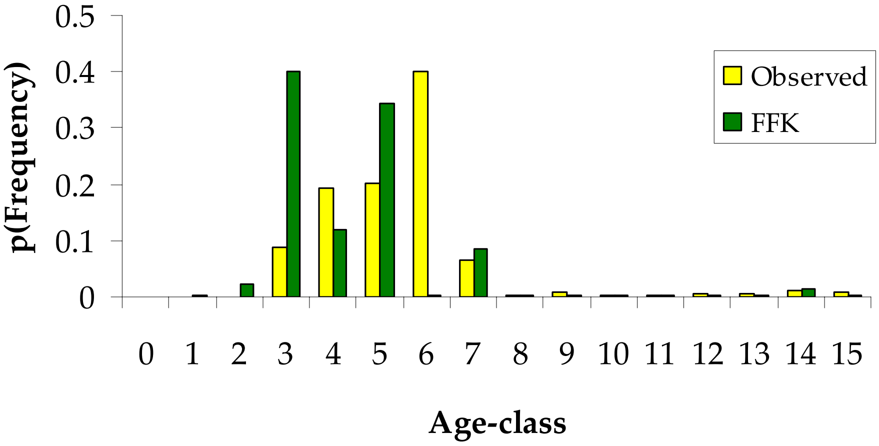


Figure 58. Reconstruction of the 2000 Western Bass Strait age composition using the 2000 and 1996 length-at-age matrix, with the addition of the length distribution from 0+ fish.

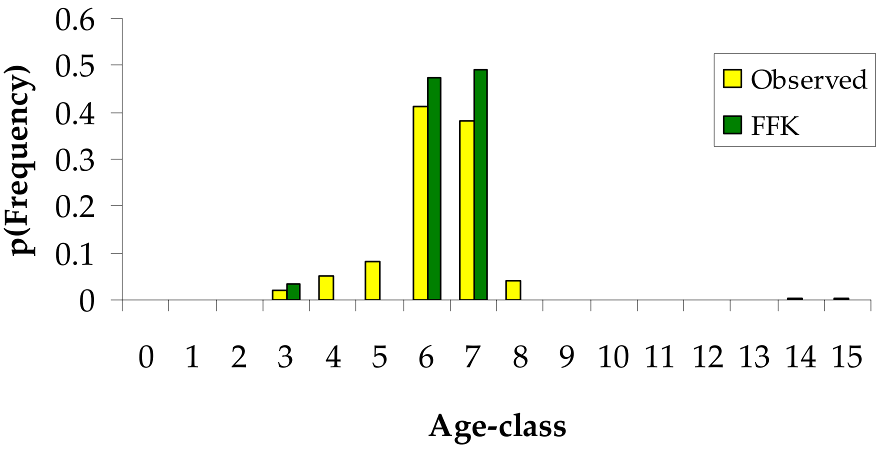


Figure 59. Reconstruction of the 2001 Western Bass Strait age composition using the 2001 and 1996 length-at-age matrix, with the addition of the length distribution from 0+ fish.

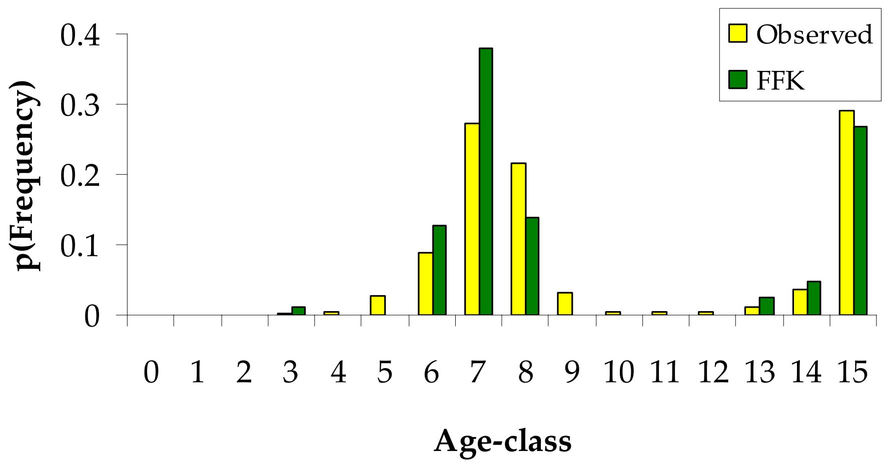


Figure 60. Reconstruction of the 2002 Western Bass Strait age composition using the 2002 and 1996 length-at-age matrix, with the addition of the length distribution from 0+ fish.

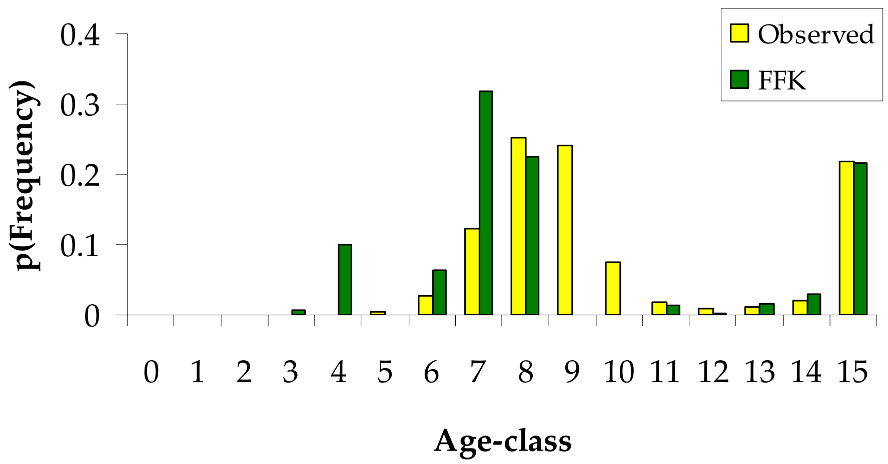


Figure 61. Reconstruction of the 2003 Western Bass Strait age composition using the 2003 and 1996 length-at-age matrix, with the addition of the length distribution from 0+ fish.

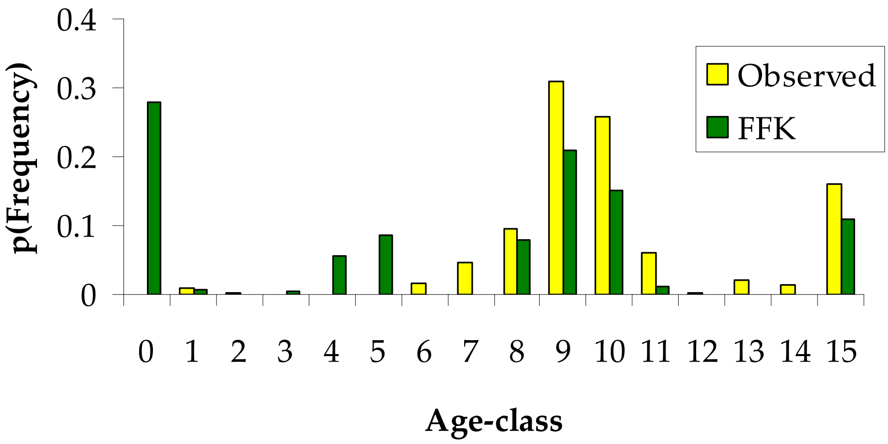


Figure 62. Reconstruction of the 2004 Western Bass Strait age composition using the 2004 and 1996 length-at-age matrix, with the addition of the length distribution from 0+ fish.

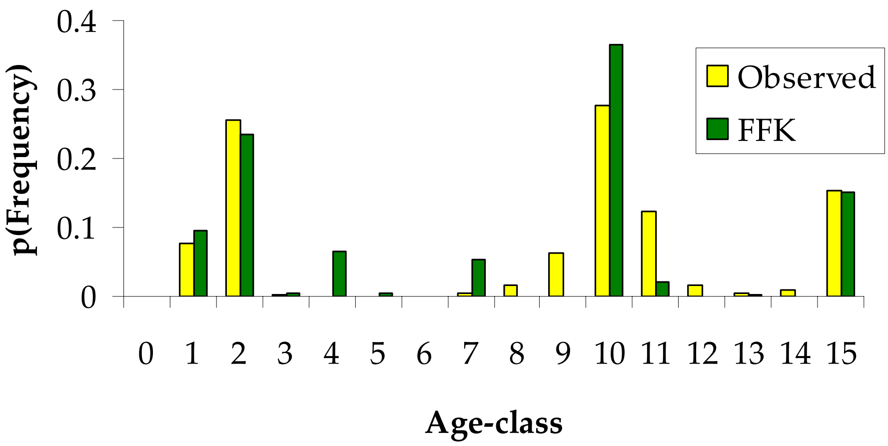


Figure 63. Reconstruction of the 2005 Western Bass Strait age composition using the 2005 and 1996 length-at-age matrix, with the addition of the length distribution from 0+ fish.

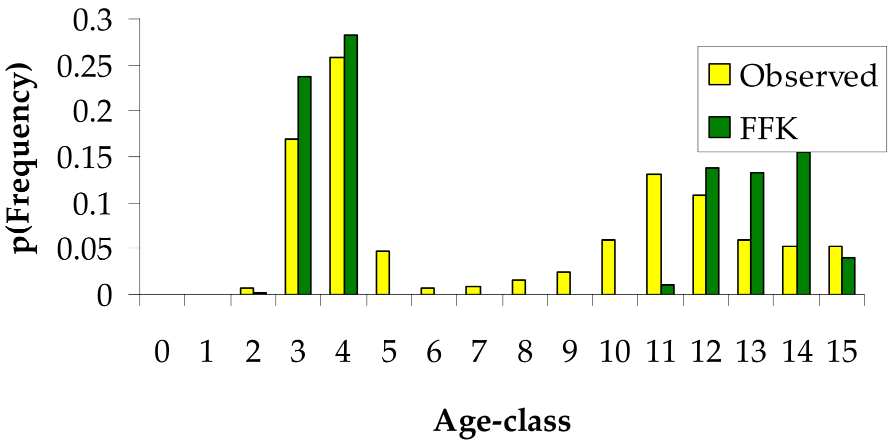


Figure 64. Reconstruction of the 1998 West Tasmania age composition using the 1998 and 1996 length-at-age matrix from West Tasmania, with the addition of the length distribution from 0+ fish.

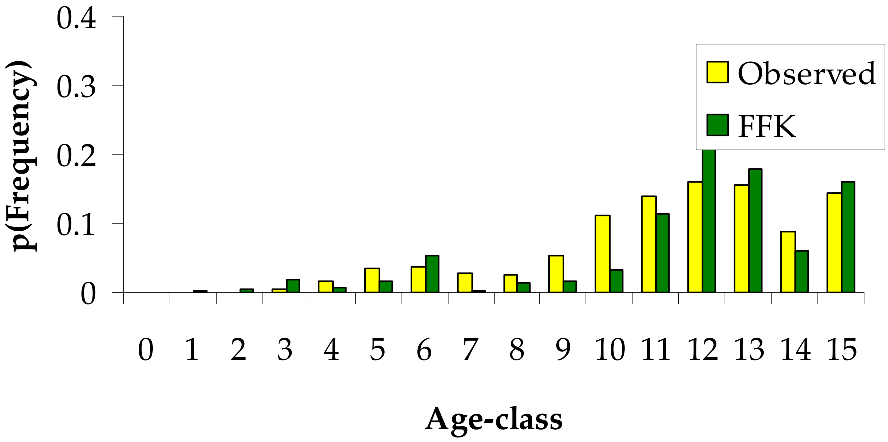


Figure 65. Reconstruction of the 1999 West Tasmania age composition using the 1999 and 1996 length-at-age matrix from West Tasmania, with the addition of the length distribution from 0+ fish.

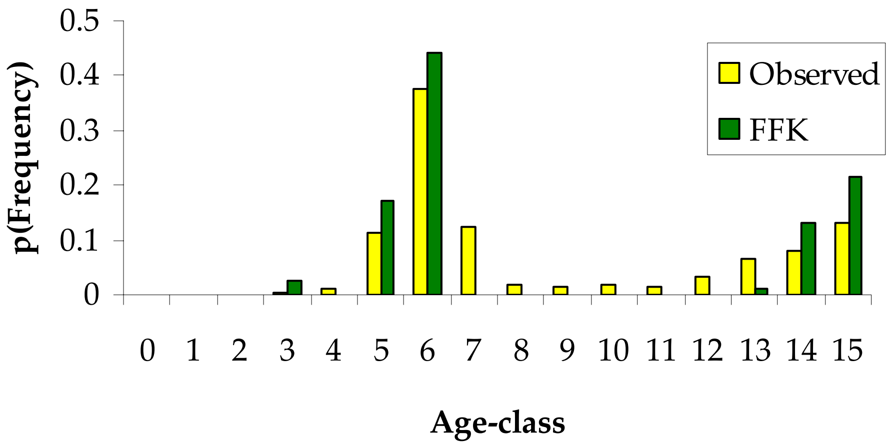


Figure 66. Reconstruction of the 2000 West Tasmania age composition using the 2000 and 1996 length-at-age matrix from West Tasmania, with the addition of the length distribution from 0+ fish.

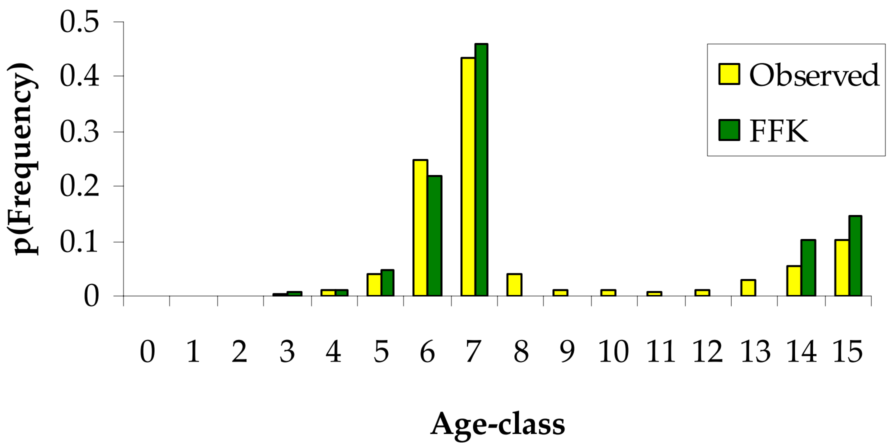


Figure 67. Reconstruction of the 2001 West Tasmania age composition using the 2001 and 1996 length-at-age matrix from West Tasmania, with the addition of the length distribution from 0+ fish.

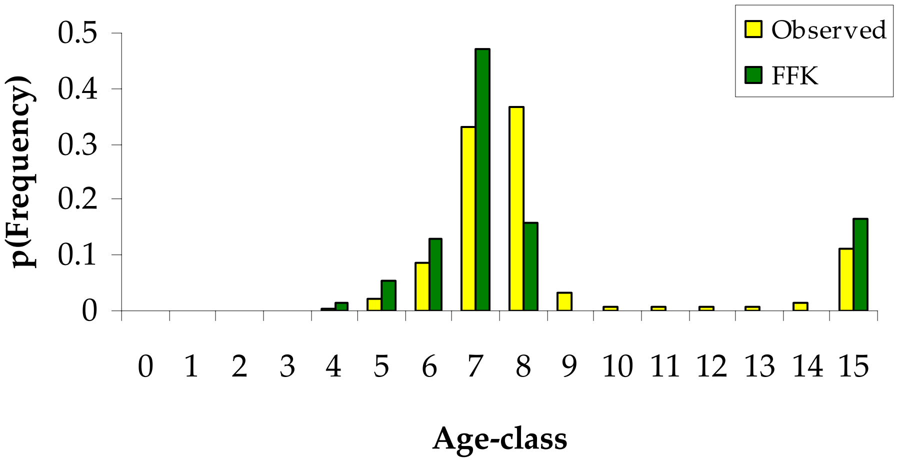


Figure 68. Reconstruction of the 2002 West Tasmania age composition using the 2002 and 1996 length-at-age matrix from West Tasmania, with the addition of the length distribution from 0+ fish.

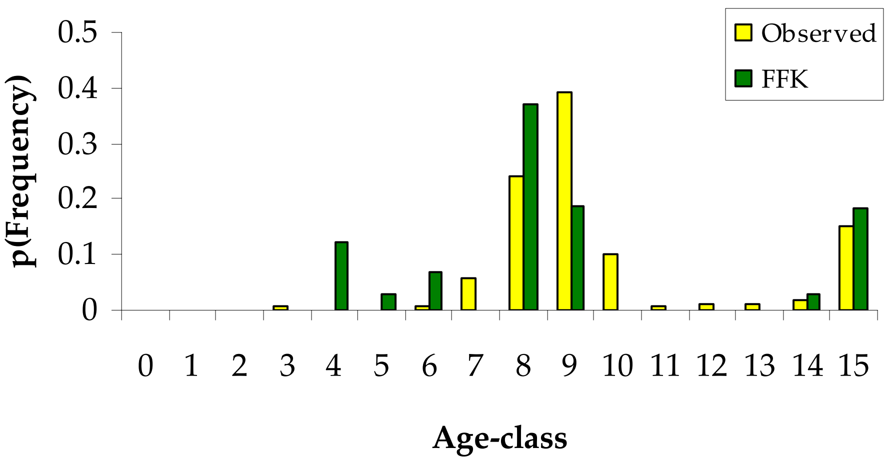


Figure 69. Reconstruction of the 2003 West Tasmania age composition using the 2003 and 1996 length-at-age matrix from West Tasmania, with the addition of the length distribution from 0+ fish.

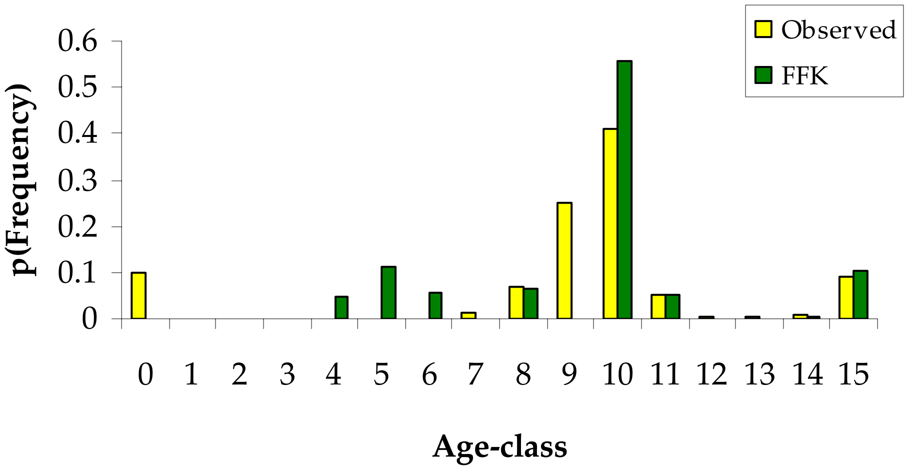


Figure 70. Reconstruction of the 2004 West Tasmania age composition using the 2004 and 1996 length-at-age matrix from West Tasmania, with the addition of the length distribution from 0+ fish.

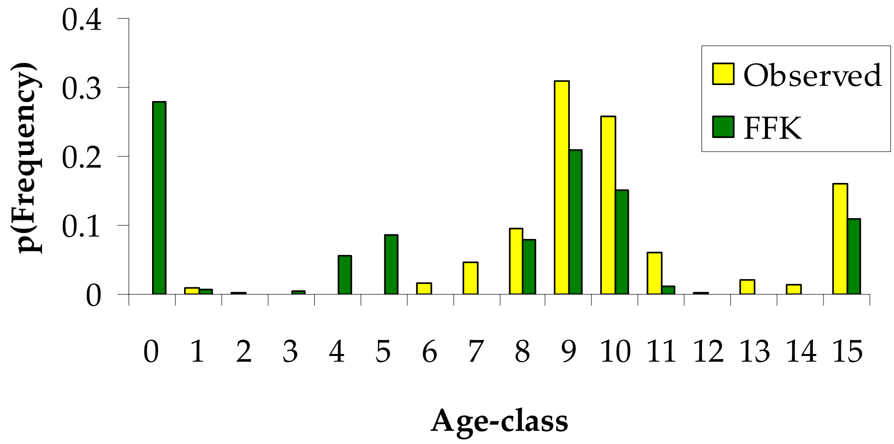


Figure 71. Reconstruction of the 2005 West Tasmania age composition using the 2005 and 1996 length-at-age matrix from West Tasmania, with the addition of the length distribution from 0+ fish.

The reconstruction of the age composition from the 1999 length frequency data using the length-at-age matrix from 1998 failed to approximate the underlying age composition (Figure 72). The model still determined the majority of the ages were between one and five-years of age, however the relative abundance of each age-class within this group was not accurate.

The quadratic errors were relatively high for these trials reflecting the failure to reconstruct the observed age composition from the length frequencies (Table 35). Before the addition of the 1996 length-at-age matrix, the FFK method produced results closely approximating the known age compositon. After the additon of the 1996 length-at-age matrix, the reconstruction of the age compositon from the length frequencies was poor.

Further trials were conducted but not reported in this document. This was because the the FFK method failed to produce results which approximated the underlying age composition from any of these trials.

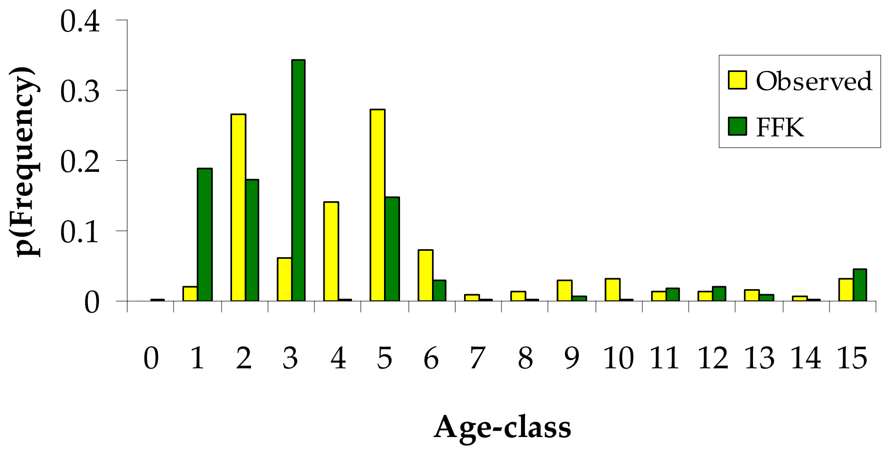


Figure 72. Reconstruction of the 1999 Western Bass Strait grenadier age composition using the 1998 length-at-age matrix, with the addition of the length distribution from 0+ fish.

Table 35. Quadratic errors for age composition reconstruction from length frequencies for Blue Grenadier from Western Bass Strait and West Tasmania fisheries.

Western Bass Strait			**West Tasmania**		
Matrix	**Length frequency**	**Q-Error**	**Matrix**	**Length frequency**	**Q-Error**
1998	1998	0.014			
1996, 1998	1998	0.932	1996, 1998	1998	1.303
1996, 1999	1999	0.920	1996, 1999	1999	1.190
1996, 2000	2000	3.326	1996, 2000	2000	1.216
1996, 2001	2001	1.015	1996, 2001	2001	0.565
1996, 2002	2002	0.911	1996, 2002	2002	1.666
1996, 2003	2003	2.101	1996, 2003	2003	1.912
1996, 2004	2004	2.155	1996, 2004	2004	2.079
1996, 2005	2005	1.094	1996, 2005	2005	2.155
1998	1999	2.458			

Discussion

The FFK method failed to reproduce accurate age composition from the length frequencies of Blue Grenadier from two SEF fisheries. The addition of all the length-at-age distributions which may be seen within the population (age-classes 0 to 15+) was necessary to cover the range of all ages as the unknown age composition of a length frequency may contain any of these classes at an unknown frequency. The addition of the length-at-age distributions from the 0+ age class did not have any effect on the model to determine the underlying age composition when only these data were applied to a matrix.

The performance of the FFK method with respect to calculating the observed age-classes within target length frequencies was similar for both summer and winter fisheries. The summery fishery samples were collected over a longer time period than the West Tasmanian winter sample (collected over a three month period). The distributions of length-at-age collected over a shorter period of time would have less protracted distributions for length-at-age, especially for the younger age-classes. It was therefore thought that this shorter span of collection may improve the ability to predict age composition from length frequencies. This was not evident; however, the lowest quadratic error was recorded from the winter fishery.

Although some of these results suggest that the FFK method can generally approximate some age-class frequencies within the length frequencies, the error was unacceptably high. This pattern was also seen with silver and blue warehou (not reported).

One additional dataset was trialled. These trials were on datasets collected under sampling regimes considered ideal for the application of the FFK method. This dataset was Atlantic cod collected by French researchers.

Chapter 6. The French connection

Introduction

After obtaining inconsistent results from the application of the FFK method to length data collected from the School Whiting and the Blue Grenadier fisheries, other data were sourced to test the FFK method.

Data were sought that were derived from a sampling strategy where the ageing sample was a sub-sample of the length frequency sample. While this is the aim of the data collection in Australia's SEF, this study demonstrates that either the ageing sample was not a sub-sample of the length frequency sample, or in these fisheries, length-at-age is not temporally stable.

Other data sets were sourced from overseas which conformed to the fundamental requirement that the ageing sample was a sub-sample from the length frequency sample. Data collected using this rigorous protocol was considered to be more appropriate to investigate the FFK method in the context of fisheries management. Contact was made with three principal researchers from three agencies, these were;

1. Dr. Peter Hagen, NOAA fisheries, Juneua, Alaska.
2. Dr Daniel Kimura, NOAA Fisheries, Seattle, Washington.
3. Dr Ronan Fablet, IFREMER, Plouzané, France.

Each of these researchers collect, or have access to fisheries data used for long-term stock assessments (length frequency and ageing data) of commercially important fisheries. The aim of the contact was to ask for access to data with a view to trialling the FFK method.

The data requested must satisfy the following criteria:

1. Consisted of a number of years of data for both ageing and length frequency.
2. Demonstrated variable age compositions between years.
3. The sample, which was used for ageing, was a sub-sample taken from the length frequency sample.

It was considered that the French data supplied by Dr Ronan Fablet and Dr Robert Bellail from IFREMER was close to ideal and provided a robust data framework to test the FFK method. Further, Dr Fablet and Dr Bellail responded quickly to the data request which made the trials with their data possible.

Data from two species, Atlantic cod and blue whiting, were supplied. Both species were sampled over a five-year (2001-2005) period for the purposes of routine ageing and stock assessment. It was decided to test the FFK method on one species – Atlantic cod (*Gadus morhua*).

Method

Data description

The cod were sampled by demersal trawlers in the Atlantic Ocean. Further catch location information for cod is unknown. One trip was sampled each month. The duration of each fishing trip was between 7 and 12 days. The catch was sampled from commercial fish markets at Lorient or Concarneau. Each catch was sorted into four size classes:

1. Category 1: very large fish.
2. Category 2: large fish.
3. Category 3: medium fish.
4. Category 4: small fish.

Each month fish were sampled according to a stratified sampling plan from the commercial fishing trips. The number of fish sampled with respect to the number of fish measured varies depending on the size category. More samples are taken from larger fish, with decreasing frequency of collection of ageing samples from successively smaller length categories.

Fish were stored in bins (equivalent to market fish bins in Australia) which have weights of between 45-60 kilograms. One box from each category was sampled according to the sampling plan. The sampling regime for otolith removal and fish measurement is detailed in Table 36.

Table 36. Stratified otolith sampling for cod and blue whiting

Category	Otolith sampled per fish measured per 1 cm length-class
1	1 otolith sample per 2 fish measured.
2	1 otolith sample per 3 fish measured.
3	1 otolith sample per 5 fish measured.
4	1 otolith sample per 5 fish measured.

For category 1, 50% of the fish were processed for otolith removal and 100% were measured while for category 4, 20% of the fish were processed for ageing and 100% of the cod were measured at each length-class. The range of lengths in the samples across all years ranged from 33 – 116 cm providing 86 length classes.

The length frequencies were adjusted by the ratio of the sample weight and the total catch weight to give the length frequency of the catch for each month. Length frequencies for each month were summed across each year to provide a length frequency of the catch-adjusted length frequency for each year. These length frequencies are shown in Figure 73 through to Figure 77 for year 2001 to 2005 respectively.

Approximately 100 fish were sampled for ageing each month giving a target equating to 1200 samples per year. From the five years of sampling, a total of 5308 otoliths were collected for age estimation. The ageing methods used for these samples are unknown. The modal age-class was two years for all sampling years except 2003 where the most abundant age-class was three years. The maximum age of the cod sampled was 10 years in 2002 and 2004; for the years 2001, 2003 and 2005, the maximum age was 9 years. Samples over the age of 6 years (7+ grouping) accounted for a relatively small proportion of the sample (1.39%) and were grouped to the one class of 7+. All the otoliths sampled were read by one experienced reader with only 7 from the 5308 samples classified as unreadable. Numbers of fish sampled for ageing from the fish markets for each year between 2001 and 2005 are shown in Table 37. Percentages of each age-class are shown in Table 37. The modal age-class accounts for up to 56% of the samples in a year (Table 38).

Table 37. Age composition of Atlantic cod samples from 2001 to 2005.

Age-class	2001	2002	2003	2004	2005
1	127	72	28	58	73
2	694	435	285	248	579
3	141	381	482	212	190
4	135	50	244	173	97
5	91	47	36	89	129
6	25	22	28	13	42
7	14	8	15	4	3
8	7	9	2	3	1
9	1	3	1	0	2
10	0	1	0	1	0
unreadable	0	2	3	0	2
Total	1235	1030	1124	801	1118

Table 38. Percentage age composition of cod samples from 2001 to 2005.

Age-class	2001	2002	2003	2004	2005
1	10.28	6.99	2.49	7.24	6.53
2	56.19	42.23	25.36	30.96	51.79
3	11.42	36.99	42.88	26.47	16.99
4	10.93	4.85	21.71	21.60	8.68
5	7.37	4.56	3.20	11.11	11.54
6	2.02	2.14	2.49	1.62	3.76
7	1.13	0.78	1.33	0.50	0.27
8	0.57	0.87	0.18	0.37	0.09
9	0.08	0.29	0.09	0.00	0.18
10	0.00	0.10	0.00	0.12	0.00
unreadable	0.00	0.19	0.27	0.00	0.18
Total	100	100	100	100	100

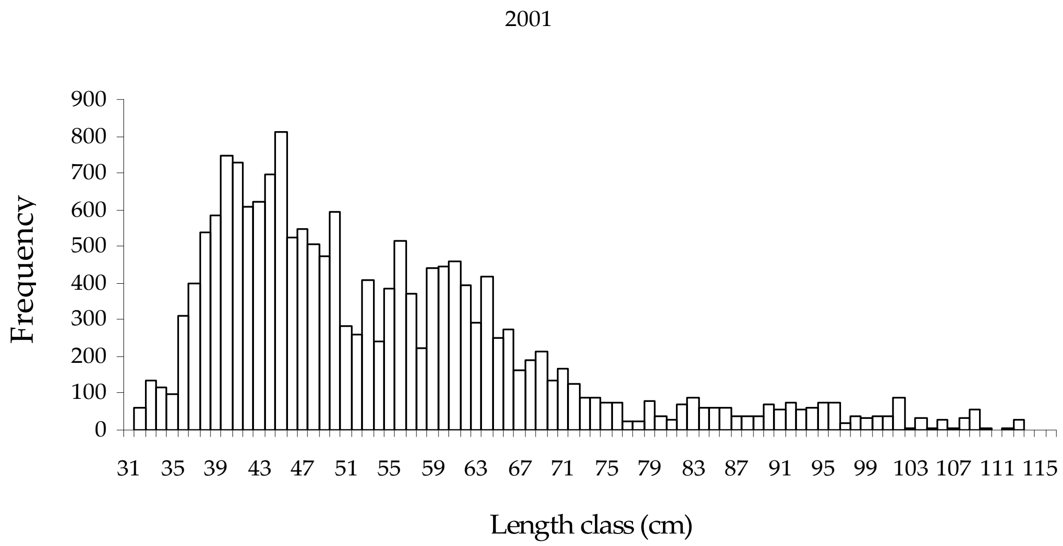


Figure 73. Length frequency for the 2001 cod sample. Length frequency has been catch adjusted.

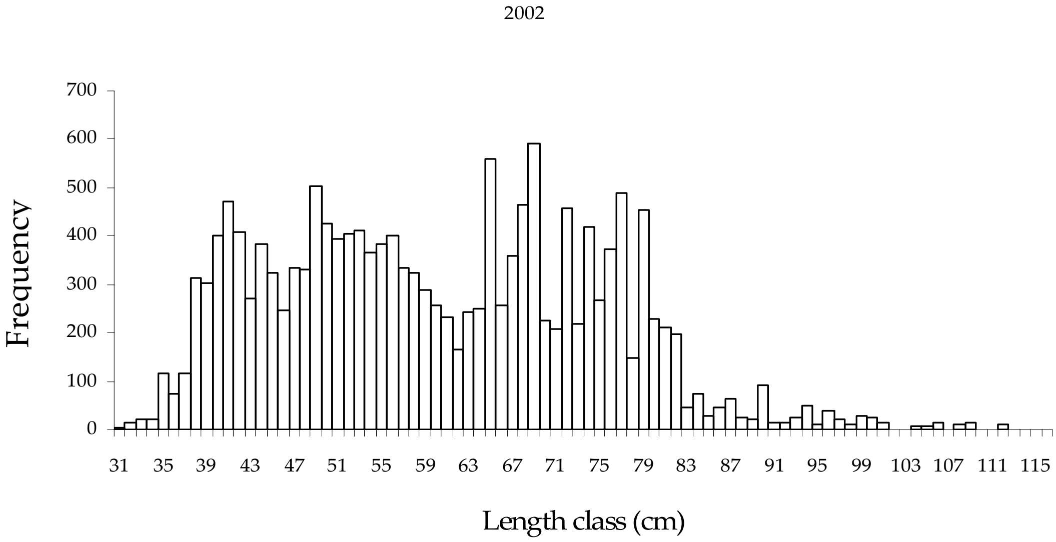


Figure 74. Length frequency for the 2002 cod sample. Length frequency has been catch adjusted.

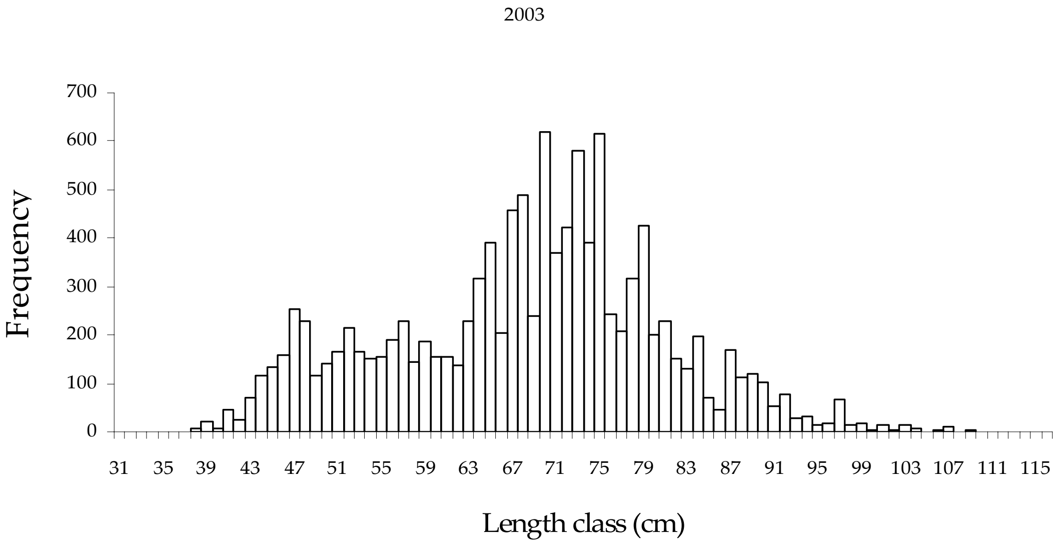


Figure 75. Length frequency for the 2003 cod sample. Length frequency has been catch adjusted.

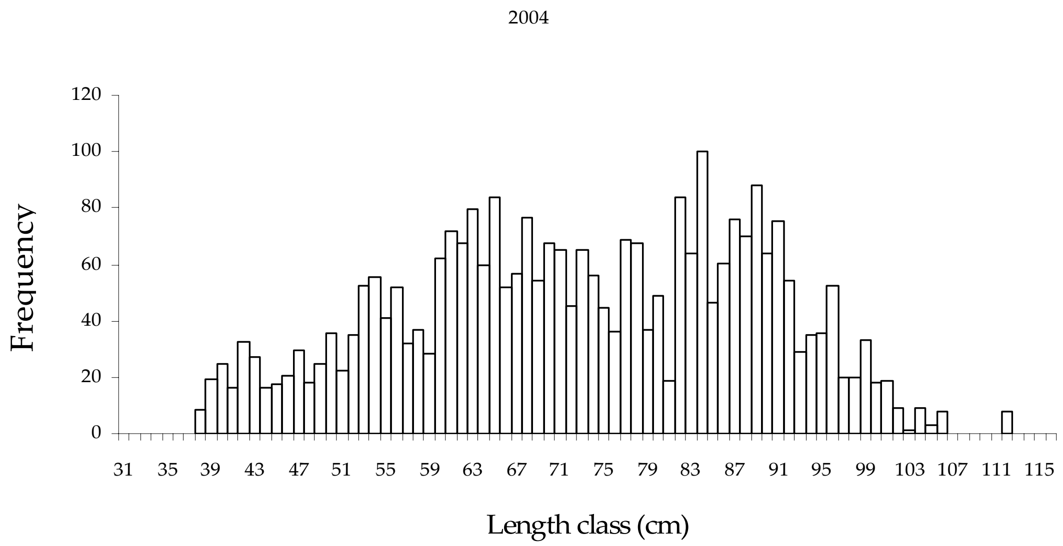


Figure 76. Length frequency for the 2004 cod sample. Length frequency has been catch adjusted.

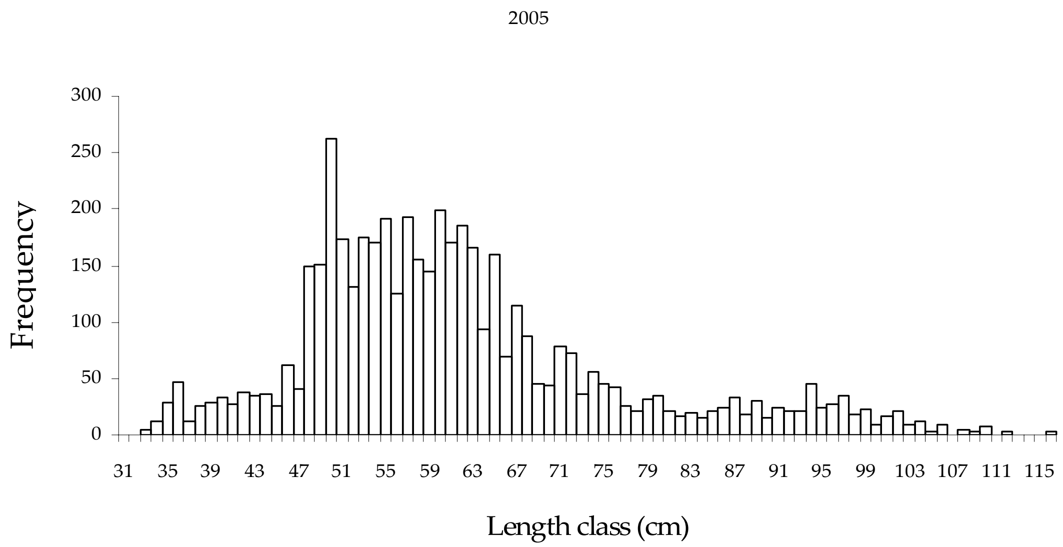


Figure 77. Length frequency for the 2005 cod sample. Length frequency has been catch adjusted.

Data preparation

The Atlantic cod data were collected for each month over the five year period. This allowed the data to be separated by quarters to increase the resolution of the FFK method. The catch-adjusted length frequencies were then calculated for each quarter of the year using the ALK method. To correct the age composition for biases as an artefact from the stratified sampling, the age composition for each quarter was Bayes corrected by the catch-adjusted length frequency from the respective quarter.

To determine the target age composition, the ALK method was used. The data were summed by quarter yearly groups for both age and length frequency before making the calculations. The proportion of age at each length-class was multiplied by the length frequency at each centimetre length class from the catch adjusted length frequency distribution. These results were summed by quarter within each year to provide the target age composition. These are the frequencies to which the results from the FFK method results were compared.

To test the FFK on these data a series of trials were used. The FFK method was initially tested in the same data environment as the ALK would be used (same year comparisons). Age compositions were predicted from the length frequency data for up to five years in advance (the maximum possible using these data i.e. this was 2001 age composition predicting the age composition in 2005 using the length frequency data from 2005). Age composition was also tested for the proposition of determining length frequencies from historical years (e.g. using the length-at-age data from 2005 and determining the age composition from the 2001 length frequency). Age compositions from either side of a target year were also tested (e.g. using the length-at-age from 2001 and 2003 to predict the 2002 age composition using the 2002 length frequency). A total of 26 trials were tested for cod, these are summarised in Table 39.

Length resolution

Decreased length resolution was introduced to the analyses. This was done to reduce the size of the matrices used to determine the target age compositions. The size of the matrix used for the cod FFK method trials was $a[7,86]$. This gave 86 length-class bins in the array. To reduce the number of length-classes, these were binned using two centimetre units. This reduced the matrices to $a[7,43]$ in size. Reduction in the length resolution had the effect of smoothing the empirical distributions of length-at-age.

As length increases with age before the asymptote of a growth function; the distribution of length-at-age for each age-class prior to reaching the asymptotic length would correspondingly change over the course of the year. These effects are more apparent with the faster growing age-classes (one-year olds and two-year olds). It was hypothesised that this effect would decrease the ability of the FFK technique to decompose the length frequencies into the underlying age compositions. This would effectively 'smear' the distributions of length-at-age. The length-at-age distributions would be affected by the period in which the samples were collected. If the samples were collected later in the year, the distribution of length-at-age for a given year class would be composed of larger fish.

To remove (or at least minimise) the impact of the shifting of length-at-age distributions on the model's ability to decompose a length frequency into the underlying age composition, the temporal distribution of length-at-age was examined at a finer resolution than a one-year increment. Length-at-age was examined in quarter year blocks. Most age-classes were collected in each month of every year; the exception was one-year-olds. No one-year olds were collected in the first quarter from some of the sample years. This effectively gave 27 age-classes with each being ¼ of a year. After analyses, these quarter year units were summed by year to make results comparable with those achieved using the ALK method.

The new size for all length-at-age matrices were $a[27,86]$ for the length resolution equal to one centimetre trials and $a[27,43]$ for the length resolution equal to two centimetre trials. The distributions of length-at-age for one centimetre resolution are shown in Figure 78 and the length-at-age distributions for length resolution of two centimetres are shown in Figure 79. The distributions of length-at-age for three centimetre length resolution are shown for reference in Figure 80 as distributions of individual age-classes can be difficult to observe at finer length units.

All matrices that were used in the trials were Bayes-corrected by the length frequency from the year from which the ageing data were taken. Where multiple years of ageing data were used to construct the length-at-age matrix, the weighted averages were used to adjust the relative abundances within each age class of the length-at-age matrix.

Table 39. Trials used to test the FFK method on the cod data. The age composition column represents the data from which the information on the distribution of length-at-age was taken and the length frequency column was the target combined distribution which was decomposed to the underlying age-classes. (*year) marks years where trials were also conducted at a length resolution of 2 centimetres.

Trial group	Trial Number	Age composition	Length frequency
Within Year	1	2001	*2001
	2	2002	*2002
	3	2003	*2003
	4	2004	*2004
	5	2005	*2005
Moving Forward	6	2001	*2002
	7	2001	*2003
	8	2001	*2004
	9	2001	*2005
	10	2002	*2003
	11	2002	*2004
	12	2002	*2005
Moving Backwards	13	2005	*2004
	14	2005	*2003
	15	2005	*2002
	16	2005	*2001
Multiple years	17	2001 - 2002	*2003
	18	2001 - 2002	*2004
	19	2001 -2002	*2005
	20	2002 - 2003	2004
	21	2002 - 2003	2005
	22	2001 - 2002 - 2003	2004
	23	2001 - 2002 - 2003 - 2004	2005
Missing year	24	2001 - 2003	2002
	25	2002 - 2004	2003
	26	2003 - 2005	2004

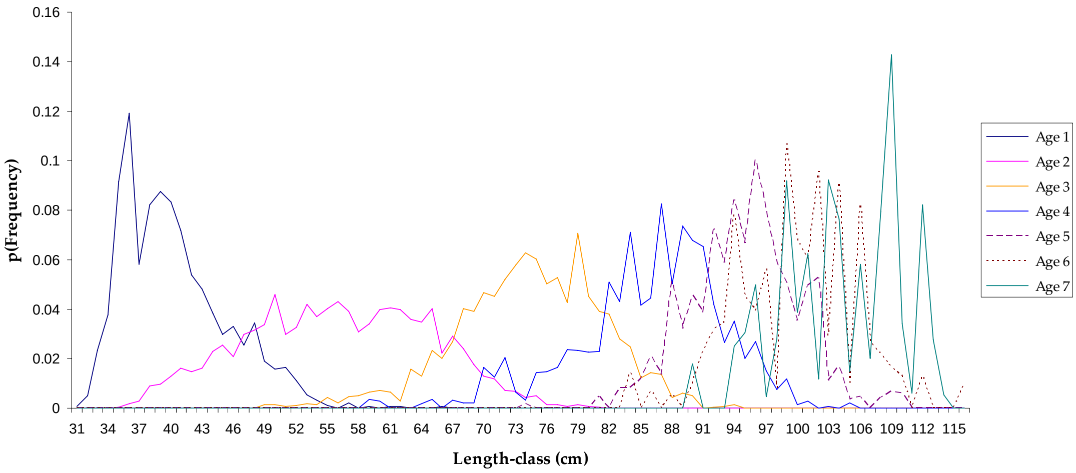


Figure 78. Length-at-age distributions at one centimetre resolution (all years).

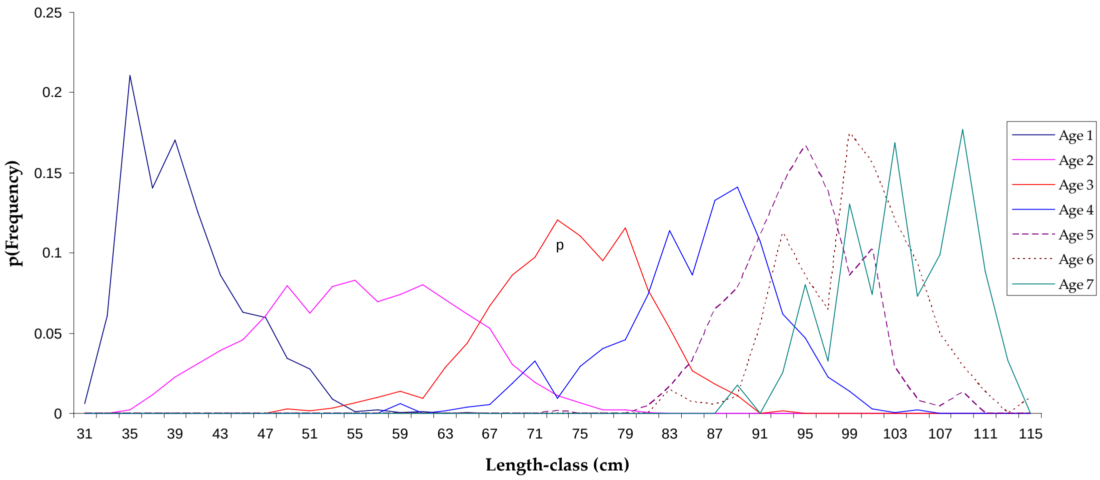


Figure 79. Length-at-age distributions at two centimetre resolution (all years).

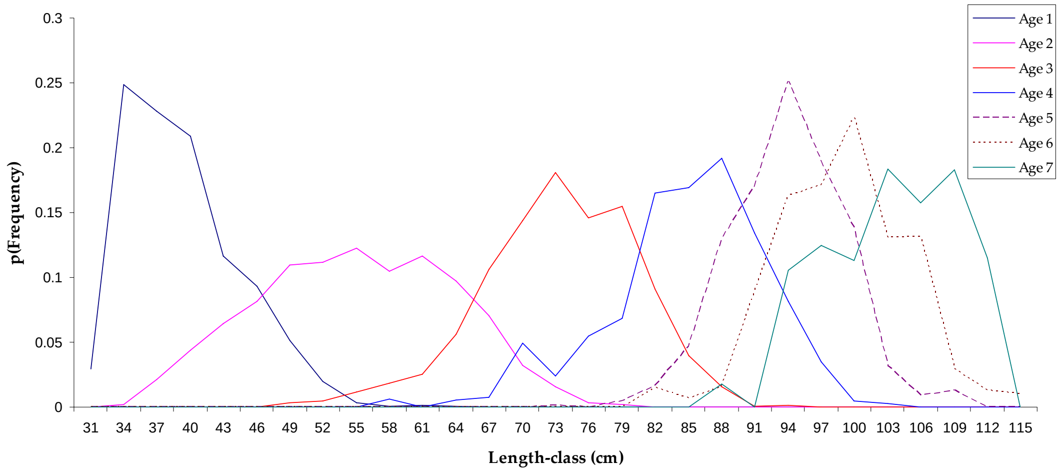


Figure 80. Length-at-age distributions at three centimetre resolution (all years - for clarity).

Results

The first trials were testing the FFK method in the same data environment as would be used with the ALK. The age composition of each year was calculated using the ageing sample from that year and the length frequency from the same year. These results are shown in Table 40.

Table 40. Observed and predicted age compositions from years 2001 to 2005. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. Total error is the quadratic error between observed and predicted age composition.

2001		2002		2003		2004		2005	
Observed	FFK	Observed	FFK	Observed	FFK	Observed	FFK	Observed	FFK
0.26	0.26	0.07	0.07	0.02	0.02	0.09	0.09	0.08	0.08
0.63	0.63	0.56	0.56	0.31	0.31	0.30	0.30	0.71	0.71
0.04	0.04	0.33	0.33	0.52	0.52	0.25	0.25	0.10	0.10
0.03	0.03	0.02	0.02	0.13	0.13	0.24	0.24	0.04	0.04
0.02	0.02	0.01	0.01	0.01	0.01	0.10	0.10	0.05	0.05
0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01	0.02	0.02
0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.01
Q-error	**0.0004**		**0.0012**		**0.0008**		**0.0014**		**0.0021**

The age composition predicted by the FFK method was very close to the target age composition. This was seen in all years (Table 40). The observed age composition and the predicted age composition (FFK) results are shown for year 2004 in Figure 81. The result from one year has been plotted to demonstrate the accuracy between observed and predicted age composition. Other years are similar in the accuracy between the observed and predicted age compositions.

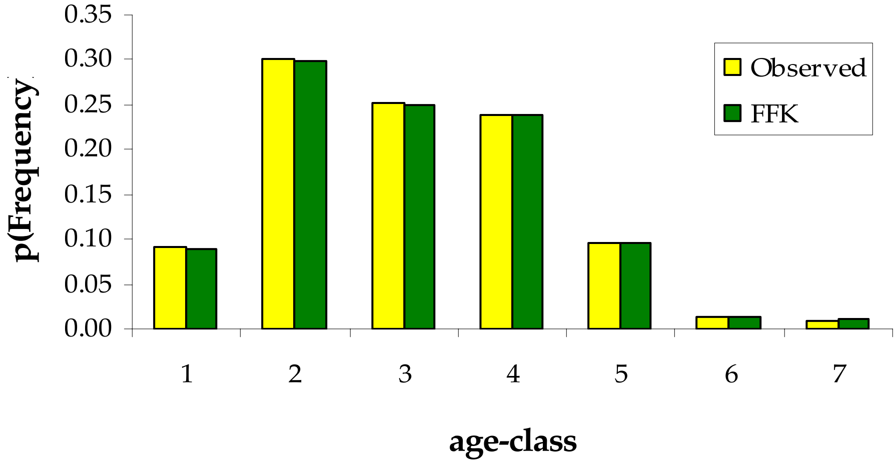


Figure 81. Age composition (Trial 4 - 2004) calculated using the FFK and the ALK method using age composition length frequency data from the same year for Atlantic cod.

The results of trials where the age composition was predicted from length frequencies in latter years are presented next. A total of seven trials were conducted using two length-at-age matrices. The base years from which the length-at-age matrices were constructed were 2001 and 2002. The 2001 length-at-age matrix was used to reconstruct the age composition from the length frequencies in the following four years. The 2002 length-at-age matrix was used to reconstruct the age composition from the length frequencies for the next consecutive three years.

Table 41. Observed and predicted age compositions from years 2002-2005 using the 2001 length-at-age matrix. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. The measure of total error is the quadratic error between observed and predicted age composition.

Using 2001 length-at-age matrix							
2002		**2003**		**2004**		**2005**	
Observed	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**
0.07	0.07	0.02	0.07	0.09	0.06	0.08	0.06
0.56	0.63	0.31	0.44	0.30	0.43	0.71	0.81
0.33	0.27	0.52	0.42	0.25	0.19	0.10	0.04
0.02	0.01	0.13	0.06	0.24	0.25	0.04	0.06
0.01	0.01	0.01	0.01	0.10	0.04	0.05	0.03
0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.01
0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Q-error	0.035		0.072		0.060		0.047

The lowest reconstruction error was seen in the reconstruction of the 2002 age composition using the Bayes-corrected 2001 length-at-age matrix (Figure 82, Table 41). The error was not increasing as the temporal variation was increased. The next lowest quadratic error was for 2005, then for 2004. The highest reconstruction error was seen using when the 2003 length frequency was decomposed to the underlying age composition (Table 41).

All of the results using the 2001 length-at-age matrix suggest that the information contained in the length-at-age distributions can be used to approximate the age composition of any of the length frequencies to a reasonable level of accuracy.

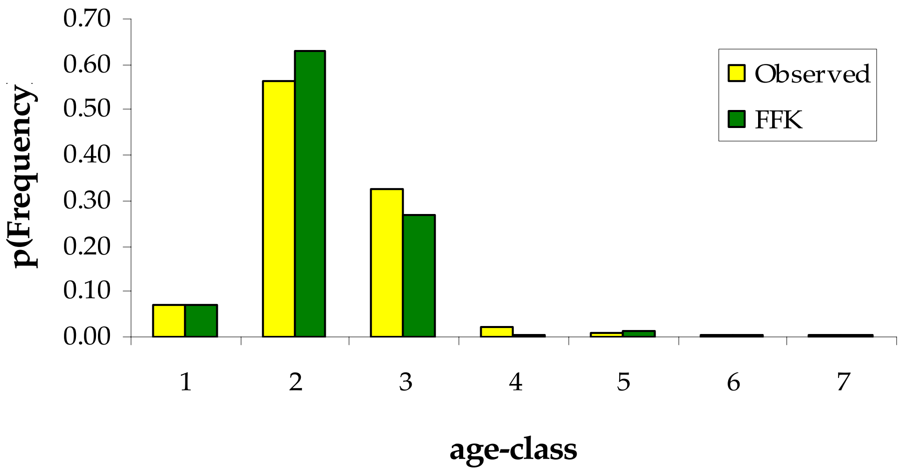


Figure 82. Age reconstruction for 2002 using the Bayes-corrected 2001 length-at-age matrix.

The reconstructions of age composition using the 2002 length-at-age matrix produced results which were similar to those produced using the 2001 length-at-age matrix (Table 42). The minimum and maximum quadratic errors for this series of trials were bracketed by those using the 2001 length-at-age matrix. All reconstructions approximated the observed age composition as determined by the ALK method using ageing data from the observed age-class for each year. Age composition from 2004 is presented in Figure 83. Age reconstruction for 2004 using the Bayes-corrected 2002 length-at-age matrix; this reconstruction uses the length-at-age information from two years previous.

Table 42. Observed and predicted age compositions from years 2003-2005 using the 2002 length-at-age matrix. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. The measure of total error is the quadratic error between observed and predicted age composition.

Using 2002 length-at-age matrix					
2003		2004		2005	
Observed	FFK	Observed	FFK	Observed	FFK
0.02	0.01	0.09	0.05	0.08	0.02
0.31	0.40	0.30	0.41	0.71	0.79
0.52	0.44	0.25	0.23	0.10	0.04
0.13	0.04	0.24	0.14	0.04	0.05
0.01	0.02	0.10	0.10	0.05	0.04
0.01	0.08	0.01	0.05	0.02	0.06
0.00	0.00	0.01	0.01	0.00	0.01
Q-error	0.066		0.060		0.048

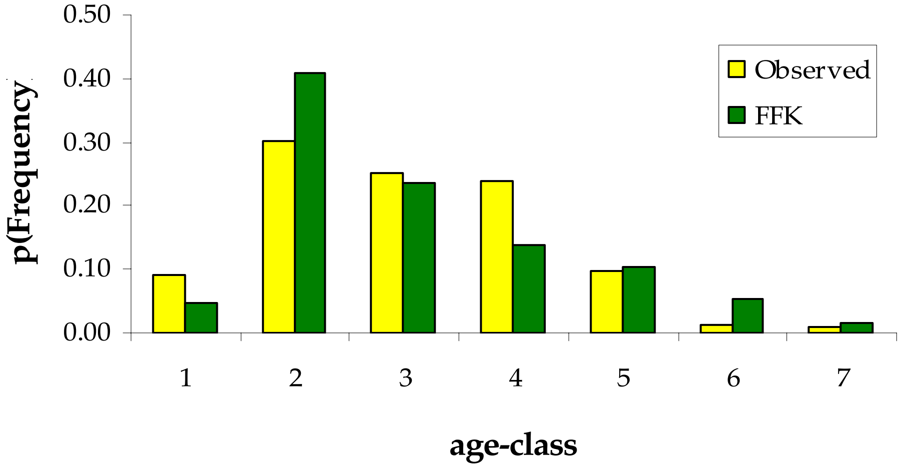


Figure 83. Age reconstruction for 2004 using the Bayes-corrected 2002 length-at-age matrix

Age compositions for trials 13 to 16 tested the utility of the technique to determine the age composition in the scenario where ageing data were available in one year and historical length frequencies were available for the previous 4 years. In this trial, the length-at-age distributions from 2005 were used to reconstruct the age composition for 2004, 2003, 2002 and 2001. The results for these trials are presented in Table 43.

Table 43. Observed and predicted age compositions from years 2004-2001 using the 2005 length-at-age matrix. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. The measure of total error is the quadratic error between observed and predicted age composition.

Using 2005 length-at-age matrix							
2004		**2003**		**2002**		**2001**	
Observed	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**
0.09	0.06	0.02	0.10	0.07	0.20	0.26	0.34
0.30	0.33	0.31	0.27	0.56	0.38	0.63	0.44
0.25	0.23	0.52	0.38	0.33	0.33	0.04	0.10
0.24	0.26	0.13	0.19	0.02	0.05	0.03	0.03
0.10	0.04	0.01	0.01	0.01	0.01	0.02	0.02
0.01	0.03	0.01	0.01	0.00	0.01	0.01	0.02
0.01	0.04	0.00	0.04	0.00	0.02	0.01	0.06
Q-error	0.033		0.068		0.085		0.085

The maximum quadratic error for these series of trials increase slightly as the temporal difference becomes greater. The closest approximation was from using the 2005 length-at-age data to reconstruct the age composition from the 2004 length frequency (Table 43, Figure 84). The maximum quadratic error was observed when the 2005 length-at-age distributions were used to reconstruct the age compositions from 2002 and 2001 using their respective length frequencies.

The age composition reconstruction for 2004 using the 2005 length-at-age distributions is shown in Figure 84.

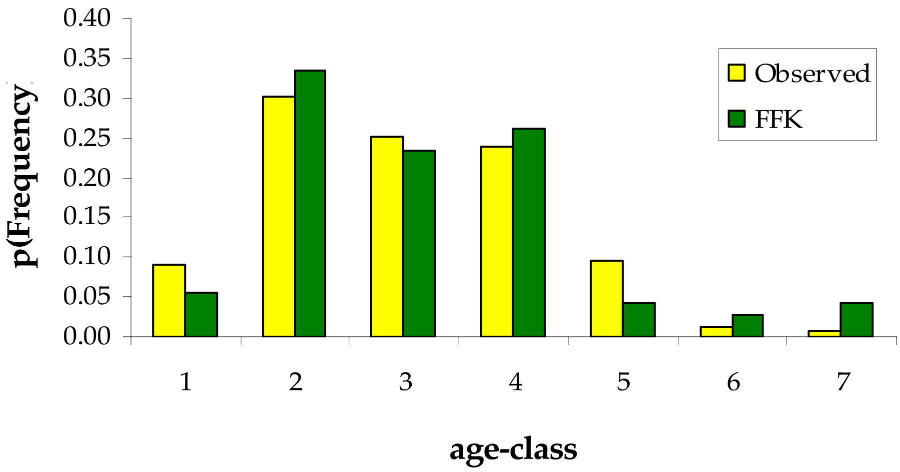


Figure 84. Age reconstruction for 2004 using the Bayes corrected 2005 length-at-age matrix

The following five trials tested combining two years of length-at-age distributions and testing these matrices to reconstruct years hence. Two groups of combined years were trialled. These were using a length-at-age matrix constructed from 2001 and 2002 length-at-age distributions and a length-at-age matrix constructed from the 2002 and 2003 length-at-age distributions. The 2001 and 2002 length-at-age matrix was trialled to determine the underlying age composition from the 2003, 2004 and 2005 length frequencies. The matrix constructed from the 2002 and 2003 length-at-age distributions was trialled to determine the age composition from the 2004 and 2005 length frequencies (Table 44).

Table 44. Observed and predicted age compositions from years 2003, 2004 and 2005 using the combined 2001 and 2002 length-at-age matrix; and observed and predicted age compositions from 2004 and 2005 using the 2002 and 2003 length-at-age matrices. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. The measure of total error is the quadratic error between observed and predicted age composition.

Using combined 2001 & 2002 matrices						Using combined 2002 & 2003 matrices			
2003		2004		2005		2004		2005	
Observed	FFK	Observed	FFK	Observed	FFK	Observed	FFK	Observed	FFK
0.02	0.00	0.09	0.01	0.08	0.02	0.09	0.02	0.08	0.02
0.31	0.39	0.30	0.45	0.71	0.84	0.30	0.40	0.71	0.79
0.52	0.52	0.25	0.18	0.10	0.03	0.25	0.14	0.10	0.06
0.13	0.06	0.24	0.25	0.04	0.06	0.24	0.35	0.04	0.06
0.01	0.01	0.10	0.06	0.05	0.03	0.10	0.04	0.05	0.02
0.01	0.01	0.01	0.04	0.02	0.02	0.01	0.03	0.02	0.03
0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.01
Q-error	0.043		0.071		0.064		0.079		0.044

The reconstructions show little improvement over using single year length-at-age distributions to reconstruct the age composition from a length frequency with an unknown age composition. This was evident from the quadratic error between the results from the FFK method and those obtained using the ALK method (Table 44). The age composition for 2003 is shown in Figure 85 This age composition was reconstructed using the length-at-age distributions from both 2001 and 2002. The assignment of the relative abundance for the modal class (3 year-olds) is close while age-class abundance on either side of the mode was slightly under or over estimated (Figure 85).

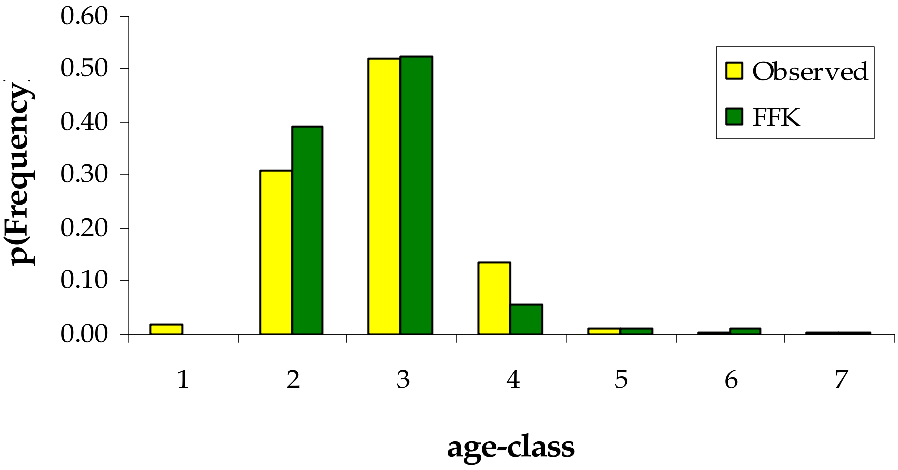


Figure 85. Age reconstruction for 2003using the Bayes-corrected 2001 and 2002 combined length-at-age matrices.

The following two trials used more than two years of length-at-age data to reconstruct the age composition from the length frequency from the target year. Here, the length-at-age distributions were constructed from;

a) the 2001, 2002, 2003 length-at-age distributions and

b) the 2001, 2002, 2003 and 2004 length-at-age distributions.

The first length-at-age matrix (a) was used to reconstruct the age composition from the 2004 length frequency and the second (b) length-at-age matrix was used to reconstruct the age composition from the 2005 length frequency. This data is presented in Table 45.

By combining more than two years of length-at-age distribution information, the FFK method again produced results which approximated the underlying age composition in both trials. The modal age-class was predicted using the FFK method in both cases. Each age-class on either side of the mode was under-estimated (Figure 86). Relatively close agreement was observed for other age-classes within each trial.

For the following three trials, scenarios where ageing data were available on either side of a year where no ageing data exists, were tested. The age composition was determined by the FFK method using a length-at-age matrix constructed from the length-at-age distributions on either side of the year where only length frequency data were available. These three trials determined age composition of 2002 (using 2001 and 2003), 2003 (using 2002 and 2004) and 2004 (using 2003 and 2005) using the target year's respective length frequency. The known or observed age composition was determined using the ALK method and ageing and catch adjusted length frequency data from the trial year. These results are presented in Figure 87.

Table 45. Observed and predicted age compositions from years 2004 using the combined 2001, 2002 and 2003 length-at-age matrix; and observed and predicted age compositions from 2005 using the 2001, 2002, 2003 and 2004 length-at-age matrices. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. The measure of total error is the quadratic error between observed and predicted age composition.

Using 2001,2002,2003 matrix		Using 2001,2002,2003,2004 matrix	
2004		2005	
OBS	FFK	OBS	FFK
0.09	0.01	0.08	0.02
0.30	0.42	0.71	0.80
0.25	0.12	0.10	0.05
0.24	0.37	0.04	0.05
0.10	0.04	0.05	0.03
0.01	0.03	0.02	0.04
0.01	0.01	0.00	0.00
Q-error	0.091		0.049

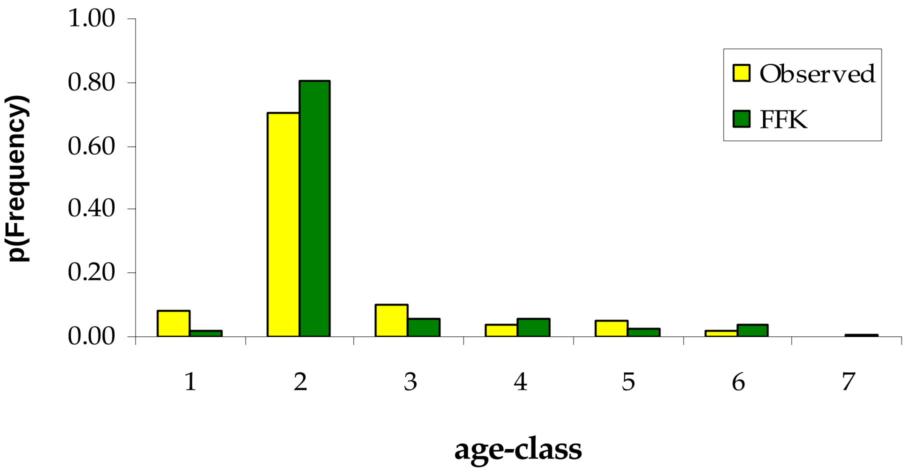


Figure 86. Age reconstruction for 2004 using the Bayes-corrected 2001, 2002 and 2003 combined length-at-age matrices.

Table 46. Observed and predicted age compositions from years 2002, 2003 and 2004 using length-at-age matrices which were combined from either side of the length frequency with the unknown age composition.

Using length-at-age matrices on either side of the target year					
2002		2003		2004	
Observed	FFK	OBS	FFK	OBS	FFK
0.07	0.13	0.02	0.00	0.09	0.05
0.56	0.50	0.31	0.44	0.30	0.35
0.33	0.34	0.52	0.47	0.25	0.20
0.02	0.03	0.13	0.06	0.24	0.32
0.01	0.00	0.01	0.01	0.10	0.04
0.00	0.00	0.01	0.00	0.01	0.01
0.00	0.00	0.00	0.00	0.01	0.03
Q-error	0.034		0.061		0.049

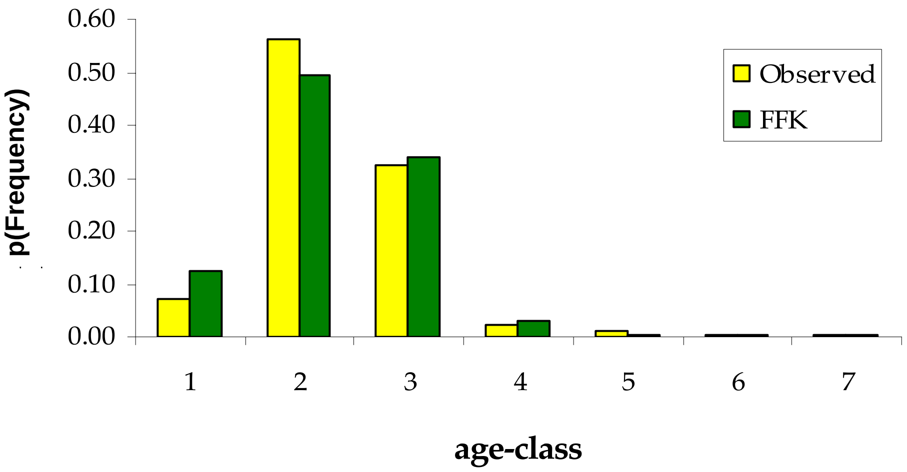


Figure 87. Age reconstruction for 2002 using the Bayes-corrected 2001 and 2003 combined length-at-age matrices.

The quadratic errors from these trials were similar to the results from the other trials (Table 52). These results demonstrate that the FFK method can be used to provide age composition data from years where no samples were collected for routine ageing but where length frequency data exists. The age composition of cod reconstructed by the FFK method using the combined length-at-age matrices from 2001 and 2003 demonstrates the close approximation of the underlying age composition decomposed from the 2002 length frequency (Figure 87).

Length resolution at two centimetres

Trials to reconstruct the age composition from a length frequency with an unknown age composition were also conducted at a length resolution of two centimetres. The following 19 trials were conducted using the same methods and data sets as the first series of trials. The difference between the two sets of trials is that all matrices and length frequencies used were at a resolution of two centimetres instead of one centimetre.

The reconstructions using the FFK method, in the same data environment as the ALK method, is currently used to demonstrate that the FFK method produces estimates of age composition that closely approximate current methods (Table 47).

The reconstruction quadratic error was lower than those trials using the one centimetre resolution in all cases except for 2003, where the error was the same. A reconstruction from 2004 is shown in Figure 88 using a length-at-age matrix and length frequency at two centimetre resolution.

Where the FFK method was used to determine the age composition from years 2002 to 2005 using their respective length frequencies and the length-at-age matrix from 2001 the resultant age compositions was similar to those obtained using the one centimetre length resolution matrix (Table 48). The reconstruction errors (Q-error) were lower for the 2002 and 2003 and higher for 2004 and 2005 length frequencies using the 2001 length-at-age distributions.

Table 47. Observed and predicted age compositions from years 2001 to 2005. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method using two centimetre length resolution. Total error is the quadratic error between observed and predicted

2001		2002		2003		2004		2005	
Observed	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**
0.26	0.26	0.07	0.07	0.02	0.02	0.09	0.09	0.08	0.08
0.63	0.63	0.56	0.56	0.31	0.31	0.30	0.30	0.71	0.71
0.04	0.04	0.33	0.33	0.52	0.52	0.25	0.25	0.10	0.10
0.03	0.03	0.02	0.02	0.13	0.13	0.24	0.24	0.04	0.04
0.02	0.02	0.01	0.01	0.01	0.01	0.10	0.10	0.05	0.05
0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01	0.02	0.02
0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.01
Q-error	0.0003		0.0011		0.0008		0.0023		0.0018

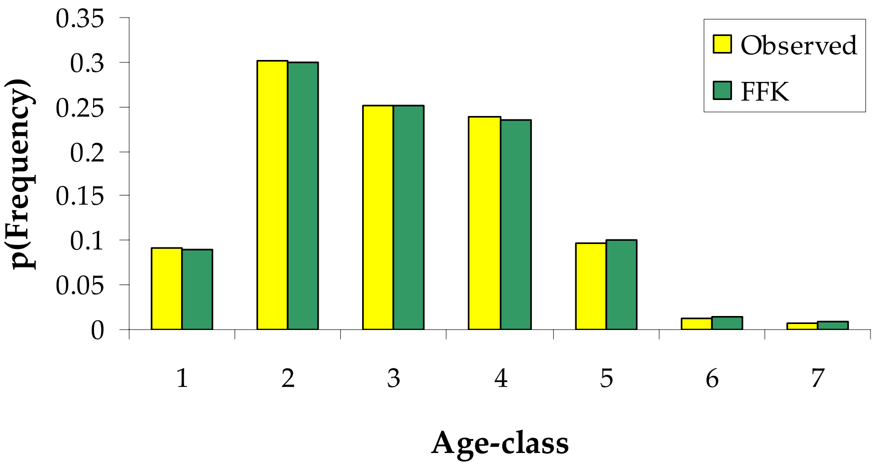


Figure 88. Age reconstruction for 2004 using the Bayes-corrected 2004 length-at-age matrix.

Table 48. Observed and predicted age compositions from years 2002-2005 using the 2001 length-at-age matrix as a length resolution of two centimetres. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. The measure of total error is the quadratic error between observed and predicted age composition.

Using 2001 length-at-age matrix							
2002		**2003**		**2004**		**2005**	
Observed	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**	**Observed**	**FFK**
0.07	0.08	0.02	0.03	0.09	0.04	0.08	0.04
0.56	0.62	0.31	0.44	0.30	0.44	0.71	0.82
0.33	0.28	0.52	0.43	0.25	0.18	0.10	0.03
0.02	0.00	0.13	0.06	0.24	0.22	0.04	0.06
0.01	0.01	0.01	0.02	0.10	0.07	0.05	0.04
0.00	0.00	0.01	0.02	0.01	0.05	0.02	0.01
0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Q-error	0.027		0.065		0.066		0.053

The age composition from 2002 is shown in Figure 89. The distributions between observed and predicted age compositions were relatively close with the FFK method approximating the general relative abundance of each age-class.

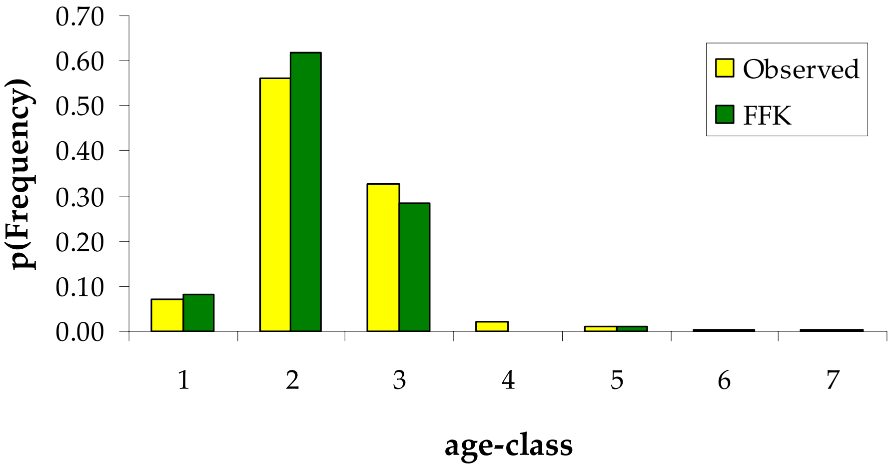


Figure 89. Age reconstruction for 2002 using the Bayes corrected 2001 length-at-age matrix at two centimetre length resolution.

The relative abundances of age classes for the trial using the 2002 length-at-age matrix to determine the age composition from the 2003, 2004 and 2005 length frequencies were also estimated using a two centimetre length resolution. The lowest quadratic error between the observed and predicted age compositions was seen for the trial that reconstructed the age composition from the 2003 length frequency using the 2002 length-at-age distribution matrix (Table 49).

The three reconstructions of the age composition using the 2002 length-at-age matrix for 2003, 2004 and 2005 show close agreement with the observed age (Table 49, Figure 90). The modal age-class was slightly higher than the observed age class (age class 2), while the relative abundance of each age class either side of the modal age was slightly under estimated (Figure 90). The general shape of the underlying age composition is however approximated using length-at-age distribution information from two years previous.

Table 49. Observed and predicted age compositions from years 2003-2005 using the 2002 length-at-age matrix at a resolution of two centimetres. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. The measure of total error is the quadratic error between observed and predicted age composition.

Using 2002 length-at-age matrix					
2003		2004		2005	
Observed	FFK	Observed	FFK	Observed	FFK
0.02	0.00	0.09	0.05	0.08	0.03
0.31	0.33	0.30	0.39	0.71	0.79
0.52	0.57	0.25	0.20	0.10	0.04
0.13	0.08	0.24	0.26	0.04	0.08
0.01	0.01	0.10	0.05	0.05	0.04
0.01	0.01	0.01	0.04	0.02	0.01
0.00	0.00	0.01	0.01	0.00	0.01
Q-error	0.028		0.049		0.045

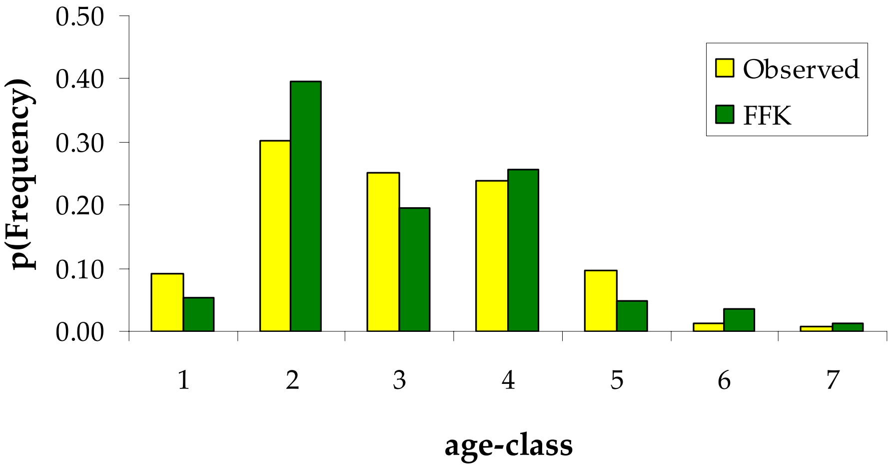


Figure 90. Age reconstruction for 2004 using the Bayes corrected 2002 length-at-age matrix at a length resolution of two centimetres.

The following four trials demonstrate the use of the FFK method to determine the age composition from historical length frequency data using the length resolution at two centimetres for all datasets (Table 50).

Age reconstructions from length-at-age matrices in the scenario where length frequencies exist but no age data exist demonstrate that, using temporally later length-at-age distributions, it is possible to approximate the age composition for the years where no age data exist. Quadratic errors exhibited using these scenarios are similar to those from trials where age compositions are predicted in later years. The age reconstruction for 2004 using the length-at-age distributions from 2005 produced the lowest quadratic error and reproduced a similar age composition to the observed age-class abundance for that year (Figure 91).

Table 50. Observed and predicted age compositions from years 2004-2001 using the 2005 length-at-age matrix at a length resolution of two centimetres. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. The measure of total error is the quadratic error between observed and predicted age composition.

Using 2005 length-at-age matrix							
2004		**2003**		**2002**		**2001**	
OBS	**FFK**	**OBS**	**FFK**	**OBS**	**FFK**	**OBS**	**FFK**
0.09	0.05	0.02	0.10	0.07	0.23	0.26	0.40
0.30	0.33	0.31	0.23	0.56	0.37	0.63	0.44
0.25	0.26	0.52	0.54	0.33	0.34	0.04	0.08
0.24	0.28	0.13	0.08	0.02	0.02	0.03	0.03
0.10	0.03	0.01	0.00	0.01	0.01	0.02	0.02
0.01	0.03	0.01	0.01	0.00	0.01	0.01	0.02
0.01	0.02	0.00	0.04	0.00	0.03	0.01	0.02
Q-error	0.034		0.049		0.094		0.091

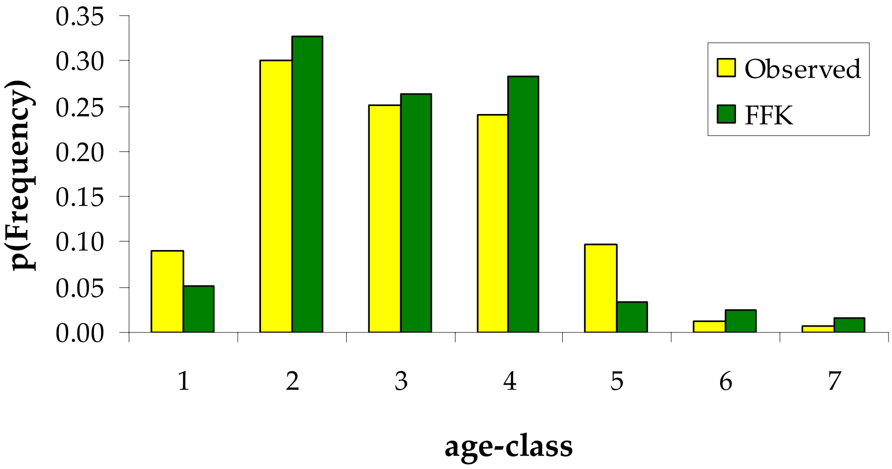


Figure 91. Age reconstruction for 2004 using the Bayes corrected 2005 length-at-age matrix

Multiple years were also trialled. Here three trials were conducted using combined length-at-age distributions from 2001 and 2002 to reconstruct the age composition from the length frequencies from 2003, 2004 and 2005.

Age reconstructions using the reduced size length-at-age resolution were slightly better than the same reconstruction using the larger one-centimetre length-at-age distribution matrix (Table 51, Table 52). The abundance of the modal age class of three-year-olds was approximated closely while that of the next most abundant age class was slightly over-estimated and that of the third most abundant age class (four-year-olds) was slightly under-estimated (Figure 92).

The errors for all reconstructions are tabulated to provide an overall view of the efficacy of the FFK method in reconstructing the age compositions from length-at-age distributions and length frequencies in the trialled scenarios (Table 52).

Table 51. Observed and predicted age compositions from years 2003, 2004 and 2005 using the combined 2001 and 2002 length-at-age matrix at a length resolution of two centimetres. Observed is the age composition calculated using the ALK method and FFK is the age composition calculated using the FFK method. The measure of total error is the quadratic error between observed and predicted age composition.

Using combined 2001 & 2002 matrices					
2003		2004		2005	
Observed	FFK	Observed	FFK	Observed	FFK
0.02	0.00	0.08	0.03	0.08	0.03
0.31	0.36	0.71	0.81	0.71	0.81
0.52	0.54	0.10	0.03	0.10	0.03
0.13	0.08	0.04	0.07	0.04	0.07
0.01	0.01	0.05	0.04	0.05	0.04
0.01	0.01	0.02	0.01	0.02	0.01
0.00	0.00	0.00	0.00	0.00	0.00
Q-error	0.030		0.053		0.053

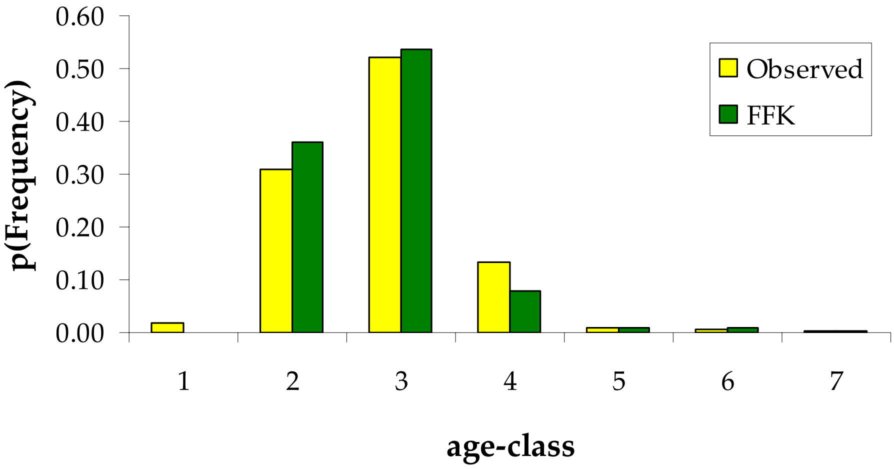


Figure 92. Age reconstruction for 2003 using the Bayes-corrected 2001 and 2002 combined length-at-age matrices at a length resolution of two centimetres.

Table 52. Trial number showing quadratic errors age composition reconstructions for length resolutions one centimetre and two centimetre length-at-age matrices. %Change is the 100-percentage difference between the quadratic errors (LR1 and LR2).

Trial	Trial Number	Age composition	Length frequency	Q-error (LR1)	Q-error (LR2)	%Change
Within Year	1	2001	2001	0.0004	0.0003	21.4
	2	2002	2002	0.0012	0.0011	7.3
	3	2003	2003	0.0008	0.0008	3.1
	4	2004	2004	0.0014	0.0023	-60.5
	5	2005	2005	0.0021	0.0018	11.9
Moving Forward	6	2001	2002	0.035	0.027	21.5
	7	2001	2003	0.072	0.065	9.0
	8	2001	2004	0.060	0.066	-9.7
	9	2001	2005	0.047	0.053	-13.7
	10	2002	2003	0.066	0.028	57.8
	11	2002	2004	0.060	0.049	19.0
	12	2002	2005	0.048	0.045	7.2
Moving Backwards	13	2005	2004	0.033	0.034	-5.4
	14	2005	2003	0.068	0.049	27.6
	15	2005	2002	0.085	0.094	-10.2
	16	2005	2001	0.085	0.091	-7.4
Multiple years	17	2001 -2002	2003	0.043	0.030	30.4
	18	2001 -2002	2004	0.071	0.053	24.6
	19	2001 -2002	2005	0.064	0.053	17.5

A closer reconstruction with respect to the observed age composition in 68% of the trials is obtained using the length resolution of two centimetres than using one centimetre length resolutions. The most consistent increase in accuracy was seen when multiple years were used to construct the length-at-age matrix. Overall, the closest reconstruction of age composition using the FFK method was obtained when using the 2001 length-at-age matrix to determine the age composition from the 2002 length frequency (Table 52).

Discussion

The trials with the Atlantic cod data demonstrated that the FFK method can decompose length-frequency data into age classes using length-at-age data which was not sampled from the same distribution. Within the same data environment where ALK method would normally be used, the FFK method produced almost identical results. The FFK could therefore be used as a replacement to the ALK method for these data. The FFK method does not suffer from the same fundamental assumption as the ALK method. The FFK method can be used to decompose length frequencies into the age composition where the samples used to construct the length-at-age distributions were not sampled. These facts make the FFK method a more flexible technique to determine the underlying age composition from a length frequency sample.

The success of the FFK method requires that the length-at-age distributions are stable over time and that the length-at-age distributions that are modified by the Bayes-correction are sampled from the same length frequency as the initial length-at-age distribution. When these sampling requirements are met, the FFK method can be used on independent length frequencies. The Bayes-corrected length-at-age distribution was successful in determining the age composition of the trialled length frequencies from the French Atlantic cod dataset. The reason for the successful application of the FFK method was that the French dataset fulfilled these requirements. As such the French data provided an ideal platform in which to test the efficacy of the FFK method.

The FFK method could be applied to length data and provide fisheries managers with rapid assessments of the age composition of a catch before the laborious task of ageing could be completed. This would allow indicators of changing age compositions to be rapidly identified. The potential that the FFK method could provide age composition information in days rather than months has been realised.

Many trials were tested in this chapter, all providing promising results in various scenarios. These included testing the FFK method to reconstruct the age compositions from the same and disparate years. The utility of using the FFK method to reconstruct historical age composition data has also been quantitatively demonstrated. This is not possible with current methods for determined age composition. One application of the FFK method may exist where species which have a low priority can be aged on a bi-yearly basis and a length frequency sampled every year. The age composition could be determined for the missing years at little cost. The benefit of this would be providing age structured models on a much greater number of species for the same cost as is currently used to provide the age composition for a limited number of species.

Chapter 7. Conclusion

This project had two objectives

- The first milestone was to validate the technique on a short-lived species with stable age composition.
- The second milestone was validating the FFK method on a longer-lived species with variable recruitment.

The two species chosen to test the FFK method were School Whiting (*Sillago flindersi*) and Blue Grenadier (*Macruronus novaezelandiae*). These species have been aged at the CAF since 1991 and robust datasets exist for these species.

School Whiting is a short-lived species (maximum age of 8 years) with a fairly stable age composition. Blue Grenadier are relatively long-lived, living to 25 years of age. The majority of the fish were under 15 years of age. Blue Grenadier is characterised by variable recruitment. A proviso for the continuation of the project was the successful demonstration of the FFK method on these species.

The FFK method was not successful in decomposing length-frequency data to age composition for either of the two selected species. As a direct result of this, the recommendation was made to terminate FRDC project 2005/023.

The FFK method could not be demonstrated on any of the species within the SEF. Although two species are reported, blue and silver warehou were also trialled in an attempt to produce accurate reconstructions of age composition from length frequencies. The FFK method failed to provide satisfactory results. There are two possible reasons for this failure:

- the length-at-age may not be stable over time
- the sampling regimes may be biasing the data.

These factors may act independently or in unison to render the FFK method ineffective. In both cases, this finding may be valuable for future monitoring and stock assessment of these species.

The FFK method was tested successfully on a fisheries dataset from France proving the FFK method does work with precision. These results show that the FFK method could be used by fisheries management to provide a rapid indication of the age distribution from a length frequency. This would allow proactive management of fish stocks before the process of ageing the samples could be undertaken.

The biological instability of length-at-age would cause the method to fail. The FFK method uses the *a priori* length-at-age to decompose a length frequency of an unknown age composition into the relative abundance of each age-class within that length distribution. If the length-at-age matrix is constructed from samples collected in (for example) February, and corrected by the length frequency from which it was sampled, then this would be valid for reconstructing the age composition from the following year's February length frequency. If, for biological or environmental reasons, the length-at-age distribution from one or more age-classes shifted, the results would be in error. The FFK method iteratively determines what combinations of length-at-age distributions at what frequencies would be required to produce the observed length frequency of unknown age composition. *It is therefore a fundamental requirement of the FFK method that the length-at-age distributions are similar between the length-at-age matrix and the length frequency which it is used to determine the age composition for each length class*. The reason for biological changes in the length-at-age distributions may include increased or decreased growth rates through factors such as climate change or changes in growth linked to density dependence. Where length-at-age distributions change over time, the FFK method could not decompose a length frequency into an age composition. In these cases, the fundamental assumption of the FFK would be violated.

The problems associated with sampling also have the capacity to introduce errors which would render the technique invalid. This is true for both the FFK method and the ALK method. A primary assumption of the ALK method is that the sample taken for ageing (to develop the age-at-length key) is a sub-sample from the length frequency sample. If this requirement is not fulfilled, then the resultant age composition calculated from the age-at-length matrix has a component of error. Although the amount of error in the age composition is not known, the basic assumptions have been violated. When an age-at-length distribution is used to determine the age composition from a series of length frequencies, the final age composition will be a result of additive errors.

With the School Whiting data, the initial trials and the previous work (Troynikov and Robertson, 2005) demonstrated that, under ideal sampling, the FFK method produced results that accurately described the underlying age composition. When 'real' fisheries data were used, the results were less than satisfactory. The sampling that was undertaken as part of the ISMP program was not suited to the application of the FFK method when used to reconstruct the age composition from length frequencies. The sampling strategy used to collect the Atlantic cod, however, was ideally suited to the application of the FFK method.

Benefits

The benefits of this project have not been realised for this project as termination was sought after failure to meet the first objective.

Further development

No further development for this project is foreseen.

Planned outcomes

The original planned outcomes for the project were not achieved.

As a result of the FFK method failing to produce consistent age composition data from length frequency and *a priori* length-at-age distribution data, the original planned outcomes of the project were not met. The original planned outcomes are listed below:

- Use the new method to determine optimal sample size that is sufficient for numerical reconstruction of cohorts and age groups for commercially important species.
- Determine appropriate yearly time schema for age estimation for given fish species. For example, age determination may only be necessary in every third or fourth year to obtain complete yearly time series of age compositions. Only length frequencies from years with missing ageing data will be necessary to obtain age composition.
- Fill missing years in time series of age compositions data where length-frequency data are available.

The project was terminated after the second milestone. The planned outcomes were to be addressed later in the project after the successful proof-of-concept stage, the first objective of the project (first two milestones).

References

Bakushinsky A., and Goncharsky A. (1994). 'Ill-posed Problems: Theory and Applications.' (Kluwer Academic Publishers: London.)

Campana, S. , and Thorrold, S. R. (2001). Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations. *Canadian Journal of Fish and Aquatic Science* **58**, 30–38.

Clark, W. G. (1981). Restricted least-squares estimates of age composition from length composition. *Canadian Journal of Fisheries and Aquatic Sciences* **38**, 297–307.

Fridriksson, A. (1934). On the calculation of age-distribution within a stock of cod by means of relatively few age determinations as a key to measurements on a large scale. Rapp. Proc.-Verb. Reun **86**, 1–14.

Gavaris S., and Eckhaute L. (1994). Application of length-at-age distribution to derive age composition of Georges Bank haddock from length frequencies. *Meeting of the ICCAT Standing Committee on Research and Statistics* **42**(1), 104–109 (SCRS9355).

Hoenig J. M., Heisey D. M., and Hanumara R. C. (1993). Using prior and current information to estimate age composition: a new kind of age-length key. (ICESCM1993D52) International Council for the Exploration of the Sea, Copenhagen.

Kimura, D. K. (1977). Statistical assessment of the age-length key. *Journal of the Fisheries Research Board of Canada* **34**, 317–326.

Kimura, D. K. , and Chikuni, S. (1987). Mixtures of empirical distributions: an iterative application of the age-length key. *Biometrics* **43**, 23–35.

Morison, A. K. , Robertson, S. G. , and Smith, D. C. (1998). An integrated production fish ageing system: quality assurance and image analysis. *North American Journal of Fisheries Management* **18**, 587–598.

Tikhonov A. N., Goncharsky A. V., Stepanov V. V., and Yagola A. G. (1995). 'Numerical Methods for the Solution of Ill-posed Problems.' (Kluwer Academic Publishers: Dordrecht.)

Troynikov, V. S. (1998). Probability density functions useful for parametrization of heterogeneity in growth and allometry data. *Bulletin of Mathematical Biology* **60**, 1099–1121.

Troynikov, V. S. (1999). Use of Bayes theorem to correct size-specific sampling bias in growth data. *Bulletin of Mathematical Biology* **61**, 355–363.

Troynikov, V. S. (2004). Use of weak quasi-solution of the Fredholm first-kind equation in problems with scarce data. *Applied Mathematics and Computations* **150**(3), 855–863.

Westrheim, S. J. , and Ricker, W. E. (1978). Bias in using an age-length key to estimate age-frequency distributions. *Journal of the Fisheries Research Board of Canada* **35**, 184–189.

Appendix 1 Intellectual Property

The intellectual property derived from this project is shared equally between Fisheries Research and Development Corporation and Fisheries Victoria.

Appendix 2: Staff

Fisheries Research Branch (Fisheries Victoria) (formerly Marine and Freshwater Resources Institute):

Simon Robertson - Principal Investigator

Vladimir Troynikov - Scientist

Appendix 3: Any other relevant material including eg raw data

Not relevant