A comparative study of new methods of quantifying scale shape for stock discrimination

By
Douglas Watkinson

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree

Master of Science

Department of Zoology
University of Manitoba
Winnipeg, Manitoba

© Copyright August 2001

National Library of Canada

Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services

Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-62863-9

Canada

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

A COMPARATIVE STUDY OF NEW METHODS OF QUANTIFYING SCALE SHAPE FOR STOCK DISCRIMINATION

BY

DOUGLAS WATKINSON

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree of MASTER OF SCIENCE

DOUGLAS WATKINSON © 2001

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

Abstract

Conventional methods of scale characterization commonly used to discriminate between stocks of different fish species have proven to be ineffective as a tool for fisheries management. This study investigated new analytical techniques for stock discrimination such as averaging of scale outline signals, wavelet signal processing methods, and computer intensive non-parametric statistics. These techniques were used to test the significance of discriminant results in a combined effort to improve the researcher's ability to discriminate between fish stocks. This study found that combining signals from several scales significantly improved the ability to discriminate between stocks. Non-parametric statistics effectively tested for significance and significant differences in analyses when assumptions of discriminant analysis are violated. Variables produced from wavelet decompositions formed significantly better discriminant functions than Fourier analysis variables for most comparisons. This research has increased our ability to discriminate between fish stocks based upon scale morphology.

ACKNOWLEDGEMENTS

Thank-you Darren Gillis, my advisor, for your patience and wisdom. I was fortunate to have you as an advisor. Also thank-you to the rest of my committee members Dr. M.V. Abrahams and A.N. Arnason for their comments and input into this thesis.

I am also extremely grateful to Walt Lysack and Steve Kendall for aiding me in obtaining the walleye scales that were used in this study. My sister Tracy was a great research assistant when I needed one. Thanks to Ken Campbell for allowing me to tag along during index netting.

This thesis was funded by the Fish Futures Scholarship awarded to myself and a research grant to D.M. Gillis from the National Science and Engineering Research Council, Canada.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: General Introduction

A stock can be defined as a group of fish that share a common environment, participate in a common gene pool and is sufficiently discrete to warrant consideration as a self-perpetuating unit which can be managed (Larkin 1972). Due to their reproductive isolation, each stock can develop phenotypic and genotypic differences over time (Waldman et al. 1988). These differences may arise from diverse environmental conditions, differential selection pressure, or random factors. However, a practical management definition of a stock often requires treating at least some groups collectively rather than separately (Larkin 1972).

It is important to be able to recognize different stocks in the catch of a mixed stock fishery. Stocks often mix at times when spawning is not occurring (Jarvis et al. 1978; Casselman et al. 1981; Cook 1982; De Pontual and Prouzet 1987; Margraf and Riley 1993; Campana and Casselman 1993; Richards and Esteves 1997). This can confound measures of growth, survival, and reproductive success; thereby, invalidating studies of fish biology, population dynamics, and most estimates of yield (Campana and Casselman 1993). When a large number of stocks of differing reproductive productivity mix in the fishery and quotas are set at the maximum sustainable yield, this may exterminate the less productive stocks, thereby reducing the productivity of the entire fishery (Ricker 1958). Identifying the stocks that compose a fishery allows their contribution to the fishery and population status to be monitored.

Walleye (*Stizostedion vitreum*) can be divided into stocks within most river systems or lakes where they are associated with a particular spawning area to which they return annually (Rawson 1957; Crowe 1962; Olson and Scidmore 1962; Forney 1963).

As a consequence, walleye stocks can be identified by their spawning area (Crowe 1962; Forney 1963; Ferguson and Derksen 1971). The post spawning movements of individual walleye can be extensive (Wolfert 1963; Forney 1963; Ferguson and Derksen 1971). In Lake Winnipeg, tagging studies conducted by the Department of Conservation confirmed some walleye travel distances in excess of 200km (Walt Lysack, pers. com., 2001). Since walleye captured during the fishing season may not be part of the local spawning population, stock identification should be considered in the management of walleye in a large system such as Lake Winnipeg.

Stock discrimination involves quantifying characteristics that differ significantly among stocks of "known" origin (Waldman et al. 1988; Margraf and Riley 1993). These characters are used to assign fish of unknown origin to one of the described stocks (Ihssen et al. 1981; Casselman et al. 1981; Riley and Carline 1982; Waldman et al. 1988). In principle, differentiation based on genotype should be used for inferences concerning distinct stocks (Campana and Casselman 1993); however, this may not always be feasible due sampling or analysis constraints. Alternatively, phenotypic differences in morphology, physiology, behaviour and biochemistry characters may be useful in identifying stocks or even preferable to genetics if these methods are better discriminators.

Fourier analysis of scale outlines has been used successfully as a method of stock identification for a number of fish species: walleye (Jarvis et al. 1978), Atlantic salmon (De Pontual and Prouzet 1987), sockeye salmon (Cook 1982), lake whitefish (Casselman et al. 1981), and striped bass (Margraf and Riley 1993; Richards and Esteves 1997; Waldman et al. 1997). Scale shape does not distinguish between genetic and environmental differences. Therefore, it can potentially distinguish among stocks or components of the same stock that have a different environmental history. In the case of walleye, established homing behaviour suggests that there may be some genetic component to any differences observed among geographically distinct spawning aggregations.

Past studies found the classification rates were considered too low for the methods to be applied in routine assessment of fisheries. It was concluded that Fourier analysis held promise, but further investigation was necessary (Casselman et al. 1981; Riley and Carline 1982; De Pontual and Prouzet 1987; Margraf and Riley 1993; Waldman et al. 1997; Richards and Esteves 1997). The objective of this study was to refine methods of signal processing to increase our ability to differentiate between stocks using scale shapes.

To distinguish between groups effectively the variability of parameters that define a group should not overlap significantly between groups. Therefore, by reducing variability within a group our ability to differentiate between groups may increase. The first goal of this study is to reduce the variability in scale shape that is used to represent an individual fish. This will be studied by using several scales to form averaged scale signals.

There are also theoretical problems with the application of Fourier analysis to irregular shapes with non-periodic signals. Fourier analysis is inefficient in dealing with localized frequency variation or discontinuities (Kaiser 1994; Hubbard 1996). Additionally, Fourier analysis is able to describe the frequency information from a signal, but very different outlines can have near identical values calculated for the coefficients if the original signal is not periodic (Bird et al. 1986). Wavelet analysis, a signal processing method that is suited to non-stationary (aperiodic) signals will be investigated as an alternative signal processing method to Fourier analysis (Bradshaw and Spies 1992; Hubbard 1996). To quantify differences between the signal analysis methods it is necessary to develop non-parametric techniques that can be used to detect differences between the signal analysis methods. Ultimately, by refining analyses based upon scale morphology, I hope to provide managers and researchers with a cost-effective tool to aid them in stock discrimination.

Literature Cited

Bird, J. L, Eppler, D.T., and Checkley, D.M., Jr. 1986. Comparisons of herring otoliths using Fourier series shape analysis. Can. J. Fish. Aquat. Sci. **43**: 1228-1234.

Bradshaw, G.A., and Spies, T.A. 1992. Characterizing canopy gap structure in forests using wavelet analysis. J. Ecol. **80**: 205-215.

Campana, S.E., and Casselman, J.M. 1993. Stock discrimination using otolith shape analysis. Can. J. Fish. Aquat. Sci. **50**: 1062-1083.

Casselman, J.M., Collins, J.J., Crossman, E.J., Ihssen, P.E., and Spangler. G.R. 1981. Lake whitefish (*Coregonus clupeaformis*) stocks of the Ontario waters of Lake Huron. Can. J. Fish. Aquat. Sci. **38**: 1772-1789.

Cook, R.C. 1982. Stock identification of sockeye salmon (*Oncorhynchus nerka*) with scale pattern recognition. Can. J. Fish. Aquat. Sci. **39**: 611-617.

Crowe, W.R. 1962. Homing behaviour in walleye. Trans. Am. Fish. Soc. **91**: 350-354.

De Pontual, H., and Prouzet, P. 1987. Atlantic salmon, *Salmo salar* L., stock discrimination by scale-shape analysis. Aquacult. Fish. Manage. **18**: 277-289.

Ferguson, R.G., and Derksen, A.J. 1971. Migrations of adult and juvenile walleyes (*Stizostedion vitreum vitreum*) in southern Lake Huron, Lake St. Clair, Lake Erie, and connecting waters. J. Fish. Res. Bd. Canada **28**: 1133-1142.

Forney, J.L. 1963. Distribution and movement of marked walleye in Oneida Lake, New York. Trans. Amer. Fish. Soc. **92**: 47-52.

Hubbard, B.B. 1996. The world according to wavelets. A.K. Peters, Ltd., Wellesly, Mass.

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

Jarvis, R.S., Klodowski, H.F., and Sheldon, S.P. 1978. New method of quantifying scale shape and an application to stock identification in walleye (*Stizostedion vitreum vitreum*). Trans. Am. Fish. Soc. **107**: 528-534.

Kaiser, G. 1994. A friendly guide to wavelets. Birkhäuser, Boston.

Larkin, P.A. 1972. The stock concept and management of Pacific salmon. H.R. MacMillan lectures in fisheries, Univ. British Columbia, Vancouver.

Margraf, F.J., and Riley, L.M. 1993. Evaluation of scale shape for identifying spawning stocks of coastal Atlantic striped bass (*Morone saxatilis*). Fish. Res. **18**: 163-172.

Olson, D.E., and Scidmore, W.J. 1962. Homing behavior of spawning walleye. Trans. Am. Fish. Soc. **91**: 355-361.

Rawson, D.S. 1957. The life history and ecology of yellow walleye, *Stizostedion vitreum*, in Lac La Ronge, Saskatchewan. Trans. Amer. Fish. Soc. **86**: 15-37.

Richards, R.A., and Esteves, C. 1997b. Use of scale morphology for discriminating wild stocks of Atlantic Striped Bass. Trans. Am. Fish. Soc. **126**: 919-925.

Ricker, W.E. 1958. Maximum sustained yields from fluctuating environments and mixed stocks. J. Fish. Res. Bd. Canada. **15**: 991-1006.

Riley, L.M., and Carline, R.F. 1982. Evaluation of scale shape for the identification of walleye stocks from western Lake Erie. Trans. Am. Fish. Soc. **111**: 736-741.

Waldman, J.R., Grossfield, J., and Wirgin, I. 1988. Review of stock discrimination techniques for striped bass. N. Am. J. Fish. Manage. **8**: 410-425.

Waldman, J.R., Richards, R.A., Schill, W.B., Wirgin, I., and Fabrizio, M.C. 1997. An empirical comparison of stock identification techniques applied to striped bass. Trans. Am. Fish. Soc. **126**: 369-385.

Wolfert, D.R. 1963. The movements of walleyes tagged as yearlings in Lake Erie. Trans. Amer. Fish. Soc. **92**: 414-420.

CHAPTER TWO: Computer intensive analysis of stock differentiation based upon the Fourier decomposition of averaged scale outline signals.

Abstract

Signals created from averaging several walleye (*Stizostedion vitreum*) scales produced variables that formed significantly better discriminant functions for stock discrimination than the single scale signals traditionally used due to the variability of scale outline shapes from individual fish. Fourier analysis was used to quantify scale outline shape; the variables produced violated the discriminant analysis assumption of multivariate normality. A non-parametric randomization procedure, new to stock discrimination studies, was developed to test the significance of the discriminant functions formed. All signal comparison discriminant functions were significant. A specialized jackknife procedure was also developed to test for significant differences between the discriminant functions formed from the single scale and averaged scale signal variables. Based on the findings of this study, multiple scale signals should be considered in future stock discrimination studies and non-parametric statistics used to test for significance in instances where assumptions of the discriminant analysis are violated.

Introduction

Fourier analysis of scale outlines has been used successfully as a fast method of stock identification in a number of fish species: walleye (*Stizostedion vitreum*) (Jarvis et al. 1978), Atlantic salmon (*Salmo salar*) (De Pontual and Prouzet 1987), sockeye salmon (*Oncorhyncus nerka*) (Cook 1982), lake whitefish (*Coregonus clupeaformis*) (Casselman et al. 1981), and striped bass (*Morone saxatilis*) (Margraf and Riley 1993; Richards and Esteves 1997a; Waldman et al. 1997). A stock can be defined as an intra-specific group of randomly mating individuals with temporal or spatial integrity (Ihssen et al. 1981). Techniques such as electrophoretic analysis of serum proteins, cytogenetics, meristic or morphometric characteristics, physiology, behaviour, mineral compositions of flesh or hard parts, and markings tend to be labour intensive and produce results similar to scale morphology (Ihssen et al. 1981; Waldman et al. 1988).

Fourier analysis describes the periodic frequencies contained within a scale's outline by partitioning the signal into a series of either sine or cosine terms corresponding to different frequencies (Bird et al. 1986; Hubbard 1996). Each term is characterized by an amplitude value that measures the contribution of the term to the overall shape and phase angle which is a measure of the amount of rotation needed by an individual sine or cosine function to maximize its contribution. The n^{th} harmonic is composed of a figure with n "lobes", thus the first harmonic is an offset circle, the second a figure eight, the third a "three-leaf clover," etc. (Younker and Ehrlich 1977; Bird et al. 1986). The harmonic amplitude measures the contribution to overall shape, so elongate shapes tend to have large amplitude values for the second harmonic, and square forms tend to produce high amplitudes for the fourth harmonic. The summation of the first harmonics largely captures the gross shape of the scale outline; however, the importance of any given harmonic is represented by its amplitude, which may be very small if the harmonic is not present in the curve being described. The addition of successive harmonics adds increasing detail to the description of the scale outline and Fourier analysis is able to describe any two-dimensional signal completely given enough harmonics and corresponding phase angles (Bird et al. 1986).

Researchers were able to differentiate between groups of fish using Fourier analysis; however, the classification rates were considered too low for the methods to be applied in a routine assessment of fisheries. It was concluded that the methods held promise, but further investigation was necessary (Casselman et al. 1981; Riley and Carline 1982; De Pontual and Prouzet 1987; Margraf and Riley 1993; Waldman et al. 1997; Richards and Esteves 1997b). A refinement of methods that increased the ability to differentiate between stocks would improve the utility of Fourier analysis for management purposes.

From visual observations it is evident variability exists amongst the outlines of scales from a single fish. This implies that the discriminant functions formed may have reduced effectiveness due to uncontrolled variability based on the scale chosen to represent a fish. All previous studies have used only one scale from each individual to create a signal that the Fourier transform could analyze (Jarvis et al. 1978; Casselman et al. 1981; Riley and Carline 1982; De Pontual and Prouzet 1987; Margraf and Riley 1993; Richards and Esteves 1997a; Waldman et al. 1997).

The prime objective of this study was to investigate the effects of scale shape variability by selecting multiple scales from each fish so that averaged signals could be formed and analyzed by Fourier methods. The variables produced from Fourier analysis were entered into a stepwise discriminant analysis. However, some variables did not meet the assumption of normality. Thus, the secondary objective of this study was to develop non-parametric methods to test the significance of the discriminant functions formed. It was also necessary to develop non-parametric techniques that could be used to detect differences between discriminant functions formed with variables produced from the Fourier analysis of one scale and averaged scale signal. This was the first time computer intensive, non-parametric methods have been used to test the significance of the discriminant functions formed for stock discrimination using scale morphology.

Methods

Scales collected from walleye in two distinct Manitoba lakes were used to create single scale and averaged scale signals appropriate for analysis with the Fourier transform. The Manitoba Department of Conservation, formerly the Manitoba Department of Natural Resources, captured the walleye in May or June in 1997 on known spawning grounds (index samples) to determine the spawning condition of walleye and to collect basic stock assessment data. Scales were removed from sampled fish on the left side, above the lateral line and below the posterior end of the anterior dorsal fin for all fish. Spawning condition, sex, length and weight data were also collected. All fish used in this study were collected in 1997 near Easterville on Cedar Lake and Dauphin River on Lake Winnipeg (Figure 2.1). These sites are in excess of 200 km from one another by water. In addition, the movement of walleye between

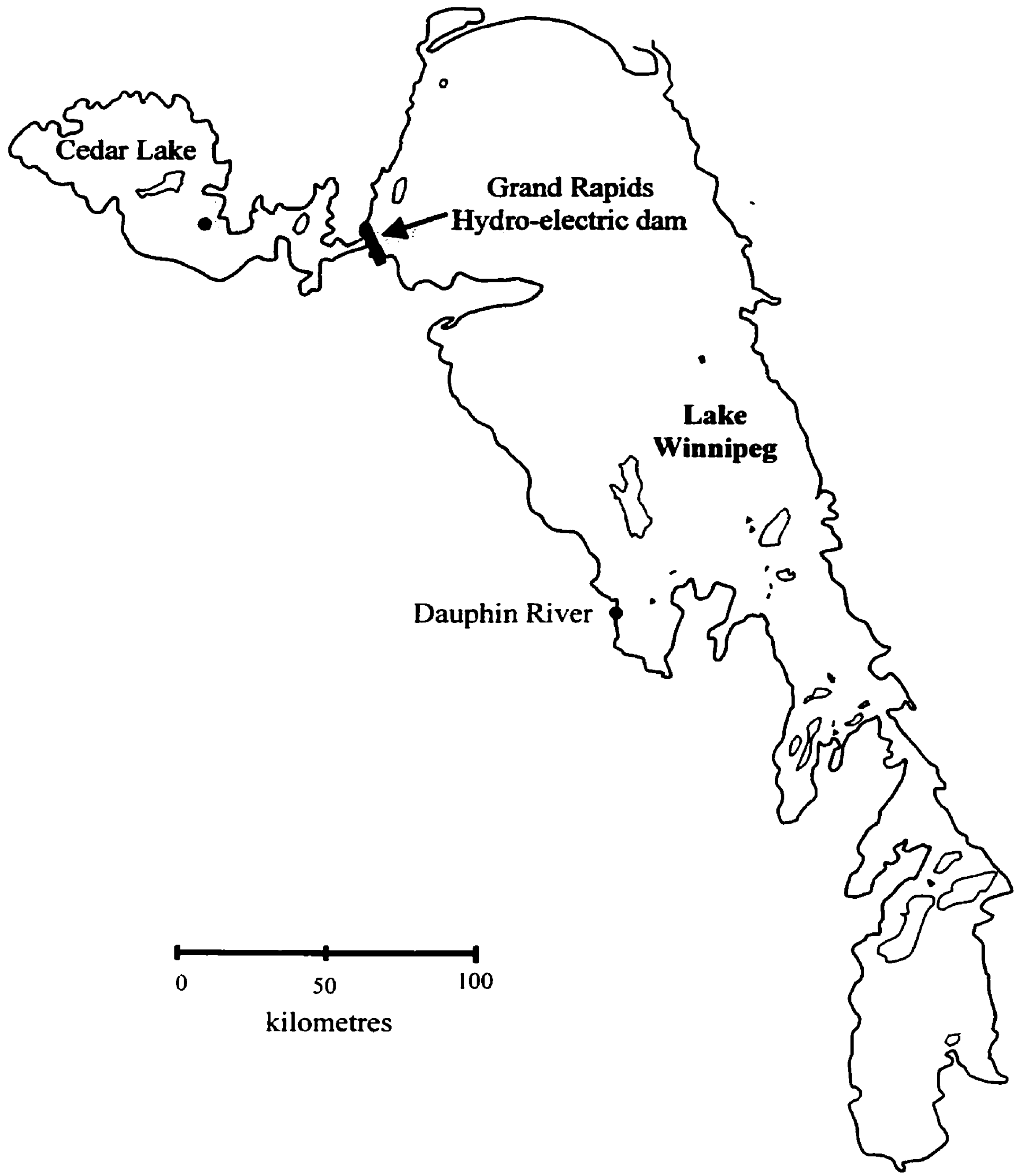


Figure 2.1. Locations of Lake Winnipeg and Cedar Lake index netting sites indicated by (•) from which scales samples were removed from walleye captured in 1997. The Grand Rapids hydroelectric generating station constructed in 1968 restricts fish movement between the lakes.

Cedar Lake and Lake Winnipeg has been restricted since 1968 when construction was completed on the Grand Rapids generating station. Walleye aged at three and four years were used in these analyses. Twenty-one fish, three years old (hatched in 1994) and 21 fish, four years old (hatched in 1993) from Cedar Lake were combined for a total of 42 fish and compared to 87 three year old fish (hatched in 1994) from Lake Winnipeg.

Three different methods were tried in an attempt to clean the scales. All three (ultrasonic jewelry cleaner, a small brush, and rubbing between the thumb and index finger) were found to damage the cteni portion of the scale, altering the outline shape. Therefore, only scales that had no visible residual tissue attached to the edges at 6.5X magnification were used in the analysis. Scales that were highly asymmetrical, damaged or regenerated were also rejected.

Appendix A provides a flowchart overview of the steps involved in this study to perform the analyses. Five scales from each fish were selected and ordered randomly exterior side down on a glass slide. A second slide was taped on top to flatten the scales and hold them in place. The prepared slides were placed on a flatbed scanner and covered by a sheet of black paper to enhance background contrast. They were then digitized as gray scale images at an optical resolution of 600 dpi (42.2 μ/pixel) (Figure 2.2a). The resulting images were digitally processed using the Scion Image® software package to produce scale outlines (Figure 2.2b) (Appendix 1). A standardized starting point was necessary for the meaningful use of phase angles from the Fourier analysis as descriptors of scale shape (Jarvis et al. 1978). The boundary between the circuli and the cteni on the left side of the scale image was used to begin the trace and convert the outline into X and Y coordinates.

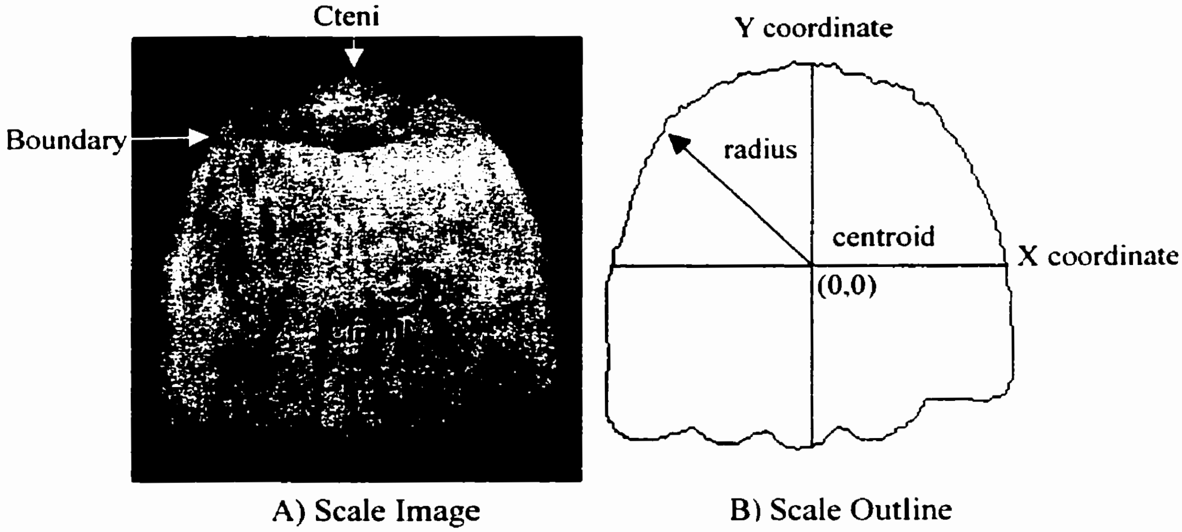


Figure 2.2. a) Grayscale bitmap scale image obtained from a 600 dpi flatbed scanner. Cteni, circuli and the boundary between them are indicated. b) Scale outline from original image that was defined in Scion Image®. The radius and centroid defined by the mean X and Y coordinates is indicated.

To investigate the effects of variability of scales from the same fish, averages of two, three, four and five individual scale signals were created that could be analyzed using the same methods as the single scale signals. The Fast Fourier Transform (FFT, Wei 1990) algorithm used in the analysis requires a data series length of an even power of two (e.g. 64, 128, 256, 512). Some scales produced outlines of less than 256 data points; therefore, all outlines were standardized to 128 data points at angular intervals of 2.8125° with a program written in Microsoft® Visual Basic for Applications (VBA) (Appendix 2). Radius measurements taken from the scale centroid (determined as the mean X and mean Y coordinates) to the scale edge generated a signal that could be analyzed by the FFT (Figure 2.3).

The signal's amplitude influences the harmonic amplitudes calculated by Fourier analysis. To correct for size differences each scale signal was divided by the mean radius length for that particular scale, removing size effects for scales from the same fish and amongst scales from different fish (Jarvis et al. 1978). An example of standardized signals from one scale and the average of five scales are displayed in Figure 2.3. A total of 64 harmonic amplitudes and 64 phase angles were calculated by the FFT algorithm and these variables were used to discriminate between the two groups of fish (Appendix 3). The total of 128 variables exceeded the number of fish (cases).

All previous studies of scale shape have limited the variables placed into the discriminant function to the first 10 to 20 harmonic amplitudes and phase angles or in some studies only the harmonic amplitudes. The criterion for using these variables was that they explained the majority of the variation in the scale shape (Jarvis et al. 1978;

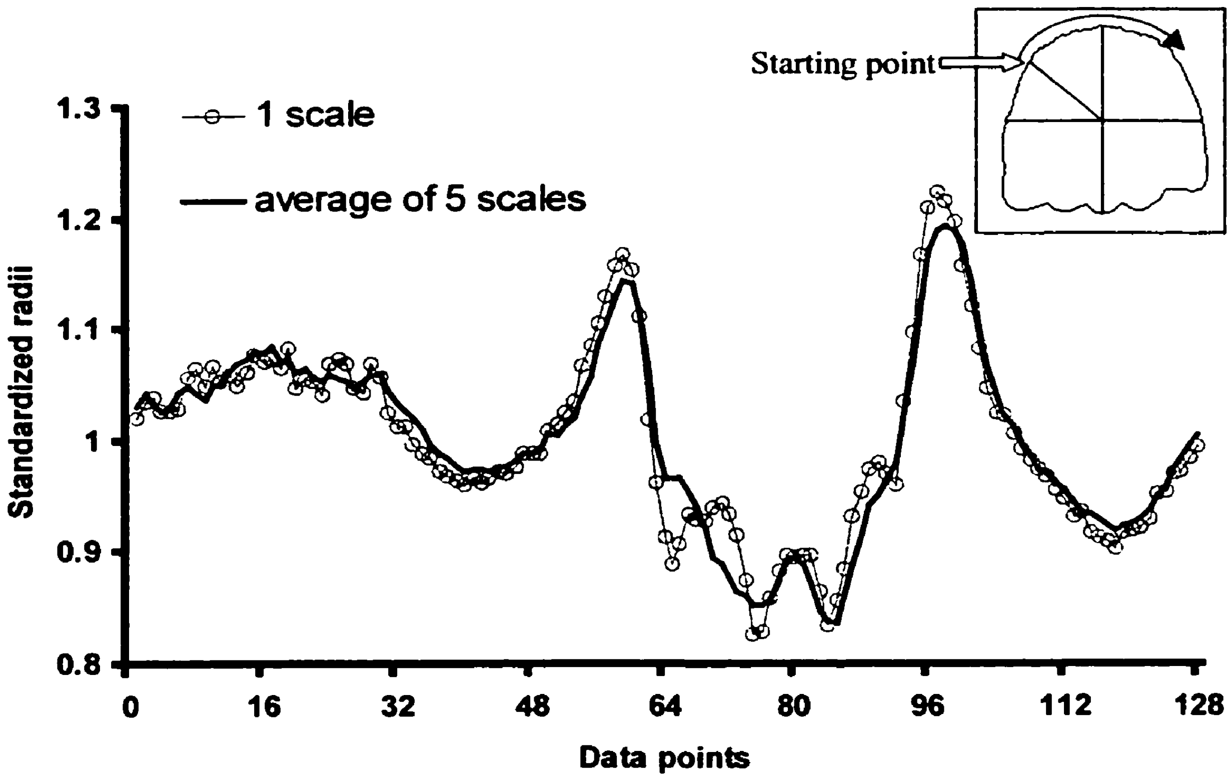


Figure 2.3. Plot of an interpolated 128 data point one scale and the average of five scale signals. Dividing by the mean radius measurement standardized the radii measurements. Included in the upper right is a scale outline displaying the point from which radii measurements are started and the direction of measurement.

Casselman et al. 1981; Riley and Carline 1982; Castonguay et al. 1991; Margraf and Riley 1993; Campana and Casselman 1993).

Ideally the number of variables used to form the discriminant function should be at least two to four times the number of cases (Lachenbruch 1975; Williams and Titus 1988). In this study, all 128 variables were considered in a stepwise discriminant analysis in SPSS® to select a subset of those variables that were best for differentiating between the sampling locations. To keep the number of variables selected to form the discriminant function equivalent to other studies (Jarvis et al. 1978; Casselman et al. 1981; Riley and Carline 1982; Castonguay et al. 1991; Campana and Casselman 1993; Kenchington and Full 1994) and two to four times greater than the number of cases, 20 steps were set as the maximum that could be performed by the procedure. At each step the variable that resulted in the smallest Wilks' lambda for the discriminant function was selected for entry, with the probability of *F*-to-enter set at 0.05 and the probability of *F*-to-remove set at 0.10. The discriminant function formed was used to classify the fish to either of the two groups, Cedar Lake or Lake Winnipeg. The discriminant technique used for classifying cases into groups was based on Bayes' theorem, where the probability that a case with a discriminant score of D belongs to groups *i* is estimated by:

$$P(G_i|D) = \frac{P(D|G_i)P(G_i)}{\sum_{i=1}^{g} P(D|G_i)P(G_i)}$$

where: G_i is the group and P is the probability of belonging to that group (Norusis 1990). The prior probability ($P(G_i)$) is an estimate of the likelihood that a case belonged to a particular group. The prior probabilities of all discriminant analyses were based on original group sample sizes. In this study, there were 42 cases in group one and 87 cases in group two, $P(G_1)$=0.3256=42/129 and $P(G_2)$=0.6744=87/129.

The discriminant function was evaluated by estimating its performance in the classification of future observations with a "leave-one-out" estimator (Lachenbruch 1975), analogous to jackknifing (Efron and Tibshirani 1993). Jackknifing deleted each case in turn from the sample and a discriminant function was calculated based on the remaining n-1 cases. The case that was left out was then classified with the new discriminant function. These steps were repeated for all cases to calculate the jackknifed classification rate. This method makes no assumptions about variable distributions and uses all available sub-samples to give an estimate for the expected actual error rate (Lachenbruch 1975), although a jackknife may underestimate the variance. All classification rates referred to in this study are jackknifed estimates.

Discriminant analysis is based on the assumptions that variables are a sample that is multivariate normal, and the covariance matrices of the variables used to form the discriminant function are homogeneous. The Lilliefors corrected Kolmogorov-Smirnov test was conducted on the individual variables for the five different scale signals. Thirty nine point seven percent of the variables were found to be non-normal. Since a large proportion of the variable sets are non-normal the assumption of multivariate normality of the combined variable sets is also violated. Box's M test was used to examine the assumption of homogeneity of covariance matrices, and is also sensitive to meeting the assumption of multivariate normality. The assumption of equality of covariance matrices was only violated for the average of five scales and low p-values were calculated for the average of three and four scale signals (Table 2.1). A discriminant

Table 2.1. Box's M test of the assumption of homogeneity of covariance matrices for the one scale and average of two, three, four and five scale signals harmonic amplitudes and phase angles. Asterisks (**) define significant values, $\alpha = 0.05$. Cedar Lake aged three and four fish (N = 42) and Lake Winnipeg aged three fish (N = 87).

	Box's M	F			
		Approximate	$d.f._1$	$d.f._2$	p-value
one scale	96.42	0.92	91	22290.52	0.688
two scale average	195.02	1.06	153	21930.70	0.299
three scale average	66.78	1.35	45	23243.19	0.059
four scale average	110.42	1.25	78	22441.50	0.070
five scale average	149.57	1.43	91	22290.52	0.005**

analysis can differentiate between groups even when the assumptions of multivariate normality are violated; however, the prediction rule will not be optimal as a discriminant analysis is based on linear relationships in the data (Lachenbruch 1975; Krzanowski 1977; Norusis 1990). In addition, the test of significance for the discriminant function, Wilks' Lambda, will be invalid.

A Monte Carlo randomization of the data was developed to test the statistical significance of the discriminant function because of the violation of multivariate normality (Appendix 4). Randomization tests make no assumption about the distribution function of the variables although they typically assume independence (Manly 1997). Cases from both of the groups in the discriminant analysis were randomly reassigned to a group, with the overall number of fish belonging to each group remaining the same as the original data (Manly 1997). Once all fish had been reassigned, a discriminant analysis was performed to form a new discriminant function and a jackknife classification rate was calculated. This classification rate could then be compared to the jackknifed classification rate of the original discriminant function. One thousand randomized classifications were generated. The results were removed from the text output file using a program written in Visual Basic for application (VBA) through Microsoft Excel® (Appendix 5). The p-value was calculated as the portion of randomizations whose classification rate was equal to or exceeded the value for the original analysis.

An additional quantitative test was developed to determine if variables produced from Fourier analysis of averaged scale signals formed discriminant functions that classify the sampled walleye from the two locations significantly better than the single scale variables. The variables for all scale signals (one scale and the average two, three, four, and five scales) were placed together in SPSS®. A VBA program had calculated a worksheet of random variables that could be used as a selection criterion in SPSS® (Appendix 6). This program randomly removed, or jackknifed, one case from both Cedar Lake and Lake Winnipeg (Appendix 7). The five different variable sets thus had the same cases removed. A separate discriminant analysis was then run for each of the five comparisons and a classification rate calculated for each. The difference in classification rates was calculated by comparing the classification rates. If the averaged classification rate was higher than the single scale classification rate the comparison scored a one. This difference was examined with a one-tailed significance test (H_o: difference ≤ 0). To obtain a precise p-value equivalent to $\propto = 0.05$, 1000 randomizations were conducted (without the repetition of the same two cases being removed) (Manly 1997). The averaged scale comparisons were also tested for significant differences between one another by comparing the classification rates for all other possible combinations.

Results

The discrimination of fish sampled from Cedar Lake and Lake Winnipeg had high classification rates for all scale signal comparisons (Table 2.2). The jackknifed classification rate generally increased with the number of scales forming the signal. The classification rates ranged from 86.82 to 94.57% with the maximum reached using the average of four and five scale signals. The classification rates for Cedar Lake were lower

Table 2.2. Stepwise discriminant analysis classification rates of Cedar Lake aged three and four fish and Lake Winnipeg aged three fish. Analyses were performed for the one scale, and the average of two, three, four and five scale signals. The harmonic amplitude and phase angles are listed in order of selection into the stepwise discriminant function. The percent (%) correct classification rate are included for both lakes separately, as well as an overall classification rate.

		Jackknife classification summary		
		Cedar Lake (N = 42)	Lake Winnipeg (N = 87)	
	Harmonic Amplitudes (A) or Phase Angles (P) in the discriminant function	%	%	Average %
one scale	A4, A6, A21, A9, P3, P5, P48, A57, A48, A31, A29, A24, A7	78.57	90.80	86.82
two scale average	A4, A6, P1, A21, A22, A54, P10, A18, P57, P51, P42, A31, A40, P58, P48, P45	85.71	95.40	92.25
three scale average	A4, A6, P1, A14, A3, A9, A20, A8, A1	83.33	93.10	89.92
four scale average	A4, A6, P3, A14, P1, A9, A3, P31, A42, P51, A19, P13	92.86	95.40	94.57
five scale average	A9, P3, A4, A6, A3, A27, A8, P7, A1, A23, A24, P57, A25	90.48	96.55	94.57

than the corresponding Lake Winnipeg classification rates for all comparisons. The difference was as high as 12.23% in the one scale comparison and as low as 2.54% for the average of four scales comparison (Table 2.2). These differences in the classification rates were consistent with the difference in the number of cases from each group, resulting in a corresponding higher prior probability of classifying fish as being from Lake Winnipeg.

As expected with high classification rates, a bimodal distribution of the discriminant function scores existed for the different signal comparisons. This is clearly seen in the discriminations based upon a single scale (Figure 2.4). Cedar Lake scores were normally distributed (Lilliefors corrected Kolmogorov-Smirnov test for normality, $p=0.657$) with discriminant scores ranging from –4.30 to 0.20. The Lake Winnipeg discriminant scores were not normally distributed (Lilliefors corrected Kolmogorov-Smirnov test for normality, $p=0.045$) and had a larger discriminant score range of –2.14 to 3.52.

The number of variables selected to form the discriminant function varied from 9 to 16 for the five different signal comparisons (Table 2.2). More than two thirds of the variables selected to form the discriminant functions were amplitudes. The number of variables selected by the stepwise procedure was unrelated to the classification rates (Spearman's correlation coefficient =0.026, $p=0.966$).

The Monte Carlo randomization provided an estimate of the significance of the discrimination results. The randomization tests found all discriminant functions were significant ($p=0.004$ to $p<0.001$, Table 2.3). Only the one scale signal comparisons had a p-value >0.001. The distributions of the 1000 classification rates from the

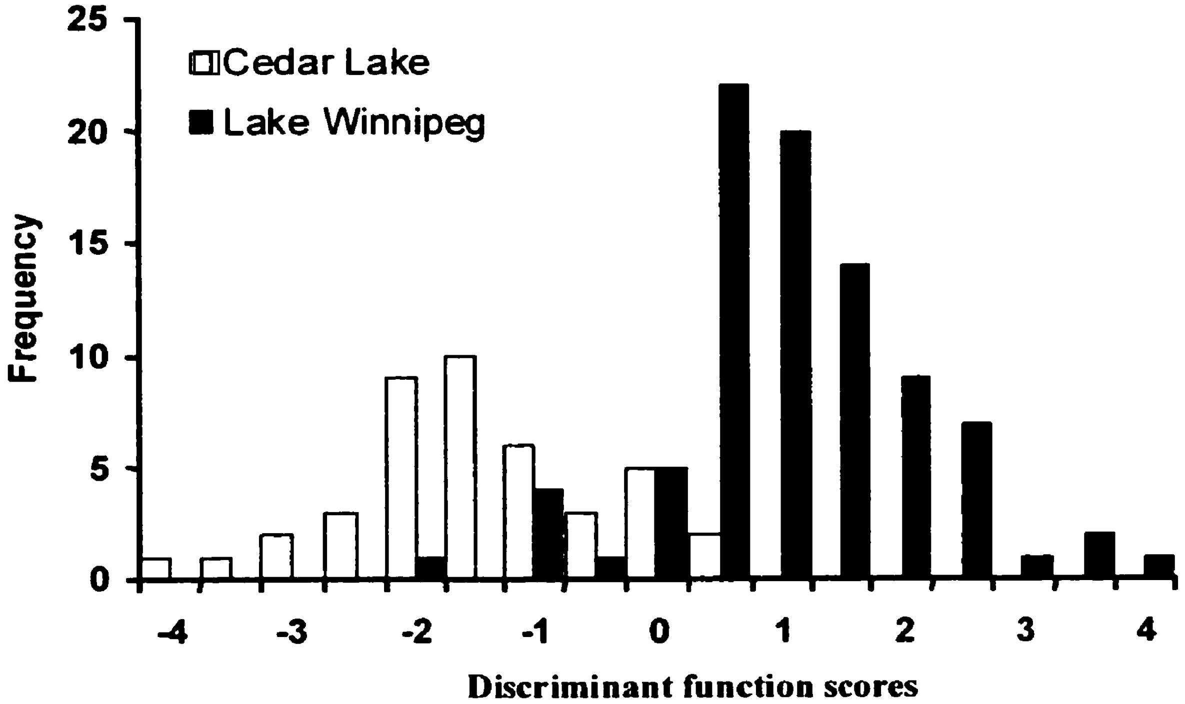


Figure 2.4. Frequency distributions of discriminant scores from the Cedar Lake (N = 42) and Lake Winnipeg (N = 87), one scale signal comparison. This comparison was chosen as it had the highest degree of overlap in discriminant scores. The lowest scores for Cedar Lake and Lake Winnipeg were –4.30 and –2.14, respectively. The highest scores for Cedar Lake and Lake Winnipeg were 0.20 and 3.52, respectively. The Cedar Lake discriminant scores are normally distributed (Lilliefors corrected Kolmogorov-Smirnov test, $p = 0.657$), but, the Lake Winnipeg discriminant scores are not (Lilliefors corrected Kolmogorov-Smirnov test for normality, $p = 0.045$).

Table 2.3. Monte Carlo randomized results of 1000 random shuffles testing the significance of the discriminant functions formed for the one scale and average of two, three, four and five scale signals. The minimum, maximum, average classification rates of the randomization's and the original data's discriminant function are included as well as the calculated p-value. Asterisks (**) define significant values, $\alpha = 0.05$. Cedar Lake aged three and four fish N = 42 and Lake Winnipeg aged three fish N = 87.

	% correct classification				
	Randomizations			Original data	
	Average	Minimum	Maximum	Actual calculated	p-value
one scale	74.51	63.57	92.25	86.80	0.004**
two scale average	74.57	63.57	89.15	92.25	<0.001**
three scale average	74.66	62.02	87.60	89.90	<0.001**
four scale average	74.46	62.02	88.37	94.57	<0.001**
five scale average	74.64	61.24	89.15	94.57	<0.001**

randomization were not normally distributed (Lilliefors corrected Kolmogorov-Smirnov test for normality, $p<0.001$). The highest correct classification rate for the Monte Carlo randomization's varied from 87.60 to 92.25% for the signals analyzed. The average classification rate varied over the narrow range of 74.46 to 74.66% for all comparisons with the lowest correct classification always above 61%.

The significance of the difference in the discriminant ability of the Fourier variables produced from one scale and averaged scale signals was tested with the specialized jackknife procedure that removed one case from both groups. The average of two, four and five scale signals had significantly higher classification rates than the one scale signal (Table 2.4). The 1000 differences in the classification rates of the average of four scales and the single scale signals comparison are displayed in Figure 2.5. None of the averaged scale comparisons were significantly different from one another. The distribution of the differences in the classification rates were not normally distributed for this and all other signal comparisons (Lilliefors corrected Kolmogorov-Smirnov $p<0.001$).

Discussion

This work represents the first time averaged scale signals have been used to create discriminant variables with Fourier analysis. Averaged scale signals provided significantly higher classification rates than the single scale traditionally used to represent an individual fish in studies discriminating between fish stocks. The results demonstrate the potential for improved discrimination of fish stocks through scale

Table 2.4. Resulting p-values calculated from the 1000 differences in the classification rates from the jackknife test for significant differences between the one scale and averaged scale signals. Sample sizes for Cedar Lake and Lake Winnipeg were N = 42 and N = 87 respectively. Asterisks (**) define significant values, one-tailed test, $\alpha = 0.05$.

	one scale	two scale average	three scale average	four scale average	five scale average
one scale		0.985**	0.879	0.983**	0.999**
two scale average			0.134	0.328	0.749
three scale average				0.175	0.948
four scale average					0.819
five scale average					

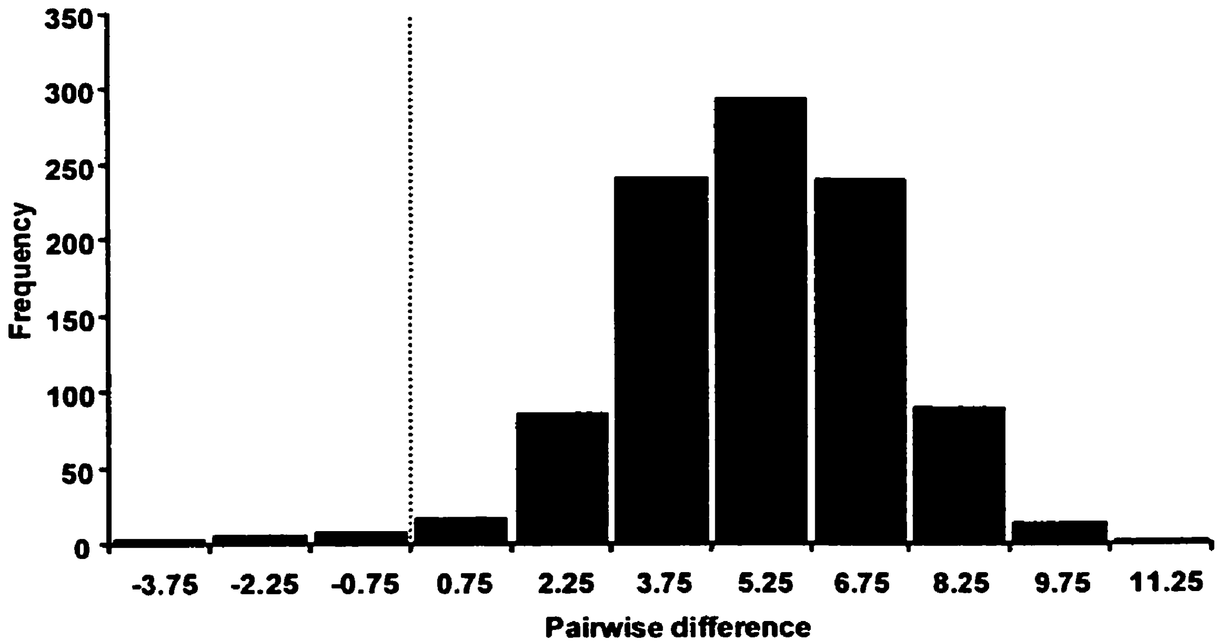


Figure 2.5. Histogram plot of the differences in classification rates for the one scale signal subtracted from the averaged four scales signal for the 1000 jackknifed classification rates. The dashed line indicates a pairwise difference of zero. For this comparison, the calculated p = 0.017.

averaging as the observed jackknifed classification rates for the average of two, three and five scales were higher than the classification rates from all previous studies. Jarvis et al. (1978) was the first to use Fourier analysis, and had discriminant function classification rates of 80% of for Lake Erie walleye. Margraf and Riley (1993) and Richards and Esteves (1997b) studied Atlantic striped bass and had classification rates as high as 78% and 84% respectively. Casselman's et al. (1981) study of lake whitefish had a classification rate as high as 85.3%. Riley and Carline (1982) obtained the highest classification rate in a similar study that compared walleye from two isolated populations (Escanaba Lake and Lake Erie), correctly classifying 92% of the fish.

The discriminant function relies on variability in the characteristics studied to define the groups. Since an increase in classification rates occurred with averaging, a reduction in the overlap of the ranges of variability between the groups must have occurred. When a scale outline is averaged, the higher amplitude harmonics representing the more detailed features tend to be cancelled out, whereas the lower frequencies that describe the general shape of the signal remain relatively intact. Therefore, averaging appears to reduce discrimination errors that can be caused by individual scale variability.

The early selection of low order harmonics for entry into the discriminant function agreed with the results of Waldman et al. (1997) and Richards and Esteves (1997b), indicating the importance of gross shape characteristics in distinguishing amongst groups of fish. The fourth harmonic amplitude was entered first and the sixth harmonic amplitude second for all stepwise discriminant analysis comparisons with the exception of the average of five scales signal comparison (Table 2.2). Both the fourth and sixth harmonic amplitudes represent gross shape of the scale outline. The fourth harmonic amplitude is a measure of the squareness of the scale signal and the sixth harmonic represents a 'six-leaf clover' (Bird et al. 1986). The early entry of these harmonics indicates that the discrimination is based heavily upon variation in the contribution of these scale forms from the different lakes.

All studies of scale shape have limited the variables placed into the discriminant function to the first 10 to 20 harmonic amplitudes and phase angles, or in some studies, to only the amplitudes. Waldman et al. (1997) and Richards and Esteves (1997b) further refined the variable selection process by using a MANOVA to identify which of the first 12 harmonics and phase angles had significant stock of origin effects and then placed those variables in a discriminant function. De Pontual and Prouzet (1987) entered the first eight harmonics and phase angles into a stepwise discriminant analysis. Analyses conducted by Kenchington and Full (1994) on sea scallop (*Placopecten magellanicus*) found the harmonics with significant stock of origin effects varied amongst the 24 harmonics amplitudes they considered ranging from the second to the twenty-fourth. Bird et al. (1986) studied herring otoliths to investigate significant harmonic amplitudes and Richard and Esteves (1997b) studied striped bass scales analyzed by Fourier analysis to discriminate between stocks, both these studies had results similar to Kenchington and Full (1994). This is not surprising as a harmonic may explain much of the shape of the scale outline, but this does not guarantee the harmonic will be a good discriminator. Only stock specific differences in the range of variation in the variables is important for differentiation.

This study was distinctly different from these previous studies in that I initially considered all 64 harmonic amplitudes and 64 phase angles produced from the Fourier transform in the stepwise procedure, limiting the final number of variables as previously described. The variables selected to form the discriminant functions varied for the different averaged signal considered. Many of the higher order harmonic amplitudes and phase angles were selected to form discriminant functions but not until later in the stepwise procedure. These variables undoubtedly refine the function but are not as vital to discrimination.

The drawback to using larger variable sets is that the probability of obtaining higher classification rates by chance increases. However, the number of variables initially available to discriminate between the groups is accounted for by the randomization test. Therefore, the test of significance for the discriminant functions is valid. The average classification rate calculated from the 1000 randomizations is equivalent to the probability of correct classification by chance. The average randomized classification rates were at least 0.184 higher than the 0.561 probability expected based on the number of cases in the groups (Table 2.3). This is because a stepwise discriminant analysis selects variables that maximize differences between the groups and will most often perform better than chance even when there is no true basis for discrimination. In spite of this trend, the randomization test indicated that effective discriminant functions were generated by my procedure. These results are specific to the number of cases from each group and the number of variables initially used will need to be reassessed with new data sets.

The jackknife test for significant differences between the one scale and averaged scale comparisons found that averaging resulted in significantly improved classification rates. These results should however, be interpreted with some caution, as the jackknife tends to underestimate the variance. This means the test will tend to overestimate the significance level of the one tailed test (Efron and Tibshirani 1993). A simple bootstrap of the cases, which is generally a less biased method, was not a valid alternative to the jackknife because it would increase the amount of similarity within the randomly generated groups. This would result from sampling cases with replacement. The increase in similarity between the groups means the discriminant analysis would produce artificially high classification rates.

The negative aspect to using averaged scale signals is an increase in scale processing time. Using multiple scales to represent a fish took approximately one hour to mount, scan, obtain the outline images, define the X and Y coordinates and calculate the radius measurements of six fish using five scales per fish compared to the 20 minutes it would take to do the same number of fish with using only one scale. However, considering that stock identification can be critically important to fisheries management, I feel that classification rates should not be sacrificed solely because of time considerations.

A number of factors could not be investigated in this study due to sample size constraints. Age-classes have been found to differ significantly for otolith shape in a number of fish species (Bird et al. 1986; Castonguay et al. 1991; Campana and Casselman 1993). Year-class and sex effects were also found in Castonguay's et al. (1991) study of Atlantic Mackerel and Kenchington and Full's (1994) study of scallop shell shape for stock discrimination. Significant year-class effects were discovered for scales from Atlantic striped bass (Richards and Esteves 1997b). No sex, age-class or year-class effects were found to be significant in Riley and Carline's (1982) evaluation of scales for stock discrimination of Lake Erie walleye. If samples sizes are large enough sex, age-class and year class effects should be investigated as they may impact discrimination results.

Non-parametric computer intensive statistics confirmed that averaged signals can produce significantly higher classification rates. Future research should focus on walleye or other fish species with overlapping distributions to determine if the averaged signals result in improved discrimination in more complex situations, where stocks are mixing. Ideally, a reevaluation of past studies with available scale samples would be the simplest way to further examine the benefits to signal averaging allowing comparisons to published results.

Literature Cited

Bird, J. L, Eppler, D.T., and Checkley, D.M., Jr. 1986. Comparisons of herring otoliths using Fourier series shape analysis. Can. J. Fish. Aquat. Sci. **43**: 1228-1234.

Campana, S.E., and Casselman, J.M. 1993. Stock discrimination using otolith shape analysis. Can. J. Fish. Aquat. Sci. **50**: 1062-1083.

Casselman, J.M., Collins, J.J., Crossman, E.J., Ihssen, P.E., and Spangler. G.R. 1981. Lake whitefish (*Coregonus clupeaformis*) stocks of the Ontario waters of Lake Huron. Can. J. Fish. Aquat. Sci. **38**: 1772-1789.

Castonguay, M., Simard, P., and Gagnon, P. 1991. Usefulness of Fourier analysis of otolith shape for Atlantic mackerel (*Scomber scombrus*) stock discrimination. Can. J. Fish. Aquat. Sci. **48**: 296-302.

Cook, R.C. 1982. Stock identification of sockeye salmon (*Oncorhynchus nerka*) with scale pattern recognition. Can. J. Fish. Aquat. Sci. **39**: 611-617.

De Pontual, H., and Prouzet, P. 1987. Atlantic salmon, *Salmo salar* L., stock discrimination by scale-shape analysis. Aquacult. Fish. Manage. **18**: 277-289.

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

Hubbard, B.B. 1996. The world according to wavelets. A.K. Peters, Ltd., Wellesly, Mass.

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

Jarvis, R.S., Klodowski, H.F., and Sheldon, S.P. 1978. New method of quantifying scale shape and an application to stock identification in walleye (*Stizostedion vitreum vitreum*). Trans. Am. Fish. Soc. **107**: 528-534.

Kenchington, E.L., and Full, W.E. 1994. Fourier analysis of sea scallop (*Placopecten magellanicus*) shells in determining population structure. Can. J. Fish. Aquat. Sci. **51**: 348-356.

Krzanowski, W.J. 1977. The performance of Fisher's linear discriminant function under non-optimal conditions. Technometrics, **19**: 191-200.

Lachenbruch, P.A. 1975. Discriminant analysis. Hafner Press, A Division of Macmillan Publishing Co., Inc., New York.

Manly, B.F.J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. 2nd ed. Chapman & Hall, London.

Margraf, F.J., and Riley, L.M. 1993. Evaluation of scale shape for identifying spawning stocks of coastal Atlantic striped bass (*Morone saxatilis*). Fish. Res. **18**: 163-172.

Norusis, M.J. 1990. SPSS advanced statistics user's guide. SPSS, Chicago.

Richards, R.A., and Esteves, C. 1997a. Stock-specific variation in scale morphology of Atlantic striped bass. Trans. Am. Fish. Soc. **126**: 908-918.

Richards, R.A., and Esteves, C. 1997b. Use of scale morphology for discriminating wild stocks of Atlantic Striped Bass. Trans. Am. Fish. Soc. **126**: 919-925.

Riley, L.M., and Carline, R.F. 1982. Evaluation of scale shape for the identification of walleye stocks from western Lake Erie. Trans. Am. Fish. Soc. **111**: 736-741.

Waldman, J.R., Grossfield, J., and Wirgin, I. 1988. Review of stock discrimination techniques for striped bass. N. Am. J. Fish. Manage. **8**: 410-425.

Waldman, J.R., Richards, R.A., Schill, W.B., Wirgin, I., and Fabrizio, M.C. 1997. An empirical comparison of stock identification techniques applied to striped bass. Trans. Am. Fish. Soc. **126**: 369-385.

Wei, W.W.S. 1990. Time series analysis. Addison-Wesley Publishing Company, New York.

Williams, B.K., and Titus, K. 1988. Assessment of sampling stability in ecological applications of discriminant analysis. Ecology. **69**: 1275-1285.

Younker, J.L., and Ehrlich, R. 1977. Fourier biometrics: harmonic amplitudes as multivariate shape descriptors. Syst. Zool. **26**: 336-342.

CHAPTER THREE: A comparative study of Fourier and wavelet signal processing methods applied to scale shape signals for stock discrimination of Lake Winnipeg walleye.

Abstract

I introduced wavelet analysis, an alternate signal processing method to Fourier analysis, to provide variables from scale morphology that discriminate among different spawning aggregations of fish. Fourier and wavelet analyses transformed averaged walleye (*Stizostedion vitreum*) scale outline signals. The variables produced violated the discriminant analysis assumption of multi-variate normality. Therefore, non-parametric computer intensive statistics were used to assess the significance of the discriminant functions. Age-class effects were significant for one of the sampling locations, restricting all subsequent analyses by age-class. Variables from wavelet decompositions formed significantly better discriminant functions than Fourier analysis variables for most comparisons. There was no detectable difference between the different orders of wavelets used to analyze the signals. Finally, I examined the ability of a discriminant function formed from the spawning aggregations to classify commercially caught walleye of the same year-class. The majority of the commercial samples were classified as fish from only one of the spawning areas, raising concerns regarding the application of stock discrimination methods in Lake Winnipeg.

Introduction

In fisheries research and management a stock can be assumed to loosely represent populations that are either temporally or spatially isolated from conspecific spawning aggregates (Waldman et al. 1988). Stocks often mix at times when spawning is not occurring, confounding measures of growth, survival, and reproductive success. Samples from mixed stocks can invalidate studies of fish biology, population dynamics, and most estimates of yield (Campana and Casselman 1993). Also, if managers do not take into account the contribution and productivity of the individual stocks being harvested, some stocks may be over harvested. This can lead to their depletion and an eventual decline in the productivity of the entire fishery (Ricker 1958). Identifying the stocks that compose a fishery allows their contribution to the fishery and population status to be monitored.

Walleye (*Stizostedion vitreum*), can be divided into stocks within most lakes or river systems they occur as individual fish are associated with a particular spawning area, to which they return annually (Crowe 1962; Olson and Scidmore 1962). The post spawning movements of individual walleye can be extensive (Ferguson and Derksen 1971). Typically, identification of individual stocks is based upon markings, behavioural characteristics, meristics, calcified structures, cytogenetic characters or biochemical characters (Ihssen et al. 1981; Casselman et al. 1981; Waldman et al. 1988). Scale outline signals transformed with Fourier analysis have proven to be a fast, cost effective, and accurate identification method successful for the stock identification of walleye (*Stizostedion vitreum*) (Jarvis et al. 1978), lake whitefish (*Coregonus clupeaformis*) (Casselman et al. 1981), sockeye salmon (*Oncorhyncus nerka*) (Cook

1982), Atlantic salmon (*Salmo salar*) (De Pontual and Prouzet 1987), and striped bass (*Morone saxatilis*) (Margraf and Riley 1993, Richards and Esteves 1997a).

Fourier analysis describes shape quantitatively by partitioning a two dimensional signal into a series of either sine or cosine terms corresponding to different periodic frequencies contained within the signal (Bird et al. 1986; Hubbard 1996). Each term is considered a harmonic, characterized by an amplitude that measures the contribution of an individual sine or cosine to the overall shape and a phase angle that measures the amount of rotation needed by the sine or cosine function to maximize its contribution. Lower order harmonics explain gross shape, higher order harmonics add detail to the description of the signal. Given enough harmonics and corresponding phase angles, Fourier analysis is able to describe a two-dimensional signal completely (Bird et al. 1986). Since, the harmonic amplitudes and phase angles are specific to a signal; similarities within groups of signals can be used to differentiate between stocks.

There are theoretical problems however, with the application of Fourier analysis to irregular, non-periodic signals. Fourier analysis is inefficient in dealing with localized frequency variation or discontinuities, yet in signal processing, these signals often carry the information that allows differentiation of signals (Kaiser 1994; Hubbard 1996). Calculating the phase angles for the harmonic amplitudes can extract some of the temporal information; however, information about one instant in the signal is dispersed throughout all the frequencies of the entire Fourier transform. This is a serious drawback, as a local characteristic of the signal becomes a global characteristic of the transform (Hubbard 1996). Additionally, Fourier analysis may be able to describe the frequency information of a signal, but very different signals can have near identical values calculated for the coefficients if the original signal is not periodic. These combined deficiencies potentially limit the ability of Fourier analysis harmonic amplitudes and phase angles to represent complex scale outlines for discrimination.

Wavelet analysis, an alternative signal processing method, is suited to describing the frequencies of highly non-stationary signals with sudden peaks or discontinuities that may be poorly represented by Fourier analysis (Bradshaw and Spies 1992; Hubbard 1996). Scientists conducting biological research have used wavelets as a method to automate the identification of individual sperm whales using images of the trailing edges of their flukes (Huele and Udo de Haes 1998). Wavelets have also been used to identify spatial structure in transect data of plant communities (Bradshaw and Spies 1992; Dale and Mah 1998).

Wavelets are localized waves, that persist for only one or a few cycles, and unlike Fourier sines and cosines that oscillate forever, drop to zero in amplitude (Figure 3.1) (Strang and Nguyen 1997). The support of a wavelet is the region over which it is non-zero (Aboufadel and Schlicker 1999). Wavelet decomposition varies the size of the wavelet so that a narrow time-window is used to examine the high frequency components and a wide time-window is used to examine the low-frequency components (Ogden 1997; Chui 1997). This is called multiresolution analysis; meaning signals are analyzed at different levels of spatial scale, and is designed so that high frequency signal components are resolved better in terms of their location, and low frequencies values are better estimated but poorly localized (Hubbard 1996; Ogden 1997). The wavelet variables measure the correlation, or agreement, between the wavelet, with its peaks and

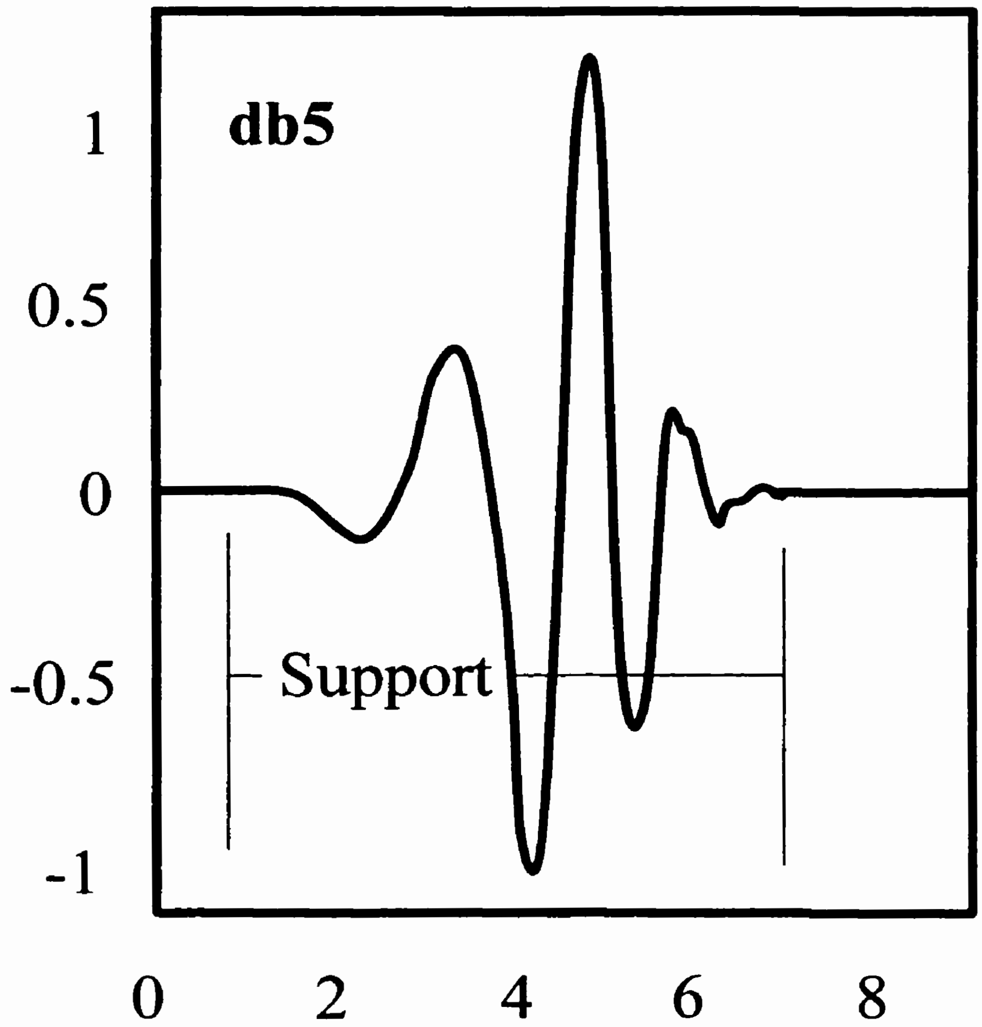


Figure 3.1. Daubechies five wavelet. The support (region over which it is nonzero) of the wavelet is indicated.

valleys and the corresponding segment of the signal. A high correlation means the segment of the signal is similar to the wavelet (Hubbard 1996). The variables produced from wavelets can be used to reconstruct the original signal by adding together wavelets of different sizes, at different positions, just as one can construct a signal by adding together sines and cosines of Fourier analysis.

My goal in this study was to compare the discriminatory ability of Fourier and wavelet variables. Based upon the limitations of the Fourier transform to describe non-periodic signals effectively, wavelets are expected to improve our ability to distinguish between walleye stocks. By refining scale morphology analysis I hope to provide managers and researchers with a cost-effective tool to aid them in stock discrimination.

Methods

Scales collected from Lake Winnipeg walleye were used to create an average of four scale signals appropriate for the subsequent analyses. The Manitoba Department of Conservation, formerly the Manitoba Department of Natural Resources, have annually collected walleye from Lake Winnipeg for the last two decades. Nets are set in May or June on known spawning grounds near Grand Rapids, Matheson Island, and Riverton (Figure 3.2) to collect index samples used to determine the spawning condition of walleye and to collect basic stock assessment data. Walleye were also sampled from the commercial catch in 1999, October 7-15, from fish delivered whole to Norway House, Frog Bay, and Selkirk and then shipped to the Freshwater Fish Marketing Corporation in Winnipeg, MB. Scales were removed from the left side, above the lateral line and below

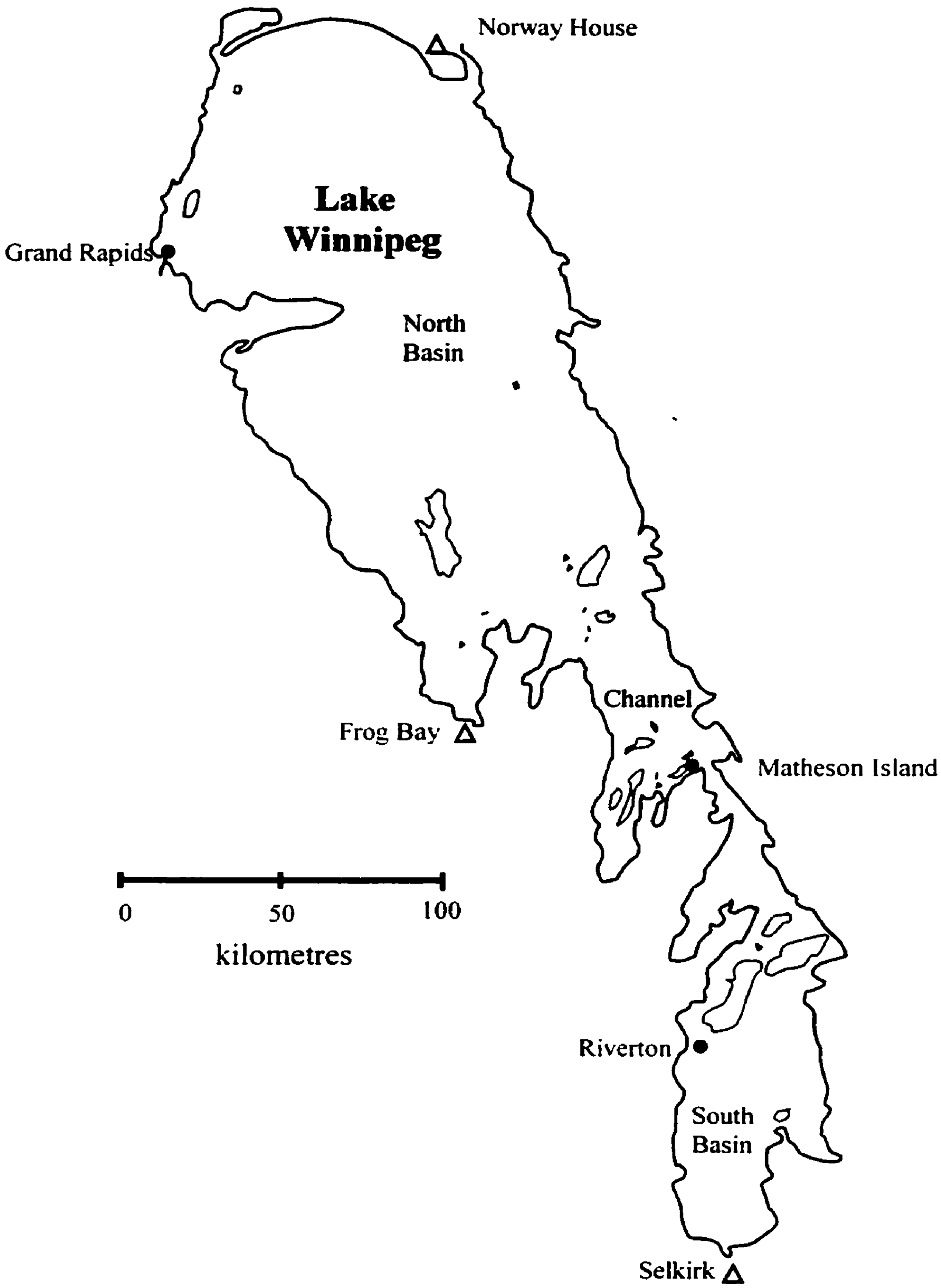


Figure 3.2. Locations of Lake Winnipeg index sampling sites (●) in 1997 and commercial delivery points (Δ) in 1999 from which walleye were sampled. The north basin, channel and south basin regions of the lake are indicated.

the posterior end of the anterior dorsal fin. Sex, length, weight, spawning condition and maturity data were also collected.

Sample sizes of fish from the index samples were not sufficient to limit the fish included in the analysis to those that were sexually mature (Table 3.1). The majority of the fish from the Matheson Island and Riverton sites that are the closest to one another, approximately 100km, were sexually mature. The Grand Rapids fish were largely immature. There are however, age at maturity differences among sample locations, an indication of stock specific differences between the sample sites (Ricker 1958).

Appendix A provides a flowchart overview of the steps involved in this analysis. Four scales from each fish were selected and placed randomly with regard to order exterior side down on a glass slide. Scales that were highly asymmetrical, damaged, regenerated or had tissue attached near the edges that was not easily removed were rejected. The scales were covered with another slide and the slides were taped together to flatten and hold the scales in place. The slides were placed on a flatbed scanner, covered by a sheet of black paper to enhance background contrast, and digitized as gray scale images at an optical resolution of 600 dpi (42.2 μ•pixel). The Scion Image® software package was used to reduce the scale images to outlines (Appendix 1). The boundary between the circuli and the cteni on the left side of the scale outline was used as a standard starting point to begin a perimeter trace that converted the outline to X and Y co-ordinates. This allowed the meaningful use of phase angle descriptors from Fourier analysis (Jarvis et al. 1978). It is also necessary for consistent analysis of wavelet coefficients (Hubbard 1996).

Table 3.1. Sample sizes of fish from the index sampling locations. Samples are differentiated by age, sex and maturity status.

	Aged four fish			Aged five fish			
	Grand Rapids	Matheson Island	Riverton	Grand Rapids	Matheson Island	Riverton	Total
Male Immature	21	8	2	12	6	0	49
Male Mature	0	39	71	1	57	223	391
Female Immature	43	21	10	59	9	9	151
Female Mature	1	2	3	12	3	8	29
Total	65	70	86	84	75	240	

The Fast Fourier Transform (Wei 1990) and the discrete wavelet transform (Hubbard 1996) algorithms used in the analyses required a data series where length was an even power of two (e.g. 32, 64, 128, 256). The scale centroid was determined as the mean X and Y coordinates and radius measurements were taken from the centroid to the scale edge. All outlines were interpolated to 128 radii measurements (Appendix 2).

Both Fourier and wavelet analyses are influenced by signal amplitude (Jarvis et al. 1978; Hubbard 1996). All scale signals were standardized prior to analysis by dividing each radius measurement by the mean radius length for the particular scale. This removed amplitude effects for both scales from the same fish and amongst scales from different fish (Jarvis et al. 1978). Based on previous work (see chapter 2), the average of four scale signals was used for all analyses.

Discrete wavelet analysis was chosen over continuous wavelet analysis to transform the scale signals. A continuous wavelet analysis would produce thousands of coefficients for a 128 data point signal. By choosing scales and positions based on powers of two the discrete wavelet analysis is more computationally efficient and just as accurate (Misiti et al. 1996). Discrete wavelet analysis passes the signal through two complementary filters and it emerges as two signals. The low frequencies represented by the approximations and the high frequencies represented by the details. These signals are then downsized, where every second data point is removed. The successive approximations are decomposed in turn to yield additional approximations and details that are downsized. One signal is broken down into many lower-resolution components. The final decomposition consists of a single data point for the approximations and details. From a 128 data point signal, seven levels of decomposition generate the wavelet variables.

There are numerous wavelet families capable of performing discrete wavelet transforms and within each family there are at least several orders of wavelets. To simplify the analysis I selected three orders (one, five, nine) within the Daubechies (db) wavelet family to perform the signal analyses. Daubechies wavelets were chosen as the different orders within the family resembled segments of the scale outline signal, an important criteria for choosing the wavelet family to use in the analyses (Hubbard 1996).

MATLAB® was used to perform both the fast Fourier and discrete wavelet transforms on the scale signals in combination with Excel Link® (Appendix3; Appendix 8). From the 128 data points, 64 harmonic amplitudes and phase angles were calculated for the Fourier transform. The number of variables that were calculated by the discrete wavelet transforms varied depending on the order of wavelet used (128 for db1, 185 for db5 and 241 for db9). This is from the filtering process, which convolutes the original signal within the filter, introducing several extra samples in some cases into the decomposition results (Misiti et al. 1996).

All variables produced from either the Fourier or wavelet transform were entered in a stepwise discriminant analysis in SPSS® to select a subset of those variables that was best for differentiating between the index sampling locations. In this paper, whenever a discriminant analysis is used, it refers to the result of a stepwise discriminant procedure. At each step, the variable that resulted in the smallest Wilks' lambda was selected for entry, with the probability of *F*-to-enter set at 0.05 and the probability of *F*-to-remove set at 0.10. Twenty was set as the maximum number of steps the procedure would perform. This kept the number of variables selected to form the discriminant function similar to other studies (Jarvis et al. 1978; Casselman et al. 1981; Riley and Carline 1982; Castonguay et al. 1991; Campana and Casselman 1993; Kenchington and Full 1994) and the number of cases at least two to four times greater than the number of variables (Lachenbruch 1975; Williams and Titus 1988).

To evaluate the discriminant functions performance in the classification of future observations the leave-one-out estimator (Lachenbruch 1975) or jackknifing (Efron and Tibshirani 1993) was used. Jackknifing estimated misclassification rates for all discriminant functions in this study, whenever a classification rate is reported or discussed, it is a jackknifed classification rate.

A large proportion of the variables produced from the Fourier and wavelet transforms of the scale outline signals were not normally distributed, based on a Lilliefors corrected Kolmogorov-Smirnov test with $\alpha = 0.05$ (Table 3.2). A discriminant analysis can differentiate between groups even when the assumptions of multivariate normality are violated, although the discriminant functions decision rule will not be optimal and the test of significance for the discriminant function is invalid (Lachenbruch 1975; Krzanowski 1977). Therefore, a Monte Carlo randomization, which makes no assumption of the underlying distribution of the data, was used to test the statistical significance of the discriminant functions formed (Appendix 4). Cases from each index sampling location were randomly reassigned to a group, with the overall number of fish belonging to each location remaining the same as the original data sets (Manly 1997). Once all fish had been reassigned, a discriminant analysis was performed to form a new discriminant function. One thousand repeated randomizations of the original data with a

Table 3.2. Proportion of Fourier, db1, db5 and db9 variables from transformations of the average of four scale signals that were non-normal based on Lilliefors corrected Kolmogorov-Smirnov tests, $\alpha = 0.05$. Grand Rapids, Matheson Island and Riverton fish are separated based on age.

	Aged four fish			Aged five fish		
	Grand Rapids	Matheson Island	Riverton	Grand Rapids	Matheson Island	Riverton
Fourier (N = 128)	0.281	0.391	0.445	0.359	0.359	0.859
db1 (N = 128)	0.125	0.125	0.094	0.164	0.141	0.211
db5 (N = 185)	0.130	0.076	0.059	0.238	0.086	0.259
db9 (N = 241)	0.158	0.058	0.062	0.224	0.050	0.357

corresponding discriminant analysis allowed the construction of a null distribution for comparison to the original discriminant results, $\alpha = 0.05$. The SPSS® text output was parsed with a Microsoft® Visual basic for Applications (VBA) program (Appendix 9). The p-value was estimated as the portion of randomization's for which the classification rate was equal to or exceeded the rate of the original analysis.

Sample sizes of index sample walleye were sufficient to investigate possible age-class effects (Table 3.1). The randomization procedure tested for significant age-class effects using the Fourier transform variables, with $\alpha = 0.05$. Other age-classes were available from these sampling locations but their sample sizes were not large enough to be considered.

The index samples were placed in a three-way discriminant analysis for each of the Fourier, db1, db5 and db9 variable sets. The significance of the discriminant functions formed was tested by the randomization procedure. A jackknife test was developed similar to the test in previous work to determine if wavelets variables form discriminant functions that had significantly higher classification rates than the Fourier harmonic amplitudes and phase angles. The discriminant function classification rates were compared in a pairwise fashion by placing the variables for all scale signals (Fourier, db1, db5 and db9) together in SPSS®. A program jackknifed, one case from each of Grand Rapids, Matheson Island, and Riverton (Appendix 10). A separate discriminant analysis was run for each of the four different signal analysis variable sets and a classification rate calculated for each. To compare the Fourier classification rates to the different orders of wavelet's classification rates, the wavelet rate (R_w) was subtracted from the Fourier rate (R_f). The difference was used in a one-tailed significance test (H_0: R_w-R_f ≤ 0). To obtain a p-value that could be compared to a α set at 0.05, 1000 jackknifes were conducted (Manly 1997). The SPSS® output text file was parsed with a VBA program (Appendix 9). The probability of repetition of the exact same cases removed for the three groups in the age four and five comparisons was 0.0026 and 0.00066 respectively. The wavelets were also compared amongst themselves to determine if one of the three orders of wavelet variables formed discriminant functions that classified the stocks significantly better than the others.

Finally, fish caught in the commercial fishery were placed as unknowns into the three-way index sample discriminant function formed from the signal analysis method with significantly higher classification rates, to observe how the commercial catch fish would be classified. The commercial catch samples from Norway House, Frog Bay and Selkirk were placed as unknowns in the discriminant analyses formed by the index samples. Samples were restricted to the same year-class as the index samples, therefore, only the commercial catch fish aged six years old (hatched in 1993) and fish aged seven years old (hatched in 1992) were used in the analyses.

A discriminant analysis will always assign a test sample to one of the known groups, no matter if the test sample actually came from that stock. It is possible that the commercial catch samples were merely classified as an index fish due to similarities. A Principal Component Analysis (PCA) was performed for both of the different age class comparisons in SPSS®. To investigate how the commercial catch fish scores differ on the first two axes the index location fish the majority of the commercial catches were classified as and those commercial catch fish classified to that index group had their PCA scores ranked. The average ranks were then compared.

Results

The discriminant functions formed to discriminate between the age-classes for Grand Rapids, Matheson Island and Riverton varied considerably in their classification rates (Table 3.3). The Riverton age-class comparison had the highest classification rate and number of variables selected to form the discriminant function. The variables selected to form the discriminant functions for the three comparisons varied amongst the lower to the higher harmonic amplitudes and phase angles considered. Unbalanced sample sizes resulted in unbalanced correct classifications in favour of the age with the larger sample size in all three comparisons.

Tests for significant age-class effects with the randomization procedure had mixed results (Table 3.4). The Grand Rapids and Matheson Island comparisons had no significant age-class effects. The Riverton discriminant function was significant, indicating different scale morphology amongst the age-classes. Based on these results all additional analyses considered the age-classes separately. The Riverton minimum and average classification rates of 70.25 and 74.25% were considerably higher than the approximately 53% minimum and 68% average classification rates of Grand Rapids and Matheson Island. The maximum classification rate of 80.98% for Riverton was smaller than Grand Rapids and Matheson Island rates of 85.91 and 83.41% respectively.

The classification rates of discriminant functions formed from wavelet variables were higher than Fourier rates for the aged four and five fish, with the exception of the age four db1 analysis (Table 3.5). The first harmonic amplitude for Fourier, the 8th coefficient for db1 and 118th coefficient for db5 were selected first for both age

Table 3.3. Stepwise discriminant analysis classification summary of age-class comparisons for Grand Rapids (GR), Matheson Island (MI) and Riverton (R) samples. Analyses were performed for the average of four scale signals. The harmonic amplitude and phase angles are listed in order of selection into the stepwise discriminant function. The number of fish, percent (%) correct classification rate for each age-class, as well as the overall classification rate are included.

		Jackknife classification summary				
		Age four GR (N = 65) MI (N = 70) R (N = 86)		Age five GR (N = 84) MI (N = 75) R (N = 240)		
	Harmonic Amplitudes (A) or Phase Angles (P) in the discriminant function	Number classified correctly	%	Number classified correctly	%	Overall %
Grand Rapids	A37, P51, P16	35	46.15	60	71.73	63.76
Matheson Island	A46	39	55.71	42	56.00	55.86
Riverton	A6, A4, P1, P9, A21, P12, A1, A9, A2, A19, A14, A58, A31, P2, A28, A38	41	47.67	222	92.50	80.67

Table 3.4. Randomization results of 1000 random shuffles testing for significant age-class effects. These tests were performed for the average of four scale signals. The minimum, maximum, average classification rates of the randomization's and the original data's classification rate are included as well as the calculated p-value. Asterisks (**) define significant values, $\alpha = 0.05$. The Grand Rapids, Matheson Island, and Riverton fish aged at four years had sample sizes of 65, 70 and 86 respectively. The Grand Rapids, Matheson Island, and Riverton fish aged at five years had sample sizes of 84, 75 and 240 respectively.

	% correct classification				
	Randomizations			Original discriminant function	p-value
	Minimum	Average	Maximum		
Grand Rapids	53.02	67.77	85.91	63.76	0.842
Matheson Island	53.79	68.54	83.45	55.86	0.994
Riverton	70.25	74.20	80.98	80.67	0.001**

Table 3.5. Stepwise discriminant analysis classification results of index sample fish aged four and five years analyzed by Fourier, db1, db5 and db9. Analyses were performed for the average of four scale signals. The harmonic amplitude and phase angles and wavelet coefficients are listed in order of selection into the stepwise discriminant function. The number of fish, and % correct classification rate are included for each age-class, as well as an overall classification rate.

		Jackknife classification summary						
		GrandRapids		Matheson Island		Riverton		
	Fourier Harmonic Amplitudes (A) or Phase Angles (P) and wavelet coefficients (C) in the discriminant function	Number classified correctly	%	Number classified correctly	%	Number classified correctly	%	Overall %
	Age four	N = 65		N = 70		N = 86		
Fourier	A1, P2, A6, P17, A40, A58, P57, A27, A8, A64	45	69.23	29	41.42	60	69.76	60.63
db1	C8, C124, C65, C64, C125, C81, C2, C33, C107, C53	45	69.23	28	40.00	60	69.76	60.18
db5	C118, C180, C7, C76, C77, C165, C125, C137, C155, C59, C61	48	73.85	34	48.57	65	75.58	66.52
db9	C121, C14, C118, C104, C142, C231, C230, C232, C208, C213, C99, C143, C101, C174, C103, C84, C187	51	78.46	32	45.71	59	68.60	64.25
	Age five	N = 84		N = 75		N = 240		
Fourier	A1, A9, P9, A2, A6, P17, A21, P7, A31, A18, A37, P47, P2, P24	50	59.52	22	29.33	211	87.91	70.93
db1	C8, C65, C3, C64, C4, C63, C48, C101, C125, C110, C106, C59, C33	59	70.23	26	34.67	208	86.67	73.43
db5	C118, C8, C41, C35, C39, C78, C113, C44, C155, C50, C180, C178, C162, C94, C104	60	71.43	25	33.33	207	86.25	73.18
db9	C73, C120, C119, C231, C11, C63, C214,C105, C153, C217, C224, C185, C160, C229, C108, C199	54	64.59	32	42.67	212	88.33	74.69

comparisons. The 73^{rd} coefficient was selected first for both db9 comparisons but it was removed in the stepwise procedure for the aged four comparison. Riverton samples in the aged five comparison had much higher correct classification rates than Grand Rapids and Matheson Island. This is attributable to the difference in the number of fish from each group, resulting in a higher prior probability of classifying fish as being from Riverton. The db9 discriminant function had the highest number of variables selected for both the aged four and five year old comparisons. The randomization procedure found all discriminant functions were highly significant ($p < 0.001$) for both the aged four and five fish comparisons (Table 3.6). There was little variation between the minimum, average and maximum classification rates of the 1000 randomizations for the different signal analysis methods.

The pairwise jackknife test for differences between discriminant functions formed from Fourier and the db1, db5 and db9 wavelets variables varied for the aged four and five fish comparisons (Table 3.7). The db9 wavelet variables formed discriminant functions that performed significantly better than Fourier variables for both the aged four and five comparisons with $p = 0.015$ and < 0.001 respectively. The discriminant functions formed from db5 variables compared to Fourier variable based functions for the age four comparison had a $p = 0.055$. The db1 and db5 wavelet variables were also significantly better than Fourier variables for discriminating the aged five fish with p-values of 0.008 and < 0.001 respectively. No significant differences were found amongst the wavelet orders for the aged four or five fish comparisons.

Table 3.6. Monte Carlo randomization results testing the significance of the discriminant functions formed for the aged four and four-year-old fish, average of four scale signals, index samples analyzed by Fourier, db1, db5 and db9. The minimum, maximum and average classification rates of the randomization's and the original data's discriminant function are included as well as the calculated p-value. Asterisks (**) define significant values, $\alpha = 0.05$. Grand Rapids, Matheson Island, and Riverton fish aged at four years had sample sizes of 65, 70 and 86 respectively. Grand Rapids, Matheson Island, and Riverton fish aged five years had sample sizes of 84, 75 and 240 respectively.

	% correct classification				
	Randomizations			Original data	
	Minimum	Average	Maximum	Actual calculated	p-value
			Age four		
Fourier	34.84	48.10	59.28	60.63	<0.001**
db1	32.58	45.11	57.01	60.18	<0.001**
db5	34.84	47.07	57.92	66.52	<0.001**
db9	37.10	47.27	59.28	64.25	<0.001**
			Age five		
Fourier	56.89	60.09	63.91	70.93	<0.001**
db1	56.89	60.02	63.16	73.43	<0.001**
db5	55.89	59.99	63.91	73.18	<0.001**
db9	56.39	60.11	64.91	74.69	<0.001**

Table 3.7. Resulting p-values calculated from the jackknife test for significant differences between Fourier, db1, db5 and db9 discriminant functions. Analyses were conducted for both the aged four and five Grand Rapids, Matheson Island and Riverton walleye. One thousand jackknifed combinations were performed. Asterisks (**) define significant values, $\alpha = 0.05$.

	Classification subtracted	p-values	
		Age four	Age five
Fourier	db1	0.183	0.008**
	db5	0.055	<0.001**
	db9	0.015**	<0.001**
db1	db5	0.289	0.183
	db9	0.144	0.109
db5	db9	0.333	0.366

Based upon the index sample discriminant results the db9 wavelet was chosen to transform the commercial catch scale signals. The commercial samples were classified with a discriminant function formed from the index sample signals, db9 variables. Over 75% of the fish aged six years old in the commercial catch fish classified as Riverton samples (Table 3.8). Ten to 20% of the commercial samples were classified as Matheson Island and less than 4.5% were classified as Grand Rapids. A chi-square test of independence performed on the classifications found no difference in the classifications for the three different commercial sampling locations with $\alpha = 0.05$ ($\chi^2 = 2.011$, $p = 0.734$).

For fish aged at seven years, 100% of Selkirk fish were classified as Riverton. Approximately 20% of Frog Bay and Norway House fish were classified as Matheson Island and approximately 80% were classified as Riverton. None of the aged seven years old, commercial catch fish were classified as Grand Rapids samples. A chi-square test of independence performed on the aged seven classifications found there was no difference in the classifications for the three different commercial sampling locations with $\alpha = 0.05$ ($\chi^2 = 1.861$, $p = 0.3944$).

All three commercial location fish for both age comparisons were different than the Riverton fish average rank for at least one of the PCA axes. For the aged four years index samples and aged six years commercial catch samples the Riverton average rank of 128.2 for the first PCA axis was larger than all commercial samples (Table 3.9). On the second PCA axis the Riverton average rank of 83.5 was lower than the commercial catch scores. Norway House had the lowest average score of 81.30 on the first axis and highest average score of 146.42 on the second axis.

Table 3.8. Stepwise discriminant analysis classification results of commercial catch fish aged six and seven years from Norway House, Frog Bay and Selkirk. Average of four scale signals, transformed with db9 were placed as unknowns into the discriminant function formed from db9 variables of the equivalent year-class index fish sampled in 1997. The number of fish, and percent (%) classification rate are included for each of the index locations the commercial catch was classified as.

	Jackknife classification summary					
	GrandRapids		Matheson Island		Riverton	
	Number classified	%	Number classified	%	Number classified	%
	Age six					
Norway House (N = 47)	1	2.13	8	17.02	38	80.85
Frog Bay (N = 49)	2	4.08	10	20.41	37	75.51
Selkirk (N = 46)	2	4.35	5	10.87	39	84.78
	Age seven					
Norway House (N = 25)	0	0	5	20	20	80
Frog Bay (N = 12)	0	0	2	16.67	10	83.33
Selkirk (N = 8)	0	0	0	0	8	100

Table 3.9. Rank results of PCA analysis scores on the first and second axes. The ranks were calculated for the Riverton age four and five samples and the age six and seven commercial samples that classified as being Riverton fish. The average and total number of fish are indicated for each sample location.

		Age four Riverton and age six commercial samples (N=200)			
			Commercial samples		
		Riverton	Frog Bay	Norway House	Selkirk
PCA1	average	128.2	92.7	70.1	75.8
	N	86	38	37	39
PCA2	average	83.5	116.1	132.6	92.2
	N	86	38	37	39

		Age five Riverton and age seven commercial samples (N=277)			
			Commercial samples		
		Riverton	Frog Bay	Norway House	Selkirk
PCA1	average	132.4	178.2	179.9	204.1
	N	240	20	10	8
PCA2	average	133.8	182.9	204.1	119.1
	N	240	20	10	8

Riverton fishes average rank of 132.4 for the first PCA axis was smaller than all commercial samples for the aged five years index samples and aged seven years commercial catch samples (Table 3.9). On the second PCA axis the Selkirk average rank of 119.1 was the lowest average rank score. Selkirk had the highest average rank score of 204.1 on the first PCA axis. Norway House had the highest average rank score of 204.1 on the second axis.

Discussion

Walleye sampled from Lake Winnipeg can be discriminated more effectively with wavelet analysis variables than Fourier analysis harmonic amplitudes and phase angles. However, not all orders of wavelet variables formed significantly better discriminant functions than Fourier analysis. This is likely due to an inability of those orders to describe the frequency information of the scale signals for discrimination anymore effectively than Fourier analysis.

More variables were selected for placement into the discriminant function for the db5 and db9 wavelets than Fourier or db1 discriminant comparisons. By random chance a variable may be able to discriminate between the groups, so with increased number of variables available, there is a possible increase in the number of variables that are spuriously good discriminators. If the number of variables available to the discriminant function were influencing the classification this would be observed in an increased average classification for the randomization procedure. The average classifications differed very little between the comparisons; thus, any differences between the classification rates were not artifacts of the number of variables initially available. I therefore did not adjust for differences in the number of variables available to the discriminant analysis, when testing for significant differences between the signal processing methods.

Similar to other studies that used scale or otolith outlines, significant age-class effects were found for Riverton samples (Casselman et al. 1981; Bird et al. 1986; Castonguay et al. 1991; Campana and Casselman 1993; Richards and Esteves 1997b). The same variables were selected first for both the aged four and five comparisons. This implies that these variables are good discriminators for the Lake Winnipeg index samples and there is a high degree of similarity between the age-class comparisons despite the significant age-class effects. This is promising for future work as similarities in the discriminations for different age-classes potentially allows them to be combined, increasing sample sizes and reducing the number of discriminant functions needed to differentiate between stocks. Castonguay et al. (1991) found year-class effects were significant and concluded that the resulting temporal instability in scale shape eliminated their interest in using Fourier harmonics to perform long-term stock discrimination.

The selection of higher order Fourier variables first into the discriminant analyses for the Grand Rapids and Matheson Island age-class comparisons differed from other studies that limited the discrimination to lower order harmonics (Jarvis et al. 1978; Casselman et al. 1981; Riley and Carline 1982; Castonguay et al. 1991; Campana and Casselman 1993; Kenchington and Full 1994). These discriminant functions were not significant however, based on the randomization test. These variables may be less important for discriminating between stocks. This was supported by the Riverton age-class comparison that selected low order harmonic amplitudes first to form the discriminant function that was significant. Lower order Fourier harmonics were also selected first to form the discriminant functions in the three-way comparisons of index samples, both of which were significant functions.

The majority of commercial catch samples were classified as being Riverton fish when placed as unknowns in the discriminant function formed from the index samples. This raised more questions than were answered. The simplest explanation would be to conclude the results are correct and the majority of the commercial catch fish are from the Riverton stock, meaning walleye undergo large migrations away from Riverton after spawning. Although possible, it is unlikely that the majority of commercial catch fish were of Riverton origin.

There are a number of other possible explanations for the classification results. The commercial catch samples may have been distinct genetically but environmental factors contributing to scale growth resulted in a scale shape that was most similar to Riverton or scale shape did not vary greatly amongst stocks (Riley and Carline 1982). Tagging studies in western Lake Erie and in other systems (Crowe 1962; Olson and Scidmore 1962; Ferguson and Derksen 1971) have showed strong tendency for adults to return to the same area to spawn year after year. Whether the adults are returning to their natal stream of reef to spawn is not known. Circumstantial work by Rawson (1957) suggested that this does occur; however, if the spawning site chosen by the adult fish is not their natal stream or reef, gene flow amongst spawning aggregations may be quite high (Riley and Carline 1982).

It is also possible that significant inter-annual effects exist and the comparisons made were not valid, in which case a new reference sample would need to be created every year to effectively classify the unknowns. Both inter-annual and year-class effects could be investigated by analyzing Lake Winnipeg index samples from 1996 or 1998 and comparing them to the 1997 samples to determine if significant differences exist. Significant results would suggest that environmental factors are having a substantial influence on the growth of the scales (Richards and Esteves 1997a). The analyses may be further complicated by temporal differences in the samples as we are classifying fish samples in October with a discriminant function formed from fish sampled in May and early June. However, this would have a limited effect if inter-annual effects do not exist. This seems likely because low order harmonics, representative of the gross shape of the signals are the most important coefficients for discrimination. Furthermore, given the short time span of five months between sampling periods, it is unlikely the gross shape would undergo major alterations.

Potentially, the index samples that formed the discriminant function were distinct from the stocks represented in the commercial catch. The commercial catch fish may have been classified as Riverton fish because these scale outlines were most similar to the commercial catch. The stock identification method used in this study does not have the ability to determine the number of actual stocks contributing to a mixed stock fishery (Waldman et al. 1997). The PCA score rank results are evidence that the commercial catch samples potentially are different stocks that were classified as Riverton fish. If the model includes fewer stocks than actually contribute to the mixed stock, then fish from the excluded stock or stocks will be misallocated to one of those included in the model.

Alternatively, including too many groups in a discriminant analysis, reduces the power of the analysis (De Pontual and Prouzet 1987). The first step in any mixed stock analysis should be to determine which stocks might be expected to constitute the mixed stock. This is best determined with independent data rather than the results of the stock composition analysis (Waldman and Fabrizio 1994).

The number of Lake Winnipeg walleye spawning locations is unknown, and could likely not be identified; certainly not all sampled. For mixed stock fisheries, stock identification is best suited to situations where there is a known, limited number, of spawning locations available to the fish. Fish that are anadromous, spawning in specific river systems or freshwater species that live in lakes and whose spawning is limited to rivers or specialized habitat within the lake itself are suitable for mixed stock identification studies. Stock identification on a lake wide basis could only be possible if scale shapes are similar in regions of the lake rather than individual spawning areas and different between regions, for example the north basin, channel and south basin (Figure 3.2). This would then allow a much-simplified sampling and analysis program; however, a more extensive sampling and analysis program would need to be conducted to confirm or reject this possibility.

If it is decided stock identification is not practical for Lake Winnipeg walleye it is important that an understanding of fish movements and stock mixing is achieved. The division of the lake into regulatory zones with distinct mesh size regulations, season dates and quotas requires an understanding of fish movements for management of the fishery. A renewed effort could be placed on tagging studies, with greater incentive for tag returns so that an understanding of seasonal fish movements and the potential for mixing can be gained. Past studies by the Department of Conservation confirmed that some walleye traveled distances in excess of 200km within the lake (Walt Lysack, pers. com., 2001). Based on these and additional tag returns, managers could perhaps effectively monitor fish populations with commercial catch and index samples.

Wavelet analysis improved our ability to discriminate between index sampling locations compared to the traditionally used Fourier analysis, but there are advantages to using the Fourier transform. Fourier transforms are not just mathematical abstractions; they have physical meanings (Hubbard 1996). Wavelets do not have a physical existence and are extremely difficult to associate with frequencies within the signal, making it difficult to interpret wavelet variables. Researchers typically use discrete wavelets for compression, and continuous wavelet transforms for analysis to gain additional information as a continuous transform allows one to see what is happening at different scales at any point of the signal. Unfortunately, continuous wavelet analysis is difficult to combine with discriminant analysis, as the coefficient set consists of thousands of variables. This limits discriminant studies to discrete wavelet analysis variables. Another advantage of Fourier analysis is that there is only one Fourier transform; whereas, wavelet analysis has numerous families and orders within those families that can complicate the appropriate selection of a wavelet for signal processing.

It is possible that for a different study species, the scale outlines will vary to the extent that different wavelets or orders of wavelet will be more effective at analyzing the scale outlines for discrimination. Further investigation should focus on the effects of scale shape variation, those frequencies within the signals that are important discriminators and establishing which wavelet family or order within that family is most effective at describing scale outline signals of walleye and other fish species.

Literature Cited

Aboufadel, E., and Schlicker, S. 1999. Discovering wavelets. John Wiley & Sons, Inc., New York.

Bird, J. L, Eppler, D.T., and Checkley, D.M., Jr. 1986. Comparisons of herring otoliths using Fourier series shape analysis. Can. J. Fish. Aquat. Sci. **43**: 1228-1234.

Bradshaw, G.A., and Spies, T.A. 1992. Characterizing canopy gap structure in forests using wavelet analysis. J. Ecol. **80**: 205-215.

Campana, S.E., and Casselman, J.M. 1993. Stock discrimination using otolith shape analysis. Can. J. Fish. Aquat. Sci. **50**: 1062-1083.

Casselman, J.M., Collins, J.J., Crossman, E.J., Ihssen, P.E., and Spangler. G.R. 1981. Lake whitefish (*Coregonus clupeaformis*) stocks of the Ontario waters of Lake Huron. Can. J. Fish. Aquat. Sci. **38**: 1772-1789.

Castonguay, M., Simard, P., and Gagnon, P. 1991. Usefulness of Fourier analysis of otolith shape for Atlantic mackerel (*Scomber scombrus*) stock discrimination. Can. J. Fish. Aquat. Sci. **48**: 296-302.

Chui, C.K. 1997. Wavelets: A mathematical tool for signal analysis. SIAM, Philadelphia.

Cook, R.C. 1982. Stock identification of sockeye salmon (*Oncorhynchus nerka*) with scale pattern recognition. Can. J. Fish. Aquat. Sci. **39**: 611-617.

Crowe, W.R. 1962. Homing behaviour in walleye. Trans. Am. Fish. Soc. **91**: 350-354.

Dale, M.R.T., and Mah, M. 1998. The use of wavelets for spatial pattern analysis in ecology. J. Veg. Sci. **9**: 805-814.

De Pontual, H., and Prouzet, P. 1987. Atlantic salmon, *Salmo salar* L., stock discrimination by scale-shape analysis. Aquacult. Fish. Manage. **18**: 277-289.

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

Ferguson, R.G., and Derksen, A.J. 1971. Migrations of adult and juvenile walleyes (*Stizostedion vitreum vitreum*) in southern Lake Huron, Lake St. Clair, Lake Erie, and connecting waters. J. Fish. Res. Bd. Canada **28**: 1133-1142.

Hubbard, B.B. 1996. The world according to wavelets. A.K. Peters, Ltd., Wellesly, Mass.

Huele, R., and Udo de Haes, H. 1998. Identification of individual sperm whales by wavelet transform of the trailing edge of the flukes. Mar. Mamm. Sci. **14**: 143-145.

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

Jarvis, R.S., Klodowski, H.F., and Sheldon, S.P. 1978. New method of quantifying scale shape and an application to stock identification in walleye (*Stizostedion vitreum vitreum*). Trans. Am. Fish. Soc. **107**: 528-534.

Kaiser, G. 1994. A friendly guide to wavelets. Birkhäuser, Boston.

Kenchington, E.L., and Full, W.E. 1994. Fourier analysis of sea scallop (*Placopecten magellanicus*) shells in determining population structure. Can. J. Fish. Aquat. Sci. **51**: 348-356.

Krzanowski, W.J. 1977. The performance of Fisher's linear discriminant function under non-optimal conditions. Technometrics, **19**: 191-200.

Lachenbruch, P.A. 1975. Discriminant analysis. Hafner Press, A Division of Macmillan Publishing Co., Inc., New York.

Manly, B.F.J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. 2nd ed. Chapman & Hall, London.

Margraf, F.J., and Riley, L.M. 1993. Evaluation of scale shape for identifying spawning stocks of coastal Atlantic striped bass (*Morone saxatilis*). Fish. Res. **18**: 163-172.

Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J. 1996. Wavelet toolbox user's guide. The Mathwork, Inc, Natick, Mass.

Ogden, R.T. 1997. Essential wavelets for statistical applications and data analysis. Birkhäuser, Boston.

Olson, D.E., and Scidmore, W.J. 1962. Homing behavior of spawning walleye. Trans. Am. Fish. Soc. **91**: 355-361.

Rawson, D.S. 1957. The life history and ecology of yellow walleye, Stizostedion vitreum, in Lac La Ronge, Saskatchewan. Trans. Amer. Fish. Soc. **86**: 15-37.

Richards, R.A., and Esteves, C. 1997a. Use of scale morphology for discriminating wild stocks of Atlantic Striped Bass. Trans. Am. Fish. Soc. **126**: 919-925.

Richards, R.A., and Esteves, C. 1997b. Stock-specific variation in scale morphology of Atlantic striped bass. Trans. Am. Fish. Soc. **126**: 908-918.

Ricker, W.E. 1958. Maximum sustained yields from fluctuating environments and mixed stocks. J. Fish. Res. Bd. Canada. **15**: 991-1006.

Riley, L.M., and Carline, R.F. 1982. Evaluation of scale shape for the identification of walleye stocks from western Lake Erie. Trans. Am. Fish. Soc. **111**: 736-741.

Strang, G., and Nguyen, T. 1997. Wavelets and filter banks. Wellesley-Cambridge Press, Wellesley, Mass.

Waldman, J.R., Grossfield, J., and Wirgin, I. 1988. Review of stock discrimination techniques for striped bass. N. Am. J. Fish. Manage. **8**: 410-425.

Waldman, J.R., and Fabrizio, M.C. 1994. Problems of stock definition in estimating relative contributions of Atlantic striped bass to the coastal fishery. Trans. Am. Fish. Soc. **123**: 766-778.

Waldman, J.R., Richards, R.A., Schill, W.B., Wirgin, I., and Fabrizio, M.C. 1997. An empirical comparison of stock identification techniques applied to striped bass. Trans. Am. Fish. Soc. **126**: 369-385.

Wei, W.W.S. 1990. Time series analysis. Addison-Wesley Publishing Company, New York.

Williams, B.K., and Titus, K. 1988. Assessment of sampling stability in ecological applications of discriminant analysis. Ecology. **69**: 1275-1285.

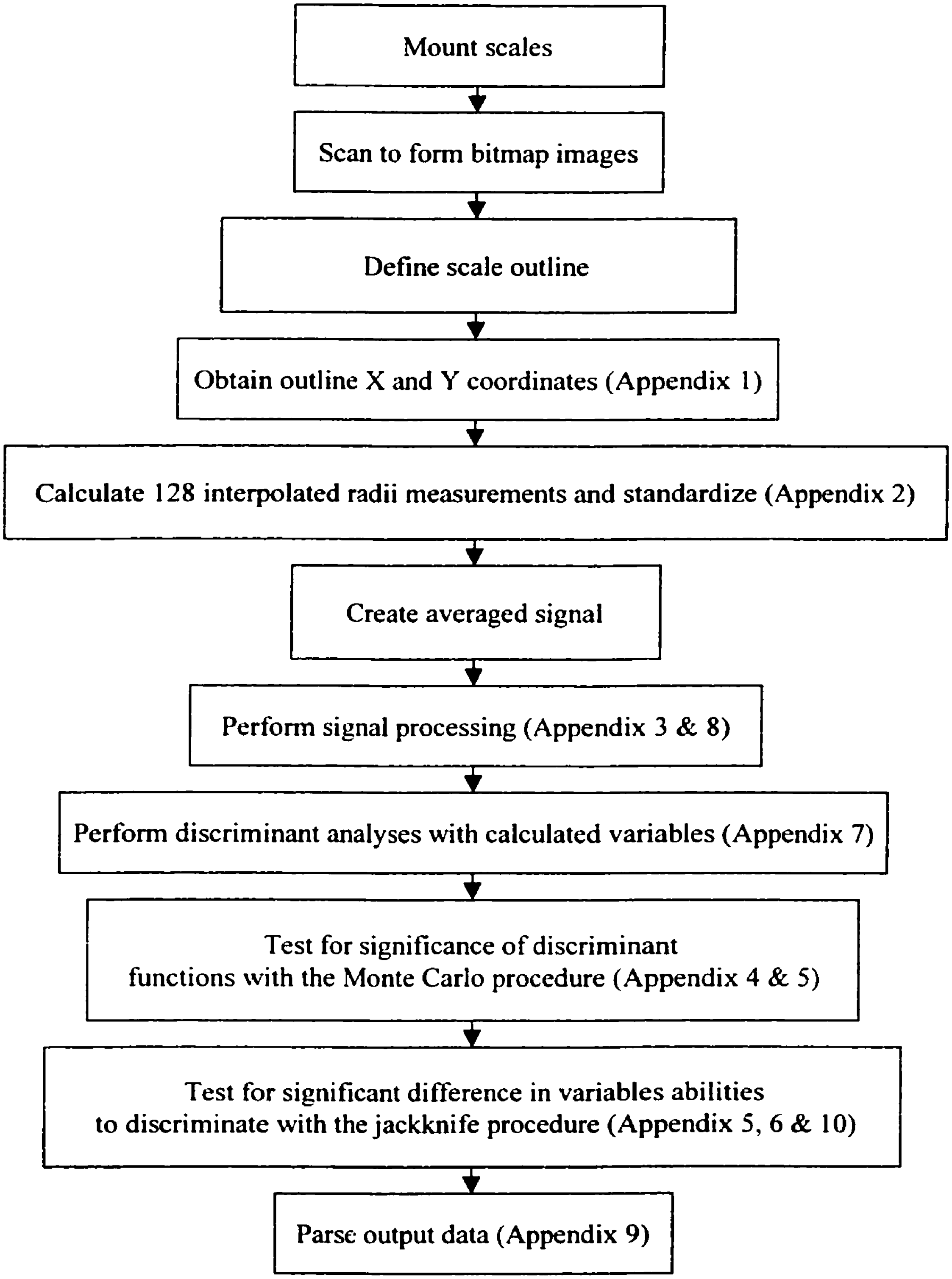
CHAPTER 4: General Summary

The combination of new techniques developed in this study improved our ability to discriminate between walleye stocks in Manitoba lakes. This study has shown that the variability of scale shapes from the same fish is detrimental to the stock discrimination. Averaging is able to remove some of this variation, resulting in better discrimination. Wavelet analysis improved our ability to discriminate between stocks compared to the traditionally used Fourier analysis, but there are still advantages in using the Fourier transform. The advantages and disadvantages of each must be weighed when making decisions regarding the signal processing method to use. Other possible sources of variation such as age-class, inter-annual and sex effects should be investigated when sample sizes permit partitioning of the samples. If necessary, partitioning will limit variability within samples and provide better discrimination results. Finally, non-parametric computer intensive statistics are effective alternatives to test the significance of the discriminant functions and test for significant differences between the methods of scale shape analysis when the assumptions of classical discriminant analysis are violated.

Stock identification is ideally suited to situations where there is a known, limited number, of spawning locations available to the fish. Since, Lake Winnipeg does not represent an ideal study population the methods developed in this study should be applied to more appropriate study populations to reconfirm their effectiveness. However, if habitat or environmental factors are influencing scale growth in what is genetically one stock in Lake Winnipeg, this is a situation in which scale morphology is preferred to genetics for stock identification.

A reevaluation of past studies would be beneficial so that the methods developed in this study could be compared to previously published results. Additional research focusing on the frequencies within scale signals that are important for stock discrimination may provide insight into the scale characteristics important for scale growth and distinguishing between stocks. Finally, further investigation could focus on the effects of scale shape variation from individual walleye and establishing which wavelet family and order within that family is most effective at describing scale shape signals.

Appendix A. Flowchart of processes involved in the analyses of scale shapes for stock discrimination.



Appendix 1. Scion Image program to define the outline of the scale image by searching for contrast in pixels along the scale images edges and set the outline pixels black and all other pixels white. The program then produces an output file of the X and Y coordinates of the outline pixels by starting a search routine from where the user initiates the search.

```
******************
macro 'Scale Outline'; {Converts Scale Image to an outline}
******************
var
 left,top,width,height,x,y:integer;
begin
  Requiresversion(1.55);
  KillROI;
  repeat
  SetCursor('arrow');
  showmessage('Click on the image');
  until button;
  GetMouse(x,y);
  AutoThreshold;
  AutoOutline(x,y);
  SetThreshold(-1);
  GetRoi(left,top,width,height);
  ShowMessage('left=',left:1,'\top=',top:1);
  if width=0 then begin
    PutMessage('Selection required.');
    exit;
  end;
  SetForeground(255); {Black}
  SetBackground(0); {White}
  SelectAll;
  Clear;
  RestoreRoi;
  Fill;
  KillRoi;
  Outline;
end;

**************************
macro 'Digitize Outline Pixels';
**************************
{Saves the X-Y Coordinates from a window with a SINGLE  Outline}
{Outline MUST be at least 1 pixel away from edge of window}
{Outline must have pixel value of 255}
var
 i,x,y,xStart,yStart,xOld,yOld,xMax,yMax,value:integer;
 finished:integer;
begin
 RequiresVersion(1.48);
 SetPalette('Grayscale');
 SetOptions('X-Y Center');
 GetPicSize(xMax,yMax);
```

```
repeat
        GetMouse(x,y);
until button;
value:=GetPixel(x,y);
while value = 0 do begin
  if x=xMax then
    begin
      x:=1;
      y:=y+1;
    end
  else x:=x+1;
  value:=GetPixel(x,y);
end;
i:=0;
finished:=0;
while finished=0 do begin
  i:=i+1;
  SetCounter(i);
  rX[i]:=x;
  rY[i]:=y;
  PutPixel(x,y,24);
  if GetPixel(x,y-1) = 255 then
    begin
      y:=y-1;
    end
  else if GetPixel(x+1,y-1) = 255 then
    begin
      x:=x+1;
      y:=y-1;
    end
  else if GetPixel(x+1,y) = 255 then
    begin
      x:=x+1;
    end
  else if GetPixel(x+1,y+1) = 255 then
    begin
      x:=x+1;
      y:=y+1;
    end
  else if GetPixel(x,y+1) = 255 then
    begin
      y:=y+1;
    end
  else if GetPixel(x-1,y+1) = 255 then
    begin
      x:=x-1;
      y:=y+1;
    end
  else if GetPixel(x-1,y) = 255 then
    begin
      x:=x-1;
    end
  else if GetPixel(x-1,y-1) = 255 then
    begin
      x:=x-1;
      y:=y-1;
```

```
      end
    else finished:=1;
  end;
  ShowResults;
end;
```

Appendix 2. VBA program to copy the text Scion® output file into Excel® and then convert the coordinates to radians. The data is then interpolated to 128 radius measurements and division by the mean radius measurement standardizes the data.

```
Option Explicit
Option Base 1
**********
Sub grab2()
*********
' grab2 Macro
' Keyboard Shortcut: Ctrl+r
  Range("A6:B3000").Select
  Application.CutCopyMode = False
  Selection.clear
  Range("h6:i3000").Select
  Selection.clear
  Range("m6:m133").Select
  Selection.Copy
  Sheets("128points").Select
End Sub

************
Sub scale128()
************
Dim oldPosition(1200) As Double
Dim oldRadii(1200) As Double
Dim newPosition(128) As Double
Dim newRadii(128) As Double
Dim iold As Integer, iNew As Integer, iOldCount As Integer, iNewCount As Integer, intercount As
Integer
Dim mean As Integer, count As Integer
Dim inputCell As Variant
Dim meanR As Double

' Read in original data
iOldCount = 0
inputCell = Cells(6, 6)
Do Until inputCell = ""
    iOldCount = iOldCount + 1
    oldRadii(iOldCount) = inputCell
    oldPosition(iOldCount) = Cells(5 + iOldCount, 7)
    inputCell = Cells(6 + iOldCount, 6)
    Cells(5 + iOldCount, 9) = oldRadii(iOldCount)
    Cells(5 + iOldCount, 8) = oldPosition(iOldCount)
Loop
For iNew = 1 To 128
  newPosition(iNew) = (iNew - 1) * 2.8125
Next iNew
For iNew = 1 To 128
  iold = 1
  Do Until oldPosition(iold) > newPosition(iNew)
    iold = iold + 1
```

```
      If oldRadii(iold) = 0 Then
         oldPosition(iold) = 360
         oldRadii(iold) = Cells(6, 9)
      End If
   Loop
   Cells(5 + iNew, 12) = newPosition(iNew)
   Rem Interpolation formula((ABS(1-ABS(C2-E2)/ABS(C1-C2))*D2)+((ABS(1-ABS(C1-E2)/ABS(C1-
C2)))*D1))
   newRadii(iNew) = ((Abs(1 - Abs(oldPosition(iold) - newPosition(iNew)) _
   / Abs(oldPosition(iold - 1) - oldPosition(iold))) * oldRadii(iold)) + _
   ((Abs(1 - Abs(oldPosition(iold - 1) - newPosition(iNew)) / Abs(oldPosition(iold - 1) _
   - oldPosition(iold)))) * oldRadii(iold - 1)))
   meanR = meanR + newRadii(iNew) / 128
Next iNew
For iNew = 1 To 128
   Cells(5 + iNew, 13) = newRadii(iNew) / meanR
Next iNew
End Sub

*********
Sub grab()
*********
' shortcut Ctrl+e
Range("A1:B1500").Select
Selection.Copy
ActiveWindow.WindowState = xlMinimized
Windows("scale128analysis.xls").Activate
ActiveWindow.WindowState = xlMaximized
Sheets("Scale 600dpi Measurements").Select
Range("A6").Select
ActiveSheet.Paste
Call scale128
End Sub

*****************
Sub clear128points()
*****************
" Keyboard Shortcut: Ctrl+t
Windows("scale128analysis.xls").Activate
ActiveWindow.WindowState = xlMaximized
Sheets("128points").Select
Range("A3:S130").Select
Selection.clear
End Sub
```

Appendix 3. VBA program to format Excel® workbook and perform Fourier transform on the 128 scale radii signals with Excel Link® to Matlab®. The transform variables are placed in summary worksheets and new workbooks that combine sampling locations are produced.

```
***************************
Sub FormatWorkbookFourier1()
***************************

'important first macro which should be run before the other macros

Dim four_avg1 As String
Dim four_avg2 As String
Dim four_avg3 As String
Dim four_avg4 As String
Dim four_avg5 As String
Dim avg1 As String
Dim avg2 As String
Dim avg3 As String
Dim avg4 As String
Dim avg5 As String

If Cells(1, 1) <> "four_avg1" Then
   Sheets.Add
   ActiveSheet.Name = "four_avg1"
   Sheets.Add
   ActiveSheet.Name = "four_avg2"
   Sheets.Add
   ActiveSheet.Name = "four_avg3"
   Sheets.Add
   ActiveSheet.Name = "four_avg4"
   Sheets.Add
   ActiveSheet.Name = "four_avg5"
   Sheets.Add
   ActiveSheet.Name = "starter"
End If
   Cells(1, 1) = "four_avg1"
   Cells(2, 1) = "four_avg2"
   Cells(3, 1) = "four_avg3"
   Cells(4, 1) = "four_avg4"
   Cells(5, 1) = "four_avg5"
   Cells(260, 1) = "avg1"
   Cells(260, 2) = "avg2"
   Cells(260, 3) = "avg3"
   Cells(260, 4) = "avg4"
   Cells(260, 5) = "avg5"
End Sub
```

```
**********************
Sub customize_sheetfour2()
**********************

Dim b As Integer
Dim q As Integer
Dim h As Integer
Dim i As Integer
Dim e As Integer
Dim sheetcount As Integer
Dim sheet As String

Sheets("starter").Select
sheetcount = 0
For b = 1 To 250
   If Cells(b, 2) <> "" Then
      sheetcount = sheetcount + 1
   End If
Next b
For q = 1 To sheetcount
   Sheets("starter").Select
   sheet = Cells(q, 2)
   Sheets(sheet).Select
   If Cells(134, 1) <> "avg1" Then
      Range("A134:A262").Select
      Selection.Insert Shift:=xlToRight
      Range("A2:A130").Select
      Selection.Copy
      Range("A134").Select
      ActiveSheet.Paste
      ActiveWindow.Zoom = 50
      Range("f134:k262").Select
      Selection.ClearContents
   End If
      Range("a270:iv480").Select
      Selection.Clear
      Range("a270:iv480").Select
      Selection.Clear
      Cells(270, 2) = 1
      For e = 1 To 127
         Cells(270, 2 + e) = Cells(270, 1 + e) + 1
      Next e
      Range("270:270").Select
      Selection.Font.Bold = True
      Range("300:300").Select
      Selection.Font.Bold = True
      Cells(270, 1) = "fftAmplitude"
      Cells(270, 1).Select
      With Selection.Font
         .Size = 14
         .Strikethrough = False
      End With
      Cells(280, 1) = "fftPhase"
      Cells(280, 1).Select
      With Selection.Font
         .Size = 14
```

```
        .Strikethrough = False
      End With
      Range("a270:a470").Select
      Selection.Font.Bold = True
      With Selection
        .HorizontalAlignment = xlCenter
        .VerticalAlignment = xlBottom
      End With
      Columns("J:J").Select
      With Selection
        .HorizontalAlignment = xlCenter
        .VerticalAlignment = xlBottom
      End With
      Rows("270:270").Select
      Selection.NumberFormat = "General"
      With Selection
        .HorizontalAlignment = xlCenter
        .VerticalAlignment = xlBottom
      End With
Next q
Sheets("starter").Select
sheetcount = 0
For h = 1 To 250
   If Cells(h, 2) <> "" Then
      sheetcount = sheetcount + 1
   End If
Next h
For i = 1 To sheetcount
   Sheets("starter").Select
   sheet = Cells(i, 2)
   Rows("260:260").Select
   Selection.Copy
   Sheets(sheet).Select
   Rows("134:134").Select
   ActiveSheet.Paste
Next i
End Sub

**********************
Sub DecompositionFour3()
**********************

Dim c As Variant, coef As Variant
Dim i As Integer, d As Integer, iCount As Integer, e As Integer, no_scales As Integer
Dim h As Integer, q As Integer, sheetcount As Integer, t As Integer
Dim vectorname As String, sheet As String
Dim cmdstring As String, Wavelet As String, cmdstring1 As String, cmdstring2 As String
Dim cmdstring3 As String, cmdstring4 As String, cmdstring5 As String

'defines the sheets which will be used
Sheets("starter").Select
sheetcount = 0
For h = 1 To 250
   If Cells(h, 2) <> "" Then
      sheetcount = sheetcount + 1
   End If
```

```
Next h
For q = 1 To sheetcount
   Sheets("starter").Select
   sheet = Cells(q, 2)
   Sheets(sheet).Select
   no_scales = 0
   'finds out the number of scales; to be used as a counter
   For e = 1 To 9
     If Cells(135, e) <> "" Then
        no_scales = no_scales + 1
      End If
   Next e
     'performs wavelet decomposition on the signals
      For d = 1 To 2
        If Cells(260 + (d * 10), 1) <> "" Then
        Wavelet = Cells(260 + (d * 10), 1)
        For i = 1 To no_scales
           coef = ""
           vectorname = Cells(134, i)
           MLPutMatrix vectorname, Range(Cells(135, i), Cells(262, i))
           cmdstring = "y=fft(" + vectorname + ")"
           MLEvalstring (cmdstring)
           If d = 1 Then
              cmdstring3 = "m=abs( y )"
              MLEvalstring (cmdstring3)
              cmdstring1 = "m"
              MLGetVar (cmdstring1), coef
           ElseIf d = 2 Then
              cmdstring4 = "p = angle(y)"
              MLEvalstring (cmdstring4)
              cmdstring5 = "p"
              MLGetVar (cmdstring5), coef
           End If
           Range("a1000:a1300") = coef
           'number of coefs. differs depending on the db wavelet used so this varies the selection size
           Range("a1000:a1127").Select
           'selects and copies over the appropriate transposed file
           Selection.Copy
           Cells(260 + (d * 10) + i, 2).Select
           Selection.PasteSpecial Paste:=xlAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=True
           Range("a1000:a1300").Select
           Selection.ClearContents
           Cells(270 + ((d - 1) * 10) + i, 1) = Cells(134, i)
           cmdstring2 = "clear"
           MLEvalstring (cmdstring2)
        Next i
      End If
   Next d
   Cells(270, 1).Select
Next q
End Sub

******************
Sub dataGatherFour4()
******************
```

```
Dim d As Integer
Dim e As Integer
Dim q As Integer
Dim sheetcount As Integer
Dim sheet As String
Dim sheet1 As String

'Change to the average number of scales being analyzed
For q = 1 To 5
  Sheets("starter").Select
  sheet1 = Cells(q, 1)
  Range("b1:b250").Select
  Selection.Copy
  Sheets(sheet1).Select
  Cells(2, 1).Select
  ActiveSheet.Paste
  ActiveWindow.Zoom = 70
  Range("b2:iv2000").Select
  Selection.Clear
  Range("a250:iv2000").Select
  Selection.Clear
  Range("a2:a250").Select
  Selection.Copy
    'finds the number of sheets and the names of the sheets
    sheetcount = 0
    For d = 1 To 248
      If Cells(1 + d, 1) <> "" Then
        sheetcount = sheetcount + 1
      End If
    Next d
    'goes to selected sheet and copies the needed coefs. and pastes them back on the summary page
    For e = 1 To sheetcount
      sheet = Cells(1 + e, 1)
      Sheets(sheet).Select
      If Cells(270 + q, 2) <> "" Then
        Range(Cells(270 + q, 3), Cells(270 + q, 66)).Select
        Selection.Copy
        Sheets(sheet1).Select
        Cells(1 + e, 2).Select
        ActiveSheet.Paste
        Sheets(sheet).Select
        Range(Cells(280 + q, 3), Cells(280 + q, 66)).Select
        Selection.Copy
        Sheets(sheet1).Select
        Cells(1 + e, 66).Select
        ActiveSheet.Paste
      ElseIf Cells(270 + q, 2) = "" Then
        Sheets(sheet1).Select
      End If
    Next e
    'does formatting of page
    Sheets("starter").Select
    Range("a258:dx258").Select
    Selection.Copy
    Sheets(sheet1).Select
    Cells(1, 2).Select
```

```
        ActiveSheet.Paste
        Cells(1, 1) = "fft"
        Range(Cells(2, 1), Cells(1 + sheetcount, 1)).Select
        Selection.Copy
        Cells(2, 1).Select
        ActiveSheet.Paste
        Cells(1, 1).Select
Next q
End Sub

********************
Sub SPSSanalysisPrep5()
********************

Dim w As Integer, i As Integer, e As Integer, y As Integer, p As Integer, u As Integer
Dim q As Integer, counter As Integer, counter1 As Integer, counter2 As Integer
Dim sheetselect As String, worksave As String, worksave1 As String

  'Keep as this workbook
  Windows("MathesonAge5Four.xls").Activate
  Sheets("starter").Select
  'Can change up to 9 if doing average of nine
  For w = 1 To 5
    sheetselect = Cells(w, 1)
    'Can change the number of wavelets
    For i = 1 To 1
      If Cells(261 + i + ((w - 1)), 1) <> "" Then
        worksave = Cells(261 + i + ((w - 1)), 1)
        worksave1 = Cells(261 + i + ((w - 1)), 10)
        Workbooks.Add
        ActiveWorkbook.SaveAs FileName:=(worksave) _
          , FileFormat:=xlExcel4, Password:="", WriteResPassword:="", _
          ReadOnlyRecommended:=False, CreateBackup:=False
        counter = 0
        For e = 1 To 248
          'change to first workbook
          Windows("GrandRapidsAge5Four.xls").Activate
          Sheets(sheetselect).Select
          If Cells(1 + e + ((i - 1) * 250), 2) <> "" Then
            counter = counter + 1
            Range(Cells(1 + e + ((i - 1) * 250), 1), Cells(1 + e + ((i - 1) * 250), 242)).Select
            Selection.Copy
            Windows(worksave1).Activate
            Cells(1 + counter, 1).Select
            ActiveSheet.Paste
          End If
        Next e
        counter2 = 0
        For y = 1 To 248
          'change to second workbook
          Windows("RivertonAge4Four.xls").Activate
          Sheets(sheetselect).Select
          If Cells(1 + y + ((i - 1) * 250), 2) <> "" Then
            counter2 = counter2 + 1
            Range(Cells(1 + y + ((i - 1) * 250), 1), Cells(1 + y + ((i - 1) * 250), 242)).Select
            Selection.Copy
```

```
                Windows(worksave1).Activate
                Cells(1 + counter2 + counter, 1).Select
                ActiveSheet.Paste
            End If
        Next y
        'Formatting of the worksheet so that its ready for SPSS
        Windows(worksave1).Activate
        Cells(1, 1) = "fish"
        Columns("B:B").Select
        Application.CutCopyMode = False
        Selection.Insert Shift:=xlToRight
        Cells(1, 2) = "Group"
        Columns("c:c").Select
        Application.CutCopyMode = False
        Selection.Insert Shift:=xlToRight
        Cells(1, 3) = "select"
        counter1 = 0
        For p = 1 To 400
            If Cells(1 + p, 1) <> "" Then
                counter1 = counter1 + 1
            End If
        Next p
        For u = 1 To counter1
            If u <= counter Then Cells(1 + u, 2) = 1
            If u > counter Then Cells(1 + u, 2) = 2
            Cells(1 + u, 3) = 1
        Next u
        'Keep as this workbook
        Windows("MathesonAge5Four.xls").Activate
        Sheets("starter").Select
        Range(Cells(258, 1), Cells(258, 128)).Select
        Selection.Copy
        Windows(worksave1).Activate
        Cells(1, 4).Select
        ActiveSheet.Paste
        Cells(1, 1).Select
        ActiveWorkbook.Save
        ActiveWorkbook.Close
    End If
    'Keep as this workbook
    Windows("MathesonAge5Four.xls").Activate
    Sheets("starter").Select
    Next i
  Next w
End Sub
```

Appendix 4. SPSS® program to perform a Monte Carlo randomization of the data and then run a stepwise discriminant analysis on the new data set. The syntax is repeated 1000 times to create a null distribution from the discriminant analysis classification rates.

```
SET MXLOOP=1000.
SET SEED=RANDOM

MATRIX.
GET A /FILE=* /variables=group select a1 a10 a11 a12 a13 a14 a15 a16 a17 a18 a19 a2 a20 a21 a22 a23
 a24 a25 a26 a27 a28 a29 a3 a30 a31 a32 a33 a34 a35 a36 a37 a38 a39 a4 a40
 a41 a42 a43 a44 a45 a46 a47 a48 a49 a5 a50 a51 a52 a53 a54 a55 a56 a57 a58
 a59 a6 a60 a61 a62 a63 a64 a7 a8 a9 p1 p10 p11 p12 p13 p14 p15 p16 p17 p18
 p19 p2 p20 p21 p22 p23 p24 p25 p26 p27 p28 p29 p3 p30 p31 p32 p33 p34 p35
 p36 p37 p38 p39 p4 p40 p41 p42 p43 p44 p45 p46 p47 p48 p49 p5 p50 p51 p52
 p53 p54 p55 p56 p57 p58 p59 p6 p60 p61 p62 p63 p64 p7 p8 p9.
COMPUTE NR = NROW(A).
LOOP RCOUNT=1 TO NR BY 1.
+ COMPUTE SWITCH = TRUNC(UNIFORM(1,1)*NR)+1.
+ COMPUTE TEMP = A(RCOUNT,1).
+ COMPUTE A(RCOUNT,1) = A(SWITCH,1).
+ COMPUTE A(SWITCH,1)=TEMP.
END LOOP.
SAVE A /OUTFILE=* /VARIABLES group select a1 a10 a11 a12 a13 a14 a15 a16 a17 a18 a19 a2 a20
a21 a22 a23
 a24 a25 a26 a27 a28 a29 a3 a30 a31 a32 a33 a34 a35 a36 a37 a38 a39 a4 a40
 a41 a42 a43 a44 a45 a46 a47 a48 a49 a5 a50 a51 a52 a53 a54 a55 a56 a57 a58
 a59 a6 a60 a61 a62 a63 a64 a7 a8 a9 p1 p10 p11 p12 p13 p14 p15 p16 p17 p18
 p19 p2 p20 p21 p22 p23 p24 p25 p26 p27 p28 p29 p3 p30 p31 p32 p33 p34 p35
 p36 p37 p38 p39 p4 p40 p41 p42 p43 p44 p45 p46 p47 p48 p49 p5 p50 p51 p52
 p53 p54 p55 p56 p57 p58 p59 p6 p60 p61 p62 p63 p64 p7 p8 p9.
END MATRIX.

DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=a1 a10 a11 a12 a13 a14 a15 a16 a17 a18 a19 a2 a20 a21 a22 a23
 a24 a25 a26 a27 a28 a29 a3 a30 a31 a32 a33 a34 a35 a36 a37 a38 a39 a4 a40
 a41 a42 a43 a44 a45 a46 a47 a48 a49 a5 a50 a51 a52 a53 a54 a55 a56 a57 a58
 a59 a6 a60 a61 a62 a63 a64 a7 a8 a9 p1 p10 p11 p12 p13 p14 p15 p16 p17 p18
 p19 p2 p20 p21 p22 p23 p24 p25 p26 p27 p28 p29 p3 p30 p31 p32 p33 p34 p35
 p36 p37 p38 p39 p4 p40 p41 p42 p43 p44 p45 p46 p47 p48 p49 p5 p50 p51 p52
 p53 p54 p55 p56 p57 p58 p59 p6 p60 p61 p62 p63 p64 p7 p8 p9
 /ANALYSIS ALL
  /METHOD=WILKS
  /MAXSTEPS = 20
  /PIN= .05
  /POUT= .10
 /PRIORS  size
  /HISTORY
```

```
/STATISTICS=TABLE CROSSVALID
/PLOT=CASES
/CLASSIFY=NONMISSING POOLED .
```

Appendix 5. VBA program to parse the text output file from SPSS® and retrieve the discriminant analysis results for the randomization and jackknife classifications. Program also calculates the p-value.

```
Option Explicit
Option Base 1
**********
Sub Parser()
**********

Dim InputData As String, CompString As String, CompString1 As String, compString2 As String,
CompStringALT As String
Dim compString3 As String, compString4 As String
Dim count As Integer, t As Integer, ii As Integer, startposition As Integer, length As Integer, count1 As
Integer
Dim y As Integer, count2 As Integer, count3 As Integer
Dim variable As Variant, variable1 As Variant

Open "G:\ output1.txt" For Input As #1   ' Open file for input.
Sheets("Raw").Select
Range("A3:D403,G3:J403,N3:gr403").Select
Selection.Clear
count1 = 0
count3 = 0
CompString = Chr(9) + Chr(9) + Chr(9) + "Statistic" + Chr(9) + "df1" + Chr(9) + "df2" + Chr(9) + "df3"
+ Chr(9) + "Exact F"
CompString1 = Chr(9) + Chr(9) + "Cross-validated" + Chr(9) + "Count"

Do While Not EOF(1) ' Check for end of file.
  Line Input #1, InputData   ' Read line of data.
  CompString = Chr(9) + Chr(9) + Chr(9) + "Statistic" + Chr(9) + "df1" + Chr(9) + "df2" + Chr(9) +
"df3" + Chr(9) + "Exact F"
  CompStringALT = Chr(9) + Chr(9) + Chr(9) + Chr(9) + "Statistic" + Chr(9) + "df1" + Chr(9) + "df2" +
Chr(9) + "df3" + Chr(9) + "Exact F"
  CompString1 = Chr(9) + Chr(9) + "Cross-validated" + Chr(9) + "Count"
  compString2 = "Cases Selected" + Chr(9) + "Original" + Chr(9) + "Count"
  compString3 = "Variables Entered/Removed"
  compString4 = "Step"
  count = 8
  count2 = 55
  If Left(InputData, 32) = CompString Then
    Line Input #1, InputData ' Read line of data.
    Line Input #1, InputData ' Read line of data.
    count1 = count1 + 1
    'If count1 = 90 Then End

    Do While InputData <> "" And Not EOF(1)
      ii = 1
      Do While Mid(InputData, ii, 1) <> Chr(9)
        ii = ii + 1
      Loop
      startposition = ii + 1
```

```
        ii = startposition
        Do While Mid(InputData, ii, 1) <> Chr(9)
          ii = ii + 1
        Loop
        length = ii - startposition

        variable = Mid(InputData, startposition, length)
        count = count + 1
        Sheets("Raw").Select
        Cells(count1 + 2, count + 5) = variable
        Line Input #1, InputData ' Read line of data.
      Loop
    End If

    If Left(InputData, 33) = CompStringALT Then
      Line Input #1, InputData ' Read line of data.
      Line Input #1, InputData ' Read line of data.

      count1 = count1 + 1
      'If count1 = 90 Then End

      Do While InputData <> "" And Not EOF(1)
        ii = 1
        Do While Mid(InputData, ii, 1) <> Chr(9)
          ii = ii + 1
        Loop
        startposition = ii + 1

        ii = startposition
        Do While Mid(InputData, ii, 1) <> Chr(9)
          ii = ii + 1
        Loop
        length = ii - startposition

        variable = Mid(InputData, startposition, length)
        count = count + 1
        Sheets("Raw").Select
        Cells(count1 + 2, count + 5) = variable
        If variable = "" Then
          ii = 2 Or 3
          Do While Mid(InputData, ii, 1) <> Chr(9)
            ii = ii + 1
          Loop
          If count >= 18 Then startposition = ii + 2
          If count < 18 Then startposition = ii + 1
           ii = startposition + 1
          Do While Mid(InputData, ii, 1) <> Chr(9)
            ii = ii + 1
          Loop
          length = ii - startposition

          variable = Mid(InputData, startposition, length)
          count = count + 1
          Sheets("Raw").Select
          Cells(count1 + 2, count2) = variable
          Cells(count1 + 2, count2).Select
```

```
            Selection.Font.Bold = True
            count = count - 1
            count2 = count2 + 1
        End If
        Line Input #1, InputData ' Read line of data.
    Loop
End If

count = 1
If Left(InputData, 23) = CompString1 Then
    For t = 1 To 2
        If t = 1 Then ii = 25 Else ii = 8

        Do While Mid(InputData, ii, 1) <> Chr(9)
            ii = ii + 1
        Loop
        startposition = ii + 1

        ii = startposition
        Do While Mid(InputData, ii, 1) <> Chr(9)
            ii = ii + 1
        Loop
        length = ii - startposition

        variable = Mid(InputData, startposition, length)
        Cells(count1 + 2, count + 6) = variable

        If t = 1 Then ii = 30 Else ii = 10
        Do While Mid(InputData, ii, 1) <> Chr(9)
            ii = ii + 1
        Loop
        startposition = ii + 1

        ii = startposition
        Do While Mid(InputData, ii, 1) <> Chr(9)
                ii = ii + 1
        Loop
        length = ii - startposition

        variable1 = Mid(InputData, startposition, length)
        Sheets("Raw").Select
        Cells(count1 + 2, count + 7) = variable1

        count = count + 2
        Line Input #1, InputData ' Read line of data.
    Next t
End If

If Left(InputData, 29) = compString2 Then
    For t = 1 To 2
        If t = 1 Then ii = 33 Else ii = 8
        Do While Mid(InputData, ii, 1) <> Chr(9)
            ii = ii + 1
        Loop
        startposition = ii + 1
        ii = startposition
```

```
        Do While Mid(InputData, ii, 1) <> Chr(9)
          ii = ii + 1
        Loop
        length = ii - startposition

        variable = Mid(InputData, startposition, length)
        Cells(count1 + 2, count) = variable

        If t = 1 Then ii = 36 Else ii = 10
        Do While Mid(InputData, ii, 1) <> Chr(9)
          ii = ii + 1
        Loop
        startposition = ii + 1
         ii = startposition
        Do While Mid(InputData, ii, 1) <> Chr(9)
            ii = ii + 1
        Loop
        length = ii - startposition
        variable1 = Mid(InputData, startposition, length)
        Sheets("Raw").Select
        Cells(count1 + 2, count + 1) = variable1

        count = count + 2
        Line Input #1, InputData ' Read line of data.
      Next t
    End If

    If Left(InputData, 25) = compString3 Then
      Line Input #1, InputData ' Read line of data.
      If Left(InputData, 4) = compString4 Then
        count3 = count3 + 1
        Cells(1, 18) = count3
      End If
    End If
Loop
Cells(1, 18) = count3
MsgBox "DONE"
Range("A3:Bv220").Select
Selection.Copy
Close #1    ' Close file.
End Sub

******************
Sub countlargevalues()
******************

Dim count As Integer, t As Integer
t = 0
count = 0
For t = 1 To 1000
  If Cells(2 + t, 12) >= Cells(1007, 6) Then count = count + 1
Next t
Cells(1008, 6) = count
End Sub
```

Appendix 6. VBA program to create a set of selection variables in Excel® used in SPSS® that do not repeat a sequence so 1000 discriminant analyses can be performed without repetition.

```
Option Explicit
Option Base 1
************************
Sub RandomTableGenerator()
************************
Range("b2:iv10000").Select
Selection.ClearContents

Count = 42
For d = 1 To 42  ' d = number of cases group 1
   For t = 1 To 87  ' t = number of combinations for group 2
      For y = 1 To 42
         If y <> Count Then
         Cells(2 + t + ((d - 1) * 87), 1 + y) = 1  ' change multiplier to number of cases in group 2
         End If
      Next y
   Next t
Count = Count - 1
Next d

For d = 1 To 42  ' d = number of cases group 1
   Count1 = 87  ' count1 = number of cases group 2
   For t = 1 To 87  ' t = number of combinations for group 2
      For y = 1 To 87  ' y = number of selected and pasted variables for group 2
         If y <> Count1 Then
         Cells(2 + t + ((d - 1) * 87), 43 + y) = 1  ' change multiplier to number of cases in group 2
         End If
      Next y
      Count1 = Count1 - 1
   Next t
Next d

Range("B3:IV30002").Select
Selection.Copy
End Sub
```

Appendix 7. SPSS® syntax to perform stepwise discriminant analyses with the selection variables generated from the appendix 6 VBA program. The selection variables are increased from "var00001(1)" to "var01000(1)".

```
DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=a1 a10 a11 a12 a13 a14 a15 a16 a17 a18 a19 a2 a20 a21 a22 a23
 a24 a25 a26 a27 a28 a29 a3 a30 a31 a32 a33 a34 a35 a36 a37 a38 a39 a4 a40
 a41 a42 a43 a44 a45 a46 a47 a48 a49 a5 a50 a51 a52 a53 a54 a55 a56 a57 a58
 a59 a6 a60 a61 a62 a63 a64 a7 a8 a9 p1 p10 p11 p12 p13 p14 p15 p16 p17 p18
 p19 p2 p20 p21 p22 p23 p24 p25 p26 p27 p28 p29 p3 p30 p31 p32 p33 p34 p35
 p36 p37 p38 p39 p4 p40 p41 p42 p43 p44 p45 p46 p47 p48 p49 p5 p50 p51 p52
 p53 p54 p55 p56 p57 p58 p59 p6 p60 p61 p62 p63 p64 p7 p8 p9
 /SELECT=var00001(1)
 /ANALYSIS ALL
  /METHOD=WILKS
  /MAXSTEPS = 20
  /PIN= .05
  /POUT= .10
 /PRIORS  size
  /HISTORY
 /STATISTICS=TABLE CROSSVALID
 /PLOT=CASES
 /CLASSIFY=NONMISSING POOLED .

DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=aa1 aa10 aa11 aa12 aa13 aa14 aa15 aa16 aa17 aa18 aa19 aa2 aa20 aa21 aa22 aa23
 aa24 aa25 aa26 aa27 aa28 aa29 aa3 aa30 aa31 aa32 aa33 aa34 aa35 aa36 aa37 aa38 aa39 aa4 aa40
 aa41 aa42 aa43 aa44 aa45 aa46 aa47 aa48 aa49 aa5 aa50 aa51 aa52 aa53 aa54 aa55 aa56 aa57 aa58
 aa59 aa6 aa60 aa61 aa62 aa63 aa64 aa7 aa8 aa9 pa1 pa10 pa11 pa12 pa13 pa14 pa15 pa16 pa17 pa18
 pa19 pa2 pa20 pa21 pa22 pa23 pa24 pa25 pa26 pa27 pa28 pa29 pa3 pa30 pa31 pa32 pa33 pa34 pa35
 pa36 pa37 pa38 pa39 pa4 pa40 pa41 pa42 pa43 pa44 pa45 pa46 pa47 pa48 pa49 pa5 pa50 pa51 pa52
 pa53 pa54 pa55 pa56 pa57 pa58 pa59 pa6 pa60 pa61 pa62 pa63 pa64 pa7 pa8 pa9
 /SELECT=var00001(1)
 /ANALYSIS ALL
  /METHOD=WILKS
  /MAXSTEPS = 20
  /PIN= .05
  /POUT= .10
 /PRIORS  size
  /HISTORY
 /STATISTICS=TABLE CROSSVALID
 /PLOT=CASES
 /CLASSIFY=NONMISSING POOLED .

DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=ab1 ab10 ab11 ab12 ab13 ab14 ab15 ab16 ab17 ab18 ab19 ab2 ab20 ab21 ab22 ab23
 ab24 ab25 ab26 ab27 ab28 ab29 ab3 ab30 ab31 ab32 ab33 ab34 ab35 ab36 ab37 ab38 ab39 ab4 ab40
 ab41 ab42 ab43 ab44 ab45 ab46 ab47 ab48 ab49 ab5 ab50 ab51 ab52 ab53 ab54 ab55 ab56 ab57 ab58
 ab59 ab6 ab60 ab61 ab62 ab63 ab64 ab7 ab8 ab9 pb1 pb10 pb11 pb12 pb13 pb14 pb15 pb16 pb17 pb18
```

```
pb19 pb2 pb20 pb21 pb22 pb23 pb24 pb25 pb26 pb27 pb28 pb29 pb3 pb30 pb31 pb32 pb33 pb34 pb35
pb36 pb37 pb38 pb39 pb4 pb40 pb41 pb42 pb43 pb44 pb45 pb46 pb47 pb48 pb49 pb5 pb50 pb51 pb52
pb53 pb54 pb55 pb56 pb57 pb58 pb59 pb6 pb60 pb61 pb62 pb63 pb64 pb7 pb8 pb9
/SELECT=var00001(1)
/ANALYSIS ALL
 /METHOD=WILKS
 /MAXSTEPS = 20
 /PIN= .05
 /POUT= .10
/PRIORS  size
 /HISTORY
/STATISTICS=TABLE CROSSVALID
/PLOT=CASES
/CLASSIFY=NONMISSING POOLED .

DISCRIMINANT
/GROUPS=group(1 3)
/VARIABLES=ac1 ac10 ac11 ac12 ac13 ac14 ac15 ac16 ac17 ac18 ac19 ac2 ac20 ac21 ac22 ac23
ac24 ac25 ac26 ac27 ac28 ac29 ac3 ac30 ac31 ac32 ac33 ac34 ac35 ac36 ac37 ac38 ac39 ac4 ac40
ac41 ac42 ac43 ac44 ac45 ac46 ac47 ac48 ac49 ac5 ac50 ac51 ac52 ac53 ac54 ac55 ac56 ac57 ac58
ac59 ac6 ac60 ac61 ac62 ac63 ac64 ac7 ac8 ac9 pc1 pc10 pc11 pc12 pc13 pc14 pc15 pc16 pc17 pc18
pc19 pc2 pc20 pc21 pc22 pc23 pc24 pc25 pc26 pc27 pc28 pc29 pc3 pc30 pc31 pc32 pc33 pc34 pc35
pc36 pc37 pc38 pc39 pc4 pc40 pc41 pc42 pc43 pc44 pc45 pc46 pc47 pc48 pc49 pc5 pc50 pc51 pc52
pc53 pc54 pc55 pc56 pc57 pc58 pc59 pc6 pc60 pc61 pc62 pc63 pc64 pc7 pc8 pc9
/SELECT=var00001(1)
/ANALYSIS ALL
 /METHOD=WILKS
 /MAXSTEPS = 20
 /PIN= .05
 /POUT= .10
/PRIORS  size
 /HISTORY
/STATISTICS=TABLE CROSSVALID
/PLOT=CASES
/CLASSIFY=NONMISSING POOLED .

DISCRIMINANT
/GROUPS=group(1 3)
/VARIABLES=ad1 ad10 ad11 ad12 ad13 ad14 ad15 ad16 ad17 ad18 ad19 ad2 ad20 ad21 ad22 ad23
ad24 ad25 ad26 ad27 ad28 ad29 ad3 ad30 ad31 ad32 ad33 ad34 ad35 ad36 ad37 ad38 ad39 ad4 ad40
ad41 ad42 ad43 ad44 ad45 ad46 ad47 ad48 ad49 ad5 ad50 ad51 ad52 ad53 ad54 ad55 ad56 ad57 ad58
ad59 ad6 ad60 ad61 ad62 ad63 ad64 ad7 ad8 ad9 pd1 pd10 pd11 pd12 pd13 pd14 pd15 pd16 pd17 pd18
pd19 pd2 pd20 pd21 pd22 pd23 pd24 pd25 pd26 pd27 pd28 pd29 pd3 pd30 pd31 pd32 pd33 pd34 pd35
pd36 pd37 pd38 pd39 pd4 pd40 pd41 pd42 pd43 pd44 pd45 pd46 pd47 pd48 pd49 pd5 pd50 pd51 pd52
pd53 pd54 pd55 pd56 pd57 pd58 pd59 pd6 pd60 pd61 pd62 pd63 pd64 pd7 pd8 pd9
/SELECT=var00001(1)
/ANALYSIS ALL
 /METHOD=WILKS
 /MAXSTEPS = 20
 /PIN= .05
 /POUT= .10
/PRIORS  size
  HISTORY
/STATISTICS=TABLE CROSSVALID
/PLOT=CASES
/CLASSIFY=NONMISSING POOLED .
```

```
DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=a1 a10 a11 a12 a13 a14 a15 a16 a17 a18 a19 a2 a20 a21 a22 a23
 a24 a25 a26 a27 a28 a29 a3 a30 a31 a32 a33 a34 a35 a36 a37 a38 a39 a4 a40
 a41 a42 a43 a44 a45 a46 a47 a48 a49 a5 a50 a51 a52 a53 a54 a55 a56 a57 a58
 a59 a6 a60 a61 a62 a63 a64 a7 a8 a9 p1 p10 p11 p12 p13 p14 p15 p16 p17 p18
 p19 p2 p20 p21 p22 p23 p24 p25 p26 p27 p28 p29 p3 p30 p31 p32 p33 p34 p35
 p36 p37 p38 p39 p4 p40 p41 p42 p43 p44 p45 p46 p47 p48 p49 p5 p50 p51 p52
 p53 p54 p55 p56 p57 p58 p59 p6 p60 p61 p62 p63 p64 p7 p8 p9
 /SELECT=var00002(1)
 /ANALYSIS ALL
  /METHOD=WILKS
  /MAXSTEPS = 20
  /PIN= .05
  /POUT= .10
 /PRIORS  size
  /HISTORY
 /STATISTICS=TABLE CROSSVALID
 /PLOT=CASES
 /CLASSIFY=NONMISSING POOLED .

DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=aa1 aa10 aa11 aa12 aa13 aa14 aa15 aa16 aa17 aa18 aa19 aa2 aa20 aa21 aa22 aa23
 aa24 aa25 aa26 aa27 aa28 aa29 aa3 aa30 aa31 aa32 aa33 aa34 aa35 aa36 aa37 aa38 aa39 aa4 aa40
 aa41 aa42 aa43 aa44 aa45 aa46 aa47 aa48 aa49 aa5 aa50 aa51 aa52 aa53 aa54 aa55 aa56 aa57 aa58
 aa59 aa6 aa60 aa61 aa62 aa63 aa64 aa7 aa8 aa9 pa1 pa10 pa11 pa12 pa13 pa14 pa15 pa16 pa17 pa18
 pa19 pa2 pa20 pa21 pa22 pa23 pa24 pa25 pa26 pa27 pa28 pa29 pa3 pa30 pa31 pa32 pa33 pa34 pa35
 pa36 pa37 pa38 pa39 pa4 pa40 pa41 pa42 pa43 pa44 pa45 pa46 pa47 pa48 pa49 pa5 pa50 pa51 pa52
 pa53 pa54 pa55 pa56 pa57 pa58 pa59 pa6 pa60 pa61 pa62 pa63 pa64 pa7 pa8 pa9
 /SELECT=var00002(1)
 /ANALYSIS ALL
  /METHOD=WILKS
  /MAXSTEPS = 20
  /PIN= .05
  /POUT= .10
 /PRIORS  size
  /HISTORY
 /STATISTICS=TABLE CROSSVALID
 /PLOT=CASES
 /CLASSIFY=NONMISSING POOLED .

DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=ab1 ab10 ab11 ab12 ab13 ab14 ab15 ab16 ab17 ab18 ab19 ab2 ab20 ab21 ab22 ab23
 ab24 ab25 ab26 ab27 ab28 ab29 ab3 ab30 ab31 ab32 ab33 ab34 ab35 ab36 ab37 ab38 ab39 ab4 ab40
 ab41 ab42 ab43 ab44 ab45 ab46 ab47 ab48 ab49 ab5 ab50 ab51 ab52 ab53 ab54 ab55 ab56 ab57 ab58
 ab59 ab6 ab60 ab61 ab62 ab63 ab64 ab7 ab8 ab9 pb1 pb10 pb11 pb12 pb13 pb14 pb15 pb16 pb17 pb18
 pb19 pb2 pb20 pb21 pb22 pb23 pb24 pb25 pb26 pb27 pb28 pb29 pb3 pb30 pb31 pb32 pb33 pb34 pb35
 pb36 pb37 pb38 pb39 pb4 pb40 pb41 pb42 pb43 pb44 pb45 pb46 pb47 pb48 pb49 pb5 pb50 pb51 pb52
 pb53 pb54 pb55 pb56 pb57 pb58 pb59 pb6 pb60 pb61 pb62 pb63 pb64 pb7 pb8 pb9
 /SELECT=var00003(1)
 /ANALYSIS ALL
  /METHOD=WILKS
```

```
 /MAXSTEPS = 20
 /PIN= .05
 /POUT= .10
/PRIORS size
 /HISTORY
/STATISTICS=TABLE CROSSVALID
/PLOT=CASES
/CLASSIFY=NONMISSING POOLED .

DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=ac1 ac10 ac11 ac12 ac13 ac14 ac15 ac16 ac17 ac18 ac19 ac2 ac20 ac21 ac22 ac23
 ac24 ac25 ac26 ac27 ac28 ac29 ac3 ac30 ac31 ac32 ac33 ac34 ac35 ac36 ac37 ac38 ac39 ac4 ac40
 ac41 ac42 ac43 ac44 ac45 ac46 ac47 ac48 ac49 ac5 ac50 ac51 ac52 ac53 ac54 ac55 ac56 ac57 ac58
 ac59 ac6 ac60 ac61 ac62 ac63 ac64 ac7 ac8 ac9 pc1 pc10 pc11 pc12 pc13 pc14 pc15 pc16 pc17 pc18
 pc19 pc2 pc20 pc21 pc22 pc23 pc24 pc25 pc26 pc27 pc28 pc29 pc3 pc30 pc31 pc32 pc33 pc34 pc35
 pc36 pc37 pc38 pc39 pc4 pc40 pc41 pc42 pc43 pc44 pc45 pc46 pc47 pc48 pc49 pc5 pc50 pc51 pc52
 pc53 pc54 pc55 pc56 pc57 pc58 pc59 pc6 pc60 pc61 pc62 pc63 pc64 pc7 pc8 pc9
 /SELECT=var00004(1)
 /ANALYSIS ALL
  /METHOD=WILKS
  /MAXSTEPS = 20
  /PIN= .05
  /POUT= .10
 /PRIORS size
  /HISTORY
 /STATISTICS=TABLE CROSSVALID
 /PLOT=CASES
 /CLASSIFY=NONMISSING POOLED .

DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=ad1 ad10 ad11 ad12 ad13 ad14 ad15 ad16 ad17 ad18 ad19 ad2 ad20 ad21 ad22 ad23
 ad24 ad25 ad26 ad27 ad28 ad29 ad3 ad30 ad31 ad32 ad33 ad34 ad35 ad36 ad37 ad38 ad39 ad4 ad40
 ad41 ad42 ad43 ad44 ad45 ad46 ad47 ad48 ad49 ad5 ad50 ad51 ad52 ad53 ad54 ad55 ad56 ad57 ad58
 ad59 ad6 ad60 ad61 ad62 ad63 ad64 ad7 ad8 ad9 pd1 pd10 pd11 pd12 pd13 pd14 pd15 pd16 pd17 pd18
 pd19 pd2 pd20 pd21 pd22 pd23 pd24 pd25 pd26 pd27 pd28 pd29 pd3 pd30 pd31 pd32 pd33 pd34 pd35
 pd36 pd37 pd38 pd39 pd4 pd40 pd41 pd42 pd43 pd44 pd45 pd46 pd47 pd48 pd49 pd5 pd50 pd51 pd52
 pd53 pd54 pd55 pd56 pd57 pd58 pd59 pd6 pd60 pd61 pd62 pd63 pd64 pd7 pd8 pd9
 /SELECT=var00005(1)
 /ANALYSIS ALL
  /METHOD=WILKS
  /MAXSTEPS = 20
  /PIN= .05
  /POUT= .10
 /PRIORS size
  /HISTORY
 /STATISTICS=TABLE CROSSVALID
 /PLOT=CASES
 /CLASSIFY=NONMISSING POOLED .
```

Note: Continue increasing "var00001(1)" to "var01000(1)".

Appendix 8. VBA program to format Excel® workbook and perform wavelet transform on the 128 scale radii signals with Excel Link® to Matlab®. The transform variables are placed in summary worksheets and new workbooks that combine sampling locations are produced.

```
Option Explicit
Option Base 1
'**********************
Sub FormatWorkbookdb1()
'**********************

'important first macro which should be run before the other macros

Dim db_avg1 As String
Dim db_avg2 As String
Dim db_avg3 As String
Dim db_avg4 As String
Dim db_avg5 As String
Dim avg1 As String
Dim avg2 As String
Dim avg3 As String
Dim avg4 As String
Dim avg5 As String

If Cells(1, 1) <> "db_avg1" Then
   Sheets.Add
   ActiveSheet.Name = "db_avg1"
   Sheets.Add
   ActiveSheet.Name = "db_avg2"
   Sheets.Add
   ActiveSheet.Name = "db_avg3"
   Sheets.Add
   ActiveSheet.Name = "db_avg4"
   Sheets.Add
   ActiveSheet.Name = "db_avg5"
   Sheets.Add

   ActiveSheet.Name = "starter"
End If
   Cells(1, 1) = "db_avg1"
   Cells(2, 1) = "db_avg2"
   Cells(3, 1) = "db_avg3"
   Cells(4, 1) = "db_avg4"
   Cells(5, 1) = "db_avg5"
   Cells(260, 1) = "avg1"
   Cells(260, 2) = "avg2"
   Cells(260, 3) = "avg3"
   Cells(260, 4) = "avg4"
   Cells(260, 5) = "avg5"
```

```
End Sub

**********************
Sub customize_sheetdb2()
**********************

Dim b As Integer
Dim q As Integer
Dim k As Integer
Dim h As Integer
Dim i As Integer
Dim e As Integer
Dim sheetcount As Integer
Dim sheet As String

Sheets("starter").Select
sheetcount = 0
For b = 1 To 250
  If Cells(b, 2) <> "" Then
    sheetcount = sheetcount + 1
  End If
Next b
For q = 1 To sheetcount
  Sheets("starter").Select
  sheet = Cells(q, 2)
  Sheets(sheet).Select
  If Cells(134, 1) <> "Average of 1" Then
    Range("A134:A262").Select
    Selection.Insert Shift:=xlToRight
    Range("A2:A130").Select
    Selection.Copy
    Range("A134").Select
    ActiveSheet.Paste
    ActiveWindow.Zoom = 50
    Range("f134:k262").Select
    Selection.ClearContents
  End If
    ActiveWindow.Zoom = 50
    Range("a270:iv480").Select
    Selection.Clear
    Cells(270, 2) = 1
    For e = 1 To 240
      Cells(270, 2 + e) = Cells(270, 1 + e) + 1
    Next e
    Range("270:270").Select
    Selection.Font.Bold = True
    Cells(270, 1) = "db1"
    Cells(270, 1).Select
    With Selection.Font
      .Size = 14
      .Strikethrough = False
    End With
    Cells(280, 1) = "db5"
    Cells(280, 1).Select
    With Selection.Font
      .Size = 14
```

```
        .Strikethrough = False
      End With
      Cells(290, 1) = "db9"
      Cells(290, 1).Select
      With Selection.Font
        .Size = 14
        .Strikethrough = False
      End With
      Range("a270:a470").Select
      Selection.Font.Bold = True
      With Selection
        .HorizontalAlignment = xlCenter
        .VerticalAlignment = xlBottom
      End With
      Rows("270:270").Select
      Selection.NumberFormat = "General"
      With Selection
        .HorizontalAlignment = xlCenter
        .VerticalAlignment = xlBottom
      End With
Next q
Sheets("starter").Select
sheetcount = 0
For h = 1 To 250
  If Cells(h, 2) <> "" Then
    sheetcount = sheetcount + 1
  End If
Next h
For i = 1 To sheetcount
  Sheets("starter").Select
  sheet = Cells(i, 2)
  Rows("260:260").Select
  Selection.Copy
  Sheets(sheet).Select
  Rows("134:134").Select
  ActiveSheet.Paste
Next i
End Sub

**************************
Sub wavlet_Decompositiondb3()
**************************

Dim c As Variant, coef As Variant
Dim i As Integer, d As Integer, iCount As Integer, e As Integer, no_scales As Integer
Dim h As Integer, q As Integer, sheetcount As Integer, t As Integer
Dim vectorname As String, sheet As String
Dim cmdstring As String, Wavelet As String, cmdstring1 As String, cmdstring2 As String

'defines the sheets which will be used
Sheets("starter").Select
sheetcount = 0
For h = 1 To 250
  If Cells(h, 2) <> "" Then
    sheetcount = sheetcount + 1
  End If
```

```
Next h
For q = 1 To sheetcount
   Sheets("starter").Select
   sheet = Cells(q, 2)
   Sheets(sheet).Select
   no_scales = 0
   'finds out the number of scales; to be used as a counter
   For e = 1 To 9
     If Cells(135, e) <> "" Then
        no_scales = no_scales + 1
     End If
   Next e
     'performs wavelet decomposition on the signals
     For d = 1 To 3
        If Cells(260 + (d * 10), 1) <> "" Then
        Wavelet = Cells(260 + (d * 10), 1)
        For i = 1 To no_scales
           coef = ""
           vectorname = Cells(134, i)
           MLPutMatrix vectorname, Range(Cells(135, i), Cells(262, i))
           cmdstring = "[c,l]=wavedec(" + vectorname + ",7,'" + Wavelet + "')"
           MLEvalstring (cmdstring)
           cmdstring1 = "c"
           MLGetVar (cmdstring1), coef
           Range("a500:a800") = coef
           'number of coefs. differs depending on the db wavelet used so this varies the selection size
           If d = 1 Then
              Range("a500:a627").Select
           ElseIf d = 2 Then
              Range("a500:a684").Select
           ElseIf d = 3 Then
              Range("a500:a740").Select
           End If
           'selects and copies over the appropriate transposed file
           Selection.Copy
           Cells(260 + (d * 10) + i, 2).Select
           Selection.PasteSpecial Paste:=xlAll, Operation:=xlNone, SkipBlanks:=False, Transpose:=True
           Range("a500:a800").Select
           Selection.ClearContents
           Cells(270 + (d * 10 - 10) + i, 1) = Cells(134, i)
           cmdstring2 = "clear"
           MLEvalstring (cmdstring2)
        Next i
     End If
   Next d
   Cells(270, 1).Select
Next q
End Sub

*****************
Sub dataGatherdb4()
*****************

Dim f As Integer
Dim d As Integer
Dim e As Integer
```

```
Dim q As Integer
Dim sheetcount As Integer
Dim sheet As String
Dim sheet1 As String

'Change to the average number of scales being analyzed
For q = 1 To 5
   Sheets("starter").Select
   sheet1 = Cells(q, 1)
   Range("b1:b250").Select
   Selection.Copy
   Sheets(sheet1).Select
   Cells(2, 1).Select
   ActiveSheet.Paste
   ActiveWindow.Zoom = 70
   Range("b2:iv2000").Select
   Selection.Clear
   Range("a250:iv2000").Select
   Selection.Clear
   Range("a2:a250").Select
   Selection.Copy
   For f = 1 To 3
      'finds the number of sheets and the names of the sheets
      sheetcount = 0
      For d = 1 To 248
         If Cells(1 + ((f - 1) * 250) + d, 1) <> "" Then
            sheetcount = sheetcount + 1
         End If
      Next d
      'goes to selected sheet and copies the needed coefs. and pastes them back on the summary page
      For e = 1 To sheetcount
         sheet = Cells(1 + ((f - 1) * 250) + e, 1)
         Sheets(sheet).Select
         If Cells(270 + q, 2) <> "" Then
            Range(Cells(270 + q + ((f - 1) * 10), 2), Cells(270 + q + ((f - 1) * 10), 242)).Select
            Selection.Copy
            Sheets(sheet1).Select
            Cells(1 + ((f - 1) * 250) + e, 2).Select
            ActiveSheet.Paste
         ElseIf Cells(270 + q, 2) = "" Then
            Sheets(sheet1).Select
         End If
      Next e
      'does formatting of page
      Sheets(sheet).Select
      Range("b270:iv270").Select
      Selection.Copy
      Sheets(sheet1).Select
      Cells(1 + ((f - 1) * 250), 2).Select
      ActiveSheet.Paste
      If f = 1 Then
         Cells(1, 1) = "db1"
      ElseIf f = 2 Then
         Cells(251, 1) = "db5"
      ElseIf f = 3 Then
         Cells(501, 1) = "db9"
```

```
      End If
      If f <> 3 Then
         Range(Cells(2, 1), Cells(1 + sheetcount, 1)).Select
         Selection.Copy
         Cells(2 + (250 * f), 1).Select
         ActiveSheet.Paste
      End If
   Next f
   Cells(1, 1).Select
Next q

End Sub

*******************************
Sub SPSSanalysisPrepALLofOneAge()
*******************************

Dim w As Integer, i As Integer, e As Integer, y As Integer, p As Integer, u As Integer
Dim q As Integer, counter As Integer, counter1 As Integer, counter2 As Integer
Dim counter3 As Integer
Dim sheetselect As String, worksave As String, worksave1 As String

   'Keep as this workbook
   Windows("MathesonAge5.xls").Activate
   Sheets("starter").Select
   'Can change up to 9 if doing average of nine
   For w = 1 To 5
      sheetselect = Cells(w, 1)
      'Can change the number of wavelets
      For i = 1 To 3
         If Cells(261 + i + ((w - 1) * 3), 1) <> "" Then
            worksave = Cells(261 + i + ((w - 1) * 3), 1)
            worksave1 = Cells(261 + i + ((w - 1) * 3), 10)
            Workbooks.Add
            ActiveWorkbook.SaveAs FileName:=(worksave) _
               , FileFormat:=xlExcel4, Password:="", WriteResPassword:="", _
               ReadOnlyRecommended:=False, CreateBackup:=False
            counter = 0
            For e = 1 To 248
               'change to first workbook
               Windows("GrandRapidsAge5.xls").Activate
               Sheets(sheetselect).Select
               If Cells(1 + e + ((i - 1) * 250), 2) <> "" Then
                  counter = counter + 1
                  Range(Cells(1 + e + ((i - 1) * 250), 1), Cells(1 + e + ((i - 1) * 250), 242)).Select
                  Selection.Copy
                  Windows(worksave1).Activate
                  Cells(1 + counter, 1).Select
                  ActiveSheet.Paste
               End If
            Next e

            counter1 = 0
            For y = 1 To 248
               'change to second workbook
               Windows("mathesonAge5.xls").Activate
```

```
    Sheets(sheetselect).Select
    If Cells(1 + y + ((i - 1) * 250), 2) <> "" Then
      counter1 = counter1 + 1
      Range(Cells(1 + y + ((i - 1) * 250), 1), Cells(1 + y + ((i - 1) * 250), 242)).Select
      Selection.Copy
      Windows(worksave1).Activate
      Cells(1 + counter1 + counter, 1).Select
      ActiveSheet.Paste
    End If
  Next y

  counter2 = 0
  For y = 1 To 248
    'change to second workbook
    Windows("RivertonAge5.xls").Activate
    Sheets(sheetselect).Select
    If Cells(1 + y + ((i - 1) * 250), 2) <> "" Then
      counter2 = counter2 + 1
      Range(Cells(1 + y + ((i - 1) * 250), 1), Cells(1 + y + ((i - 1) * 250), 242)).Select
      Selection.Copy
      Windows(worksave1).Activate
      Cells(1 + counter2 + counter1 + counter, 1).Select
      ActiveSheet.Paste
    End If
  Next y

  'Formatting of the worksheet so that its ready for SPSS
  Windows(worksave1).Activate
  Cells(1, 1) = "fish"
  Columns("B:B").Select
  Application.CutCopyMode = False
  Selection.Insert Shift:=xlToRight
  Cells(1, 2) = "Group"
  Columns("c:c").Select
  Application.CutCopyMode = False
  Selection.Insert Shift:=xlToRight
  Cells(1, 3) = "select"
  counter3 = 0
  For p = 1 To 800
    If Cells(1 + p, 1) <> "" Then
      counter3 = counter3 + 1
    End If
  Next p
  For u = 1 To counter3
    If u <= counter Then Cells(1 + u, 2) = 1
    If u > counter Then Cells(1 + u, 2) = 2
    If u > (counter + counter1) Then Cells(1 + u, 2) = 3
    Cells(1 + u, 3) = 1
  Next u
  'Keep as this workbook
  Windows("MathesonAge5.xls").Activate
  Sheets("starter").Select
  Range(Cells(259, 1), Cells(259, 241)).Select
  Selection.Copy
  Windows(worksave1).Activate
  Cells(1, 4).Select
```

```
            ActiveSheet.Paste
            ' Change q to the possible number of coefficients
            For q = 1 To 250
               If Cells(2, 2 + q) = "" Then Cells(1, 2 + q) = ""
            Next q
            Cells(1, 1).Select
            ActiveWorkbook.Save
            ActiveWorkbook.Close
         End If
      'Keep as this workbook
      Windows("MathesonAge5.xls").Activate
      Sheets("starter").Select
      Next i
   Next w
End Sub
```

Appendix 9. VBA program to parse the text output file from SPSS® of the three-way discriminant analysis for the randomization and jackknife classifications. Program also calculates the p-value.

```
Option Explicit
Option Base 1

**********
Sub Parser()
**********

Dim InputData As String, CompString As String, CompString1 As String, compString2 As String,
CompStringALT As String
Dim compString3 As String, compString4 As String
Dim count As Integer, t As Integer, i As Integer, ii As Integer, startposition As Integer, lengthi As Integer,
count1 As Integer
Dim y As Integer, b As Integer, k As Integer, count2 As Integer, count3 As Integer, h As Integer, length
As Integer
Dim variable As Variant, variable1 As Variant, variable2 As Variant
Open "G:\ output1.txt" For Input As #1   ' Open file for input.
Sheets("Raw").Select
Cells(1, 11).Select
Selection.Clear
Range("A3:i403,l3:t403,x3:gx403").Select
Selection.Clear
count1 = 0
count3 = 0
CompString = Chr(9) + Chr(9) + Chr(9) + "Statistic" + Chr(9) + "df1" + Chr(9) + "df2" + Chr(9) + "df3"
+ Chr(9) + "Exact F"
CompString1 = "Cross-validated" + Chr(9) + "Count"

Do While Not EOF(1) ' Check for end of file.
   Line Input #1, InputData   ' Read line of data.
   CompString = Chr(9) + Chr(9) + Chr(9) + "Statistic" + Chr(9) + "df1" + Chr(9) + "df2" + Chr(9) +
"df3" + Chr(9) + "Exact F"
   CompStringALT = Chr(9) + Chr(9) + Chr(9) + Chr(9) + "Statistic" + Chr(9) + "df1" + Chr(9) + "df2" +
Chr(9) + "df3" + Chr(9) + "Exact F"
   CompString1 = "Cross-validated" + Chr(9) + "Count"
   compString2 = "Original" + Chr(9) + "Count"
   compString3 = "Variables Entered/Removed"
   compString4 = "Step"
   count = 8
   count2 = 55
   If Left(InputData, 32) = CompString Then
      Line Input #1, InputData ' Read line of data.
      Line Input #1, InputData ' Read line of data.

      count1 = count1 + 1
      'If count1 = 43 Then Cells(1, 11) = count3
```

```
  'If count1 = 43 Then End

  Do While InputData <> "" And Not EOF(1) 'Coefficients
    ii = 1
    Do While Mid(InputData, ii, 1) <> Chr(9)
      ii = ii + 1
    Loop
    startposition = ii + 1

    ii = startposition
    Do While Mid(InputData, ii, 1) <> Chr(9)
      ii = ii + 1
    Loop
    length = ii - startposition

    variable = Mid(InputData, startposition, length)
    count = count + 1
    Sheets("Raw").Select
    Cells(count1 + 2, count + 15) = variable
    Line Input #1, InputData ' Read line of data.
  Loop
End If

If Left(InputData, 33) = CompStringALT Then 'Coefficients
  Line Input #1, InputData ' Read line of data.
  Line Input #1, InputData ' Read line of data.

  count1 = count1 + 1

  Do While InputData <> "" And Not EOF(1)
    ii = 1
    Do While Mid(InputData, ii, 1) <> Chr(9)
      ii = ii + 1
    Loop
    startposition = ii + 1

    ii = startposition
    Do While Mid(InputData, ii, 1) <> Chr(9)
      ii = ii + 1
    Loop
    length = ii - startposition

    variable = Mid(InputData, startposition, length)
    count = count + 1
    Sheets("Raw").Select
    Cells(count1 + 2, count + 15) = variable

    If variable = "" Then
      ii = 2 Or 3
      Do While Mid(InputData, ii, 1) <> Chr(9)
        ii = ii + 1
      Loop
      If count >= 18 Then startposition = ii + 2
      If count < 18 Then startposition = ii + 1

      ii = startposition + 1
```

```
            Do While Mid(InputData, ii, 1) <> Chr(9)
               ii = ii + 1
            Loop
            length = ii - startposition

            variable = Mid(InputData, startposition, length)
            count = count + 1
            Sheets("Raw").Select
            Cells(count1 + 2, count2) = variable
            Cells(count1 + 2, count2).Select
            Selection.Font.Bold = True
            count = count - 1
            count2 = count2 + 1
         End If

         Line Input #1, InputData ' Read line of data.

      Loop
   End If

   count = 1
   If Left(InputData, 14) = compString2 Then
      For h = 1 To 3
         If h = 1 Then i = 20 Else i = 8
         For t = 1 To 3
            Do While Mid(InputData, i, 1) <> Chr(9)
               i = i + 1
            Loop
            startposition = i + 1

            i = startposition
            Do While Mid(InputData, i, 1) <> Chr(9)
               i = i + 1
            Loop
            lengthi = i - startposition

            variable1 = Mid(InputData, startposition, lengthi)
            Cells(count1 + 2, count + (t - 1)) = variable1

         Next t
         count = count + 3
         Line Input #1, InputData ' Read line of data.
      Next h
   End If

   count = 1
   If Left(InputData, 21) = CompString1 Then
      For b = 1 To 3
         If b = 1 Then i = 27 Else i = 8
         For k = 1 To 3
            Do While Mid(InputData, i, 1) <> Chr(9)
               i = i + 1
            Loop
            startposition = i + 1

            i = startposition
```

```
            Do While Mid(InputData, i, 1) <> Chr(9)
               i = i + 1
            Loop
            lengthi = i - startposition

            variable2 = Mid(InputData, startposition, lengthi)
            Cells(count1 + 2, count + (k - 1) + 11) = variable2

         Next k
         count = count + 3
         Line Input #1, InputData ' Read line of data.
      Next b
   End If

   If Left(InputData, 25) = compString3 Then
      Line Input #1, InputData ' Read line of data.
      If Left(InputData, 4) = compString4 Then
         count3 = count3 + 1
      End If
   End If
Loop
Cells(1, 11) = count3
MsgBox "DONE"
Range("A3:Bz220").Select
Selection.Copy
Close #1    ' Close file.
End Sub


******************
Sub countlargevalues()
******************

Dim count As Integer, t As Integer
t = 0
count = 0
For t = 1 To 1000
   If Cells(2 + t, 22) >= Cells(1007, 11) Then count = count + 1
Next t
Cells(1008, 11) = count

End Sub
```

Appendix 10. SPSS® program to perform a jackknife that tests for significant differences between the discriminant function formed by Fourier analysis variables and the db1, db5 and db9 variables. The syntax is repeated 1000 times to allow a null distribution of the difference in the classification results to be calculated.

```
SET MXLOOP=1000.
SET SEED=RANDOM

MATRIX.
GET A /FILE= "G:/PairedJack/G.R.&M.I.&R.db1_5_9.Avg4.Age4.sav" /variables=  group select c1 c10
c100 c101 c102 c103 c104 c105 c106 c107 c108 c109 c11
 c110 c111 c112 c113 c114 c115 c116 c117 c118 c119 c12 c120 c121 c122 c123
 c124 c125 c126 c127 c128 c13 c14 c15 c16 c17 c18 c19 c2 c20 c21 c22 c23 c24
 c25 c26 c27 c28 c29 c3 c30 c31 c32 c33 c34 c35 c36 c37 c38 c39 c4 c40 c41
 c42 c43 c44 c45 c46 c47 c48 c49 c5 c50 c51 c52 c53 c54 c55 c56 c57 c58 c59
 c6 c60 c61 c62 c63 c64 c65 c66 c67 c68 c69 c7 c70 c71 c72 c73 c74 c75 c76
 c77 c78 c79 c8 c80 c81 c82 c83 c84 c85 c86 c87 c88 c89 c9 c90 c91 c92 c93
 c94 c95 c96 c97 c98 c99 c100b c101b c102b c103b c104b c105b c106b c107b c108b c109b c10b
 c110b c111b c112b c113b c114b c115b c116b c117b c118b c119b c11b c120b c121b
 c122b c123b c124b c125b c126b c127b c128b c129b c12b c130b c131b c132b c133b
 c134b c135b c136b c137b c138b c139b c13b c140b c141b c142b c143b c144b c145b
 c146b c147b c148b c149b c14b c150b c151b c152b c153b c154b c155b c156b c157b
 c158b c159b c15b c160b c161b c162b c163b c164b c165b c166b c167b c168b c169b
 c16b c170b c171b c172b c173b c174b c175b c176b c177b c178b c179b c17b c180b
 c181b c182b c183b c184b c185b c18b c19b c1b c20b c21b c22b c23b c24b c25b
 c26b c27b c28b c29b c2b c30b c31b c32b c33b c34b c35b c36b c37b c38b c39b
 c3b c40b c41b c42b c43b c44b c45b c46b c47b c48b c49b c4b c50b c51b c52b
 c53b c54b c55b c56b c57b c58b c59b c5b c60b c61b c62b c63b c64b c65b c66b
 c67b c68b c69b c6b c70b c71b c72b c73b c74b c75b c76b c77b c78b c79b c7b
 c80b c81b c82b c83b c84b c85b c86b c87b c88b c89b c8b c90b c91b c92b c93b
 c94b c95b c96b c97b c98b c99b c9b var00001 var00002 var00003 var00004 var00005 var00006
var00007
 var00008 var00009 var00010 var00011 var00012 var00013 var00014 var00015
 var00016 var00017 var00018 var00019 var00020 var00021 var00022 var00023
 var00024 var00025 var00026 var00027 var00028 var00029 var00030 var00031
 var00032 var00033 var00034 var00035 var00036 var00037 var00038 var00039
 var00040 var00041 var00042 var00043 var00044 var00045 var00046 var00047
 var00048 var00049 var00050 var00051 var00052 var00053 var00054 var00055
 var00056 var00057 var00058 var00059 var00060 var00061 var00062 var00063
 var00064 var00065 var00066 var00067 var00068 var00069 var00070 var00071
 var00072 var00073 var00074 var00075 var00076 var00077 var00078 var00079
 var00080 var00081 var00082 var00083 var00084 var00085 var00086 var00087
 var00088 var00089 var00090 var00091 var00092 var00093 var00094 var00095
 var00096 var00097 var00098 var00099 var00100 var00101 var00102 var00103
 var00104 var00105 var00106 var00107 var00108 var00109 var00110 var00111
 var00112 var00113 var00114 var00115 var00116 var00117 var00118 var00119
 var00120 var00121 var00122 var00123 var00124 var00125 var00126 var00127
 var00128 var00129 var00130 var00131 var00132 var00133 var00134 var00135
```

```
var00136 var00137 var00138 var00139 var00140 var00141 var00142 var00143
var00144 var00145 var00146 var00147 var00148 var00149 var00150 var00151
var00152 var00153 var00154 var00155 var00156 var00157 var00158 var00159
var00160 var00161 var00162 var00163 var00164 var00165 var00166 var00167
var00168 var00169 var00170 var00171 var00172 var00173 var00174 var00175
var00176 var00177 var00178 var00179 var00180 var00181 var00182 var00183
var00184 var00185 var00186 var00187 var00188 var00189 var00190 var00191
var00192 var00193 var00194 var00195 var00196 var00197 var00198 var00199
var00200 var00201 var00202 var00203 var00204 var00205 var00206 var00207
var00208 var00209 var00210 var00211 var00212 var00213 var00214 var00215
var00216 var00217 var00218 var00219 var00220 var00221 var00222 var00223
var00224 var00225 var00226 var00227 var00228 var00229 var00230 var00231
var00232 var00233 var00234 var00235 var00236 var00237 var00238 var00239
var00240 var00241 a1 a10 a11 a12 a13 a14 a15 a16 a17 a18 a19 a2 a20 a21 a22 a23
a24 a25 a26 a27 a28 a29 a3 a30 a31 a32 a33 a34 a35 a36 a37 a38 a39 a4 a40
a41 a42 a43 a44 a45 a46 a47 a48 a49 a5 a50 a51 a52 a53 a54 a55 a56 a57 a58
a59 a6 a60 a61 a62 a63 a64 a7 a8 a9 p1 p10 p11 p12 p13 p14 p15 p16 p17 p18
p19 p2 p20 p21 p22 p23 p24 p25 p26 p27 p28 p29 p3 p30 p31 p32 p33 p34 p35
p36 p37 p38 p39 p4 p40 p41 p42 p43 p44 p45 p46 p47 p48 p49 p5 p50 p51 p52
p53 p54 p55 p56 p57 p58 p59 p6 p60 p61 p62 p63 p64 p7 p8 p9.

COMPUTE NEW=MAKE(218,684,0).

LOOP d=1 TO 65 BY 1.
+ COMPUTE SWITCH = TRUNC(UNIFORM(1,1)*65)+1.
+ COMPUTE   TEMP=A(d,:).
+ COMPUTE A(d,:) = A(SWITCH,:).
+ COMPUTE A(SWITCH,:)=TEMP.
END LOOP.

LOOP d=66 TO 135 BY 1.
+ COMPUTE SWITCH = TRUNC(UNIFORM(1,1)*70)+1.
+ COMPUTE   TEMP=A(d,:).
+ COMPUTE A(d,:) = A(SWITCH+65,:).
+ COMPUTE A(SWITCH+65,:)=TEMP.
END LOOP.

LOOP d=136 TO 221 BY 1.
+ COMPUTE SWITCH = TRUNC(UNIFORM(1,1)*86)+1.
+ COMPUTE   TEMP=A(d,:).
+ COMPUTE A(d,:) = A(SWITCH+135,:).
+ COMPUTE A(SWITCH+135,:)=TEMP.
END LOOP.

LOOP d=1 TO 64 BY 1.
+ COMPUTE NEW(d,:) = A(d,:).
END LOOP.

LOOP d=65 TO 133 BY 1.
+ COMPUTE NEW(d,:) = A(d+1,:).
END LOOP.

LOOP d=134 TO 218 BY 1.
+ COMPUTE NEW(d,:) = A(d+2,:).
END LOOP.
```

```
SAVE NEW /OUTFILE=* /VARIABLES group select c1 c10 c100 c101 c102 c103 c104 c105 c106 c107
c108 c109 c11
 c110 c111 c112 c113 c114 c115 c116 c117 c118 c119 c12 c120 c121 c122 c123
 c124 c125 c126 c127 c128 c13 c14 c15 c16 c17 c18 c19 c2 c20 c21 c22 c23 c24
 c25 c26 c27 c28 c29 c3 c30 c31 c32 c33 c34 c35 c36 c37 c38 c39 c4 c40 c41
 c42 c43 c44 c45 c46 c47 c48 c49 c5 c50 c51 c52 c53 c54 c55 c56 c57 c58 c59
 c6 c60 c61 c62 c63 c64 c65 c66 c67 c68 c69 c7 c70 c71 c72 c73 c74 c75 c76
 c77 c78 c79 c8 c80 c81 c82 c83 c84 c85 c86 c87 c88 c89 c9 c90 c91 c92 c93
 c94 c95 c96 c97 c98 c99 c100b c101b c102b c103b c104b c105b c106b c107b c108b c109b c10b
 c110b c111b c112b c113b c114b c115b c116b c117b c118b c119b c11b c120b c121b
 c122b c123b c124b c125b c126b c127b c128b c129b c12b c130b c131b c132b c133b
 c134b c135b c136b c137b c138b c139b c13b c140b c141b c142b c143b c144b c145b
 c146b c147b c148b c149b c14b c150b c151b c152b c153b c154b c155b c156b c157b
 c158b c159b c15b c160b c161b c162b c163b c164b c165b c166b c167b c168b c169b
 c16b c170b c171b c172b c173b c174b c175b c176b c177b c178b c179b c17b c180b
 c181b c182b c183b c184b c185b c18b c19b c1b c20b c21b c22b c23b c24b c25b
 c26b c27b c28b c29b c2b c30b c31b c32b c33b c34b c35b c36b c37b c38b c39b
 c3b c40b c41b c42b c43b c44b c45b c46b c47b c48b c49b c4b c50b c51b c52b
 c53b c54b c55b c56b c57b c58b c59b c5b c60b c61b c62b c63b c64b c65b c66b
 c67b c68b c69b c6b c70b c71b c72b c73b c74b c75b c76b c77b c78b c79b c7b
 c80b c81b c82b c83b c84b c85b c86b c87b c88b c89b c8b c90b c91b c92b c93b
 c94b c95b c96b c97b c98b c99b c9b var00001 var00002 var00003 var00004 var00005 var00006
var00007
 var00008 var00009 var00010 var00011 var00012 var00013 var00014 var00015
 var00016 var00017 var00018 var00019 var00020 var00021 var00022 var00023
 var00024 var00025 var00026 var00027 var00028 var00029 var00030 var00031
 var00032 var00033 var00034 var00035 var00036 var00037 var00038 var00039
 var00040 var00041 var00042 var00043 var00044 var00045 var00046 var00047
 var00048 var00049 var00050 var00051 var00052 var00053 var00054 var00055
 var00056 var00057 var00058 var00059 var00060 var00061 var00062 var00063
 var00064 var00065 var00066 var00067 var00068 var00069 var00070 var00071
 var00072 var00073 var00074 var00075 var00076 var00077 var00078 var00079
 var00080 var00081 var00082 var00083 var00084 var00085 var00086 var00087
 var00088 var00089 var00090 var00091 var00092 var00093 var00094 var00095
 var00096 var00097 var00098 var00099 var00100 var00101 var00102 var00103
 var00104 var00105 var00106 var00107 var00108 var00109 var00110 var00111
 var00112 var00113 var00114 var00115 var00116 var00117 var00118 var00119
 var00120 var00121 var00122 var00123 var00124 var00125 var00126 var00127
 var00128 var00129 var00130 var00131 var00132 var00133 var00134 var00135
 var00136 var00137 var00138 var00139 var00140 var00141 var00142 var00143
 var00144 var00145 var00146 var00147 var00148 var00149 var00150 var00151
 var00152 var00153 var00154 var00155 var00156 var00157 var00158 var00159
 var00160 var00161 var00162 var00163 var00164 var00165 var00166 var00167
 var00168 var00169 var00170 var00171 var00172 var00173 var00174 var00175
 var00176 var00177 var00178 var00179 var00180 var00181 var00182 var00183
 var00184 var00185 var00186 var00187 var00188 var00189 var00190 var00191
 var00192 var00193 var00194 var00195 var00196 var00197 var00198 var00199
 var00200 var00201 var00202 var00203 var00204 var00205 var00206 var00207
 var00208 var00209 var00210 var00211 var00212 var00213 var00214 var00215
 var00216 var00217 var00218 var00219 var00220 var00221 var00222 var00223
 var00224 var00225 var00226 var00227 var00228 var00229 var00230 var00231
 var00232 var00233 var00234 var00235 var00236 var00237 var00238 var00239
 var00240 var00241 a1 a10 a11 a12 a13 a14 a15 a16 a17 a18 a19 a2 a20 a21 a22 a23
 a24 a25 a26 a27 a28 a29 a3 a30 a31 a32 a33 a34 a35 a36 a37 a38 a39 a4 a40
```

```
 a41 a42 a43 a44 a45 a46 a47 a48 a49 a5 a50 a51 a52 a53 a54 a55 a56 a57 a58
 a59 a6 a60 a61 a62 a63 a64 a7 a8 a9 p1 p10 p11 p12 p13 p14 p15 p16 p17 p18
 p19 p2 p20 p21 p22 p23 p24 p25 p26 p27 p28 p29 p3 p30 p31 p32 p33 p34 p35
 p36 p37 p38 p39 p4 p40 p41 p42 p43 p44 p45 p46 p47 p48 p49 p5 p50 p51 p52
 p53 p54 p55 p56 p57 p58 p59 p6 p60 p61 p62 p63 p64 p7 p8 p9.
END MATRIX.

DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=c1 c10 c100 c101 c102 c103 c104 c105 c106 c107 c108 c109 c11
 c110 c111 c112 c113 c114 c115 c116 c117 c118 c119 c12 c120 c121 c122 c123
 c124 c125 c126 c127 c128 c13 c14 c15 c16 c17 c18 c19 c2 c20 c21 c22 c23 c24
 c25 c26 c27 c28 c29 c3 c30 c31 c32 c33 c34 c35 c36 c37 c38 c39 c4 c40 c41
 c42 c43 c44 c45 c46 c47 c48 c49 c5 c50 c51 c52 c53 c54 c55 c56 c57 c58 c59
 c6 c60 c61 c62 c63 c64 c65 c66 c67 c68 c69 c7 c70 c71 c72 c73 c74 c75 c76
 c77 c78 c79 c8 c80 c81 c82 c83 c84 c85 c86 c87 c88 c89 c9 c90 c91 c92 c93
 c94 c95 c96 c97 c98 c99
 /ANALYSIS ALL
  /METHOD=WILKS
 /MAXSTEPS=20
  /PIN= .05
  /POUT= .10
 /PRIORS  SIZE
  /HISTORY
 /STATISTICS=TABLE CROSSVALID
 /PLOT=CASES
 /CLASSIFY=NONMISSING POOLED .

DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=c100b c101b c102b c103b c104b c105b c106b c107b c108b c109b c10b
 c110b c111b c112b c113b c114b c115b c116b c117b c118b c119b c11b c120b c121b
 c122b c123b c124b c125b c126b c127b c128b c129b c12b c130b c131b c132b c133b
 c134b c135b c136b c137b c138b c139b c13b c140b c141b c142b c143b c144b c145b
 c146b c147b c148b c149b c14b c150b c151b c152b c153b c154b c155b c156b c157b
 c158b c159b c15b c160b c161b c162b c163b c164b c165b c166b c167b c168b c169b
 c16b c170b c171b c172b c173b c174b c175b c176b c177b c178b c179b c17b c180b
 c181b c182b c183b c184b c185b c18b c19b c1b c20b c21b c22b c23b c24b c25b
 c26b c27b c28b c29b c2b c30b c31b c32b c33b c34b c35b c36b c37b c38b c39b
 c3b c40b c41b c42b c43b c44b c45b c46b c47b c48b c49b c4b c50b c51b c52b
 c53b c54b c55b c56b c57b c58b c59b c5b c60b c61b c62b c63b c64b c65b c66b
 c67b c68b c69b c6b c70b c71b c72b c73b c74b c75b c76b c77b c78b c79b c7b
 c80b c81b c82b c83b c84b c85b c86b c87b c88b c89b c8b c90b c91b c92b c93b
 c94b c95b c96b c97b c98b c99b c9b
 /ANALYSIS ALL
  /METHOD=WILKS
  /MAXSTEPS=20
  /PIN= .05
  /POUT= .10
 /PRIORS  SIZE
  /HISTORY
 /STATISTICS=TABLE CROSSVALID
 /PLOT=CASES
 /CLASSIFY=NONMISSING POOLED .
```

```
DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=var00001 var00002 var00003 var00004 var00005 var00006 var00007
 var00008 var00009 var00010 var00011 var00012 var00013 var00014 var00015
 var00016 var00017 var00018 var00019 var00020 var00021 var00022 var00023
 var00024 var00025 var00026 var00027 var00028 var00029 var00030 var00031
 var00032 var00033 var00034 var00035 var00036 var00037 var00038 var00039
 var00040 var00041 var00042 var00043 var00044 var00045 var00046 var00047
 var00048 var00049 var00050 var00051 var00052 var00053 var00054 var00055
 var00056 var00057 var00058 var00059 var00060 var00061 var00062 var00063
 var00064 var00065 var00066 var00067 var00068 var00069 var00070 var00071
 var00072 var00073 var00074 var00075 var00076 var00077 var00078 var00079
 var00080 var00081 var00082 var00083 var00084 var00085 var00086 var00087
 var00088 var00089 var00090 var00091 var00092 var00093 var00094 var00095
 var00096 var00097 var00098 var00099 var00100 var00101 var00102 var00103
 var00104 var00105 var00106 var00107 var00108 var00109 var00110 var00111
 var00112 var00113 var00114 var00115 var00116 var00117 var00118 var00119
 var00120 var00121 var00122 var00123 var00124 var00125 var00126 var00127
 var00128 var00129 var00130 var00131 var00132 var00133 var00134 var00135
 var00136 var00137 var00138 var00139 var00140 var00141 var00142 var00143
 var00144 var00145 var00146 var00147 var00148 var00149 var00150 var00151
 var00152 var00153 var00154 var00155 var00156 var00157 var00158 var00159
 var00160 var00161 var00162 var00163 var00164 var00165 var00166 var00167
 var00168 var00169 var00170 var00171 var00172 var00173 var00174 var00175
 var00176 var00177 var00178 var00179 var00180 var00181 var00182 var00183
 var00184 var00185 var00186 var00187 var00188 var00189 var00190 var00191
 var00192 var00193 var00194 var00195 var00196 var00197 var00198 var00199
 var00200 var00201 var00202 var00203 var00204 var00205 var00206 var00207
 var00208 var00209 var00210 var00211 var00212 var00213 var00214 var00215
 var00216 var00217 var00218 var00219 var00220 var00221 var00222 var00223
 var00224 var00225 var00226 var00227 var00228 var00229 var00230 var00231
 var00232 var00233 var00234 var00235 var00236 var00237 var00238 var00239
 var00240 var00241
 /ANALYSIS ALL
 /METHOD=WILKS
 /MAXSTEPS=20
 /PIN= .05
 /POUT= .10
 /PRIORS SIZE
 /HISTORY
 /STATISTICS=TABLE CROSSVALID
 /PLOT=CASES
 /CLASSIFY=NONMISSING POOLED .


DISCRIMINANT
 /GROUPS=group(1 3)
 /VARIABLES=a1 a10 a11 a12 a13 a14 a15 a16 a17 a18 a19 a2 a20 a21 a22 a23
 a24 a25 a26 a27 a28 a29 a3 a30 a31 a32 a33 a34 a35 a36 a37 a38 a39 a4 a40
 a41 a42 a43 a44 a45 a46 a47 a48 a49 a5 a50 a51 a52 a53 a54 a55 a56 a57 a58
 a59 a6 a60 a61 a62 a63 a64 a7 a8 a9 p1 p10 p11 p12 p13 p14 p15 p16 p17 p18
 p19 p2 p20 p21 p22 p23 p24 p25 p26 p27 p28 p29 p3 p30 p31 p32 p33 p34 p35
 p36 p37 p38 p39 p4 p40 p41 p42 p43 p44 p45 p46 p47 p48 p49 p5 p50 p51 p52
 p53 p54 p55 p56 p57 p58 p59 p6 p60 p61 p62 p63 p64 p7 p8 p9
 /ANALYSIS ALL
 /METHOD=WILKS
```

```
/MAXSTEPS=20
 /PIN= .05
 /POUT= .10
/PRIORS  SIZE
 /HISTORY
/STATISTICS=TABLE CROSSVALID
/PLOT=CASES
/CLASSIFY=NONMISSING POOLED .
```