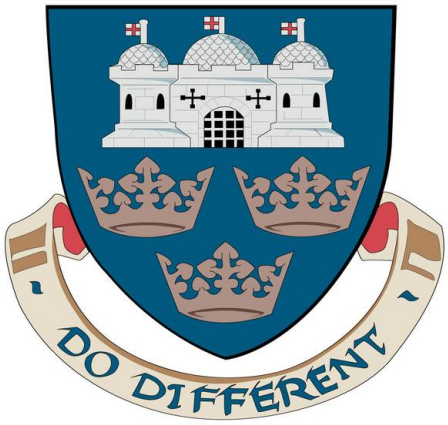
MORPHOMETRIC OTOLITH ANALYSIS

James Jonathan Iestyn Mapp

A Thesis Submitted for the
Degree of Doctor of Philosophy



University of East Anglia
School of Computing Sciences

September 2015

Abstract

Fish otoliths have long played an important role in sustainable fisheries management. Stock assessment models currently used rely on species specific age profiles obtained from the seasonal patterns of growth marks that otoliths exhibit. We compare methods widely used in fisheries science (elliptical Fourier) with an industry standardised encoding method (MPEG7 - Curvature-Scale-Space) and with a recent addition to shape modelling techniques (time-series shapelets) to determine which performs best.

An investigation is carried out into transform methods that retain size-information, and whether the boundary encoding method is impacted be otolith age, performing tests over three 2-class otolith datasets across six discrete and concurrent age groups. Impact of segmentation methods are assessed to determine whether automated or expert segmented methods of boundary extraction are more advantageous, and whether constructed classifiers can be used at different institutions.

Tests show that neither time-series shaplets nor Curvature-Scale-Space methods offer any real advantage over Fourier transform methods given mixed age datasets. However, we show that size indices are most indicative of fisheries stock in younger single-age datasets, with shape holding more discriminatory potential in older samples. Whilst commonly used Fourier transform methods generally return best results; we show that classification of otolith boundaries is impacted by the method of boundary segmentation. Hand traced boundaries produce classifiers more robust to test data segmentation methods and are more suited to distributed classifiers.

Additionally we present a proof of concept study showing that high energy synchrotron scans are a new, non-invasive method of modelling internal otolith structure, allowing comparison of slices along near infinite numbers of virtual complex planes.

Acknowledgements

First and foremost I would like to acknowledge my supervisor Dr Mark Fisher. Without his knowledge and perseverance this thesis would not have materialised. Thanks also to Dr Ewan hunter at Cefas for his knowledge of the fisheries sector, and for sanity-checking work throughout the later half of my studies. Thanks of course to Dr Tony Bagnall as a further supervisor to this work and for helping draw words off the tip of my tongue on many occasions. Combined thanks to teams at both Cefas and GMIT for help constructing the datasets used and for ground-truthing samples, and to Bruker-MicroCT and Diamond Light Ltd for MicroCT and beamline assistance. Thanks to Dr Jason Lines, Dr Jon Hills for help with shapelet testing and to Dr Joe-Scutt-Phillips for essentially starting a study that continues here.

Of course I would like to thank my whole family, in particular my beautiful wife for support and understanding, my Mother for endless financial support in the later stages, and my perfect daughter Kitty who served as my early morning alarm clock and kept me smiling through the darkest times. Thanks to all my friends for putting up with me, and to my gaming groups for serving as rest-bite when I needed some R&R.

Finally I would like to thank all the recording artists whose music has kept me sane when my mind was slipping into a sea of otolith based nightmares, particularly on dark winter nights walking home.

"Like a crazy singer in a band that's lost the words"

Virgin Prunes

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Otoliths are the calcium carbonate structures forming the inner ear of many vertebrates. Teleost fish have three otoliths in each ear-chamber (left and right ears) of the fish: sagittae, lapilli and asterisci. However, the larger sagittal otoliths are the most commonly used for classification studies as they are easier to prepare, observe and measure [24]. Otolith morphology varies markedly between species (Figure 1.1), however separate stocks of the same species, where fish are often physically similar (Figure 1.3), can sometimes be discriminated through subtle differences in otolith morphometrics.

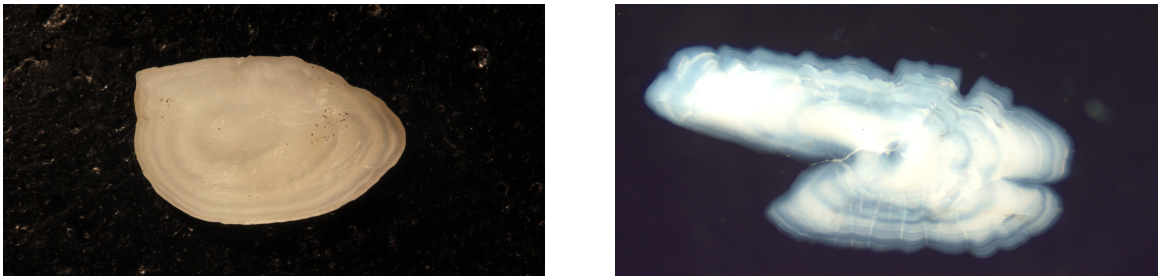


Figure 1.1: Otoliths from Plaice (left) and Herring (right) showing interspecies variation in morphology.

Figure 1.2 shows an otolith (sagittal) that is used during the studies presented in this thesis, and was taken from the left ear-chamber of a North-Sea Herring. In some species the left and right otoliths of individual fish are significantly different, a trait common in flatfish such as Plaice. In these instances the larger otolith is often used for study. However, for the species studied in this thesis the left and right sagittal otoliths may be used for study.

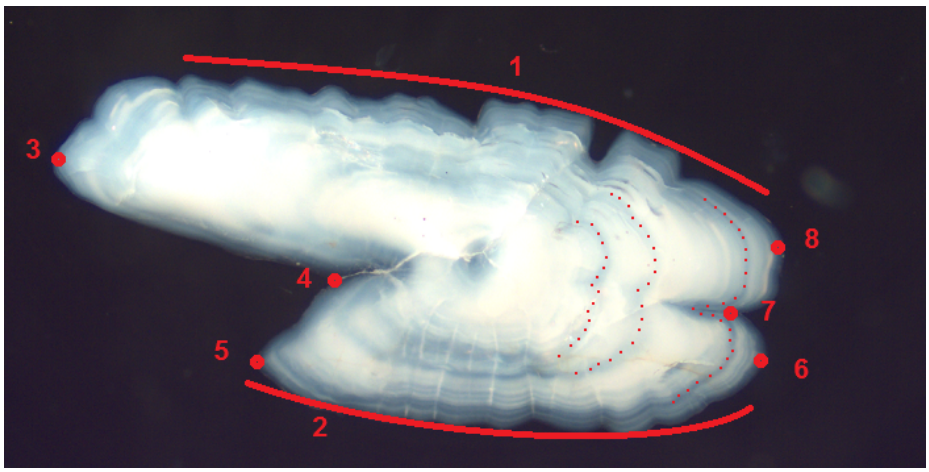


Figure 1.2: Example of a otolith (sagittal) taken from the left ear of a North-Sea Herring. Marked on the otolith are the ventral (1) and dorsal (2) edges, the rostrum (3), excisura major (4), antirostrum (5), pararostrum (6), excisura minor (7) and postrostrum (8). Dotted arcs annotate sections of three annuli; rings that show yearly growth.

Expert otolith readers have drawn on otoliths to discriminate between: different ages or cohorts [14, 21, 23]; sex [25]; diet [34] and of course stock [11, 12, 19, 31, 63]. Some of these distinctions are more complex and more important to fisheries management [13], which requires accurate measurements of stock composition/mixing or stock movement to inform decision making [90].

Figure 1.3: Three North-Sea (top) and Thames (bottom) herring otoliths, showing morphological similarity between two stocks of the same species.

Shape analysis forms a major part of otolith science, and many studies have analysed shape with a view to separating stocks. Methods include statistical analysis of general shape parameters such as circularity, eccentricity, area, perimeter length, form-factor, and annular growth increments [21, 22, 46, 76]. In some cases these measurements are supplemented with, or normalised by, measures such as fish length or weight [30, 63, 90].

Otolith boundaries are also extracted and represented or encoded in different ways (transformed) prior to analysis with methods such as Fourier transforms [10, 11, 35]; or Elliptical Fourier transforms [22, 24, 31]. Other methods of otolith boundary representation include Wavelets [73], Curvature-Scale-Space [13, 73] and the more recent Shapelet transform method [40, 54].

Otoliths also bear patterns, in the form of concentric rings (annuli) around the nucleus. Whilst these rings are commonly used to determine the age of the sample, and can be used to age the otolith, and hence the individual, in years or even days [19], these annuli are also used during stock discrimination tasks.

Stock discrimination using annuli can be carried out using microstructure analysis [19] whereby measurements of daily growth increments are taken allowing spawning

dates to be approximated, with spawning dates themselves suggesting the spawning stock. In this process the otolith must (usually) first be sectioned, a process that is time consuming and requires specialist methods and equipment. This process is undertaken when annuli or daily annuli are not easily visible on the exterior surface of the otolith and involves setting the specimen in a resin mount, sectioning it through the core and polishing the surface with a fine-grit. Magnification levels needed to clearly view/measure the daily increments vary by species and task, however whilst daily increment measurements can sometimes be performed using optical microscopy, magnification levels are often in excess of 100x magnification [19] or make use of scanning electron microscopy.

Analysis of annuli (yearly increments) typically involves viewing the otolith under magnification, although in some cases it is possible to analyse larger specimens with the naked eye. Annuli may in some instances be used as an approximation of the otolith shape at previous ages and have been used for stock discrimination [22].

Most studies focus on standard statistical methods to normalise or prepare data (eg. Burke et al. [22]), and compare them using statistical software. Whilst a number of differing techniques are used to formulate data from otolith boundaries a number of common methods have emerged. Fourier transforms are often used for analysis; usually derived using Fast Fourier Transforms (FFTs) or Elliptical Fourier Transforms (EFTs) performed on perimeter data of the otolith, or on internal features such as annual growth rings.

Additionally (or alternatively), shape indices are formulated from otolith morphometrics such as major/minor axis lengths, increment lengths, circularity, eccentricity, rectangularity, etc. These can either be manually derived, or obtained through

computer-aided derivation using software such as OPTIMAS ([12, 27]) or MATLAB [62]. These metrics are often used to supplement Fourier descriptors, but are also used as stand-alone features.

Many studies center on stock classification as the primary goal, and use a number of methods to distinguish between classes. However the aim of this study is *not* focused on the absolute classification accuracy of stock identification per se. Rather the objectives of this thesis are to critically compare the viability of morphometric methods used to represent otoliths for stock classification in an attempt to assess their potential efficacy for application in fish stock management.

To this end we perform a number of 2-class stock separation tasks using a number of pre-constructed otolith image sets (datasets) obtained from the Centre for Environment, Fisheries and Aquaculture Science (Cefas) and the Galway-Mayo Institute of Technology (GMIT), whereby individual samples are re-assigned to their source stocks using classifiers constructed with remaining samples. The details of the datasets obtained can be seen in Section 1.7) and details of how each dataset (or portion of) was used to explore our research objectives can been seen in the research chapters of this thesis (Chapters 3 to 6).

1.1 Motivation

Understanding the variance in otolith morphometrics is a basis for further understanding of population dynamics [11] and stock structure [11, 22] in that it can give information through deduction about many aspects in individual, population or species specific life-cycles. Fisheries management requires that stocks be accurately determined and separated for management [91] so that decisions on their management can be made. Whilst stocks may be managed as separate stocks as part of one

management group [20, 30], it is often the case that separating stocks into separate management areas [23] does not reflect the underlaying population structure of the fish.

Devries [30] discusses how management areas can be incorrectly separated due to lack of understanding of eco-dynamics in the areas managed, which can lead to poor management among several other issues. However, incorrect separation may also be caused by traditional methods of otolith classification as they have certain levels of ambiguity or human error [96] and may be reader-dependent in some cases. Whist there are some common methodologies emerging for morphological landmark/metric selection, there is little consensus on which methods are more successful for classification [13, 19, 20].

Otolith analysis allows estimation of stock composition to determine whether the samples obtained from an area or areas are all in fact from one stock, or from multiple stocks mixed together [24, 30, 31]. The same methods may be used to determine whether multiple known stocks are mixing [22, 23] in a specific area or areas and if so the estimated composition ratios of the mixed stock. It can also be used to determine stock movements or migration from one location to another [22, 35] or gradual dispersal of a fish stock or nursery ground.

With increased accuracy in separation of stocks or populations, a better understanding of a number of key management factors can be achieved and where possible stocks can be separated (on paper) so that they are each self-sustaining [90]. Where separate stocks have been modelled using otolith analysis, it is possible to take otolith sample(s) from a separate, un-modelled area, and compare them to the existing models. This could give an indication as to whether the samples taken from the separate

area are of individuals of known stock or stocks [24], or if samples may include individuals from other (potentially unknown) stock(s). Conversely, samples could be taken from modelled stock areas and composition checked for consistency with known data. This could indicate shifts in individual spawning components or recruitment index [11, 21], important considerations for many management decisions.

Understanding recruitment dynamics [19] assists management units in giving better protection to specific spawning grounds [78] where recruitment has been hindered by exploitation or environmental impact. Analysing samples from surveyed areas may be used as a method for establishing the current state of fisheries resources [13]. In heavily exploited areas, or in areas impacted by environmental catastrophes etc, methods can be employed to assess stock recovery or how severely stocks have been impacted. This may be beyond unit count, as otoliths allow discrimination of cohort or stock and may show if stock replenishment is due to migration, or decreased mortality due to reduced competition.

Over exploitation is an issue that can severely impact fisheries stocks of all sizes [31, 78] and can readily happen without correct management in place. Correct measurement of stock composition or recruitment dynamics, and its correct comparison to historical records is important to understand if and how a stock is being exploited and how it responds to exploitation [11]. Maintaining genetic diversity [78] can be an important factor to how well a stock can react to environmental change or other impacts and whilst genetic diversity or population richness [35] at a genetic level may not be discernible using otoliths alone. They only give indication of stock mixing or lack of mixing which are important factors in maintaining diversity. Without proper precautionary approaches, over exploitation through fishing or severe impact of environmental change/disaster due to decreased diversity can in the worst cases lead

to severe stock depletion or resource collapse [11], to the point that stocks become non-sustainable.

Fish stocks however hold high commercial value [31, 63, 78] and restricting the fishing of stock areas is often met with objection from those that rely on those area for both food or income and therefore a total ban of fishing in those areas are often socially infeasible. Many now understand however the critical importance of sustainable exploitation [16] through correct management, so that total catches taken do not exceed the maximum sustainable yield [36]. This ensures that stocks persist [35] and can continue to benefit areas where the fishing industry holds high commercial importance [96]. Fisheries management can make assessments based on current stock level, diversity and recruitment dynamics to set conservative quotas [92] that ensure stock persistence whilst at the same time allowing those communities that rely on the stock to survive. This of course extends to sport-fishing or areas where additional recreational value is placed on a stock. Additional management can be in place to allow those activities to continue with limited impact to the stock itself, or where any impact is taken into account when making management decisions.

1.2 Project Objectives

Here we list the initial objectives of the research presented in this thesis, with a brief justification for each, and methods by which they are performed.

1. To critically compare recent techniques taken from the fields of computer vision and time-series analysis to methods traditionally used for otolith classification.

We investigate methods of otolith analysis with a view to separation of fisheries stock. Current tools from the field of image and time-series analysis are compared to methods widely used in otolith analysis with a view to increasing accuracy. The initial task is to determine whether recent methods from the field of shape (Curvature Scale Space, [15]) and time-series analysis (Time-series Shapelets, [70]) compare favourably with established methods of otolith boundary analysis (Fourier analysis, [48]) when used to separate fish stocks.

2. To establish whether otolith age impacts classification accuracies dependent on scale-invariance of transformed boundaries.

Studies have found that separation of samples into discrete age categories can impact classification accuracies. Galley et al. [35] found that samples of different age are classified with different accuracy. Likewise, Begg and Brown [11] found differences in classification rates among ages when a multi-age model was used, additionally noting that classification success is more variable when samples are analysed in separate age tests. However, both studies combine Fourier descriptors with gross shape morphological metrics when analysing otoliths. We investigate shape and size metrics individually with regard to separate otolith ages to determine whether age composition of otolith datasets dictates choice of boundary representation for stock separation.

3. To determine whether the methods used for otolith boundary determination impact classification accuracies.

There are two common methods of otolith segmentation used in boundary modelling studies: Hand tracing of the boundary or internal annuli [11, 12, 22, 30], and automatic detection through intensity thresholding and/or edge detection [18, 44, 63, 92]. We assess whether the distribution of pre-built classifiers is feasible, given the non-standardised practice of otolith image capture and segmentation. We investigate whether the two common methods of outline extraction, hand tracing and intensity thresholding, affect the classification accuracies returned when these boundaries are used for stock separation.

4. To establish whether the choice of machine learning algorithm affects classification accuracies for otolith stock separation.

Whilst many of the studies use modern computing to process the complex statistical tasks, many of them do not use, or at least do not mention the use of, computer aided modelling techniques to process and classify samples. A range of learning algorithms are explored during the study and appropriate statistical techniques are employed to determine whether any particular algorithm or family of algorithms, available within commonly used and freely available machine learning tool-kits, perform better than others for otolith stock classification tasks.

5. To determine whether three dimensional modelling of otoliths is possible using non-invasive methods, and whether complex plane slices give clearer indication of internal otolith structure.

Currently methods of otolith sectioning are invasive and generally destructive. Previous attempts to produce three-dimensional models using x-ray scanning have been unsuccessful [75]. We investigate whether modelling of internal otolith features is

feasible using higher energy scanning methods employing the UK's only synchrotron particle accelerator.

6. To establish whether complex plane slices give clearer indication of internal otolith structure than traditional flat plane slicing.

We investigate whether three-dimensional reconstructions could allow virtual slicing along complex curves, such as estimated plane of maximal growth, and compare the suggested technique to traditional flat-slicing methods.

1.3 Publications

Here we list the publications that have been printed, submitted, or are in preparation for submission. Each publication is listed with a shortened description and reference to the thesis objectives.

As First Author

James Mapp, Mark Fisher, Anthony Bagnall, Jason Lines, Sally Warne, and Joe Scutt Phillips. Clupea harengus: Intraspecies distinction using curvature scale space and shapelets. In *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, pages 138–143, SciTePress, 2013

The publication presents a study comparing Curvature Scale Space (CSS) representation with Shapelet transformed data with a view to discriminating between sagittal otoliths of North-Sea and Thames Herring using otolith boundary and boundary metrics. CSS transformed boundaries combined with measures of their circularity, eccentricity and aspect-ratio are used to classify using nearest-neighbour selections with distance being computed using CSS matching methods. Shapelet transformed data are classified using a number of techniques (Nearest-Neighbour, Naive-Bayes, C4.5, Support Vector Machines, Random and Rotation Forest) and compared to CSS classification results. The work published was a portion of our investigation into whether computer vision and time-series methods offered improvement on industry used otolith classification methods (Objective 1), and whether choice of learning algorithm affects accuracies (Objective 4). It was truncated (to remove Fourier analysis) on request of the journal reviewers.

James Mapp, Mark Fisher, Richard Atwood, Duncan Bell, Mark Greco, Sally Songer, and Ewan Hunter. Three-dimensional analysis of otolith growth using phase contrast synchrotron tomography. In *Journal of Fish Biology*. Wiley Online Library, 2016

A three-dimensional computer reconstruction of a plaice Pleuronectes platessa otolith is presented from data acquired by the Diamond Light synchrotron, beamline I12, X-ray source, a high energy (53150 keV) source particularly well suited to the study of dense objects. The data allowed non-destructive rendering of otolith structure, and for the first time allows otolith annuli (internal ring structures) to be analysed in X-ray tomographic images. The publication is a short proof of concept that relates to our objective of whether three dimensional modelling of otoliths is possible using non-invasive methods (Objective 5).

James Mapp, Mark Fisher, and Ewan Hunter. Boundary based stock classification: Expert otolith readers outperform automated outlining methods. In *ICES Annual Science Conference (submitted)*, 2016

Here we examine whether the method chosen for otolith boundary extraction affects the accuracies of stock discrimination when using elliptical Fourier based classification of otolith boundaries. We compare two methods of boundary extraction: Outlines derived by two expert readers, traced by hand, and outlines derived by intensity thresholding of otolith images using bottom-up and top-down approaches. Outlines from each method are transformed using elliptical Fourier methods to create a set of harmonics for each of the outlining methods, which are in turn used to construct and test classifiers, each fully cross validated, using the WEKA machine

learning suite. This work has been submitted to be presented at the ICES Annual Science Conference 2016, and represents our investigation regarding Objectives 3 and 4.

James Mapp, Ewan Hunter, Sally Songer, Jeroen Van Der Kooij, and Mark Fisher. Operational viability of stock-separation using shape indices derived from the otolith morphometric outline. an example using sprat and herring. In *In preparation*, 2016

This publication presents a study concerning the viability of stock-separation of sprat and herring using otolith morphometrics, within the context of potential application in discrete stock management. Analysis focused on three stock discrimination problems with the aim of reassigning individual fish otoliths to source populations. Six feature sets encoding combinations of size and shape together with nine learning algorithms were explored. To assess saliency of size/shape features half of the feature sets included size indices, the remainder encoded only shape. Otolith sample sets were partitioned by age so that the impact on fish age on classification accuracy could be assessed for each encoding method. This work is being prepared for submission to a relevant machine learning journal, and was presented at the 2014 International Otolith Symposium (IOS2014). Assessing the impact of scale metrics, otolith age and learning algorithm, this work represents an intensive study, focused on Objectives 2 and 4.

As Co-Author

J Hills, J Lines, E Baranauskas, J Mapp, and A Bagnall. Classification of time series by shapelet transformation. *Data Mining and Knowledge Discovery*, 28 (4):851–881, 2014

This publication present a study comparing performance of time-series shapelets over a range of classification tasks, including separation of North-Sea and Thames herring stocks using sagittal otoliths. A number of learning algorithms availible though the WEKA machine learning suite were employed by which to build shapelet based classifiers, and shapelet quality measures are discussed and tested. This work was presented in the PhD thesis of the primary author, and relates to Objectives 1 and 4 of this work.

1.4 Contributions

Here we list the novel contributions of the work presented in this thesis. Contributions are arranged with regards to location in the thesis where possible.

1. First study on the use of time-series shapelets for stock-separation of otoliths In chapter 3 we propose the use of the MPEG7 boundary encoding standard (Curvature-Scale-Space), and the recent Shapelet transform methods for use in fisheries stock discrimination using boundary images derived from herring sagittal otoliths. We compare results of classification using a suite of learning algorithms with Elliptical Fourier encoding methods which are widely used within otolith boundary classification.

We believe this to be the first study that has been carried out using time-series shapelets as a method of otolith boundary representation, and the first to compare shapelet and curvature scale space representations to elliptical Fourier methods with regards to stock classification. We determine that neither of the newer methods provide more accurate results that the current methods used within the industry, regardless of the learning algorithms used for classifier construction. Results of this study were published in shortened form in [61], and formed part of the study presented in [40].

2. Most extensive investigation of otolith age impact on stock classification methods Chapter 4 presents an extensive investigation into whether size-inclusive or size-exclusive boundary encoding methods hold more potential for stock classification. We show using multiple discrete age classification tasks across three separate mixed stock fisheries, that the methods of boundary transform (size-inclusive

or size-exclusive) should be selected based on otolith age composition of the datasets to be classified.

We show that for younger sample sets (age 0-1) encoding methods that retain otolith size information return greater accuracies across tests than encoding methods that remove size indices during the transform process. For older specimens (age 2+) methods that remove size indices return greater accuracies (than size inclusive methods) although results are not statistically different. Whilst previous reports have shown that separating classification accuracies by sample ages yields different result ranges, we believe this to be the first study that shows clear differences between transform methods to be used for differing dataset age compositions.

Findings of the study presented in chapter 4 were presented at the International Otolith Symposium 2014, and are in preparation for publication.

3. First comparison of otolith outlining techniques and their impact of classification Chapter 5 tests whether Fourier based boundary classification accuracies (of mixed stock herring otoliths) is impacted by choice of outlining methodology. We compare four methods of boundary acquisition from digital images: two sets of hand-traced outlines by different experts and two sets of boundaries determined using two different thresholding methods. Whilst there are many studies performing boundary classification using hand-traced outlines, outlines obtained through intensity thresholding, or a combination of both; we believe that this is the first study that compares outlining methods and whether they impact classification accuracies. The work has recently been submitted for presentation at the ICES Annual Science Conference 2016.

We show that hand traced boundaries return higher accuracies overall, and that classifiers built using hand traced outlines are robust to changes in outlining methods of test data. We show in this work that classifiers constructed to discriminate between two spawning stocks may be exported and used elsewhere with little regard for outlining methods used at different facilities. Thresholded methods however, are shown to be dependent on the outlining method used for testing data, and therefore, whilst thresholding may remove the requirement for expert input, we show that the resulting classifiers require industry standardisation for thresholding methods.

4. First successful Three-dimensional modelling of internal otolith structures using non-invasive methods Chapter 6 presents a proof of concept study regarding three-dimensional modelling of internal otolith structures using non invasive methods, recently published by the Journal of Fish Biology. We show that use of high-energy x-rays with propagation phase contrast are suitable for internal imaging of plaice otoliths. All previous attempts to recover internal growth features using micro-CT x-ray sources have failed. This study has successfully shown,for the first time, that high-energy scans are capable of viewing internal structures and highlights many potential applications within otolith sciences.

We show detailed three-dimensional internal reconstructions of the otolith segment scanned, and perform virtual spline-plane slicing through the otolith along estimated plane of growth, neither of which were practically feasible using previous industry techniques. Additionally we show significant differences between visible internal structures on flat-plane slices and on growth-plane slices of the reconstructed model, which may have severe implications for shape based analysis of visible growth structures using conventional sectioning methods.

The methods of scanning presented in this concept work allows three-dimensional modelling and classification techniques based on yearly accretion volume, a task not previously possible. Internal structure modelling also has the potential to allow removal of yearly growth increments without invasive procedures, which may assist scale based classification, given findings of age-based tests presented in chapter 4 of this work.

1.5 Thesis Organisation

The following chapter (Chapter 2) gives details of methods used throughout this thesis, including: details of outlining methods; the transform methods used (in order of appearance) and learning algorithms used for classifier construction. Chapters 3 to 5 contain descriptions of research undertaken and results thereof.

Chapter 3 contains work that has previously been published in conference proceedings and was presented at the International Conference on Pattern Recognition Applications and Methods 2013. The Study has been extended as part of this thesis. We show in this chapter that established methods of otolith boundary classification are not surpassed by classification of Shapelet or Curvature-Scale-Space transformed data (Objectives 1 and 4).

Chapter 4 contains a study concerning viability of different boundary transform methods where the encoded contours are used to classify otoliths in three stock classification tasks. Tests are carried out over discrete age categories to determine whether age composition of datasets impacts accuracies using each particular method (Objectives 2 and 4). The study was presented at the International Otolith Symposium 2015 (IOS 2015) and a section of the study is currently in preparation for submission.

Chapter 5 contains a study of segmentation methods with regards to otolith classification. We conduct classification tests given a restricted dataset to determine whether classification accuracies are dependent on methods of otolith boundary extraction when encoded using elliptical Fourier methods. Tests are conducted using multiple learning algorithms and varying harmonic content to determine if classification using these factors are dependent on outlining methods (Objectives 3 and 4).

This work has recently been submitted for presentation at the ICES Annual Science Conference.

Chapter 6 investigates whether high-energy x-ray scans are capable of determining internal otolith structures using non-invasive methods and whether such scans can be used to create three-dimensional models of otolith structures (Objective 5). This work has recently been published in the Journal of Fish Biology.

In Chapter 7 we summarise discussions for each study before concluding the work and suggesting further research goals relevant to the topics presented in this Thesis.

1.6 Research Methodology

A number of datasets were obtained from external institutes for use in this project. Details of the sets and their sources are given in Section 1.7. Datasets were obtained as a collection of images, combined with information of individual samples. Ground truthing of samples, including fisheries stock and otolith age, was carried out by expert readers at the contributing institutes and were considered accurate for use in this study.

The focus of this research is otolith boundary shape analysis, and so each otolith is segmented using one of two previously used methods (or a combination of both): Hand-tracing methods, and intensity thresholding. Both methods are described in Chapter 2, and otoliths are segmented using a combination of the two methods in research presented in Chapters 3 and 4. Chapter 5, which investigates differences between segmentation methods, also details two different intensity thresholding approaches.

A number of different data transformations are used to transform the extracted boundaries and are described in detail in Chapter 2, with Chapters 3 and 4 comparing transform methods across datasets to establish which method returns best results given a number of factors, including age of otolith samples. For the study presented in Chapter 4 otoliths are separated by age using ground truth data provided by expert readers prior to classification to investigate age effects on classification accuracies.

Each of the transforms used in this research require a different method of boundary encoding, so prior to transformation the boundary must be extracted from the segmented images using different processes. The processes used are described in Chapter 2 and are referenced from the transform descriptions.

Multiple learning algorithms, included in the WEKA machine learning suite, are used to construct classifiers using transformed boundaries. The classifiers constructed are used to assign further samples a predicted class, using leave one out cross validation procedures, and results are checked against existing ground truth information provided with the datasets. The algorithms used in the research presented in this thesis are described in Chapter 2, Section 2.6. Individual research chapters list algorithms used for analysis, where the full selection has not been employed.

Results returned by cross-validated classifiers are shown in relevant chapters, and assessed using N-way analysis of variance testing (ANOVAN) and, combined with, associated post-hoc testing diagrams, are used to determine which factors impact classification results, and which choice of method returns the better results. Results are compared across: Boundary transform methods (Chapters 3 and 4); Age-classes (Chapter 4); Boundary segmentation methods (Chapter 5) and learning algorithm (Chapters 3 to 5.

Figure 1.4 shows the general flow of data through our research methods. This process is modified for each of our studies shown in Chapters 3 to 5 so that comparing of results between tests can be carried out. The relevant chapters of this thesis describe the changes that are made to the generalised model.

In addition to our generalised model we carry out three dimensional reconstruction of a single Plaice otolith. The sample was sent to the UK's national synchrotron science facility, Diamond Light synchrotron, and scanned using the beamline I12 X-ray source. Images returned with the sample are modelled using techniques described in Chapter 6.

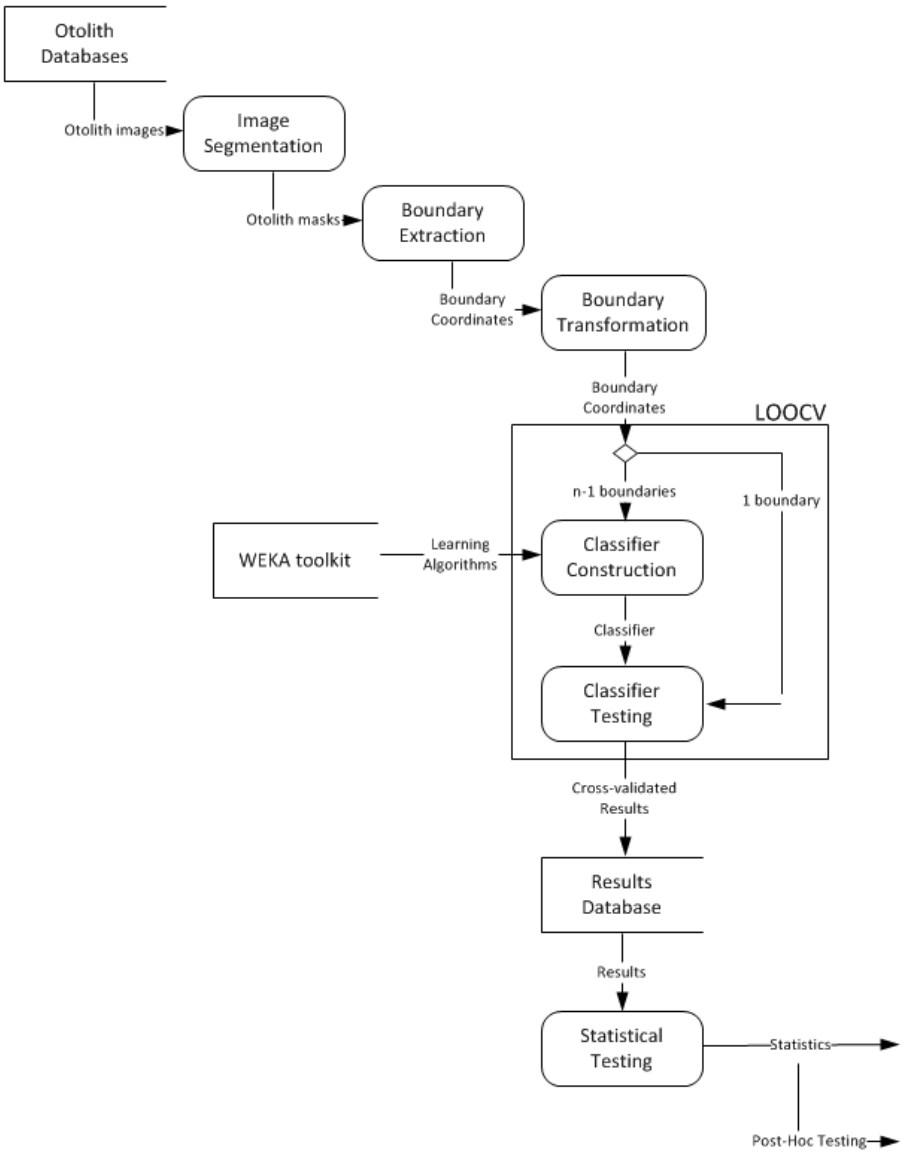


Figure 1.4: Showing the general flow of data through our research, from the acquired otolith datasets, to return of classification results, testing of results, and post-hoc testing.

Complex plane slicing was carried out using double-spline methods described in Chapter 6, and compared visually to virtual flat plane slices by expert otolith readers to determine whether complex slices better describe internal otolith structure.

1.7 Materials

This Section gives details of the datasets collected for the studies presented in this thesis. None of the samples were themselves captured by the thesis author, rather, images of samples were obtained from other institutions that had previously captured and processed the samples for use in other research or government reports.

We were able to source three matched datasets of Clupeid otolith images (one sprat, and two herring), each representing discrete paired spawning stocks which cannot easily be discriminated visually (e.g. by visual inspection of external morphology). Clupeid is the name of a family of fishes which includes species such as Herring, Sprat, Sardines etc., with the term itself the Latin name for 'sardine'. The family forms a large portion of the fishing industry and as such is of commercial interest.

The datasets obtained can be seen in Table 1.1. Herring set 1 was obtained from the Centre for Environment, Fisheries and Aquaculture Science (cefas) in two parts, the first used for our comparative study shown in Chapter 3. The remaining portion (1b) was obtained later in the study, and used during research presented in Chapter 4. Cefas also supplied our Sprat dataset, also used in Chapter 4. The remaining Dataset, Herring 2, was provided by the Galway-Mayo Institute of Technology (GMIT) and was used for two studies, shown in Chapters 4 and 5.

Each dataset obtained contains samples from two different fisheries stocks. These are dependent on the dataset and can be seen, along with sample count for each class,

in Table 1.1. The work presented in this thesis focusses on classification of fisheries stock, by assigning individual samples to the correct class using classifiers constructed using remaining samples. Ground truths were determined by expert otolith readers at the donating institutions and supplied with the donated sets. The Herring 1b, Sprat and Herring 2 sets were also separated into individual age categories using standard industry methods. Herring 1a was not subject to age categorization as it was not intended to be used in single age tests.

The full datasets are not used in any study presented in this thesis. Each chapter describes how sets were sub-sampled for use in the relevant chapters. A number of samples were not suitable for use, with a some images being of poor quality, otolith boundaries being obstructed, or otoliths being broken prior to image capture.

Additionally, a small number of Plaice otoliths (not images) were supplied by Cefas, given without any information on catchment area or potential fisheries stock. Whilst not detailed in Table 1.1, we mention them here as one of these samples was used for the tomographic reconstruction study seen in Chapter 6.

1.8 Summary

This Chapter has introduced the work presented in this thesis. We have discussed the background and motivations for the research, and stated clear objectives to be addressed. The publications and submissions of work presented have been summarised with shortened descriptions of the work, and the unique contributions that the work represents have been stated. Finally, we have looked at how the remaining texts have been organised, and the research methods and materials used throughout have been detailed.

Dataset	Abbreviation used	Class 1 (quantity)	Class 2 (quantity)	Use in thesis	Reference and Notes
Herring 1a	H	North-Sea 51	Thames 77	Chapter 3	ICES survey 2008/2009
Herring 1b	H	North-Sea 174	Thames 195	Chapter 4	
Herring 2	BB	Celtic-Sea 69	Irish-Sea 103	Chapters 4, 5	Trawls conducted by the Agri-Food and BioSciences Institute (AFBI) & the Marine Institute (MI) [22]
Sprat	S	North-Sea 263	English Channel 163	Chapter 4	ICES Surveys 2013/2014. [41, 42]. Age-0 set partially created from age-1 samples using trace of the age-0 annuli

Table 1.1: Table showing the datasets obtained for use in the study presented in this thesis.

Chapter 2

Review Of Methods

This Chapter gives details of methods used throughout this thesis. We describe in detail, techniques previously used in the fields of otolith analysis, computer vision and machine learning, and how each method is used throughout this work. The main focus of this chapter are four boundary transformation methods used to encode closed-curve boundaries: Elliptical Fourier Analysis (Section 2.2), Curvature Scale Space (Section 2.3), Time-series Shapelets (Section 2.4), and Shape indices (Section 2.5). We also describe the otolith segmentation techniques used, and the methods by which the boundary is extracted and rotationally normalised in Section 2.1, and the machine learning techniques used to build and test classifiers (Section 2.6)

2.1 Outlining Methods

The first stage in any boundary based classification is the extraction of the otolith boundary itself. In previous work two predominant methods of image segmentation are used: Image thresholding [11, 12, 22, 30] and hand tracing [25, 63, 74, 90, 92]. Both methods create a foreground (otolith) mask from which an outline/boundary can be extracted. Due to the variability of otolith images across image-sets it was

not possible to construct a uniform process that would perform adequately on all sets without supervision. The sample sets used for these studies have been captured and processed by different research groups, and stored/preserved in a variety of ways. Image sets were captured under variable lighting conditions and using different equipment, as this is currently not standardised between otolith laboratories. This resulted in a significant variability in sample or image quality.

During the work presented in this thesis both thresholding and hand-tracing methods of boundary segmentation were used, with the method of segmentation dependent on the underlying dataset to be segmented. Whilst thresholding methods appear more frequently in the literature, it was found to be incompatible with a number of the sets used in this work. Additionally, hand tracing is often employed where segmenting samples at selected annuli is required [22].

2.1.1 Hand Tracing

Hand tracing was explicitly used in work presented in Chapter 3 as the archived images used in the study presented in chapter 3 have large amounts of background noise and lighting artefacts (example shown in figure 2.1). Chapter 4 used multiple datasets from different sources, a number of which needed to be traced manually due to severe artefacts present in the images (Figure 2.1). Additionally, hand-tracing methods were used by different expert readers as part of the study presented in chapter 5 where we investigate the impact of the outlining methodology on Fourier based classification.

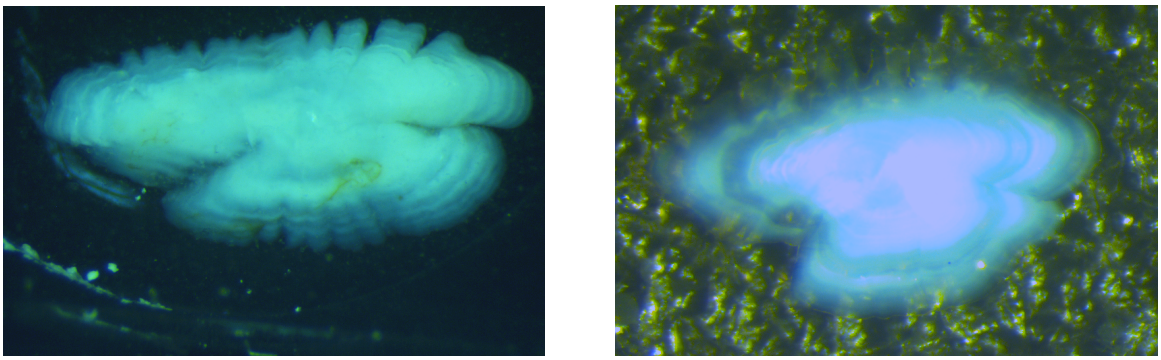


Figure 2.1: Otoliths from Herring (left) and Sprat (right) showing lighting artefacts.

All hand tracing for these studies was undertaken in a custom built system constructed using MATLAB [62]. Images were, in turn, displayed to the user and outlined by hand using a mouse. For the study presented in chapter 3 boundaries were traced using a steelseries®'RIVAL' mouse, a high performance/precision mouse paired with precision mouse-mat; whilst for the study presented in chapter 5 one expert user performed the outlining task with the high precision mouse, whilst the second expert used a standard DELL®mouse with no mouse-mat. Amendments to the segmentation were allowed, until the user was satisfied with their outlining accuracy. After each amendment (or initial outline) the boundary of the segmentation was calculated (discussed is Section 2.1.3) and overlayed on the original image so that the user may assess the accuracy.

2.1.2 Image Thresholding

Image thresholding as a method of otolith segmentation was employed in Chapters 4 and 5. As mentioned above, Chapter 4 presents a study comprising multiple datasets, one of which was of high enough image quality to segment using intensity thresholding (Figure 2.2).

Image segmentation using intensity thresholding was also carried out in MATLAB. Before images were shown to the user, they were each converted to grayscale images using existing MATLAB function 'rgb2gray'. This method converts the images original RGB values to grayscale values using a weighted sum of its individual colour components (Equation 2.1.1). Where P_{xy} is a pixel at coordinates x,y in the image; $P^{i/r/g/b}$ are the pixel values at the coordinate in the intensity/red/green/blue plane:

$$P^{i}_{xy} = 0.2989(P^{r}_{xy}) + 0.5870(P^{g}_{xy}) + 0.1140(P^{b}_{xy}) \tag{2.1.1}$$

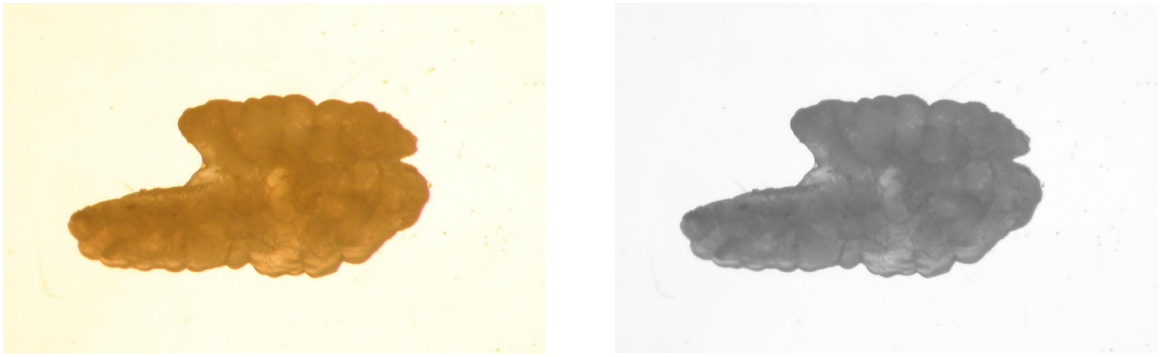


Figure 2.2: High image quality sample of (Celtic-sea) Herring, and resulting image after converting to grayscale.

The intensity range of each pixel in the resulting image is 0 (black) to 255 (white). Greyscale images are first binarised using an automatically chosen threshold [71]. Further refinement is attempted using a form of histogram back-projection [93] with strict confidence thresholds. The resulting mask is then displayed and the supervisor may make interactive adjustments, eg: change the initial threshold; discard the histogram refinement; or manually amend the mask. In extreme cases where the outline is poorly defined, the supervisor may choose to outline the entire otolith manually.

In this instance (example shown in Figure 2.2) any pixel with intensity under the selected threshold is assigned as foreground/otolith.

As with hand tracing methods, after each modification or initial threshold selection the boundary of the segmentation is calculated and overlayed on the original image (not the grayscale image) and is shown to the user. Thresholding intensity in this manner causes areas of foreground that have higher intensities than the selected threshold to be designated background and vice-versa. This can be due to noise, minor shading variance or lighting artefacts in the image. Therefore minor corrections are made to the resulting segmentation before the boundary is calculated.

- Background noise - any foreground pixel that is not member of the largest grouping of foreground pixels, is converted to background. (Morphological Opening)
- Foreground noise - any background pixel (or group of) that are completely surrounded by foreground pixels, are converted to foreground. (Morphological Closing)

2.1.3 Boundary Extraction and Rotational Normalisation

To create a visual depiction of the boundary, morphological erosion [37] is used to remove interior pixels from the otolith mask. This visual depiction is used to create an overlay image that is shown to the user during segmentation (as seen in Figure 2.3). Boundaries are flipped and/or rotated so that the ventral edge is uppermost, the otolith rostrum to the left of the image. The angle between the major axis of the mask and the horizontal axis is calculated using MATLAB's 'regionprops' function and used to align images so that all are normalised for rotation. Whilst this is the standard method of normalising outlines for Elliptical Fourier analysis, performing the normalisation prior to boundary extraction ensured that the same boundary orientation was used for each of the transform methods. This avoids any potential transform bias based on orientation of the boundary. Once rotation normalisation is complete the boundary pixels were recorded for further processing.

The different transform methods used as part of this study require the boundary to be encoded in different ways. Each boundary is therefore potentially (dependent on the study) encoded multiple times using differing methods. In all instances the first/initial boundary point was designated as the upper-leftmost foreground pixel, and the boundary was encoded in counter-clockwise order. Three different encodings were used, dependent on further transformation to be applied. Each method is described in further detail in the remainder of this subsection:

- Chain Code - Used during elliptical Fourier transformation (Section 2.2).
- Point Coordinates - Used by Curvature Scale Space (CSS) methods (Section 2.3).
- Distance Vector - Used during Shapelet transformation. (Section 2.4).

Figure 2.3: L-R: Original image; Initial segmentation using thresholding; segmentation after corrections; boundary overlayed on original image (boundary exaggerated for clarity)

Chain Code

The concept of a Chain Code was first described by Freeman [32] and is the boundary encoding method used for Elliptical Fourier Analysis (Described in Section 2.2). To encode the boundary we proceed from the upper-leftmost start pixel and proceed counter-clockwise along the boundary of the image, recording the *direction* travelled to the next boundary pixel, until we return to the initial pixel. Each pixel itself has eight pixel neighbours, the direction to each of these pixels can therefore be described

using digits 0-7. Zero is therefore designated as *right*, one as *up-and-right* etc. A Simplified example of chain-coding can be seen in Figure 2.4.

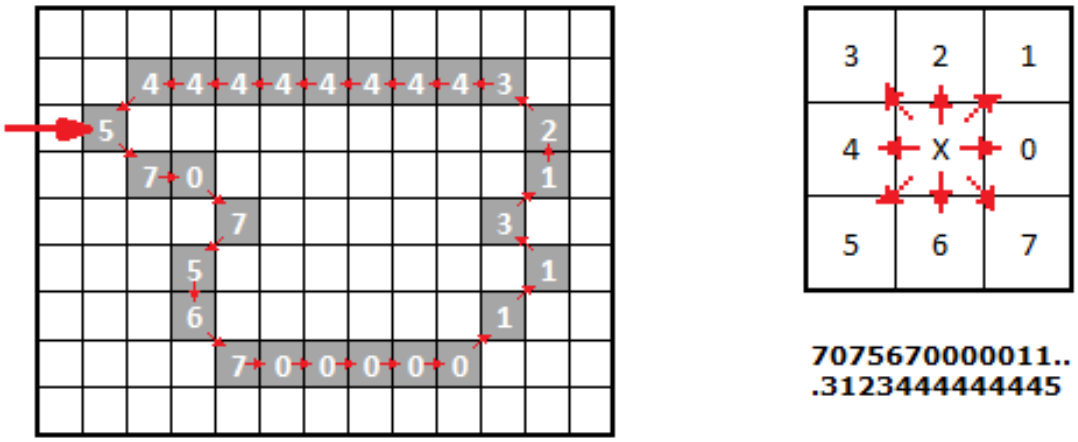


Figure 2.4: Left: A simplified example boundary to be encoded, showing the start point, next boundary pixel and direction. Right: Chart showing code for each possible direction. Encoded example shown bottom-right.

Point Coordinates

Curvature Scale Space (Section 2.3) requires boundaries be encoded in matched-length coordinate arrays; where one array holds pixel x-coordinates, and the other array holds matched y-coordinates. To encode in this manner, boundary pixels are followed the same as when chain-encoding. However, instead of returning 'directions' to the next pixel, the current pixels x and y coordinates are recorded in the two arrays. A brief example of this method is shown in Figure 2.5. Note that in the software used for processing, the vertical axis is indexed with zero at the top most edge, rather than as a standard Cartesian coordinates system.

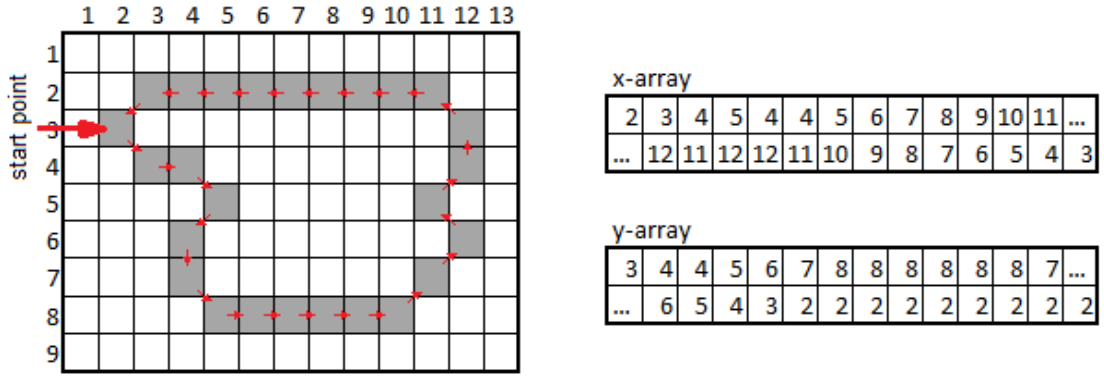


Figure 2.5: Left: A simplified example boundary to be encoded, showing the start point and next boundary pixel. Right: Paired arrays showing coordinate storage of each pixel.

Distance Vector

The Shapelet method described in Section 2.4 requires boundary data-points to be encoded as a univariate series. To achieve this we first calculate the centroid of each boundary by taking the mean x and y coordinates from coordinate arrays X and Y and compute distance to the centroid for each boundary point. We use the following Pythagorean equation to create a new array D, holding the distance to centroid for each boundary pixel. Where D_i is the ith element in the new distance-to-centroid array; k is the number of boundary points and X/Y are the coordinate arrays:

$$D_i = \sqrt[2]{(X_i - \frac{1}{k}\sum_{i=1}^{k} X_i)^2 + (Y_i - \frac{1}{k}\sum_{i=1}^{k} X_i)^2} \tag{2.1.2}$$

2.2 Elliptical Fourier Analysis

Elliptical Fourier Analysis was first discussed in 1981 in a joint work between Fairleigh Dickinson University and the U.S. Army Research and Development Command [48]. The work presents a method of accurately describing and reconstructing two-dimensional closed curves using Fourier coefficients with a calculable degree of error.

Kuhl and Giardina [48] outlines a procedure for calculating Fourier coefficients for a closed curve encoded using a chain code representation (see Section 2.1.3). The resulting descriptors are robust to rotation, scale, translation and initial boundary point designation. Kuhl's method is an extension of techniques that requires no integration and therefore lends itself to an algorithmic approach. We therefore follow Kuhl's method closely in order to calculate elliptical Fourier coefficients.

Elliptical Fourier features/descriptors (EFDs) are widely used in the (otolith) industry, being used for many otolith classification studies [21, 22, 31, 63, 81, 90, 91, 92], a number of which report elliptical Fourier based classification results at 90%+ accuracy. As such Elliptical Fourier analysis forms a starting point for the work presented in this thesis.

The remainder of this section describes the methods used during the Fourier transformation process. We build on an example of boundary encoding originally seen in Section 2.1.3 (chain coding) and used throughout this section, to detail how Fourier coefficients are calculated (Sections 2.2.1 to 2.2.4), Normalised for position (Section 2.2.5) and scale (Section 2.2.6) and how, once Fourier harmonics have been calculated, boundaries can be reconstructed (Section 2.2.7).

2.2.1 Boundary Traversal and Period

The *period* (T) of the signal is equal to the time taken to traverse the boundary. Whilst the *length* of the boundary (k) is simply equal to the number of pixels that make up the boundary; the time taken to traverse the boundary is equal to the sum of the distances between each pixel. As the direction to each pixel is recorded in a chain-code series (a), direction is either orthogonal or on a 45° diagonal. The distance/time

for each transition (i) can therefore be given by:

$$\Delta t_i = 1 + (\frac{\sqrt{2}-1}{2})(1-(-1)^{a_i}) \tag{2.2.1}$$

The time to traverse the boundary up to point p can therefore be given by the following equation. Note that the basic period (T) of the chain-encoded boundary is calculated using the same equation, where $p = k$:

$$\begin{aligned} t_p &= \sum_{i=1}^{p} \Delta t_i \\ \therefore T &= \sum_{i=1}^{k} \Delta t_i \end{aligned} \tag{2.2.2}$$

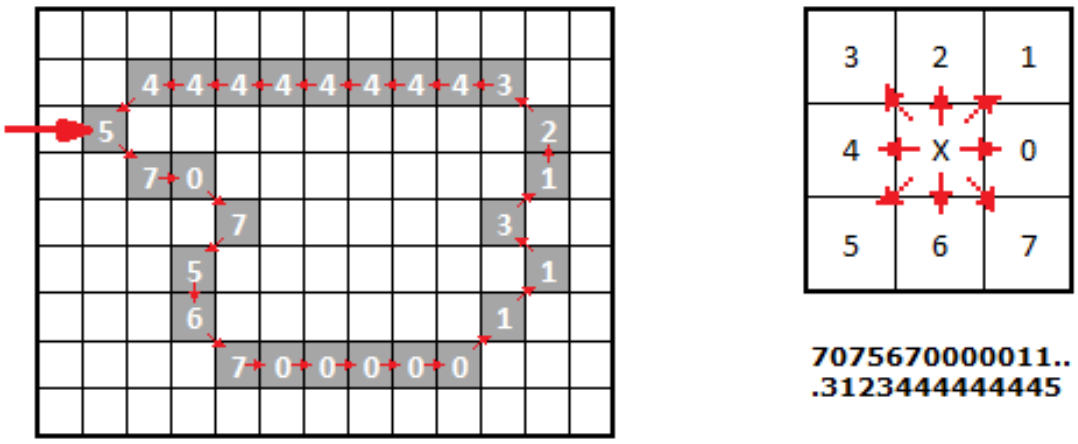


Figure 2.6: Left: A simplified example boundary to be encoded, showing the start point, next boundary pixel and direction. Right: Chart showing code for each possible direction. Encoded example shown bottom-right.

Figure 2.6 shows the same chain code example as in Section 2.1.3. Given the chain-encoded boundary in the example we may calculate arrays for both Δt and t, shown in Table 2.1. The basic period (T) for the contour is equal to t_k or in this instance $16 + 10\sqrt{2}$

	$p=1$							$p=26$
a_p	7	0	7	5	...	4	4	5
Δt_p	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{2}$	...	1	1	$\sqrt{2}$
t_p	$\sqrt{2}$	$1+\sqrt{2}$	$1+2\sqrt{2}$	$1+3\sqrt{2}$	...	$15+9\sqrt{2}$	$16+9\sqrt{2}$	$16+10\sqrt{2}$

Table 2.1: Table showing chain code (a) from the simplified example, with calculated transition times for each step (Δt_p) and 'time to t_p' (t_p)

2.2.2 Point by Point Projection

We also calculate the position change as we move from one boundary point to another. Given that it is not possible for the next pixel to be the same as the current pixel, the *next* pixel must have a different x and/or y position. The change in x/y positions for each transition is therefore given as:

$$\begin{aligned} \Delta x_i &= sign(6 - a_i) * sign(2 - a_i) \\ \Delta y_i &= sign(4 - a_i) * sign(0 - a_i) \end{aligned} \tag{2.2.3}$$

Where:

$$sign(Z) \begin{cases} 1 & Z > 0 \\ 0 & Z = 0 \\ -1 & Z < 0 \end{cases} \tag{2.2.4}$$

This ensures that when a_i is either 6 or 2, Δx_i is zero (no change in x-axis position). Similarly, when a_i is 4 or 0, the result shows no change in the y-axis. As when calculating the distance along the first p boundary pixels, we can calculate the projection (from the initial boundary point) to any pixel (p) using the following summations:

$$\begin{aligned} x_p &= \sum_{i=1}^{p} \Delta x_i \\ y_p &= \sum_{i=1}^{p} \Delta y_i \end{aligned} \tag{2.2.5}$$

Table 2.2 shows the calculated projections for our simplified example, given the formulae shown (Formulae 2.2.3 – 2.2.5). It can be seen that the last transition (where

	$p=1$							$p=26$
a_p	7	0	7	5	...	4	4	5
Δx_p	+1	+1	+1	−1	...	−1	−1	−1
Δy_p	−1	0	−1	−1	...	0	0	−1
x_p	+1	+2	+3	+2	...	+2	+1	0
y_p	−1	−1	−2	−3	...	+1	+1	0

Table 2.2: Table showing chain code (a) from the simplified example, with calculated change in projections (Δx_p and Δy_p) and 'projections to p'

$p = k$) ensures that the projections from the initial point to the p^{th} point (the starting pixel) are both zero.

2.2.3 Fourier Expansion and Summation

It is important to note that elliptical Fourier methods require cosine and sine components for both the x and y contour projections. In total the projections are calculated using two DC components (positional components), plus four coefficients (a_n, b_n, c_n and d_n) for each harmonic [n]. DC is a term taken from electrical signal analysis, referring to 'Direct Current'. It represents the average coefficient (or position) along the described axis.

$$\begin{aligned} x(t) = A_0 + &\left[a_1 Cos\frac{2\pi t}{T} + b_1 Sin\frac{2\pi t}{T}\right] + \left[a_2 Cos\frac{4\pi t}{T} + b_2 Sin\frac{4\pi t}{T}\right] + ... \\ &\left[a_3 Cos\frac{6\pi t}{T} + b_3 Sin\frac{6\pi t}{T}\right] + \left[a_4 Cos\frac{8\pi t}{T} + b_4 Sin\frac{8\pi t}{T}\right] + ... \end{aligned} \tag{2.2.6}$$

$$\begin{aligned} y(t) = C_0 + &\left[c_1 Cos\frac{2\pi t}{T} + d_1 Sin\frac{2\pi t}{T}\right] + \left[c_2 Cos\frac{4\pi t}{T} + d_2 Sin\frac{4\pi t}{T}\right] + ... \\ &\left[c_3 Cos\frac{6\pi t}{T} + d_3 Sin\frac{6\pi t}{T}\right] + \left[c_4 Cos\frac{8\pi t}{T} + d_4 Sin\frac{8\pi t}{T}\right] + ... \end{aligned}$$

Both x and y projections can theoretically be written as summations of an infinite number of harmonics, adjusted using DC components. This requires infinite series of a,b,c and d coefficients, plus the DC terms.

$$\begin{aligned} x(t) &= A_0 + \sum_{n=1}^{\infty} a_n Cos\frac{2n\pi t}{T} + b_n Sin\frac{2n\pi t}{T} \\ y(t) &= C_0 + \sum_{n=1}^{\infty} c_n Cos\frac{2n\pi t}{T} + d_n Sin\frac{2n\pi t}{T} \end{aligned} \tag{2.2.7}$$

In real terms however, since the data is discretely sampled, the number of harmonics that can be calculated is limited by the Nyquist frequency as chain-coded boundaries are discretely sampled boundary points. Thus the maximum number of harmonics that can be calculated is approximately equal to half the length of the chain encoded boundary ($k/2$).

2.2.4 Coefficient Calculation

Methods described by Kuhl and Giardina [48] show that the four coefficients for each harmonic can be calculated given the following four equations. Where a_n and b_n are the cosine and sine coefficients (respectively) of the n^{th} harmonic of the x projection; c_n and d_n are the cosine/sine coefficients of the y projection (n^{th} harmonic):

$$\begin{aligned} a_n &= \frac{T}{2n^2\pi^2} \sum_{p=1}^{k} \frac{\Delta x_p}{\Delta t_p} \left[Cos\frac{2n\pi t_p}{T} - Cos\frac{2n\pi t_{p-1}}{T} \right] \\ b_n &= \frac{T}{2n^2\pi^2} \sum_{p=1}^{k} \frac{\Delta x_p}{\Delta t_p} \left[Sin\frac{2n\pi t_p}{T} - Sin\frac{2n\pi t_{p-1}}{T} \right] \\ c_n &= \frac{T}{2n^2\pi^2} \sum_{p=1}^{k} \frac{\Delta y_p}{\Delta t_p} \left[Cos\frac{2n\pi t_p}{T} - Cos\frac{2n\pi t_{p-1}}{T} \right] \\ d_n &= \frac{T}{2n^2\pi^2} \sum_{p=1}^{k} \frac{\Delta y_p}{\Delta t_p} \left[Sin\frac{2n\pi t_p}{T} - Sin\frac{2n\pi t_{p-1}}{T} \right] \end{aligned} \tag{2.2.8}$$

It is important to note that the t_p where $p = 1$ is the first transition along the boundary. Therefore for the first summation t_{p-1} is forced to zero.

2.2.5 DC Terms

For the studies undertaken in this thesis, the position of the boundary in the image is normalised. However we give the DC terms here as further studies may wish to include them. Note that again t_{p-1} is forced to zero when performing the calculations. To normalise for boundary translation we simply force both DC terms A_0 and C_0 to zero.

$$\begin{aligned} A_0 &= \frac{1}{T}\sum_{p=1}^{k}\frac{\Delta x_p}{2\Delta t_p}(t_p^2 - t_{p-1}^2) + \varepsilon_p(t_p - t_{p-1}) \\ C_0 &= \frac{1}{T}\sum_{p=1}^{k}\frac{\Delta y_p}{2\Delta t_p}(t_p^2 - t_{p-1}^2) + \delta_p(t_p - t_{p-1}) \end{aligned} \tag{2.2.9}$$

Where:

$$\begin{aligned} \varepsilon_p &= \sum_{j=1}^{p-1}\Delta x_j - \frac{\Delta x_p}{\Delta t_p}\sum_{j=1}^{p-1}\Delta t_j \\ \delta_p &= \sum_{j=1}^{p-1}\Delta y_j - \frac{\Delta y_p}{\Delta t_p}\sum_{j=1}^{p-1}\Delta t_j \end{aligned} \tag{2.2.10}$$

2.2.6 Fourier Normalisation

There are three forms of normalisation that can be applied to Fourier descriptors: Normalisation of Rotation; Translation; and Scale. Rotational normalisation is a simple procedure described in Section 2.1.3 whereby the otolith boundary is rotated until the semi-major axis runs horizontal. Translation is normalised simply by forcing the DC Fourier components to zero which ensures the centroid of each reconstructed boundary is at point [0,0].

The only normalisation that requires further discussion here is that for scale. This ensures that the semi-major axis of the first Fourier ellipse is of length 2 for every Fourier reconstruction (from -1 to 1 on the x-axis). Further harmonics may alter

the total length of the reconstructed otolith through subsequent harmonics, but each coefficient of the $2^{nd} - n^{th}$ harmonics are scaled by the same factor.

The first step in rotation normalisation of Fourier descriptors is to calculate the angle between the first semi-major axis and the horizontal. Equation 2.2.11 calculates this from the coefficients of the first Fourier harmonic (the first major ellipse):

$$\theta_1 = \frac{1}{2} arctan\left[\frac{2(a_1 b_1 + c_1 d_1}{a_1^2 + c_1^2 - b_1^2 - d_1^2}\right] \tag{2.2.11}$$

Once θ has been calculated the terms a_1^* and c_1^*, the coefficients correct for the displaced (by rotation) starting point, and normally used during normalisation of rotation, can be calculated using equation 2.2.12:

$$\begin{bmatrix} a_1^* & c_1^* \\ b_1^* & d_1^* \end{bmatrix} = \begin{bmatrix} cos\theta_1 & sin\theta_1 \\ -sin\theta_1 & cos\theta_1 \end{bmatrix} \begin{bmatrix} a_1 & c_1 \\ b_1 & d_1 \end{bmatrix} \tag{2.2.12}$$

However during scale normalisation the a_1^* and c_1^* terms are used to calculate a scaling component ϵ (equation 2.2.13):

$$\begin{aligned} \epsilon =& (a_1^{*2} + c_1^{*2})^{\frac{1}{2}} \\ \therefore \epsilon =& ((a_1 cos\theta_1 + b_1 sin\theta_1)^2 + (c_1 cos\theta_1 + d_1 sin\theta_1)^2)^{\frac{1}{2}} \end{aligned} \tag{2.2.13}$$

However, As we have already performed rotational normalisation θ, the angle calculated between the semi-major axis and the horizontal, is already zero. Therefore, as *sin 0 = 0* and *cos 0 = 1*, ϵ is given by equation 2.2.14:

$$\therefore \epsilon = (a_1^2 + c_1^2)^{\frac{1}{2}} \tag{2.2.14}$$

Scale normalised Fourier descriptors are then calculated by division of each harmonics coefficients by the scaling factor. Where $a_i{}'$ denotes the i^{th} harmonic's scale normalised 'a' coefficient:

$$\begin{aligned} a_i{}' =& a_i/\epsilon \\ b_i{}' =& b_i/\epsilon \\ c_i{}' =& c_i/\epsilon \end{aligned} \tag{2.2.15}$$

After scale normalisation is complete the first three coefficients become redundant as $a'_1 = 1, b'_1 = 0, c'_1 = 0$, although in practice these values are approximate ($\pm 10^{-20}$) . In order to remove influence of these coefficients we remove them from all instances before training/testing classifiers.

2.2.7 Contour Reconstruction

Each boundary contour can be approximated using the coefficients calculated. The x and y projections are calculated as a summation of projections using each harmonic. Approximation allows us to reconstruct the contour using a differing number of 'sample points' to the contour that was transformed. The number of new points is given as T, the t in each equation is an array of length T holding values equal to the cell index [1,2,3,4,5,...,T]. Each element of the summation therefore equates to an array also of length T. All arrays are summed and modified with the projections DC term.

The resulting arrays are the x and y (X_n/Y_n) projections for the reconstructed contour. The projections for x and y can be used to plot ellipses corresponding to each individual harmonic, and the summed X_n and Y_n can be used to plot a contour using the first n harmonics. Where N is the number of harmonics to be used; n is the specific harmonic used:

$$\begin{aligned} X_N &= A_0 + \sum_{n=1}^{N} a_n Cos\frac{2n\pi t}{T} + b_n Sin\frac{2n\pi t}{T} \\ Y_N &= C_0 + \sum_{n=1}^{N} c_n Cos\frac{2n\pi t}{T} + d_n Sin\frac{2n\pi t}{T} \end{aligned} \tag{2.2.16}$$

Figure 2.7 shows reconstructions of the previous example using the first two harmonics. The first ellipse reconstructed using only the first harmonic and 360 calculated points ($T = 360$) is shown in black. Also shown is the ellipse reconstructed using only the second harmonic (blue) with its centroid positioned at various points on the

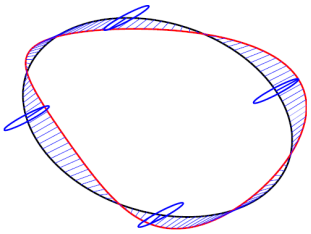


Figure 2.7: Showing first projected ellipse (black), second ellipse (blue), and summation thereof. (see text)

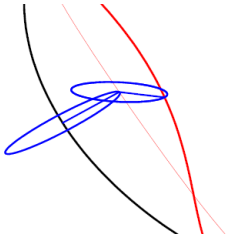


Figure 2.8: Showing three harmonics being summed to form the projection of the first three projections. (see text)

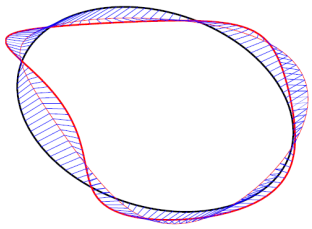


Figure 2.9: Showing first projected ellipse (black), and summation of first two ellipses (thin-red) and first three (thick-red). (see text)

reconstructed first ellipse. The second ellipse, whilst reconstructed using 360 calculated points, has only 180 discrete points ($T/2$), as the sample frequency is twice that of the first ellipse (the third ellipse will have $T/3$ discrete points etc). The Red contour shown is the summation of the first and second reconstructed ellipses, and blue 'transitions' show the vector summed with every third point on the first ellipse (every third point shown for clarity).

Figure 2.8 shows how the harmonic projections are summed. Each of the projections to be summed contains a vector of length T, with a series of T/n points repeated n times. Shown in the figure are vectors calculated from the x and y projection arrays. The point where $t = 45$ is shown in Figure 2.8, the projection for the second harmonic is centred at point t along the first projection (thick black line) and its x/y projection for point t is shown as the blue radial. The resulting point is at point t along the summation of the two harmonics (thin red line). At this point we center the third harmonic projection, find the projection at point t and this gives point t along the third summation (thick red line). This process continues using projections for each harmonic for each t value.

Figure 2.9 shows the same example reconstructed using the firsts three harmonics. Again the first ellipse is shown in black using 360 calculated points. The thinner of the two red contours represents the summation of the first and second ellipses as in Figure 2.7. The thicker of the red contours represents the summation of the first three harmonic projections. Also marked are transition lines showing the addition of the second harmonic (between black and thin-red contours) and addition of the third harmonic (between thin-red and thick-red contours).

This process can be used to create projections using any number of the calculated harmonic coefficients. Each additional harmonic (in theory) enables recreation of a contour that is more similar to the original chain-coded contour. Figure 2.10 shows reconstruction of our previous example using varying numbers of harmonics ($n = 1ton = 13$). Given the low resolution of our example contour (26 samples) only thirteen harmonics could be accurately calculated. In this example the limited number of harmonics available did not allow reconstruction to be overly accurate. Where the original boundary has larger number of boundary points, the Nyquist frequency (the maximum frequency at which boundary points can be sampled without introducing error) is much larger. This allows much more 'fine tuning' of the reconstructed contour for the samples used throughout this work.

2.3 Curvature Scale Space

Curvature Scale-Space (CSS) [65] forms the basis for contour-based shape descriptors as part of the MPEG-7 standard [15, 101, 102]. CSS transforms a boundary from its coordinate representation to one which encodes the position and magnitude of the concavities on the boundary curve, producing an ordered set of *maxima* from a boundary which are used as descriptors for the shape in question. These descriptors

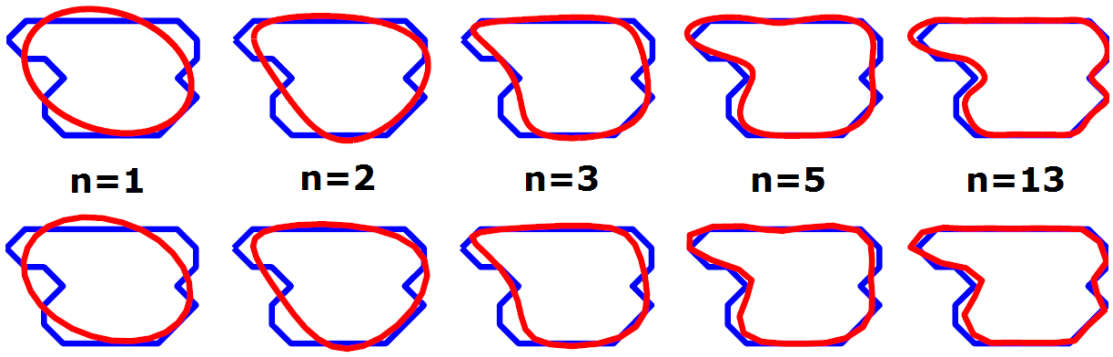


Figure 2.10: Showing example boundary reconstruction overlayed on original chain coded boundary, at (l-r) $n = 1, 2, 3, 5, 13$; reconstructed using 360 boundary points (top row) and 26 points (bottom row).

may then be used to determine similarity or dissimilarity to another image. This in turn may be used to find 'most similar' objects (or boundaries thereof) in a database.

Research has shown that CSS encoding and its attributed matching algorithm can be an effective and robust (to noise, scale, rotation) method of matching query images to database models when combined with global parameters such as Circularity and Eccentricity [1, 5]. Bober [15] states that CSS is also robust to perspective transforms.

Given these strengths, CSS forms an ideal starting point for boundary based shape classification of otoliths, and has been used for several other studies in the field of marine biology [1, 2, 66, 67, 89, 98] including the analysis of otoliths [73]. As a standard method it is also well documented in other studies of shape representation [5, 49, 64, 85, 100].

This section details the process by which boundaries are transformed to curvature scale space, including: The extraction of curvature descriptors by recursive smoothing of the boundary curve (Section 2.3.1); The standard method of comparing CSS

representations, and how to include mirrored images (Section 2.3.2), and how measures such as Circularity and Eccentricity may allow pre-indexing database images to increase accuracy and reduce run-time of matching algorithms (Section 2.3.3).

2.3.1 Extraction of Descriptors

Descriptors (or maxima) are extracted from a boundary by iteratively smoothing that boundary, until it becomes a convex curve. At each iteration of this process the curvature of the boundary is calculated using the following formula [73], where the boundary contour is represented by its coordinates $\Gamma(u) = (x(u), y(u))$.

$$Curvature(u) = \frac{\dot{x}(u)\ddot{y}(u) - \ddot{x}(u)\dot{y}(u)}{(\dot{x}(u)^2 + \dot{y}(u)^2)^{3/2}} \tag{2.3.1}$$

To smooth the boundary one filters the contour with a Gaussian function using a linearly increasing scaling parameter σ.

$$g(u, \sigma) = \frac{1}{\sqrt{2\sigma^2}} e^{-u/2\sigma^2} \tag{2.3.2}$$

Therefore one can compute the curvature of a smoothed contour at each increment using the formula:

$$Curvature(u, \sigma) = \frac{\dot{x}(u, \sigma)\ddot{y}(u, \sigma) - \ddot{x}(u, \sigma)\dot{y}(u, \sigma)}{(\dot{x}(u, \sigma)^2 + \dot{y}(u, \sigma)^2)^{3/2}} \tag{2.3.3}$$

Once curvature of the boundary is determined, inflection points (zero-crossings) are determined and noted. As iterations progress and the scaling parameter increases the boundary will approach a fully convex curve, and in the process inflection points will converge as concavities are smoothed out of the boundary. It is these convergence points that are recorded as descriptors for the boundary.

Figure 2.11 shows a CSS representation of the otolith boundary shown in Figure 2.12. The three largest structures on the image denote the three largest curvature features of the boundary. For clarity Figure 2.13 shows the boundary part

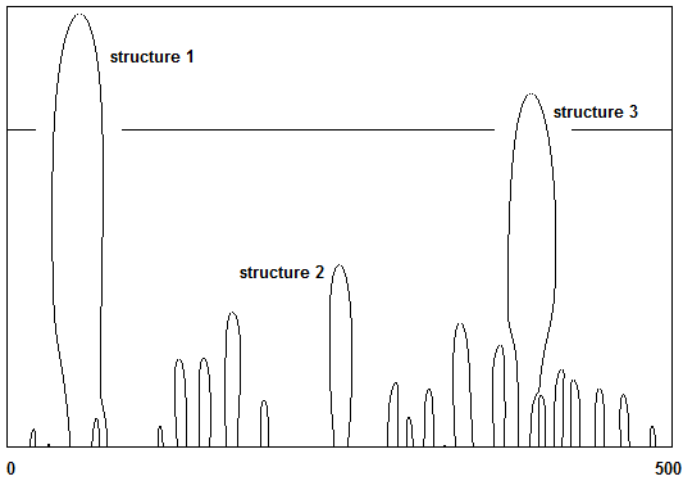


Figure 2.11: CSS representation image of a herring otolith. Horizontal axis represents point along the boundary, Vertical axis represents an arbitrary smoothing increment. The horizontal line across the image is the point at which the smoothed boundary shown in figure 2.13 was produced.

way through the smoothing process. It can be seen that a shallow but long concavity is still present on the boundary (structure three) whilst structure two, another prominent concavity, has already been smoothed to convexity.

Once the smoothing process is complete and the boundary has been reduced to a fully convex curve, we extract from the CSS image the *maxima*. These are extracted as coordinate pairs for each concavity, where each pair consists of a point/distance along the boundary, and number of smoothing increments to convergence (the point at which the concavity disappears). Point along the boundary is a comparative figure as CSS representations are invariant to rotation, or to start point on boundary. A boundary can be rotated simply by performing a circular shift on all the boundary points/distances simultaneously. Figure 2.14 shows four different rotations of the same otolith boundary. Marked on each boundary is the start point from which coordinates are extracted from the boundary (following a counter-clockwise direction).

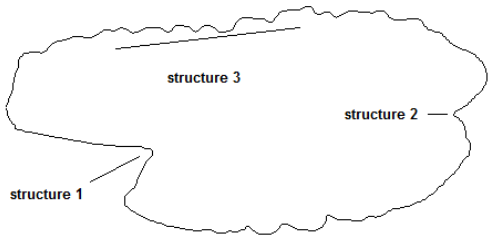


Figure 2.12: Initial boundary of a herring otolith. Three major features/structures are marked on the image: structures 1 and 2 are obvious large concavities, however structure 3 is a long concavity that, despite its shallow depth, takes a large number of smoothing increments to remove.

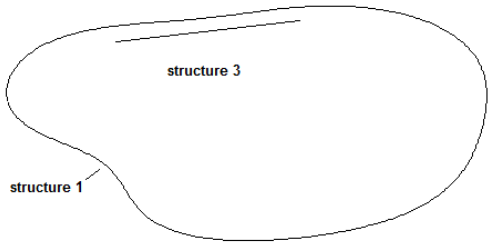


Figure 2.13: The same boundary as in figure 2.12 after several smoothing increments. Produced at the same point as marked on figure 2.11 it shows that structure 3 (as well as structure 1) is still present in the boundary.

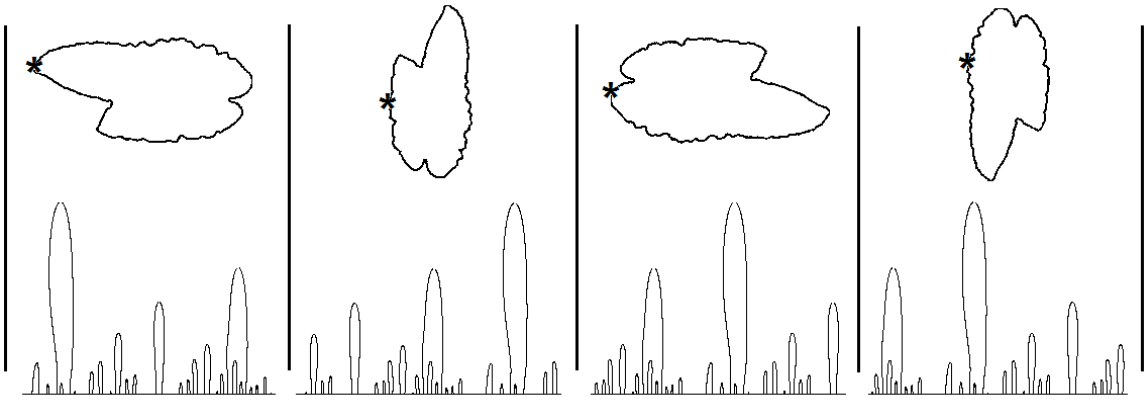


Figure 2.14: Four (clockwise) rotations of the same otolith boundary. Left to Right: 0° rotation, 90°, 180°, 270°. Under each is depicted the resulting CSS image given the marked boundary start points (stars).

It can be seen that a clockwise rotation of the image results in a (circular) shift of the CSS image to the left. It is also the case (not depicted) that designating a differing boundary start position on the non-rotated otolith boundary results in a similar shift of the CSS image.

2.3.2 Matching Algorithm

With each boundary stored as the locations of its curvature-crossing maxima, images can be compared to one another by finding the Euclidean distance between respective points. This results in a measure of similarity or dissimilarity between the two sets of maxima and can be used to compare one *image* to a number of *model*s in a database and find the *model* with the greatest similarity (or lease dissimilarity) by finding the error between respective points.

Alignment Points

The current literature (including Abbasi et al. [1]) suggests that the CSS images should be aligned by their greatest (filter level) maximum to do this. It is also said that maxima within a certain ratio of said maximum should also be considered as the alignment point. This is to account for possible boundary variation where there may be multiple points of similar significant curvature. Previous works [1, 2, 66] suggest that this ratio be eighty percent of the greatest evolution (of the individual boundary), and any maxima of greater magnitude than this be considered as an alignment point.

Each set of coordinates are shifted so that their greatest maxima are at the 'start' of the curve to make the first alignment complete. Further alignments are identified by searching for any of the maxima with a filter level greater than the recommended 80% level. This is done for both image and model, and error must be calculated for each possible alignment between the two sets of flagged maxima. Figure 2.15 shows

two highly simplified sets of maxima, alongside the four alignments that would be processed given an 80% alignment cut-off. From the figure it can be seen that, for this simplified example, there are four possible alignments of maxima given the 80% cut-off.

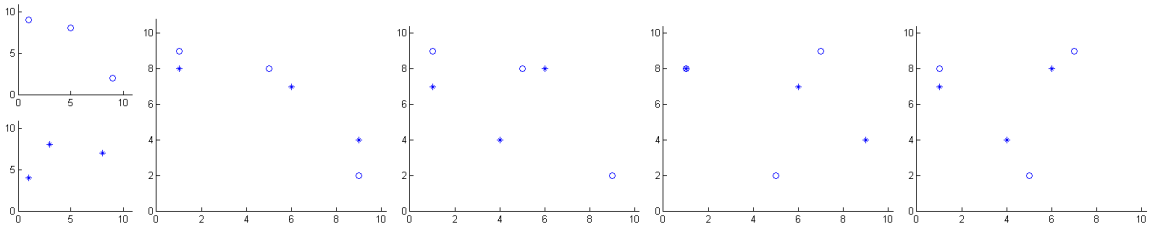


Figure 2.15: Two sets of maxima from hypothetical boundaries (top-left and bottom-left) alongside four possible alignments of maxima greater than 80% of the maximum.

Matching Order

To calculate the error of an alignment we process each of the maxima in the image, from greatest filter level to the smallest. We find the nearest of the maxima in the model to each image point in turn and check whether it is within a predefined (Euclidean) distance of our image point. This maximum distance is defined as 20% [66, 73] of the maximum distance possible (remembering that the CSS image itself wraps around). If the distance is within the 20% threshold then it (the Euclidean distance) is added to a cumulative error score for that alignment and the 'matching' point is removed from the model so as to not match it to further image points. If it is not within the threshold then the filter level (or evolution number) of the image point is added to the error. Figure 2.16 shows the order in which maxima are matched with this process, in a simplified theoretical image. Point 1 in the image is matched to A from the model, Point 2 is matched to B. Note that no model maxima are within the distance threshold from image point 3 (shown by the large circle) therefore its

magnitude (2) is added as error. Once all points in the image are processed then the filter levels of any remaining model maxima (point C in our example) are added to the error for that alignment.

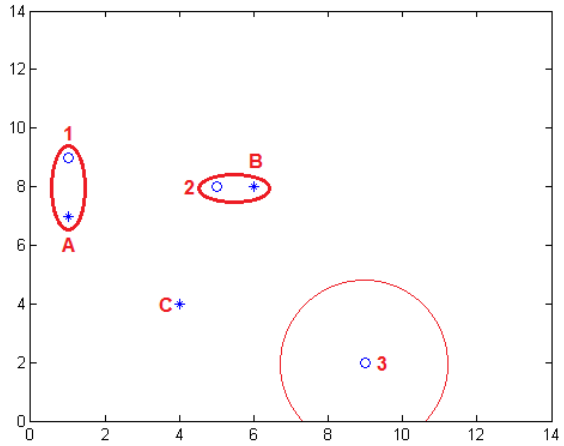


Figure 2.16: Second alignment of maxima from Figure 2.15 showing the order in which image maxima (circles) are matched to model maxima (stars).

The process must then be repeated for the same alignment but with the roles of image and model reversed (points are matched in order of model point magnitudes). This is required as when matched in order of image magnitude, point matches may occur where the model point is closer to a separate image point. Reversing the image and model roles solves this issue. Figure 2.17 shows the differencing between matching in order of image-maxima magnitude and model-maxima magnitude. It can be seen that when matched in image order image point 1 is matched to model point B, which itself is better matched to image point 2. However when image point 2 is processed point B has already been matched and removed, so 2 is matched to the much further away point A (outside the 20% limit). It can be seen that in this simplified example, the model order match results in a lower error than the image order match.

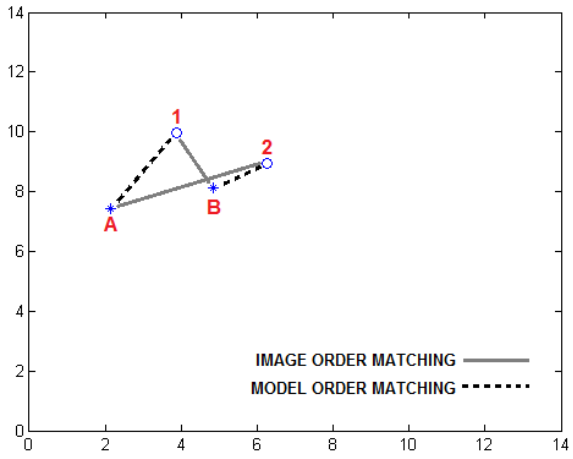


Figure 2.17: Showing difference in matching between image maxima order and model maxima order.

Mirror Images

To account for reflected boundaries, when calculating alignments we simply flip the CSS image that we are finding matches for. To 'reflect' the image we need only amend the maxima array by subtracting the 'non-reflected' boundary point value from the total boundary length plus one. The reflection array is then processed in the same way as the non-reflected array. This ensures that were a query image to be in the reverse position, it would still be accurately matched to models it is compared to. Figure 2.18 shows a CSS image being flipped in this manner; marked are two maxima in the image and mirrored image and the coordinate pairs for each are shown.

Once each possible alignment has been processed, the alignment with the least error is selected and the error returned. The alignment itself is unimportant for the process but the error is used as a measure of similarity/dissimilarity for that model from the database. Once the image has been compared to all models in the database and their respective errors logged, we are able to determine which boundaries in the database (or reflection of) are most similar to the search image.

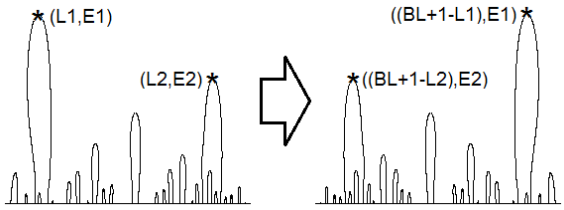


Figure 2.18: Mirroring a CSS image of boundary length *BL*. Two maxima are marked (stars) and their stored coordinate pairs shown. BL - image boundary length; L1/L2 and E1/E2 - respective point along boundary and evolution number for the marked (unmirrored) maxima.

2.3.3 Pre-Indexing

Rather than searching the entire database for a model that best matches the query image; it is possible to search a subset of those models by dismissing those that are globally dissimilar. As suggested by Abbasi et al. [1] and Mokhtarian et al. [66] we are able to dismiss models by comparing the CSS image aspect-ratios of the image and model as well as the circularity and eccentricity metrics recorded for each boundary [1]. If any of these metrics are 'significantly' different then it can be assumed that the boundaries themselves are significantly different. We calculate the difference between the eccentricity *(e)*, circularity *(c)* and aspect-ratio *(ar)*; of images *(i)* and models *(m)* using equations 2.3.4 to 2.3.6:

$$\alpha_e = \frac{|i_e - m_e|}{max(i_e, m_e)} \tag{2.3.4}$$

$$\alpha_c = \frac{|i_c - m_c|}{max(i_c, m_c)} \tag{2.3.5}$$

$$\alpha_{ar} = \frac{|i_{ar} - m_{ar}|}{max(i_{ar}, m_{ar})} \tag{2.3.6}$$

In this instance aspect-ratio is that of the CSS images rather than the boundaries. This metric allows some measure of global curvature for the boundary. As aspect ratio (max iteration/boundary length) increases it signifies more smoothing increments needed to smooth the boundary to a convex arc.

Using these metrics we can set a threshold T to determine which models in the database should be checked against the query image. A lower threshold restricts the number of models that are matched as the dissimilarity of the global metrics allowed is reduced; and were T to be set to zero, all three metrics for the model must match the image metrics exactly. At the opposite end of the scale, as T tends towards 1 the number of models that are assessed increases to a point where all models are included either before or as T reaches 1, and the indexing method becomes defunct.

Abbasi et al. [1] sets a threshold of 0.3 for all three metrics. This figure allows α_e, α_c and α_{ar} to vary within that threshold and the model still be assessed for similarity. All three metrics however are assigned the same T rather than each being set individually. This is as testing of performance increase given each metric individually shows large ranges of T where performance increase remains stable; for example, a threshold between 0.28 and 0.42 for α_c (equation 2.3.5) yields the same 10% performance enhancement. Were there to be a common value for each metric within its peak performance range, that figure can be used as a static metric for all three (in this instance 0.3).

The increase in performance is partially related to the dismissal of globally dissimilar models (using the indexing system) whose CSS images may be similar to that of the image. Figure 2.19 shows the CSS image similarity between two markedly different silhouettes. Despite their difference the CSS images may be seen as similar when matched without prior indexing. By using the metrics described above when matching the database models, dissimilar boundaries like this may be skipped during pre-indexing. Abbasi et al. [2] extends the indexing method by constructing a new global parameter for shallow concavities, showing a modest increase in matching performance when the threshold for that parameter is around 0.8. Note that whilst the

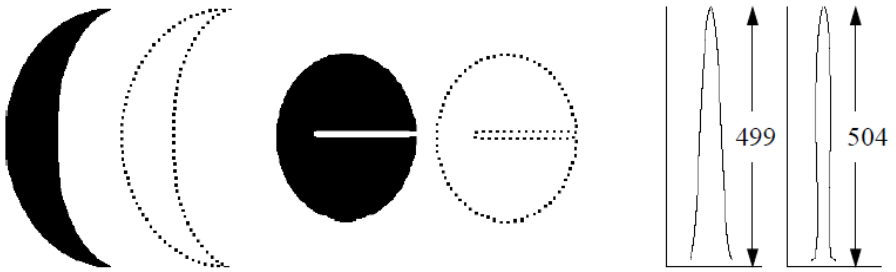


Figure 2.19: Taken from [85]. Two simple silhouettes and their extracted boundaries. Showing the similarity between their respective CSS images (right).

CSS images show a difference in width of the concavity, the maxima is stored only as a distance/magnitude coordinate pair. Richter et al. [85] suggests that width of the CSS curve also be used in calculations.

2.4 Time-Series Shapelets

A Shapelet is a time series data mining primitive that can be used to measure similarity between series based on small common shapes that occur at any point in the series. Shapelets provide interpretable results and can be used to notate which particular features of a boundary are discriminatory of class. This is a significant advantage over methods such as Fourier transforms when used to display results or 'workings' to human readers.

The original work on Shapelets was presented in Ye and Keogh [99], where shapelets are used to classify leaves from two different (but similar looking) species using portions of the boundary. In this method a recursive decision tree algorithm is formulated with Shapelets at the branching criteria. The original work is influenced by Pierre [79] where the most descriptive subsequence is extracted with which to classify. However, advancements in digital computation speeds allow consideration of all sub-sequences in a dataset; a method that was declared to be impossible at the time [79].

Mueen et al. [70] extends this method by considering conjunctions and disjunctions of Shapelets to construct multi-variate decision trees that are then used to classify further signals. Both Ye and Keogh [99] and Mueen et al. [70] use information gain as a method of assessing a shapelets *quality* (its ability to separate classes). Whilst this method is adept at calculating this quality, Lines et al. [54] suggests that the simplest approach for assessing quality is to use the F-statistic for the difference of means (ANOVA testing). Whilst this method is itself not as robust as other methods it is simple to compute; and as it is not being used to perform any hypothesis test is sufficient for the task.

Further research on Shapelet quality measures is presented in Lines and Bagnall [53]. In this work the author compares three methods of shapelet quality testing (Information Gain, Kruskal-Wallis test and Mood's Median) and tests them over a number of classification problems. This methods of quality analysis are briefly described in this section.

Of particular interest is the work in Lines et al. [54], where it is proposed that Shapelets can be used to construct a filter for transforming time series data. Transforming data in this manner moves away from the previous emphasis of tree-based classification described by Lines et al. [54], allowing any traditional classification algorithm to be used with Shapelets. The later data-transform approach forms the basis for methods used in this thesis and is described in this section.

The remainder of this section details the shapelet transform method, from a set of otolith boundaries stored as contour coordinates, to how boundaries are transformed to their shapelet representation.

2.4.1 Pre-processing of boundaries - Details how boundaries are transformed from coordinate pairs to a one-dimensional time series.

2.4.2 Candidate extraction - Describes how shapelet candidates are extracted from each time series, how many candidates can be extracted, and how closed boundaries are processed.

2.4.3 Distance Measurements - Explains how distances between candidates and time-series are calculated, and how subsequent tests *potentially* employ part of previous calculations.

2.4.4 Quality Measures - Lists possible measures for quality of shapelet candidates: Information gain, Kruskal-Wallis, F-statistic, and Mood's median.

2.4.5 Estimating min and max shaplet lengths - Gives details of how shapelet sizes can be pre-calculated, allowing pre-pruning of candidates.

2.4.6 Early Abandon Methods - Describes how certain measures can decrease potential run-time by abandoning certain calculations.

2.4.7 Shapelet Pruning - Explains how shapelets can be discarded based on where they fall within single time-series.

2.4.8 Shapelet Clustering - Explains how shaplets can be clustered between multiple time-series

2.4.9 Transforming Images - Finally, describes how a set of selected shapelets can be used to transform further time series for use in classification tests.

2.4.1 Pre-processing of boundaries

Shapelets with which to transform series must first be extracted from the initial data. Each shapelet is a contiguous sequence of points of any given length, taken from

a time-series in the dataset. As such the first step in determining shapelets is to transform the original set of boundary points to a set of one-dimensional time series. To create the series the Euclidean distance from the centroid to each boundary point is calculated in turn and recorded for use. Figure 2.20 shows such a series extracted from the original boundary. Once all boundary points have been transformed in this

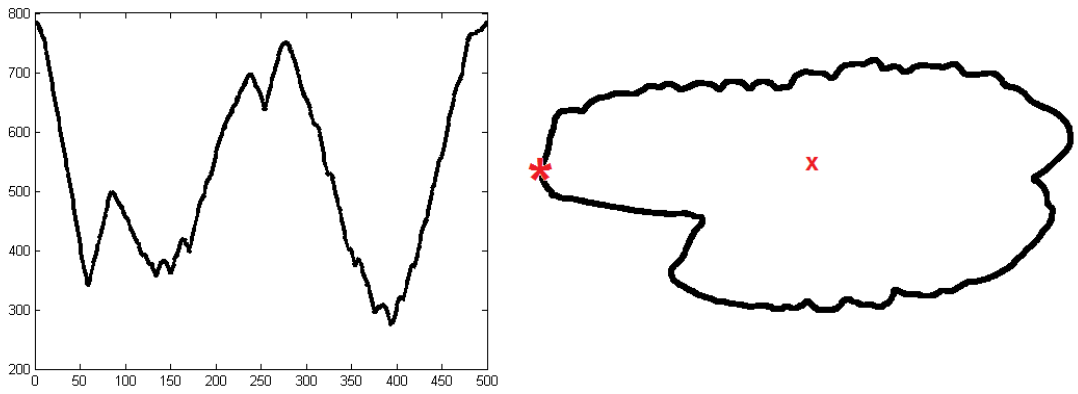


Figure 2.20: Showing a one-dimensional time-series (left) extracted from the original boundary (right). The time series is shown as Euclidean distance to centroid against point along boundary. Marked on the boundary image are the designated boundary start point (star) and the approximate centroid of the boundary (cross). Both plot and boundary are emphasized for clarity.

manner they must be normalised for length. This process ensures that the resulting shapelets are invariant to the length of the original boundaries. In this section we refer to the length of a series as m.

2.4.2 Candidate Extraction

Once all time-series have been transformed and normalised shapelet extraction can begin. Every subsequence (S) in every time-series (T) in the dataset (SET) is a potential shapelet and must be assessed. Each candidate (c) is extracted from the dataset in turn and its quality (Q) calculated and stored. Whilst each subsequence of

length $3 \leq l \leq S_m$ is a potential candidate, an exhaustive search of candidates of length in this range is often impractical due to high time-complexity. Lines et al. [54] suggests a method of reducing this range by setting minimum (min) and maximum (max) shapelet lengths to be assessed; we discuss this method later in this chapter.

A generic algorithm for extracting the best shapelet can be seen in Lines et al. [54], Ye and Keogh [99] and Hills et al. [40], the latter defines three main procedures in the generic extraction method: extraction of each candidate from the dataset; calculation of distance measurements between each shapelet and each time-series in the set (D_S); and a method of calculating class-separation or quality (Q) of each shapelet. The algorithm, taken from Hills et al. [40] is presented as Algorithm 1.

Algorithm 1 ShapeletSelection(**SET**, *min, max*)

```
1: best ← 0
2: bestShapelet ← ∅
3: for l ← min to max do
4:     C_l ← generateCandidates(SET, l)
5:     for all candidates c in C_l do
6:         D_S ← findDistances(S, SET)
7:         Q ← assessCandidate(c, D_S)
8:         if Q > best then
9:             best ← Q
10:            bestShapelet ← c
11:        end if
12:    end for
13: end for
14: return bestShapelet
```

Algorithm 1 generates candidates for each length (min to max) in turn, creating a set of candidates of length l (C_l). All candidates in C_l are then assessed before progressing to the next set of candidate lengths. The set of candidates of a length l

in a single series S_i is defined as $C_{i,l}$. Therefore the set of candidates of a given length in a set containing n series is defined as:

$$C_l = \{C_{1,l} \cup C_{2,l} \cup ... \cup C_{n,l}\} \tag{2.4.1}$$

The set of candidates of all lengths in the set is defined as:

$$C = \{C_{min} \cup C_{min+1} \cup ... \cup C_{max}\} \tag{2.4.2}$$

It is states that each series contains $(m - l) + 1$ candidates of length l [40, 54, 99]. Using this method the set C_l contains $n(m - l + 1)$ candidates of given length to be assessed; and each length in the range min to max will produce differing numbers of candidates. The total number of candidates to be assessed ($|C|$) where $min \geq 3$ and $max \leq m$ is calculated using:

$$|C| = \sum_{l=min}^{max} n(m - l + 1) \tag{2.4.3}$$

This declaration however does not allow for a closed curve series. When extracting candidates from a series produced from a closed boundary, the number of candidates of each length is simply equal to the size of the series itself. For this method the series must be treated as being 'circular'; treating the first element in the series as immediately following the last. Figure 2.21 shows an example of such circular candidate extraction. Using this method the set of candidates of length l for the dataset contains nm candidates. Using the closed series method the total number of candidates to be assessed is defined as:

$$|C| = nm(max - min + 1) \tag{2.4.4}$$

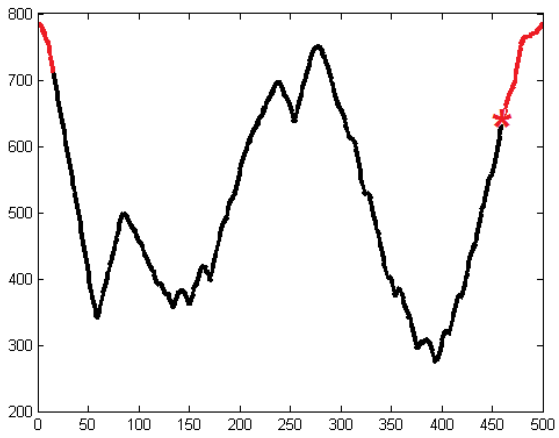


Figure 2.21: One-dimensional time-series created from a closed boundary. Marked on the curve is a potential shapelet using circular candidate extraction; its start point is noted on the curve.

In the generic method candidates are extracted from the entire set each size at a time: Each candidate of length *min* being extracted from all images/series; then each candidate of length *min+1* extracted from all images/series, etc. In practice however candidates of every size are extracted for each image/series in turn to allow for candidate pruning by series (discussed later in this section). Each time-series contains candidates (C_S) of lengths *min* to *max*. For a non-circular series S of length m there are a number of candidates for each series $|C_S|$ equal to:

$$|C_S| = \sum_{l=min}^{max} (m - l) + 1 \tag{2.4.5}$$

For circular series such as a one-dimensional 'distance to centroid' boundary shown in figure 2.21 the number of candidates for a given series is:

$$|C_S| = m(max - min + 1) \tag{2.4.6}$$

2.4.3 Distance Measurements

In order to assess each candidates quality we must measure its minimum distance to all series in the dataset, including the series from which it has been extracted

(although this is simply set as zero as an exact match will be found). With all series normalised for length each candidate must only be normalised for 'amplitude' to remove distance from centroid as a matching factor. To normalise candidates a standard z-normalisation method is used using the following formula:

$$c_i' = \frac{c_i - \mu}{\sigma} \tag{2.4.7}$$

Figure 2.22 shows a 1-D series from a hypothetical boundary. Marked on the curve are two sub-sequences that may be identified as candidates. The sub-sequences themselves are identical in shape; however, given their difference in amplitude or difference from the centroid (y-value on the figure) the two non-normalised sections would not be seen as similar. However also marked in the figure are the two candidates after z-normalisation. It can be seen from the figure that the candidates are indeed similar. To compare a candidate to a series it is compared to each subsequence of the same

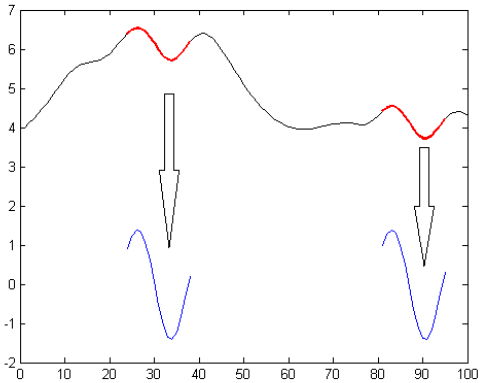


Figure 2.22: One-dimensional time-series created from a closed boundary. Marked in two positions on the curve are two identically shaped sub-sequences. Below are the two sub-sequences after z-normalisation

length from the target series. Each subsequence must be normalised in the same manner that the one being assessed. The Euclidean distance between the candidate (c)

and the subsequence (s) is then computed, where c_i denotes the ith element/value of the candidate:

$$dist(c, s) = \sum_{i=1}^{l} (c_i - s_i)^2 \tag{2.4.8}$$

Using this formula the distance between a candidate and a time-series (S_i) is calculated by taking the minimum distance between c and each of the sub-sequences in S:

$$d_{c,i} = \min_{s \epsilon C_{i,l}} dist(c, s) \tag{2.4.9}$$

We use this to calculate the distances between the candidate that we are assessing, and each other time-series in our dataset to generate a set of distance for the candidate. It is this list of distances that are used when calculating the quality (Q) of the candidate.

$$D_c = \{d_{c,1}, d_{c,2}, d_{c,3}, ..., d_{c,n}\} \tag{2.4.10}$$

Calculating the distances between a candidate and each other series carries a high time complexity $O(nml)$ where n is the number of instances in the set; m is the length of each series (uniform after length normalisation); and l is the length of the shapelet. In order to minimise the impact of this order magnitude we may implement early abandon techniques discussed later in this section.

Distance Pre-calculations

As each candidate is measured to each possible alignment of a series, distance measurements between points are duplicated. Consider two time series: S_1 and S_2 of length m and a designated shapelet length $l = 5$. Candidate c_1 is extracted from S_1 and contains values $\{S_1(1), S_1(2), S_1(3), S_1(4), S_1(5)\}$. This is then aligned with each $l = 5$ sub-series of S_2 and distance measurements between each aligned element are cumulated. After c_1 calculations are complete c_2 is extracted and contains elements $S_1(2 : 6)$, it is then aligned and calculated as c_1. Figure 2.23 shows the

first two candidates of $l = 5$ extracted from the first series, their alignments with the second series are shown. It can be seen that on the second alignment of c_2 the distance between several of the elements were calculated during the first alignment of c_1: $((S_1(2) - S_2(2))^2, \quad (S_1(3) - S_2(3))^2, \quad etc)$.

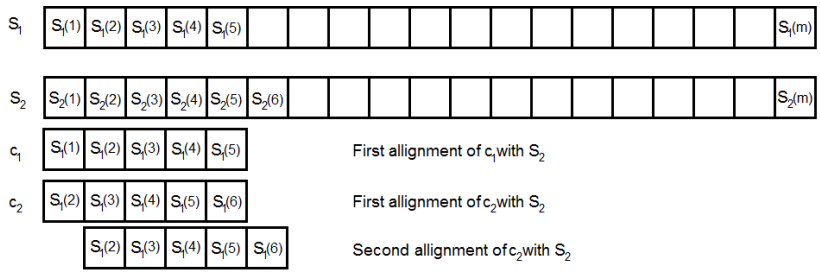


Figure 2.23: Showing alignments of two candidates extracted from S_1 and compared with S_2. Duplication of alignments can be seen; causing calculation redundancy.

2.4.4 Quality Measures

There are a number of quality assessment techniques discussed in recent literature. The quality of a given candidate is described as how well the classes of the dataset are separated by the distances from that candidate (D_c). Lines et al. [54] and Lines and Bagnall [53] investigate several methods of quality assessment for shapelet extraction. The later of these works compares additional methods to those used in previous work and concludes that these methods can offer time saving between 14 and 18% whilst maintaining classification accuracy. All four methods used are described in brief below.

Information Gain

The method first used by Ye and Keogh [99] and in further work [70] utilised information gain [87] to assess quality. Information gain methods require that the set of distances for the candidate be ordered by size, and the gain calculated given each possible split point. The split giving the highest information gain is described as the 'optimal split point' for that candidate, and its information gain is returned as a measure of its quality. Information gain is described as the entropy of a set of instances ($I(SET)$) minus the summed entropy of each subset after splitting ($\hat{I}(SET)$) [99]. Given that the proportion of class A in the original SET is $p(A)$ and the proportion of class B is $p(B)$:

$$I(SET) = -p(A)log(p(A)) - p(B)log(p(B)) \tag{2.4.11}$$

After splitting the set by each possible split point, each subsets' entropy must be calculated and weighted by its fraction of the set it was split from($f(A)$,$f(B)$). The entropy of each subset are then summed and returned as the entropy of set after splitting. Therefore the entropy of the set post split is shown as:

$$\hat{I}(SET) = f(S_1)I(S_1) + f(S_2)I(S_2) \tag{2.4.12}$$

Subtracting the entropy of the split from the original set gives us the information gained from that split; where S_1 and S_2 are subsets of SET created by the split (sp):

$$InformationGain(c, sp) = I(SET) - (f(S_1)I(S_1) + f(S_2)I(S_2)) \tag{2.4.13}$$

Figure 2.24 shows an example D_c after class instances are ordered by magnitude of distance. The set contains five instances of class A (circles) and seven of class B (squares). Also marked on the line are possible split points a to k. These are defined as the average points between consecutive instances on the line. Theoretically every

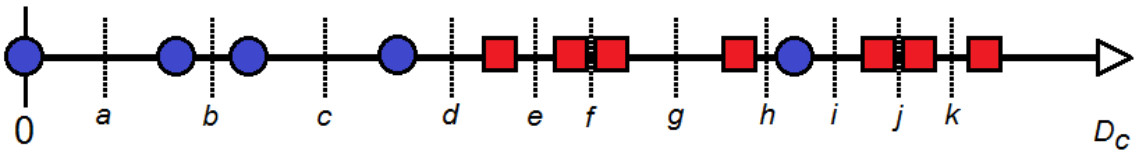


Figure 2.24: A set of class-headers in a theoretical two-class problem. Headers are ordered by distance from a hypothetical candidate c.

value on the number-line is a possible split point; however as all possible figures between two instance distances return the same gain, only one point between each is necessary. The following is an example of information gain calculation given split point d on D_c.

$$I(SET) = -p(A)\log(p(A)) - p(B)\log(p(B))$$

$$I(SET) = -(\tfrac{5}{12}\log(\tfrac{5}{12})) - (\tfrac{7}{12}\log(\tfrac{7}{12}))$$

$$I(SET) = 0.1584 + 0.1365 = 0.2949$$

$$\hat{I}(SET) = f(S_1)I(S_1) + f(S_2)I(S_2)$$

$$I(S_1) = -(\tfrac{4}{4}\log(\tfrac{4}{4})) - (\tfrac{0}{4}\log(\tfrac{0}{4}))$$

$$I(S_1) = 0 \ (0.log(0) \text{ set to zero})$$

$$I(S_2) = -(\tfrac{1}{8}\log(\tfrac{1}{8})) - (\tfrac{7}{8}\log(\tfrac{7}{8}))$$

$$I(S_2) = 0.1129 + 0.0507 = 0.1636$$

$$\therefore \hat{I}(SET) = (\tfrac{4}{12} \times 0) + (\tfrac{8}{12} \times 0.1636) = 0.1090$$

$$InformationGain(c, d) = I(SET) - \hat{I}(SET)$$

$$\therefore InformationGain(c, d) = 0.2949 - 0.1090 = 0.1859$$

To calculate the quality of any given candidate (c) we find the information gain by each possible split point on the line. The split which returns the highest information gain is the optimal split point, and the level of gain is returned as the quality of the candidate Q_c:

$$Q_c = \max_{sp \epsilon D_c} InformationGain(c, sp) \tag{2.4.14}$$

The information gain method of candidate assessment adds additional complexity to the shapelet extraction methods, of $O(n\ \log\ n)$. Hills et al. [40] states that whilst this overhead is minor compared to the inherent complexity of the shapelet methods, other methods of assessment can return comparable assessments whilst reducing complexity (of the assessment methods).

Kruskal-Wallis

First discussed in 1952 the Kruskal-Wallis test [47] is a non-parametric test used to determine whether multiple samples originate from the same distribution. It is an extension of the $Mann - Whitney\ U$ test [55] to allow testing with more than two samples; and a non-parametric equivalent of a one-way analysis of variance ($ANOVA$) that is used in many of the otolith studies referenced in this thesis.

As with the information gain method the distance measurements for the candidate (D_c) must first be sorted by magnitude. Each distance measure is then assigned a rank, and the ranks summed for each class in the set. However, unlike the information gain method, we then need only make one calculation in order to measure the candidates quality:

$$KW_c = \frac{12}{|D_c|\left(|D_c|+1\right)} \sum_{i=1}^{k} \frac{R_i}{n_i} - 3(|D_c|+1) \tag{2.4.15}$$

Where KW_c is the Kruskal-Wallis score for the candidate c; $|D_c|$ is the cardinality of D_c (in practice this is equal to the number of time-series in our set); k is the number of samples being tested (number of classes in the shapelet method); R_i and n_i are the sum of ranks for class i and the number of instances in that class.

It should be noted that whilst Kruskal-Wallis tests are normally used during hypothesis testing; candidate quality testing does not require results to be tested using critical value tables. Therefore this section has not discussed the use of α or degrees of freedom.

F-Statistic

Recent work by Lines et al. [54] uses the f-statistic of one-way ANOVA as a measure of quality. The f-stat measures the ratio of variance of within-class means to the mean of within-class variances. For this method of candidate quality assessment we need only sort the candidate distance measures D_c into separate sets $(D_i...D_C)$ by class label. We then compute the f-statistic (F) using the following formula:

$$F = \frac{\sum_{i=1}^{C}(\bar{D}_i - \bar{D})^2/(C-1)}{\sum_{i=1}^{C}(\sum_{d_j \epsilon D_i}(d_j - \bar{D}_i)^2/(n_i - 1))/C} \tag{2.4.16}$$

Where: C is the number of classes; $\bar{D}_i$ is the mean distance to the candidate for instances of class i; $\bar{D}$ is the overall mean; and n_i is the number of instances in class i.

A higher result using the above formula in indicative of higher ratio of inter-class variability to intra-class variability. A candidate of high quality shows smaller distances to instances of one class and larger distances to instances of other classes. Figure 2.25 shows two sets of hypothetical candidate distances; marked on the figure are each class' means (vertical lines) and variances with respect to the mean (horizontal). Marked at the crossing of these two descriptors is the class label (circle, triangle or square). The thicker vertical line shows the overall mean of the distances; two horizontal lines above the classes show variance of class means $v(M_{i \to C})$ (top)

and the the mean of within-class variances $M(v_{i \to C})$ (bottom) both with respect to the overall mean. It can be seen that the right-hand set has a far higher ratio of 'variance of means' to 'mean of variances'; it is therefore the higher quality shapelet of the two.

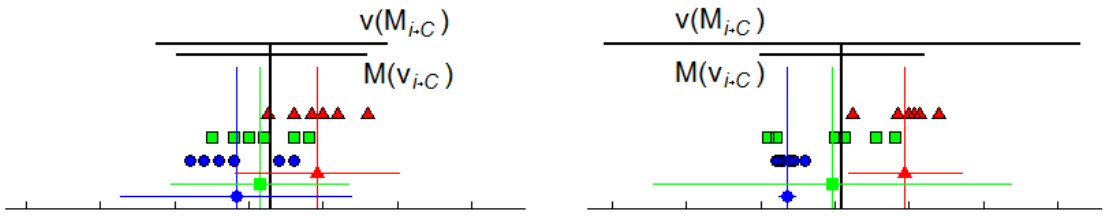


Figure 2.25: Two sets of distances to hypothetical candidates. Marked on each are three classes with associated means and variances (with respect to class mean). also marked are the overall mean; variance of class means (top bar) and mean of class variances (second from top)

Mood's Median

Mood's Median [68] is a non-parametric test to determine whether two sets of values come from the same distribution. Mood's Median requires only that we compute the median of the combined set of distances D_c rather than sort the values as done during Information-gain and Kruskal-Wallis test. Lines and Bagnall [53] and Hills et al. [40] describe the construction of a contingency table from D_c by noting how many instances from each class fall above and below (or equal to) the set median $(\tilde{D}_c)$.

Figure 2.26 shows distances of series to a hypothetical candidate. Contingency Table 2.3 is constructed by counting how many instance of each class fall above and below (or equal to) the set median. Once the contingency table has been constructed

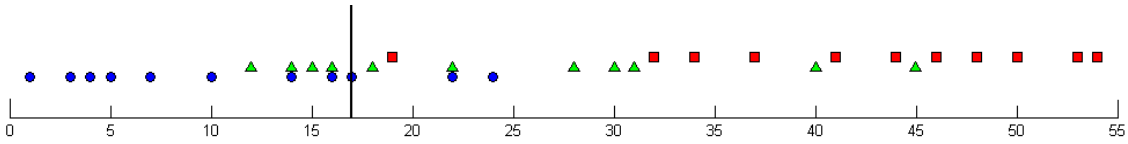


Figure 2.26: Distances from a candidate for instances from three classes shown on a distance line. Marked with a vertical line is the set median distance.

Table 2.3: Contingency table for Mood's Median test: Showing how many instance of each class in figure 2.26 fall above/below the set median

	○	△	□
$> \tilde{D}_c$	2	7	11
$\leq \tilde{D}_c$	9	4	0

the Chi-squared statistic of the table is calculated using:

$$\chi^2 = \sum_{c=1}^{C} \sum_{r=1}^{2} \frac{(o_{cr} - e_{cr})^2}{e_{cr}} \tag{2.4.17}$$

Where c is the indices of table columns; r indices of rows; o_{cr} is the observed count for element cr; and e_{cr} is the expected value of the element. The expected value is defined as:

$$\frac{(sum\ of\ column\ c)(sum\ of\ row\ r)}{total\ sum} \tag{2.4.18}$$

The Chi-squared value is returned from the process as a measure of the candidates quality. Lines and Bagnall [53] Found that using Mood's median offered a potential time-saving of around 18% over the original information-gain method. It was found that this method returned the lowest classification results of the tested methods; although the study did not find significant difference between the methods. It was a combination of Mood's Median and information-gain methods, however, that returned the best overall results.

2.4.5 Estimating min and max Shapelet lengths

Whilst the methods of Mueen et al. [70] do not set a minimum and maximum candidate size during shapelet extraction, Ye and Keogh [99],Hills et al. [40] and Lines

et al. [54] do set these parameters. Whilst Ye and Keogh [99] arbitrarily set *min* to 3 and *max* to the length of the series (m), the later studies implement a method of determining the parameters based on extraction of shapelets from randomly selected series in the dataset. In this method ten series are selected randomly and the top ten shapelets that classify that set are extracted using algorithm 4. This process is repeated tenfold and resulting shapelets from all ten folds are merged and sorted by length. The length of the 25th and 75th shapelets are then used as the *min* and *max* parameters for further shapelet extraction from the set. Algorithm 2 taken from Lines et al. [54] outlines this method of parameter setting.

Algorithm 2 MinMaxEstimation(**SET**)

```
1: shapelets ← ∅
2: for i ← 1 to 10 do
3:     randomiseOrder(SET)
4:     SET' ← SET(1 : 10)
5:     cShapelets = kShapeletSelection(SET', 1, m, 10)
6:     shapelets ← merge(shapelets, cShapelets)
7: end for
8: shapelets ← sortByLength(shapelets)
9: min ← min(shapelets_25.length, shapelets_75.length)
10: max ← max(shapelets_25.length, shapelets_75.length)
11: return min, max
```

2.4.6 Early Abandon Methods

There are two methods of early abandon of distance calculation mentioned in prior literature. These methods are concerned with halting calculations in order to reduce computation time and are described in the distance calculation description below. Additionally, Early entropy pruning allows candidates to be discarded during quality assessment using optimistic assumptions, and is described following the distance calculation methods.

Distance Calculations.

Already discussed in section 2.4.3 is the calculation of distance between a candidate c and a time-series s noted again here for convenience:

$$d_{c,i} = \min_{s \epsilon C_{i,l}} dist(c, s) \tag{2.4.19}$$

This method will indeed find the minimum distance; it potentially adds unneeded calculations during assessment of a candidate. Whilst we are unable to reduce the worst case complexity as each possible position along the series must be considered, an early abandon method for rejecting positions that exceed a previously calculated minimum are discussed in the literature [40, 99]. In addition to this distance abandon, Hills et al. [40] implements reordering and online-normalisation methods proposed previously [83] in order to reduce calculation times further. Algorithm 3 (taken from Hills et al. [40]) outlines the process of calculating the minimum distance with these measures included.

The early abandon methods add three time-saving measures: Firstly distance is calculated by summing the distance between elements in turn. This allows calculations for an alignment to be abandoned if the distance up to that point exceeds the current 'best match'. Secondly, The aligned portion of the time-series ($S_{i:i+l}$) is normalised one value at a time as the distance is calculated; reducing time spent normalising where alignments are abandoned. Finally the normalisation and subsequent distance calculations for each element are performed in order of magnitude (of the candidate). This final measure offers potential time savings as the element with the greatest magnitude has greatest potential for larger distances to paired element in the subsequence of the series.

Algorithm 3 Distance Measure with early abandon($S_{(1:m)}, c_{(1:l)}$)

1: $c' \leftarrow normalise(c, 1, l)$ {normalise the candidate}
2: $A \leftarrow sortIndexes(c')$ {A_i is the index of the i^{th} largest absolute value in c'}
3: $s \leftarrow normalise(S, 1, l)$
4: $p \leftarrow 0, q \leftarrow l$ {p stores running sum, q the running sum of squares}
5: $best \leftarrow dist(c, s)$ {find initial distance}
6: **for** $i \leftarrow 1 : m$ **do**
7: $\quad p \leftarrow p + S_{(i+l)} - S_i$ {updating running sums}
8: $\quad q \leftarrow q + S^2_{(i+l)} - S^2_i$
9: $\quad \bar{x} \leftarrow p/l$
10: $\quad t \leftarrow (q/l) - \bar{x}^2$
11: $\quad j \leftarrow 1, dist \leftarrow 0$
12: $\quad$ **while** $j \leq l \& dist < best$ **do**
13: $\quad\quad dist \leftarrow dist + (c_{A_j} - (S_{(i+A_j)} - \bar{x})/t)^2$
14: $\quad\quad j \leftarrow j + 1$
15: $\quad$ **end while**
16: $\quad$ **if** $j = l$ & $dist < best$ **then**
17: $\quad\quad best \leftarrow dist$
18: $\quad$ **end if**
19: **return** *best*
20: **end for**

Figure 2.27 shows an alignment between a candidate and a subsequence of a series. Shown on the figure are distances (vertical lines) between the two. These are cumulated in order until the total exceeds the best distance so far (the point at which distance lines cease). Figure 2.28 shows the same alignment, however in this figure the order of calculations is indexed by the magnitude of the candidate (the black line in the figures). Again these distances are marked on the figure up to the point at which the cumulative distance exceeds the best so far. It can be seen that in this simplified example, we calculate fewer than half the distances of the non-indexed measurements.

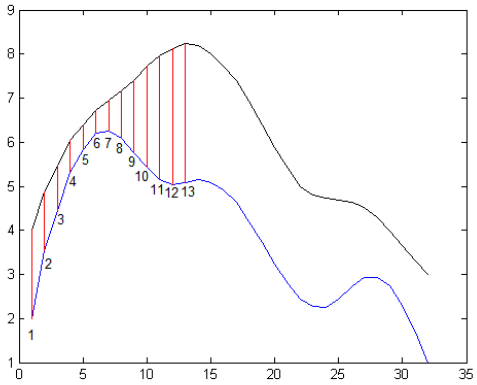


Figure 2.27: Distance calculations between a candidate and a subsequence of a series. Noted are the matching distances up to the point of abandonment.

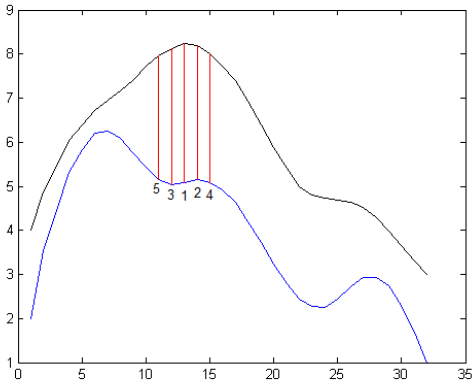


Figure 2.28: Distance calculations after re-indexing by magnitude of candidate elements. Fewer distances are calculated as total distance exceeds current best fit far sooner.

Early Entropy Pruning

In the original work by Ye and Keogh [99] information gain is used as the quality measure of a candidate. As such the study implements a method of pruning entropy calculations using optimistic prediction of information gain. After distance is calculated between a candidate and a series, its distance is then added to the distance line (as shown in 2.29). At this stage the method optimistically predicts the position/distances of the remaining series to the candidate and computes the information gain. In short this is predicted as all members of each class being at opposite ends of the distance line.

Placing unmeasured class instances at each end of the distance line allows the calculation of the best possible information gain (the upper bound) given the distances known. After each distance calculation, if maximum possible gain is lower that the k^{th} best candidate so far (where k is the number of shapelets to extract) then the candidate can be discarded. If the upper bound is in excess of the k^{th} best then

measurements continue until either the bound drops below the k^{th} (at which point we abandon the candidate) or all distances are calculated and the final gain produced.

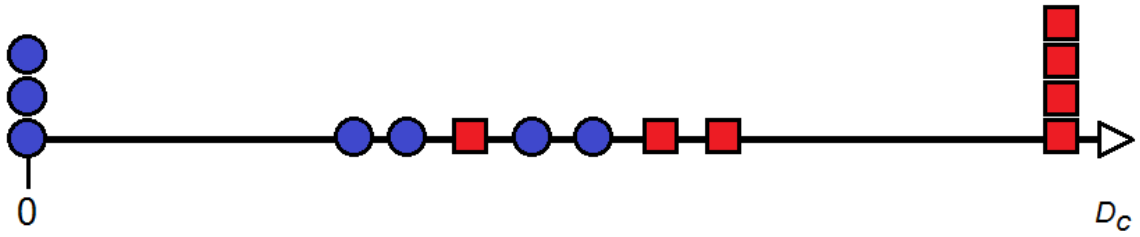


Figure 2.29: Distance line for instances of two classes from a candidate. Seven instances have been calculated and placed on the line, the remaining seven have been optimistically assumed to be at either end of the distance line.

In order to maximise the potential of this pruning, it is suggested that distances are calculated to instances of each class alternately. If distances are measured to all instances of one class then the next class etc., then the upper bound will remain at maximum until (at least) the first instance of the second class is inserted. Using this method it is possible to abandon calculations before the distance from the candidate to every series in computed.

2.4.7 Shapelet Pruning

Lines et al. [54] suggests that a number of shapelets can be discarded during the discovery stages, after assessment of all candidates from a time series. Once all candidates have been assessed they are ordered by quality. Shapelets are then discarded if they share indices with any shapelet of higher quality than themselves. Figure 2.30 shows a set of shapelets extracted from a hypothetical series and ordered by quality;

several of the shapelets are discarded due to lower quality than those they overlap. Once self-similar shapelets have been discarded in this manner the remaining shapelets are merged with any shapelets extracted for series prior.

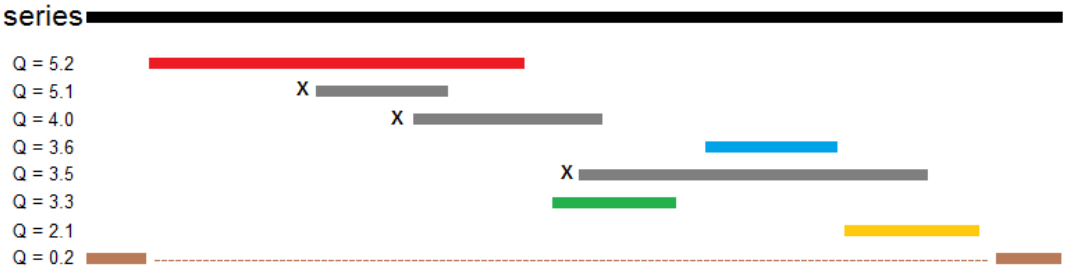


Figure 2.30: A time series and assessed shapelets from that series shown in order (top to bottom) of quality. Shapelets discarded using the self-similar discarding method are marked with a cross.

In theory the maximum number of shapelets that can be extracted from a set of series and kept during this process is equal to:

$$MaxShapelets = floor(n(m/min)) \tag{2.4.20}$$

Where n is the number of series; m is the length of each series and min is the minimum candidate length. This can quickly lead to a vast number of shapelets with which to transform series in the set by.

Whilst Hills et al. [40] limits the number of shapelets stored for use to $m/2$, Lines et al. [54] describes an alternate method for setting the number of shapelets (k) to store. This method estimates k by building classifiers (of which every type is to be used on the finally transformed data) using series transformed by varying numbers of shapelets (range 1-m) using 5-fold cross validation. The classifier that returns the best classification accuracies indicates the number of shapelets to store and transform series

by. This method attempts to limit shapelets in order to avoid over-fitting classifiers, whilst not limiting the number (of shapelets) so much as to reduce classification accuracies. Algorithm 4 describes the shapelet extraction process after the addition of the shapelet pruning methods.

Algorithm 4 kShapeletSelection(**SET**, *min, max, k*)

```
1: kShapelets ← ∅
2: for all time series S in SET do
3:     shapelets ← ∅
4:     for l ← min to max do
5:         C_l ← generateCandidates(S, l)
6:         for all candidates c in C_l do
7:             D_S ← findDistances(S, SET)
8:             Q ← assessCandidate(c, D_S)
9:             shapelets.add(c, Q)
10:        end for
11:    end for
12:    shapelets ← removeSelfSimilar(shapelets)
13:    kShapelets ← merge(kShapelets, shapelets)
14:    kShapelets ← sortByQuality(kShapelets)
15:    kShapelets ← kShapelets(1 : k)
16: end for
17: return kShapelets
```

2.4.8 Clustering of Shapelets

Hills et al. [40] outlines a method of clustering shapelets after extraction is complete. This measure is introduced as, whilst we remove overlapping shapelets on a series level, multiple series may hold shapelets that are very similar. Indeed this is indicative of a 'high-quality' shapelet, as a good candidate is likely to be similar to points on other time series, each of which may also have been selected as quality candidates. In theory two series may hold the exact same shapelet, both of which may be selected for the shapelet transform, so a method of clustering similar candidates is introduced.

A k by k distance matrix is formed by computing each shapelets distance to every other shapelet. This matrix is diagonally-reflective with diagonal holding zero-values (as each shapelet is a perfect match to itself). The minimally distanced pairs (excluding self-matches) are then extracted and clustered together. Hills et al. [40] repeats this process of matrix calculation until a 'user-specified' number of shapelets are formed. With each iteration, the size of the matrix is reduced as pairs of shapelets are removed and replaced by a cluster. The distance between two clusters is defined as the average distance between each member of each cluster. Hills et al. [40] then represents each cluster by the shapelet of highest quality within that cluster, others being matches of that shapelet.

This method of clustering can reduce the number of shapelets with which to transform further, and is likely to result in fewer than k shapelets. There is also no mention of a maximal distance under which to cluster shapelets. Rather, the clustering algorithm may cluster shapelets that, whilst the closest match in the set, are not similar in shape. The addition of a threshold for clustering by distance may be an important ammendment to this method.

2.4.9 Transforming Images

Once the final set of shapelets has been returned, they are used to transform the series in the dataset (or another dataset if shapelet extraction was cross-folded). Each series in the set is transformed by calculating its 'distance' from each of the k shapelets. The transformed series (S) will be a set of ordered distances to the set of shapelets (s); where n is the number of series transformed:

$$S'_1 = \{dist(S_1, s_1), dist(S_1, s_2), ..., dist(S_1, s_k)\}$$
$$S'_2 = \{dist(S_2, s_1), dist(S_2, s_2), ..., dist(S_2, s_k)\}$$

...

$$S'_n = \{dist(S_n, s_1), dist(S_n, s_2), ..., dist(S_n, s_k)\}$$

These transformed series can then be used to construct classifiers of choice. Each set of series to be classified using the shapelet method must then be transformed using the same shapelet set to be classified using this system.

2.5 Shape Indices

Perhaps the simplest methods used during this work are those regarding general shape of otoliths [43, chap. 2]. A number of metrics can be calculated using Matlab [62] as part of the 'regionprops' function, and a number of these are used in different studies either discretely or in addition to other measures. Burke et al. [22] for example measures area, perimeter and otolith length and width and additionally calculates further shape indices such as rectangularity, form-factor etc, used in addition to elliptical Fourier descriptors; the CSS methods described in Section 2.3 are complimented with *circularity* and *eccentricity*, used as pre-indexing metrics.

Whilst the orientation of each otolith is not to be considered as indicative of class for any tests done throughout this study, orientation is nonetheless calculated and used as a method of normalising each object for rotation. Each otolith mask has its orientation calculated (as described below) and is rotated so that its new position reduces the angle of orientation to zero, or as close as possible given the pixelation. All other measurements are taken whilst in the normalised position. A number of the measurements (*solidity, extent, eccentricity, circularity*) are scale-invariant.

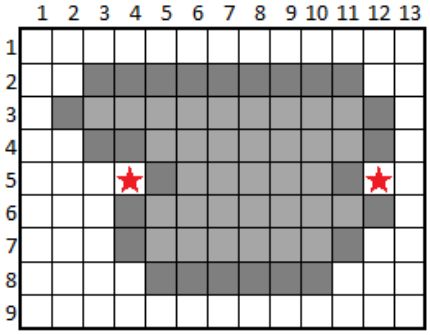


Figure 2.31: Showing the perimeter (darker shading) and area (total shaded area) of our previous example. Also showing addition of pixels (starred) to construct the convex hull of the mask.

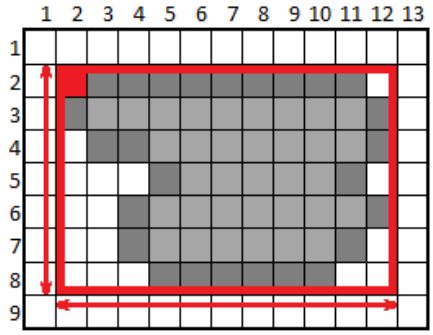


Figure 2.32: Showing the coordinates calculated for the bounding box (red), which is used to calculate the *height* and *width* of the segmented area.

Perimiter Length - Simple measurement of the length of the otolith boundary, measured in pixels. Equal to the length of the chain code used for Fourier transformation, and to the length of x/y coordinate arrays and univariate series arrays described in Section 2.1. Shown as darker grey pixels in Figure 2.31 ($perimeter = 26$).

Area - The area of the otolith mask, measured in pixels. Calculated using the mask created during the segmentation process. Shown as total shaded area in Figure 2.31 ($area = 60$).

Convex Area - The area of the convex hull of the segmented mask. To construct the convex hull, the boundary is extended so that there are no concavities around the contour. Shown in Figure 2.31 as the addition of areas marked with a red star. ($convexArea = 62$).

Solidity - Calculated as the *area* of the segmented mask, divided by the resulting *convex area*; Given as a scale invariant index. In our example shown this is $solidity = 60/62 = 0.9677$ (*4dcp*).

Height and **Width** - taken from the 'bounding box'. The bounding box is calculated by the Matlab function as a box that will totally encompass the segmented area. Recorded as the upper leftmost coordinate minus 1/2, and the height and width of the box measured in pixels. Figure 2.32 shows our example with the vertex pixel shown in red at coordinates *(2,2)*. The height and width of the segmented area (in pixels) is shown by the red bounded area in the image. In our example the top-left vertex is given as $(1.5, 1.5)$ the height as 7 and the width as 11. The bounding box therefore encompasses any pixels with in both ranges $1.5 < x < 12.5$ AND $1.5 < y < 8.5$.

Extent - A scale invariant index calculated as the division of the segmented area by the area of the bounding box ($extent = area/(height * width)$). A measure of how rectangular an area is; in our example $extent = 60/77 = 0.7792$ (*4dcp*).

Axis Lengths - Are calculated from the normalised second order moments of the segmented area. The covariance matrix for the x and y coordinates of foreground pixels is first calculated. Minor adjustments are made to the variances of x/y to account for the pixelated nature of the mask. A modified Eigenvalue calculation is used to return the major and minor axes lengths. Where V_x is the variance of x and C_{xy} is the covariance of x, y:

$$2\sqrt{2}\sqrt{V_x + V_y \pm \sqrt{(V_x - V_y)^2 - 4C_{xy}}} \tag{2.5.1}$$

or, where E is the Eigenvalue Matrix, axis lengths are calculated using:

$$\mathbf{AxisLengths} = 2\sqrt{4\mathbf{E}} \tag{2.5.2}$$

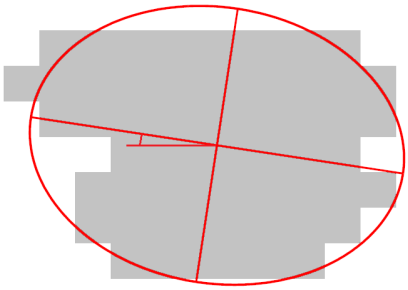


Figure 2.33: Showing the major and minor axis of the ellipse. Also shown is the orientation given as an angle of the major axis to the horizontal.

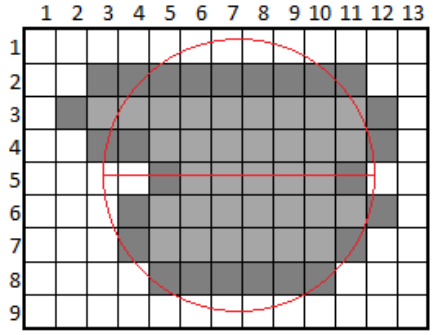


Figure 2.34: Showing a circle of equal area to the segmentation, with common centroid. Line shows the measured diameter.

Figure 2.33 shows the calculated ellipse with major and minor axes marked. It is the length of these axes that are recorded. Note that not all foreground points fall within the ellipsoid. The methods used were described by Haralock and Shapiro [39] and whilst the extensive proof in that work is not shown here, the Matlab functions used to calculate axis lengths use the resulting formulae for axis lengths.

Orientation - While not recorded for use in classification (as object orientation should not be indicative) the orientation of the segmented area is used to normalise the mask. The angle of orientation is given as the angle between the major axis of the foreground ellipse and the horizontal axis. After calculation each image is rotated by the appropriate amount to align the major ellipse with the horizontal axis.

Eccentricity - scale invariant index calculated using the major and minor axis lengths. Where L_M and L_m are the major and minor axis lengths:

$$eccentricity = 2\frac{\sqrt{(\frac{L_M}{2})^2 - (\frac{L_m}{2})^2}}{L_M} \tag{2.5.3}$$

Equivalent Diameter - Calculated as the diameter of a circle (in pixels) with the same total area as the segmented area. $EqDiameter = 2\sqrt{area/\pi}$. In our example where $area = 60$, the $eqDiameter = 2\sqrt{60/\pi} = 8.74$ ($2dcp$). Figure 2.34 shows our original segmentation, with a circle with same total area overlayed, and aligned with a common centroid.

Circularity - Whilst not calculated with the 'regionprops' function, circularity is one of the metrics used for pre-indexing CSS classification. It is a similar metric to 'extent' in that it is the division of the object area by the area of a circle which could encompass the object. Where L_M is the major axis length calculated previously: $circularity = area/\pi L_M^2$

2.6 Classification

We anticipate that many readers will have read an introductory text (eg. Alpaydin [4]) and have a broad understanding of the machine learning and transformation principles we use. Nevertheless, in-depth knowledge of the methods is not required with regards to learning algorithms used in this work. Each of the classification algorithms we employ are used 'off-the-shelf' as part of the WEKA machine learning suite. With the exception of cross validation parameters (to ensure validation is uniform across tests), all parameters used remain as the default for each respective learning algorithm.

2.6.1 WEKA

The Waikato Environment for Knowledge Analysis (WEKA, [38]) is a freely available library of machine learning tools which are widely adopted by the machine learning community, and has been used for otolith classification [40]. The library holds many statistical and modelling tools together with learning algorithms. WEKA provides a

framework allowing the construction of large numbers of classifiers or models, using different learning algorithms that can be used alone or can be integrated into other environments (or other software packages). Whilst WEKA has an application that can be used on a stand-alone basis, during the entirety of this study we interface WEKA with MATLAB in order to automate the procedure. For the studies in this work we make extensive use of WEKA's library to build multiple classifiers of otolith shape, using the learning algorithms listed below:

2.6.2 Classification Algorithms

- NB → Naïve Bayes [51]
- BN → Bayesian Networks [33]
- Log → Logistic [50]
- HP → HyperPipes [97]
- J48 → J48/C4.5 [82]
- RaF → Random Forest [17]
- IBk → k-Nearest Neighbours [3]
- SMO → Support Vector machine [80]
- RoF → Rotation Forest [86]
- NNDTW → 1-Nearest Neighbour with dynamic time warping distance [84]

Classification algorithms are further grouped according to similarity of methods used to construct classifiers. The algorithms above were chosen to cover a range of the classifier groups: Both Naïve Bayes and Bayesian Networks algorithms belong

to the *Bayesian* group of classifiers, which construct classifiers by explicitly applying Bayes' Theorem. Logistic Regression belongs to the *regression* group which iteratively refines the classifier measures of error in predictions made by the classifier. Hyperpipes and k-Nearest Neighbour algorithms are *instance based methods*, also called *memory-based* as they compare new instances to those within the training data to calculate similarities with 'previous' cases. J48, Random forest and Rotation Forest are *decision tree* algorithms. Each constructs classifiers that make series of 'decisions' based on new instances' variables (harmonic coefficients, CSS maxima etc.). Additionally, random and rotation forest algorithms also qualify as *Ensemble Methods*, where multiple (potentially) weaker classifiers are used independently, and where a further decision based on their predictions is used to classify the new instance. Finally Support Vector Machines are a *kernel method* where input data is used to create a higher dimensional information space which is then used to classify new instances.

2.7 Summary of Methods

A summary of the methods described in this chapter is shown here. A brief description of each section is given, alongside the section number (in bold) where that topic can be found.

- Otolith segmentation and boundary extraction methods commonly used in the field of otolith analysis **2.1**:
    - Hand tracing methods carried out by expert readers to create binary masks **2.1.1**.
    - How intensity thresholding can be used to create binary masks from otolith images **2.1.2**.
    - Extraction and rotation of otolith boundaries from masks **2.1.3**.

- The focus of this chapter has been the methods by which extracted boundaries may be transformed before use in building classifiers:
    - Elliptical Fourier analysis; the method widely used in the otolith industry **2.2**,
    - Curvature Scale Space; a computer vision method designed to describe the shape of objects in images **2.3**.
    - Time-series Shapelets; a newer method from the field of time-series analysis that has performed well in previous studies **2.4**.
    - How we use basic shape and scale measurements to form a feature-set suitable for building and testing classifiers **2.5**.
- The classification techniques used through this work and the commonly used machine learning suite we employ **2.6**.

Chapter 3

Comparison of Recent Boundary Encoding Techniques to Common Industry Methods

This chapter presents a study comparing three existing contour representation methods with a view to discriminating between two populations/stocks of Herring (Thames and North-Sea) using boundary contours extracted from their sagittal otoliths. We compare elliptical Fourier, Curvature scale space (CSS) and Shapelet transform methods in this chapter, to determine whether the commonly used method of otolith boundary description (Fourier harmonics) can be surpassed by computer vision standards and recent additions to time-series analysis (Objective 1)

Specimens used in this study were prepared by the Centre for Environment, Fisheries and Aquaculture Science (Cefas) and donated for use in this study. The collection comprises 128 samples from two distinct populations of Herring: North-Sea (51) and Thames (77) and have been hand labelled by experts at Cefas Lowestoft. The images obtained for use in this study were received with only population information.

Both Shapelet and Fourier transformed contours are classified using a number of algorithms provided by the Waikato Environment for Knowledge Analysis (WEKA):

Bayesian networks, Naïve-Bayes, C4.5 (J48), Random and Rotation Forest. We evaluate all classification methods using leave-one-out cross validation (LOOCV) during classification, with Shapelet parameter selection being performed within an additional cross-fold (double-cross validation). Boundaries transformed using the Curvature Scale Space method are combined with morphological measures of their boundaries and CSS image: eccentricity, circularity and aspect-ratio. They are classified using nearest-neighbour selections where distance is calculated using the CSS matching algorithm described as part of the MPEG7 standard. Additionally we also classify the maxima extracted from the CSS image using the cohort of WEKA learning algorithms listed above, in order to compare to CCS own matching algorithm. During this classification pairs are ordered both by boundary-point and by curvature intensity.

Boundary data are also transformed into individual univariate series (distance to centroid) and classified using using nearest neighbour dynamic time warping. Use of both univariate series and shapelets allows comparison of simple (univariate) and complex (shaplets) transforms of the same boundary data. We show that while shapelets are a new addition to contour classification methods, neither they nor MPEG7 standard CSS methods surpass Fourier based classification in this particular problem.

3.1 Materials and Methods

3.1.1 Sampling

With the exception of 1990 and 2007, Cefas has carried out trawl surveys in the Thames estuary and Rivers Crouch, Blackwater, Roach and Medway every year since 1989. The study area is 51° 20' N to 51° 48' N; 0° 10' W to 1° 06' W in ICES (International Centre for the Exploration of the Seas) division IVC. Surveys are carried out with the primary aim of calculating the relative abundance and distribution of

Herring in those areas. Surveys are always carried out in November, falling at the beginning of the commercial fishing season (for Herring) and the same commercial fishing vessel (MFV Ina K) has been used for every survey carried out during this period. Surveys have all used trawl nets of a similar design, Larson sprat trawl with steel doors and 16mm cod end towed at a speed of 3Kts. Herring otoliths are taken in accordance with two tallies, inside and outside the drift area. The targets for each of these areas is 480 fish, 2 per half cm up to a length of 14.5cm, 5 per half cm between lengths 15cm and 19.5cm and 10 per half cm at lengths 20cm and over. The length, sex and maturity stage of each fish is recorded, however this information was not required during this study and so was withheld at the time. For the present study, we used otoliths from fish caught on the survey in 2008 and 2009. Samples were chosen from each year to provide as many numbers of Thames and North Sea individuals from each age group 0-3 as possible (77 Thames, 51 North-Sea).

3.1.2 Image Capture

The otoliths were set whole in clear polyester resin, under a glass cover slip in black trays. These were then read for the purpose of age-determination and stock discrimination by expert readers at Cefas using a stereo microscope with reflected light so the annuli of the otolith could be clearly observed. Samples were individually inspected under magnification to determine whether they were suitable for use in this study. For the purpose of this research it was decided that only the left sagittal otolith would be examined and that to be accepted the otolith must be undamaged and have a boundary that was free from obstruction by debris or air-bubbles set in the clear resin mount. Once an otolith was identified as a suitable candidate it was imaged as 2.5x magnification using a Jenoptik C5 digital camera mounted on a Leica

DM2000 microscope and connected to a PC. All images were captured using ProgRes CapturePro 2.7 software and stored as uncompressed TIFF files for later processing.

Otoliths were imaged at the same approximate orientation and under the same magnification through the image-capture process. Minor adjustments were made to the microscope focus and the angle of reflected light in order to achieve the best possible clarity of the otolith edge. Otolith images were capture at two levels of exposure. A low exposure image of each was captured in order to allow the annuli to be more easily examined for ageing; a high exposure image of each was captured so that the boundary contour could be more easily determined by automatic image analysis.

After Image capture the number of samples per class was unbalanced. Fifty samples from each class were selected for further processing based on image quality and clarity of the boundary. Table 3.1 shows the number of otoliths of each age class in this set. Whilst age of sample was not a consideration when selecting otoliths for this comparative study, the table shows that age distribution is reasonably balanced.

Approximate Age	Count (North-Sea)	Count (Thames)
0	4	13
1	22	17
2	15	15
3	8	5
4	1	0

Table 3.1: Showing approximate age distribution (by class) of otolith samples used in this study.

3.1.3 Outlining

The majority of images collected could be outlined using thresholding methods discussed in chapter 2 section 2.1. However, a number of otolith images contained anomalies that, whilst not occluding the boundary, prohibited threshold based segmentation. The set was initially segmented automatically and failed attempts were detected by manual visual sweep of the results. Incorrect segmentation was corrected by expert boundary tracing and all resulting images were additionally checked by a second expert.

3.1.4 Fourier Transform

We use a modified version of MATLAB scripts available on the file exchange website [56] that implement the Fourier methods described in chapter 2, section 2.2. The implementation is used to calculate elliptical Fourier coefficients for n harmonics: a_n, b_n, c_n, d_n. For the study presented in this chapter we extract and record the first one-hundred fourier harmonics from the boundary. Boundaries reconstructed using the first fifty harmonics, when compared to the original boundary, shown a mean pixel disparity < 1; reconstructions using all one-hundred harmonics have a maximum Hausdorff distance < 2. It was deemed that further harmonic extractions were not required as reconstruction using up to one-thousand harmonics did not reduce the Hausdorff distance to less than one. Classification accuracies are recorded using the first 5, 10, 20, 30, 50 and 100 harmonics; both prior to and after scale normalisation.

All boundaries were normalised for rotation prior to harmonic calculation by calculating the angle between the major elliptical axis of the boundary and the horizontal axis. The boundary images were then rotated so that their major axes were parallel

with each other and with the horizontal axis. After extraction, all harmonics were normalised for scale using the procedure laid out in chapter 2.2.

3.1.5 Curvature Scale Space

We follow the general procedure described in chapter 2 section 2.3 using a custom built MATLAB implementation. In order to test our implementation of the CSS algorithms we use the SHAPE dataset [69], which is available on-line and required no image capture prior to use in this study. The SHAPE database comprises 70 classes of shape with 20 images in each class, totalling 1400 binary shape images. The dataset was constructed in order to test shape based recognition/classification systems and is used to benchmark methods such as curvature scale space. It contains classes such as: *bird*, *chicken*, *dog*, *deer*; which are included specifically for their general shape similarity. The set also includes simpler object shapes eg *fork* and a number of abstract shapes. For use in this study we remove a small selection of classes: all *device* classes are removed as they hold deliberate within class varience; we also remove the *spring* class as the shape itself (a coil) takes a prohibitively long time to process using CSS methods.

As in the work on fish contour classification by Abbasi et al. [1] we initialise the smoothing kernal width (σ) at 1, increasing by 0.1 after each iteration or *evolution*. After each boundary has been transformed to a set of CSS-maxima, we discard any maxima with an evolution or magnitude of 1. This removes possible maxima noise caused by pixel discretisation of a continuous curve.

For use in pre-indexing (Described in Section 2.3.3), we record each boundaries eccentricity, circularity, and calculate the aspect-ratio of the constructed CSS image,

which is a good indicator of the maximum concavity in the boundary. These global parameters are recorded alongside each set of CSS maxima for each sample.

3.1.6 Shapelets

For this study we use the implementation of Lines et al. [54] to perform shapelet transformation. This implementation utilises information gain to test the discriminatory power of each shapelet between the lengths of twenty and forty, and selects the best one-hundred shapelets for the transformation process.

Shapelet extraction is performed behind a second cross-validation 'layer' in order to minimise over-fitting to the dataset. The data is split into five equal subsets (ten instances of each class per set) and each fold in turn is used to calculate the shapelets by which to transform the remaining (combined) sets.

3.1.7 Classification

We use *Leave one out cross validation* for classification using each of our transform methods. Each transformed instance is withheld in turn to create a single sample test 'set', the remaining instances form the training set. Once all instances are classified in this manner the average result is returned as the classification accuracy for each process.

Curvature Scale Space

For curvature scale space we use the matching process detailed in chapter 2 to calculate *dissimilarities* between each test boundary and each instance in the training set. Alignments were carried out with maxima with magnitude $> 80\%$ of the maximum as in previous literature. However, due to the variability of the maximum evolution

magnitude, the maximum search distance was defined at 40%, the maximum error between primary maxima across all samples.

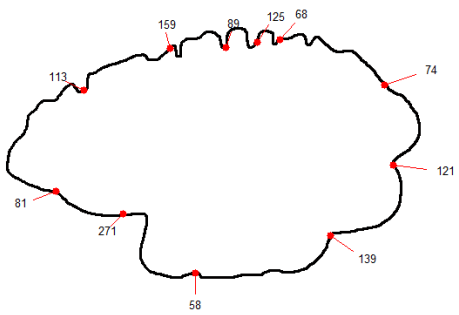
As recommended by Abbasi et al. [1] pre-indexing was carried out using the recorded eccentricity, circularity and aspect-ratio. Multiple tests were carried out with the threshold T set between 0-1 (inclusive) at intervals of 0.05. It was originally hoped that T may be determined using the SHAPE dataset results and that peak-performance T would be transferable to the Herring set. The results show however that peak-performance threshold values are significantly different between the two test sets, and use of the best threshold from one set would result in significant under-performance when indexing the other.

Otolith images were all captured with the same side of the otolith uppermost, therefore it was not necessary to perform maxima matching for mirror images when processing the otolith dataset. Conversely the SHAPE set does contain deliberate mirror images, and so the implementation used for both SHAPE and Herring testing does perform mirror matching.

Once dissimilarity is measured between a test instance and each in the training set; classification is carried out using Nearest-Neighbour (NN) selection. The result is returned as a percentage of selections that resulted in the correct class. We show results using 1, 3 and 5-NN classification.

Additionally, CSS maxima are classified using each of the classification algorithms listed in section 3.1.7 in order to compare stock algorithms with the MPREG7 matching algorithm. Four datasets are created for this testing; *bou*, *bouG*, *evo* and *evoG*. *bou* has the maxima coordinates sorted by distance along the boundary; *bouG* is the

maxima ordered as pairs in the same order as *bou* but prepended with eccentricity, circularity and aspect-ratio (the global parameters). *evo* is sorted by evolution magnitude and *evoG* is as evo prepended with the global parameters. A Simplified example of this ordering can be seen in Figure 3.1. Where the CSS matching implementation processes all maxima in the CSS images, only the eighteen maxima with greatest evolution are entered into the CSS datasets as this is the minimum number of maxima in all images. This truncation of data causes on average 4.6 maxima to be dropped from each transformed boundary, averaging 10.4 evolution magnitude per cut maxima.



Set of maxima coordinates
(27,81), (57,271), (108,58), (169,139), (218,121), (265,74), (323,68), (340,125), (365,89), (401,159), (447,113) (89,93)
method: bou **27**, 81, **57**, 271, **108**, 58, **169**, 139 ... method: evo 57, **271**, 401, **159**, 169, **139**, 340, **125**...

Figure 3.1: Left: Image showing the CSS feature points extracted; marked as points along boundary, and annotated with evolution (curvature) magnitude. Right: Showing the CSS feature pairs (point along boundary, evolution magnitude) for the boundary (*top*), and the ordering of the pairs using ***bou****ndary order*, and ***evo****lution magnitude order* methods (*bottom*). Only the ten points with largest evolution magnitude are shown in this example.

Fourier, Shapelet and Univariate Boundaries

The Univariate Boundaries (UV-Bs) and Shapelet transform thereof, and the Fourier transformed data are classified using a variety of available classifiers so that results of transformed and non-transformed boundaries can be compared. Results of these tests can be seen in Section 3.2. The classifiers used on these sets are implemented in WEKA and are;

- BN - Bayesian networks
- NB - Naïve Bayes
- C4.5 - C4.5 Decision Tree
- RaF - Random Forest ensemble
- RoF - Rotation Forest ensemble
- NNDTW - 1-Nearest Neighbour with dynamic time warping distance, performed on UV-B data only.

3.2 Results

3.2.1 CSS Matching Algorithm

Results of classification using MPEG7 CSS matching implementation can be seen in Figures 3.2 and 3.3. Results shown are for eccentricity, circularity and aspect-ratio thresholding in the range 0.00 to 0.40 only, as results in the range 0.50 to 1.00 show no improvement on $T = 0.40$ in either test. The thresholds returning peak performance in these cases are seen to be different to one another, and use of one problem's peak threshold value for the other would result in significant under-performance.

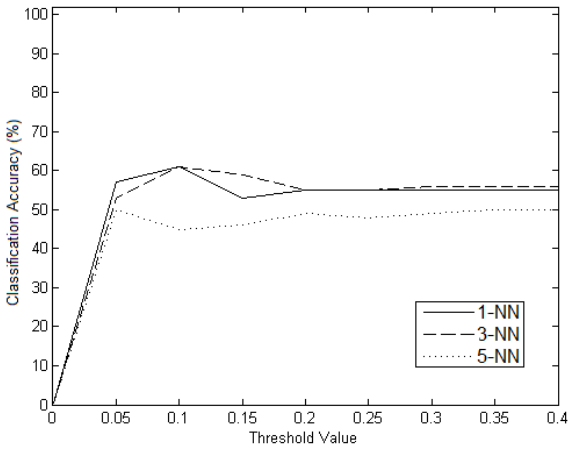


Figure 3.2: Results of North-Sea/Thames Herring classification using 1, 3 and 5-NN. Showing classification accuracies (y-axis) for varying over varying threshold values (x-axis).

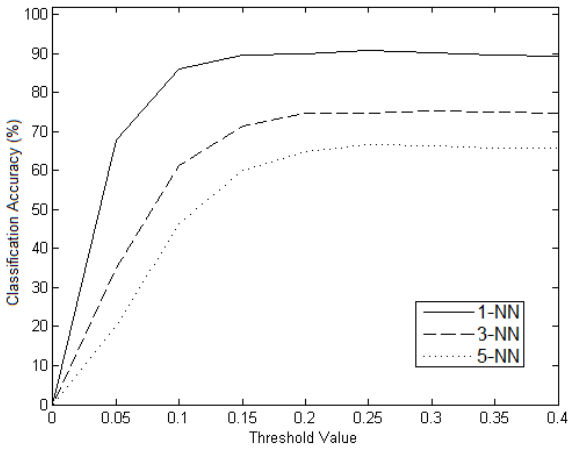


Figure 3.3: Results of SHAPE database image classification using 1, 3 and 5-NN. Showing classification accuracies (y-axis) for varying over varying threshold values (x-axis).

In all cases the inclusion of a threshold for the global parameters improve, or at least does not hinder, accuracy. Table 3.2 shows results of the two classifications using no threshold (i.e. *T=1*), and using the peak performance global threshold (value for each case given in table).

		1-NN	3-NN	5-NN
NSvTh	T = 1	55%	56%	50%
	peak T = 0.01	61%	61%	45%
SHAPE	T = 1	87%	69%	59%
	peak T = 0.25	91%	74%	67%

Table 3.2: The LOOCV classification accuracies using our CSS matching implementation on all three tests, North-Sea Vs Thames (NSvTh), and Classes from the SHAPE database (SHAPE)

Tables 3.3 and 3.4 show sensitivity and specificity (or Sensitivity to North-Sea and Sensitivity to Thames) of the CSS matching algorithm. The figures show that the CSS matching technique is generally more sensitive to Thames Herring instances than to those from the North Sea. Table 3.3 shows results using peak performance threshold; Table 3.4 shows those where no threshold of global parameters was used for pre-indexing. The increase in accuracy, sensitivity (North-Sea classification accuracy) and specificity (Thames classification accuracy) when global thresholds are used for CSS-image pre-indexing, supports the idea that the two classes of boundary have significant overlap in scale space, where boundaries that show differences while unprocessed may have much the same representation after processing.

T = 0.10		North Sea	Thames	
1-NN	North-Sea	26	24	Sensitivity: 0.52
	Thames	15	35	Specificity: 0.70
3-NN	North-Sea	29	21	Sensitivity: 0.58
	Thames	18	32	Specificity: 0.64
5-NN	North-Sea	24	26	Sensitivity: 0.48
	Thames	29	21	Specificity: 0.42

Table 3.3: Confusion matrices including sensitivity and specificity for NSvTh classification using 1,3 and 5-NN selection (Rows - Query, Columns - Result), using peak T value $T = 0.10$

3.2.2 WEKA Classification

Table 3.5 shows the classification results for Thames/North-Sea Herring stock separation using WEKA classification suite for shapelet and CSS transformed data (bou, bouG, evo, evoG). Whilst it can be seen that the overall best performing classification algorithm was Random Forest at 68% when classifying univariate boundaries, significance testing using N-way analysis of variance tests (ANOVAN [7]) showed

T = 1.00		North Sea	Thames	
1-NN	North-Sea	21	29	Sensitivity: 0.42
	Thames	16	34	Specificity: 0.68
3-NN	North-Sea	25	25	Sensitivity: 0.50
	Thames	19	31	Specificity: 0.62
5-NN	North-Sea	25	25	Sensitivity: 0.50
	Thames	25	25	Specificity: 0.50

Table 3.4: Confusion matrices including sensitivity and Specificity for NSvTh classification using 1,3 and 5-NN selection (Rows - Query, Columns - Result), with no preindexing using global parameters ($T = 1.00$)

no marked difference between transform methods (Table 3.6). Tests do show, however, that algorithm selection is an important consideration in this task. Post-hoc ANOVAN testing (Figure 3.4) shows that, whilst not significantly different to Naïve Bayes, Random Forest is the best performing algorithm in these tests. It is interesting to note that whilst shapelet methods have been shown to be a promising tool [40], in this application they afford results for the most part lower than classification of the unprocessed boundary (UV-B) using the same classification suite. However, post-hoc tests show no significant difference between boundary transform methods (Figure 3.5).

Results from the Fourier classification can be seen in Tables 3.7 and 3.8. For normalised and non-normalised transforms, results appear to peak where the first twenty Fourier harmonics were used for classification, at 73% on average across algorithms when using non-normalised harmonics. For scale normalised harmonics, however, classification using higher harmonic content gave a general improvement of results. Variance tests show no significant difference either for classification algorithm (Table 3.9) or for the majority of harmonic tests. Note that whilst the standard variance testing shows significant difference between harmonic content results (in Table 3.9),

Classifier	UV-B	Shapelet	bou	bouG	evo	evoG
BN	58%	62%	49%	53%	56%	60%
NB	63%	65%	67%	64%	59%	55%
C4.5	56%	55%	59%	50%	52%	51%
RaF	68%	58%	63%	54%	61%	64%
RoF	61%	58%	52%	62%	64%	59%
Average	61%	60%	58%	57%	58%	58%
NN - DTW	65%	N/A	N/A	N/A	N/A	N/A

Table 3.5: The LOOCV classification accuracies of a range of algorithms using Univariate Boundary data (*UV-B*), Shapelet transformed data (*Shapelet*) and four CSS maxima sets: sorted by point along boundary (*bou*); as *bou* but including global parameters (*bouG*); sorted by evolution magnitude (*evo*); as *evo* but including global parameters (*evoG*).

	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
algorithm	291.53	4.00	72.88	3.54	0.02
transform	64.00	5.00	12.80	0.62	0.68
Error	411.67	50.00	20.58		
Total	767.20	29.00			

Table 3.6: Results of N-way analysis of variance (ANOVAN) of algorithm and transform selection results.

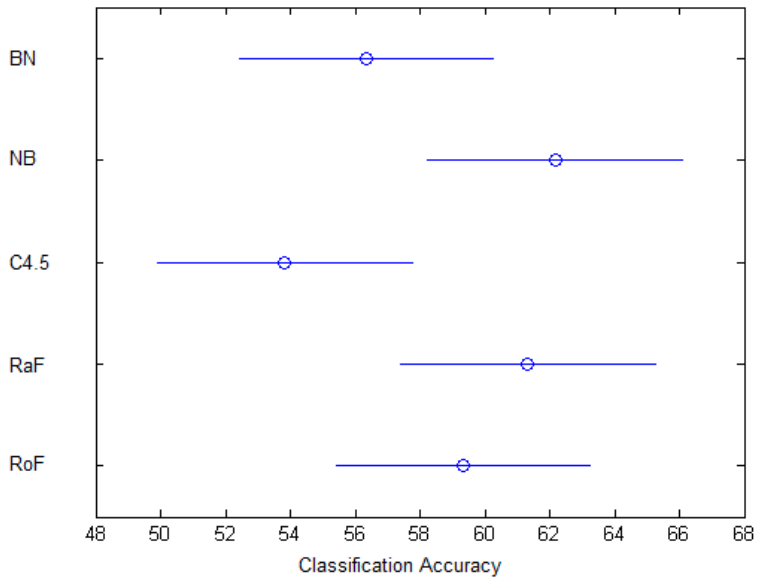


Figure 3.4: Results of post-hoc ANOVAN testing showing classification accuracies for Herring stock separation given different classification algorithms.

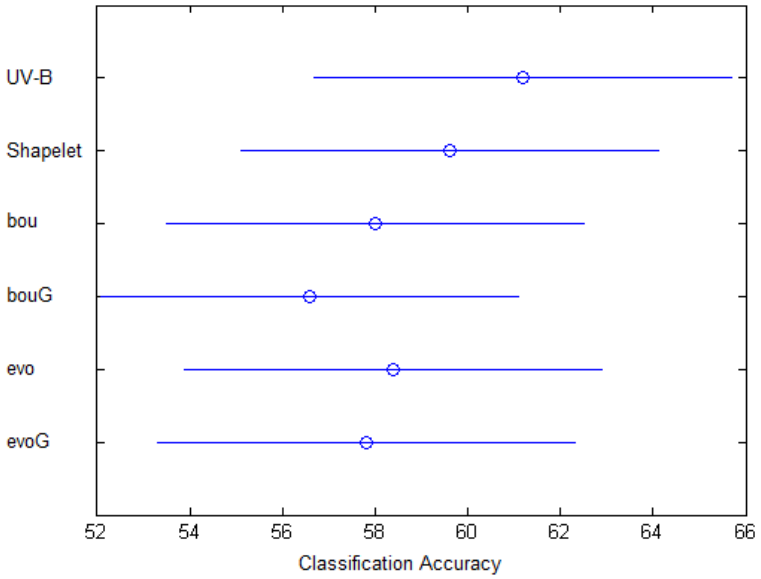


Figure 3.5: Results of post-hoc ANOVAN testing showing classification accuracies for Herring stock separation given different boundary transform methods.

post-hoc testing (Figure 3.6) shows no significant difference between results other than those for 30 harmonics.

Classifier	Harmonic Content					
	5H	10H	20H	30H	50H	100H
BN	69%	74%	73%	61%	71%	67%
NB	72%	67%	73%	51%	68%	68%
C4.5	56%	69%	72%	64%	66%	72%
RaF	76%	68%	76%	53%	69%	70%
RoF	68%	68%	72%	60%	60%	62%
Average	66%	69%	73%	58%	67%	68%

Table 3.7: Table showing accuracy of **size sensitive** harmonic classification. Columns represent harmonic content from 5 harmonics (5H) up to 100 harmonics (100H)

When tested alongside the previously shown methods, classification using the first twenty Fourier harmonics (non-normalised) does show as a significant improvement on all other boundary transform methods. shown in Table 3.10 and in post-hoc tests

Classifier	Harmonic Content					
	5H	10H	20H	30H	50H	100H
BN	45%	52%	44%	67%	63%	71%
NB	56%	52%	62%	50%	62%	69%
C4.5	57%	53%	60%	59%	58%	65%
RaF	59%	55%	57%	49%	62%	60%
RoF	64%	57%	56%	50%	49%	52%
Average	56%	54%	56%	55%	59%	63%

Table 3.8: Table showing results of **size invariant** harmonic classification. Columns represent harmonic content from 5 harmonics (5H) up to 100 harmonics (100H)

	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
algorithm	70.00	4.00	17.75	0.80	0.54
content	649.37	5.00	129.87	5.83	0.00
Error	445.80	20	22.29		
Total	1166.17	29.00			

Table 3.9: Results of N-way analysis of variance (ANOVAN) of algorithm and size invariant harmonic classification results.

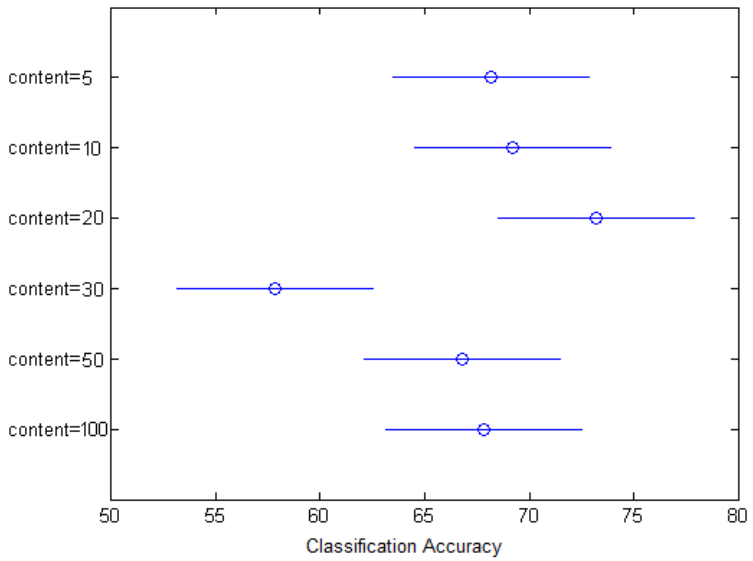


Figure 3.6: Results of post-hoc ANOVAN testing showing classification accuracies for Herring stock separation given different amounts of harmonic content.

(Figure 3.7). As when performing variance tests for univariate boundaries, shapelets and CSS results alone (Table 3.6), the Random Forest algorithm performs best in tests, although again indistinct from Naïve Bayes results.

	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
algorithm	272.40	4.00	68.10	3.70	0.02
transform	977.54	6.00	162.92	8.85	0.00
Error	441.60	24	18.40		
Total	1691.54	34.00			

Table 3.10: Results of N-way analysis of variance (ANOVAN) of algorithm and boundary transform method (including twenty non-normalised Fourier harmonics).

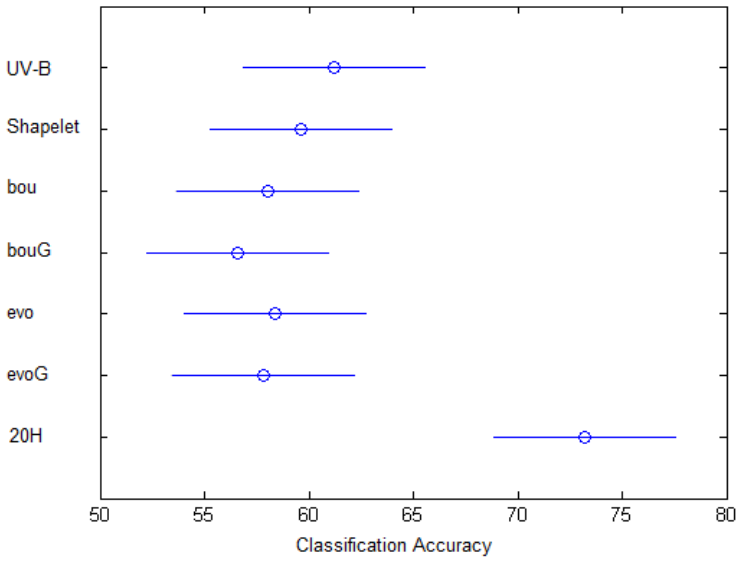


Figure 3.7: Results of post-hoc ANOVAN testing showing classification accuracies for Herring stock separation given different transform methods (including twenty non-normalised Fourier harmonics).

Confusion matrices for WEKA classification using all six transformed sets can be seen in table 3.11. Shapelet and Fourier methods appear generally more sensitive

to North-sea samples when using the classification suite. Classification of CSS maxima using Bayesian algorithms show more sensitivity to Thames than to North-sea samples, as when using CSSs own matching methods (Tables 3.3 and 3.4). However when CSS maxima are classified using decision tree methods (C4.5/Random Forest/Rotation Forest) classifiers are more sensitive to North-Sea samples.

3.3 Discussion

Curvature scale space, paired with the CSS matching algorithm, appears to work well for the SHAPE dataset, where (in our view) inter-class differences are significantly different. However, when the method was used for stock separation of the Herring dataset the results returned were not outstanding. Results of classification of CSS maxima using traditional classification algorithms show similar results. However in contrast with the CSS own matching algorithm, the addition of global metrics show no real impact on results when using the classification suite.

Whilst our shapelet implementation performed slightly better on average, results were not significant for this task. Some improvement may be possible with a larger dataset as the additional cross validation performed in order to extract shapelets may have impacted the classification accuracy. However the results returned by our study indicate that there are no areas of the boundary (within the size limits we have imposed on our shapelet extraction) that are indicative of boundary class.

Fourier methods return classification accuracies comparable to our CSS/shapelet results when using size invariant harmonics; it is interesting to note however that size sensitive harmonics return significantly increased classification results. It could be hypothesised that as all otoliths were imaged under the same magnification at the

		Shapelet			bou			bouG			evo			evoG			Fou20		
		NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp
BN	NS	19.4	20.6	0.485	20	30	0.40	24	26	0.48	28	22	0.56	32	18	0.64	36	14	0.72
	T	15.2	24.8	0.620	21	29	0.58	21	29	0.58	22	28	0.56	22	28	0.56	19	31	0.62
		NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp
NB	NS	27.2	12.8	0.680	29	21	0.58	27	23	0.54	25	25	0.50	27	23	0.54	38	12	0.76
	T	20.6	19.4	0.0485	12	38	0.76	13	37	0.74	16	34	0.68	22	28	0.56	15	35	0.70
		NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp
C4.5	NS	29.2	10.8	0.730	29	21	0.58	28	22	0.56	27	23	0.54	27	23	0.54	28	22	0.56
	T	17.6	22.4	0.560	20	30	0.60	28	22	0.44	25	25	0.50	26	24	0.48	27	23	0.46
		NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp
RaF	NS	30.0	10.0	0.750	37	13	0.74	30	20	0.60	36	14	0.72	37	13	0.74	44	6	0.88
	T	20.4	19.6	0.415	24	26	0.52	26	24	0.48	25	25	0.50	23	27	0.54	26	24	0.48
		NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp	NS	T	Se/Sp
RoF	NS	23.6	16.4	0.590	26	24	0.52	32	18	0.64	32	18	0.64	32	18	0.64	35	15	0.70
	T	17.0	23.0	0.575	24	26	0.52	20	30	0.60	19	31	0.62	23	27	0.54	17	33	

Table 3.11: Table showing confusion matrices for North-Sea Vs Thames classification, including sensitivity (Se) and specificity (Sp). Table shows five classification algorithms (rows of matrices) and six transform sets (columns of matrices). bouG and evoG are transformed set inclusive of global metrics. Shapelets hold (possible) non-integer values as the table shows the average over cross validation of shapelet extraction.

same distance from the imaging lens (ignoring minor focus adjustments), size may be a useful discriminatory factor when separating herring for the sampled stocks.

Average Fourier results show little increase in classification accuracy when including higher frequency size-sensitive harmonic content, and for some classifiers shows reductions with higher content. When using size-invariant harmonics for classification the addition of higher frequency harmonics does marginally increase accuracy but is unable to compete with size sensitive methods. Given the increased results when global metrics are included in the CSS method, this suggests that gross boundary shape or otolith size may be more discriminatory than boundary complexity with regards to herring otoliths. We investigate the viability of gross shape and scale metrics as discriminatory factors in further chapters in this thesis.

Our results compare reasonably with previous studies of stock discrimination using otoliths. Studies of dolphinfish otoliths [31] using Fourier descriptors of the boundary show results in the region 57-70% which is comparable to each of our implementations. Campana and Casselman [24] produce results of 67% using otolith boundary alone which is comparable to our Fourier implementation which also uses boundary alone, however other results in the same study fall far below those discussed in this work. Results from mackerel classification [30] show 80-86% accuracy which is significantly above our implementations; However the same study reported application study results (rather than feasibility results) nearer to our results (71.5-77.5%).

Likewise results from tests on the SHAPE database compare well with previous work [1, 49]. Whilst our results have shown a moderate (15%) improvement on those results it should be noted that we have excluded several classes from our SHAPE set, and some previous results are given with restricted sets themselves.

It is as yet unclear whether the poor performance of CSS/shapelet methods for intra-species distinction is a general issue, or restricted to this one species. Previous studies [30, 31] show that similar classification techniques used on different species can have significantly different accuracies. Further work in this thesis looks at classification of different species of fish as well as different Herring stocks to determine whether the issue is restricted to our classification specifically.

3.3.1 Summary

The study presented in this chapter shows that neither computer vision shape encoding (CSS) nor time-series shapelets, methods that have proven successful for object classification in other fields, offer any great advantage over methods that are already widely used in the otolith shape industry (Objective 1).

We do find, however, that Fourier harmonics that have not been scale-normalised perform considerably better than when scale normalised, suggesting that scale information, when retained, may increase accuracies of stock classification systems (Objective 2).

Chapter 4

Viability Of Transform Methods Over Discrete Age Categories

Chapter 4 presents a study concerning the viability of stock-separation of highly mobile Clupeids (sprat *Sprattus sprattus* and herring *Clupea harengus*) using otolith morphometrics. The aim of this study was *not* focused on overall classification accuracy, but to determine whether otolith age determines which class of transform methods have most potential for classification (Objective 2). This study also furthers research from the previous chapter, aiming to determine whether choice of learning algorithm impacts classification accuracies (Objective 4).

Analysis focuses on three 2-class stock discrimination problems (Chapter 1, Section 1.7) with the aim of reassigning individuals to source populations: Herring 1b, separated into North-Sea/Thames stocks; Herring 2, separated into Celtic/Irish-Sea stocks; Sprat, separated into North-Sea/English Channel stocks. Not all samples described in Chapter 1 were used in this study, however, the numbers of samples of each age and class can be seen in Table 4.1. As in the previous chapter, experiments are undertaken using MATLAB and employing learning algorithms from the WEKA machine learning suite.

Previous work suggests that the retention of scale information during Fourier transformation may increase classification accuracies. We therefore compare the viability of morphometric methods that include, and those that exclude, size information as representations of the otolith. To do this we use three methods of otolith boundary representation (or 'data transforms'): Elliptical Fourier Descriptors; Curvature Scale Space; and basic Shape Indices.

Burke et al. [22] reports accuracies for classification of herring stocks significantly higher that our previous accuracies of the same species in Chapter 3. The dataset used by Burke et al. [22] was obtained for use in this study. It was noted that the set used in that study contained only age-0 and age-1 samples; whereas our previously used set was comprised of samples aged 2+. Additionally, previous studies [11, 35] have reported accuracies which fluctuate depending on age composition of tests. We therefore also assess whether otolith age may affect classification accuracies by separating datasets into single age subsets.

Within this framework 14 feature sets derived using the three 'static' transformations and encoding combinations of size and shape are explored. By applying 9 state-of-the-art learning algorithms, each fully cross validated, we assessed how each system generalises to age-independent data sets, performing 1260 cross-validated experiments on the sets. Each dataset was partitioned by age using expert reader ground-truths, which were provided by Cefas for two of the datasets, and by the Galway-Mayo Institute of Technology (GMIT). To assess the saliency of size/shape features within each of the age partitions, half of the feature sets include size indices, the remainder encode only shape.

We find that for juveniles, feature sets that encode only shape perform well, but those that retain size indices return higher classification accuracies. However as fish approach maturity, transforms that exclude size indices appear more robust to ageing. This study suggests that methods of stock discrimination based on early incremental growth are likely to be effective, and that automated classification techniques will show little benefit in supplementing early growth information with shape indices derived from mature outlines.

4.1 Materials and Methods

This section details the datasets used for the study presented in this chapter, and how the datasets were divided into single age subsets (Section 4.1.1). Section 4.1.2 describes the boundary transform methods used and clarifies why they have been selected for this investigation, and gives specific details of the three boundary transform methods used, with generic methods previously described in Chapter 2, Sections 2.2, 2.3 and 2.5. Section 4.1.3 lists the learning algorithms used to construct classifiers using transformed data, selected from the WEKA machine learning suite. Finally Section 4.1.4 gives details of how the results were statistically tested to determine which factors (age, boundary transform method or learning algorithm) impact classification accuracies.

4.1.1 Datasets

For this study three pre-existing *Clupeid* datasets were used; each set containing samples of multiple ages. The North-Sea/Thames herring set collected during the ICES 2008/2009 survey, part of which was used in the previous chapter (3), contained otoliths from herring aged 0 and up. Using age data provided by expert readers at Cefas the set was subdivided into single age subsets. After division the sets deemed

of suitable size for classification studies we ages 2, 3, 4 and 5; each holding eleven or more samples per class.

Cefas supplied a further set for this study; sprat samples from the ICES 2013/2014 study in the North-sea and Western (English) channel. Details of the sample collection are described in the ICES report and not repeated here; however, North-Sea/Western channel division was carried out by Cefas expert readers before imaging, and the class details supplied with the set. Imaging of the dataset was carried out using the same equipment and under the same conditions as the previous herring set detailed in chapter 3. After division into single age subsets the subsets deemed usable were those for ages 0, 1, 2, and 3.

The Galway Marine institute supplied the final image set for this study; a portion of the dataset used for a previous study [22]. The set comprises age-0 and age-1 samples of Celtic-Sea and Irish-Sea Atlantic herring. Samples were collected in the Irish-Sea in 2006 using mid-water trawls as part of an AFBI (Agri-Food Biosciences Institute) acoustic survey. The collection was carried out in eight areas, four both to east and west of the Isle of Man. Images were captured at 20x magnification using transmitted light. Expert classification was carried out using otolith microstructure analysis for both age and stock, and class details were supplied for this study in addition to sample images. Both age-0 and age-1 sets were deemed to be of suitable size for use in this study.

Partitioned Image Sets

Table 4.1 shows the details of the sets used in the study and the number of instances per class of each set. After divisions we have concurrent single age datasets in the age-range 0 to 5 years with two sets for ages 0 through 3. Each age set is given a tag

that is both descriptive of the original set of which it is a subdivision, and the age of that subdivision, so that it can be referred to in this chapter. The number refers to the age of the set (in years) and the preceding letters pertain to the dataset from which the subdivision originates: *'BB'* for Celtic/Irish Sea Herring originally used in research by Burke et al. [22], *'S'* for North-Sea/English Channel Sprat collected during the 2013/14 ICES surveys [41, 42], and *'H'* for North-Sea/Thames Herring collected during during the 2008/09 ICES survey and previously used in Chapter 3.

4.1.2 Transform Selection and Methods

Three 'static' or 'non-data-adaptive' [8] transformation methods previously used for otolith shape analysis were selected: Elliptical Fourier transforms, Curvature Scale-Space transforms, and Shape Parameter transforms. By 'static' we mean that the transformed version of each boundary is dependent only on that instance of transformation, and not on other samples in the dataset. For example, Principle Component analysis (PCA [77]) and the more recent Shapelet transform [54, 99] (used in the previous chapter) both rely on knowledge of other samples and their classes in order to set parameters by which to transform each individual instance, and are therefore 'data-adaptive' or 'non-static'.

In machine learning parameter calculations should be performed using a separate training set [6, 26]. Failure to use separate training sets risks building a model that only represents the training set itself, rather than building a generic model, a phenomenon known as 'over fitting'. For example; when constructing a model using PCA, the principle components should be calculated using a single set of data; the model is then built using a second set (using the calculated components) and model testing is carried out using a third set; all of which must be cross validated.

Age-Set	Age	Species and Stock/Class	Samples	Reference & Notes
BB0	age-0	Herring Celtic/Irish Sea	40/class	Trawls conducted by the Agri-Food and BioSciences Institute (AFBI) & the Marine Institute (MI) [22]
BB1	age-1	Herring Celtic/Irish Sea	13/class	
S0	age-0	Sprat. North-sea/English Channel	40/class	ICES Surveys 2013/2014. [41, 42]. Age-0 set partially created from age-1 samples using trace of the age-0 annuli
S1	age-1	Sprat. North-sea/English Channel	25/class	
S2	age-2	Sprat. North-sea/English Channel	11/class	
S3	age-3	Sprat. North-sea/English Channel	13/class	
H2	age-2	Herring. North-sea/Thames	11/class	ICES survey 2008/2009. (Chapter 3)
H3	age-3	Herring. North-sea/Thames North-sea/Thames	27/class	
H4	age-4	Herring. North-sea/Thames North-sea/Thames	21/class	
H5	age-5	Herring. North-sea/Thames North-sea/Thames	15/class	

Table 4.1: Table showing species, provenance and number of each 'age-set' and their sources. 'Age-set' is the descriptive identifier given to each set, e.g. BB1 refers to the age-1 partition of the Celtic-Sea/Irish-Sea Herring. See text for further explanations.

By using only static transform methods, we avoid the need to calculate how parameters should be set. This is an important factor in our study since our age partitioned datasets are small. Further division required during validation of non-static methods results in smaller sets with which to build models, and would therefore further reduce robustness of classifiers.

The application of individual transform methods to the data creates multiple 'train/test' sets which are used to build classifiers using a selection of learning algorithms. In total 140 train/test sets were created using these methods. This constitutes fourteen sets for each of the ten single-age 'age-sets'. Half of the train/test sets hold size or size and shape data (size-inclusive sets); the remaining half contain only shape data (size-exclusive sets); these sets are described in the remainder of this section.

Figure 4.1 shows a graphical representation of data-flow through the transform and classification system. This shows the flow from each dataset (image-set sources), through boundary transformation, model construction and testing.

Curvature Scale Space

For the purpose of this study we use the built-in-house CSS transform system used in previous work (Chapter 3). Boundaries are first sub-sampled to five-hundred points and transformed by the system using methods described previously to produce the maxima coordinate-pairs (distance along boundary, evolution magnitude) which are fully invariant to scale (size-exclusive). We construct four datasets from the resulting transformed boundaries using only the largest seventeen (by evolution magnitude) maxima. As learning algorithms used in this study require an equal number of indices per instance we only use the minimum number of CSS maxima (17) held by any

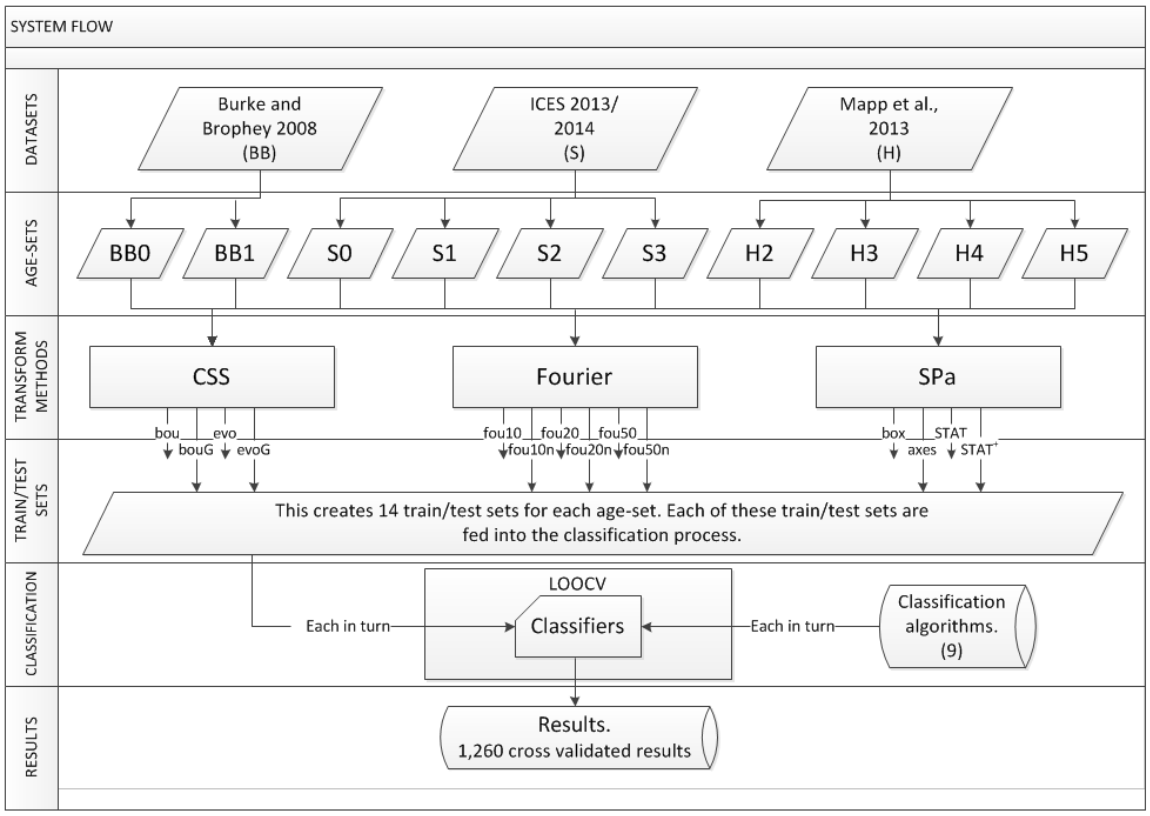
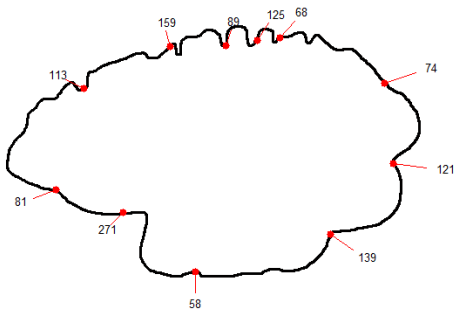


Figure 4.1: Experimental data flow through, from normalised otolith outlines (datasets) to classification results. Each age-set was transformed using each transform method to create individual train/test sets for each age-set (140 in total). Each train/test was used in turn with each of the algorithms to build and test classifiers using leave one out cross validation (LOOCV).

transformed boundaries in that image-set. The methods of ordering the data in the train/test sets are given below and an example of this ordering is seen in Figure 4.2, shown again here for clarity:

- bou → CSS maxima co-ordinate pairs; ordered according to point along the boundary. (see Table 4.2)
- evo → CSS maxima co-ordinate pairs; ordered according to evolution magnitude.
- bouG/evoG → As bou/evo; prepended by shape metrics (Circularity, Eccentricity and AspectRatio).



Set of maxima coordinates
(27,81), (57,271), (108,58), (169,139), (218,121), (265,74), (323,68), (340,125), (365,89), (401,159), (447,113) (89,93)
method: bou **27**, 81, **57**, 271, **108**, 58, **169**, 139 ... method: evo 57, **271**, 401, **159**, 169, **139**, 340, **125**...

Figure 4.2: Left: Image showing the CSS feature points extracted; marked as point along boundary, and annotated with evolution (curvature) magnitude. Right: The CSS feature pairs (point along boundary, evolution magnitude) for the boundary (*top*), and the ordering of the pairs using ***bou****ndary order*, and ***evo****lution magnitude order* methods (*bottom*). Only the ten points with largest evolution magnitude are shown in this example.

Elliptical Fourier Descriptors (EFDs)

Elliptical Fourier descriptors are commonly used for class-separation according to otolith boundary and many studies use this method for stock discrimination. Burke et al. [22] successfully discriminate two populations of herring using selected elliptical Fourier coefficients. However, in contrast to Burke et al. [22] we do not perform any statistical selection of harmonics/coefficients prior to training classifiers.

To transform the boundaries we use the same implementation used in previous work (Chapter 3) to generate fifty harmonics for each boundary. As before boundaries must be normalised first by rotation and translation to avoid building models with orientation and location. As boundaries are pre-normalised for rotation at the extraction phase we need only remove the DC components (to normalise for translation) and normalise coefficients for scale. We record the transformed boundaries both before and after scale-normalisation so that we can assess both size-variant and size-invariant harmonics. We use the generated harmonics to construct six train/test sets (per age-set) for classification, denoted as:

- Fou10 → First ten Fourier Harmonics, non-normalised for scale (size-inclusive) ordered $H_1C_4, H_2C_1, H_2C_2, H_2C_3..., H_{10}C_4$ where H = Harmonic; C = Coefficient.
- Fou20/Fou50 → First twenty/fifty non-normalised Harmonics (size-inclusive).
- Fou10n/Fou20n/Fou50n → First ten/twenty/fifty Fourier harmonics; ordered as above but normalised for scale (size-exclusive).

Shape Parameter Transformation (SPa)

Each age-set was transformed using simple shape/size measurements of the boundary and its enclosed region (SPa transform). The majority of this study is conducted using MATLAB and so we use the built-in region properties function ('regionprops') to generate the transformed 'image' from the otolith mask (the 'filled' boundary). We discard any variables generated by the regionprops method not mentioned below, as they are either unusable for classification or hold redundant data.

The transformed boundaries have the following metrics: Perimeter Length; Area (of the enclosed region, inclusive of boundary pixels); Convex Area (the area of the

convex hull); Solidity (Area/Convex Area); Height (of the calculated bounding box); Width (of the calculated bounding box); Extent (Area/(Height*Width)); Major Axis Length; Minor Axis Length; Eccentricity; Equivalent Diameter (of a circle with the same area).

We create four train/test sets (per age set) using the transformed series; Each set contains size information for the boundaries (size-inclusive) and are denoted:

- Box → Sub-selection of the SPa transform; keeping only Height and Width .
- Axes → Sub-selection of the SPa transform; keeping only Major Axis Length/Minor Axis Length.
- STAT → Full Statistic block: Perimeter, Area, ConvexArea, Solidity, Height, Width, Extent, MajorAxisLength, MinorAxisLength, Eccentricity, EquivDiameter.
- STAT^{+} → As STAT but appended with the aspect ratio of the CSS image.

4.1.3 Learning Algorithms

The Train/test sets were used to build and test classifiers using a number of available algorithms to compare and assess the transformation methods. Whilst Curvature Scale Space has a unique matching algorithm we proceeded to classify CSS transformed otoliths using the same algorithms as for the other train/test sets (created using EFDs/SPa) so that we directly compared transform methods and not the associated matching algorithm. The parameters: Circularity, Eccentricity and Aspect-Ratio, which would be used for pre-indexing were we classifying using CSSs own matching algorithm [1] are instead added to two CSS train/test sets.

The train/test sets are used to train classifiers using algorithms selected from WEKA [38] and implemented using MATLAB's java interface. We build 1,260 classifiers using each of the 140 datasets described in conjunction with each of the following nine learning algorithms (using each algorithms default settings), the classifiers used for this process (and annotation) are:

- NB → Naïve Bayes
- BN → Bayesian Networks
- Log → Logistic
- HP → HyperPipes
- J48 → J48/C4.5
- RaF → Random Forest
- IBk → k-Nearest Neighbours
- SMO → Support Vector machine
- RoF → Rotation Forest

4.1.4 Statistical Testing

To test whether otolith age, or retaining size information during transformation returns distinct results, we performed statistical testing on our results using two methods.

Mann-Whitney U-tests were performed for results for each age-set ($p>0.05$). Size-inclusive transform results are assigned to one group, and size-exclusive results to the other group. Tests that reject the null hypothesis show significant difference between size-inclusive and size-exclusive methods for that age-set.

N-way analysis of variance tests (ANOVAN [7]) are performed to determine which factors (learning algorithm, otolith age, boundary transformation) significantly impact classification accuracies. Graphical representations of ANOVAN results are shown using associated post-hoc testing.

4.2 Results

This section shows results of tests for the 1,260 classifiers constructed using combinations of different otolith age-sets, boundary transform methods and learning algorithms. Further Subsections 4.2.1–4.2.3 show results of further testing and post-hoc testing, partitioned by factor: Relative performance of learning algorithms is shown in Section 4.2.1; Impact of retention of scale information and interaction with sample age is addressed in Section 4.2.2; Otolith age effects, including partitioning by dataset, are shown in Section 4.2.3.

Whilst the large number of classifiers built for this study preclude the full results being shown here, full results are presented in Appendix A. Results from all tests were combined and subjected to variance testing, results of which can be seen in Table 4.2, which is discussed further in Subsections 4.2.1–4.2.3.

	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
algorithm	1910.30	8.00	238.80	1.58	0.13
transform	12765.0	13.00	982.00	6.50	0.00
age	149843.40	5.00	29968.70	198.27	0.00
Error	186371.70	1233.00	151.20		
Total	350891.20	31258.00			

Table 4.2: Results of N-way analysis of variance (ANOVAN) of all results across: learning algorithm, boundary transform method, and otolith age.

4.2.1 Relative Performance of Learning Algorithms

When comparing learning algorithms across ages, no recognisable patterns were found, and results of variance testing (Table 4.2) show only minor statistical difference between learning algorithms used for modelling the data using any of our age-sets ($p = 0.13$). Figure 4.3 shows post-hoc testing for the learning algorithms across all age categories, and the average result (across learning algorithms) for classification of each train/test set is shown in Table 4.4. Variance testing including interaction tests shows no interaction between algorithm selection and either transform method ($p = 0.96$), or sample age ($p = 0.32$).

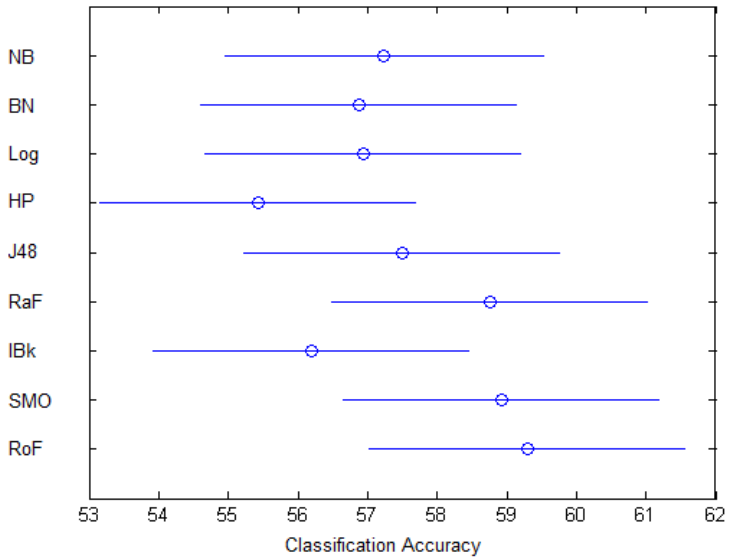


Figure 4.3: Results of post-hoc ANOVAN testing showing stock separation given different learning algorithms.

Whilst there were statistical differences between learning algorithms when comparing across single ages, no obvious choice of algorithm emerges. Figure 4.4 shows

the average rank of algorithms across results separated by age. The best performing algorithm by each individual age set can be seen in Table 4.3.

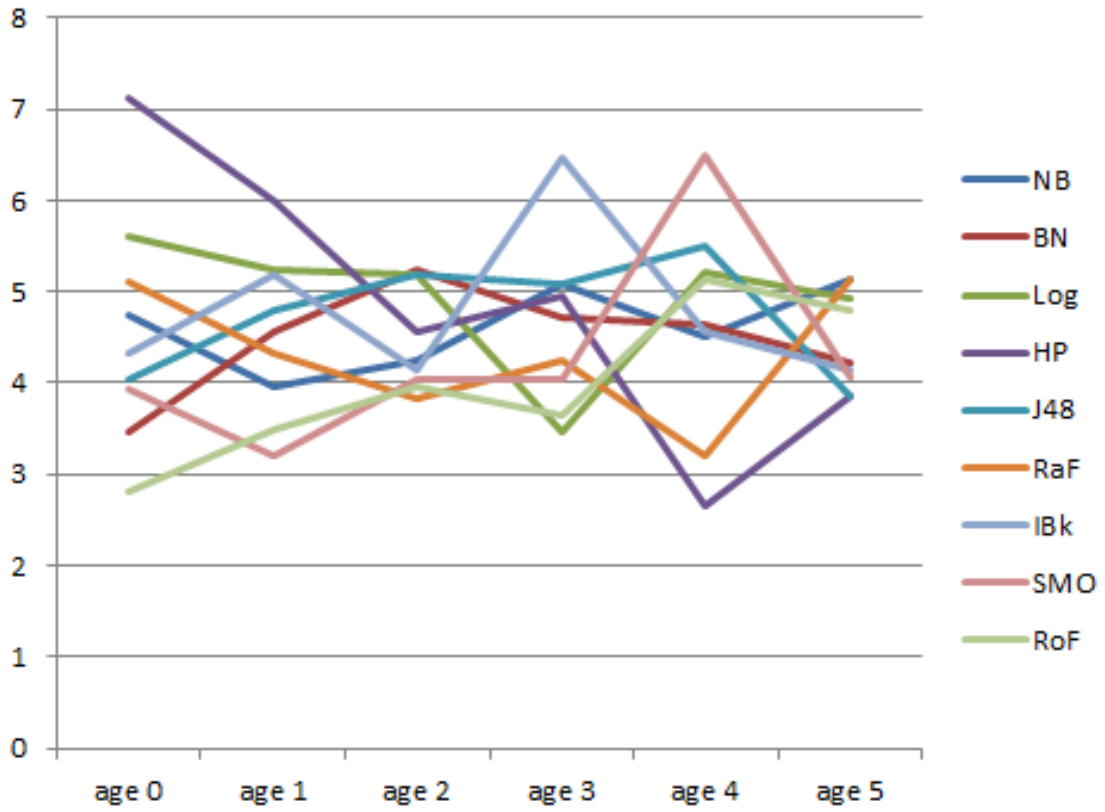


Figure 4.4: Showing average rank of learning algorithms for each age category.

4.2.2 Size-inclusive Vs Size-exclusive Transforms

Table 4.2 shows that different boundary transform methods produce classifiers which return significantly different stock separation accuracies (p = 0.00). Further variance testing of results shows that groupings of size-inclusive and size-exclusive transform results are also statistically distinct across all age categories (p = 0.00). The mean result (across learning algorithms) for each train/test set is shown in Table 4.4. Check-marks in Table 4.4 show age-sets that reject the null hypothesis (p>0.05)

Age	Celtic/Irish Sea herring	North-Sea/Channel sprat	North-Sea/Thames herring
0	RoF	BN	
1	N/A	SMO	
2		IBk	N/A
3		SMO	N/A
4			HP
5			N/A

Table 4.3: Peak performing learning algorithms for each age set. *N/A* denotes times where no algorithms performs (statistically) better than others.

during Mann-Whitney U Tests, and therefore show significant difference between size-inclusive and size-exclusive transform results when grouped by age. These sets are plotted in Figures 4.6 and 4.7 using solid markers; whilst those that accept the null hypothesis (that results between groups are indistinct) are shown using 'hollow' markers (both at $p>0.05$).

N-way analysis of variance including interaction tests between factors, shows strong interaction between grouped (size inclusive/exclusive) transform methods and sample age ($p = 0.00$). Results of post-hoc testing can be seen in Figure 4.5 (modified for clarity) where it can be seen that as age increases, the relative positions of size-inclusive and size-exclusive transforms reverse.

Figure 4.6 shows the average rank of size-inclusive and size-exclusive methods by age, separated for each of the single-age image sets. As with post-hoc ANOVAN testing, size-inclusive transform methods produce significantly better results than size-exclusive for the younger age-sets. Again, as the age of the samples increases the average rank of size-inclusive transforms increases while the rank of exclusive methods decreases; this proceeds until exclusive (shape only) methods perform better on average (have lower ranks) than size-inclusive methods.

	Age-Set									
	BB0	**BB1**	**S0**	**S1**	**S2**	**S3**	**H2**	**H3**	**H4**	**H5**
bou	64.3	46.2	56.9	57.3	49.5	47.9	49.0	49.0	54.2	54.4
bouG	83.8	47.0	55.1	64.0	47.5	50.9	46.5	47.5	52.4	54.1
evo	68.8	56.4	64.4	64.2	50.0	53.0	52.5	52.3	46.6	54.1
evoG	83.3	57.3	65.0	68.0	46.0	50.9	52.0	54.5	47.9	52.2
fou10	97.4	70.1	66.3	83.3	48.5	64.1	53.5	47.9	52.9	52.6
fou20	96.7	66.2	69.7	79.1	44.9	59.0	52.0	42.8	45.0	50.0
fou50	81.9	61.1	61.7	76.7	34.8	53.4	50.5	45.5	43.9	52.6
fou10n	81.8	53.4	63.1	74.4	42.4	64.1	47.0	48.4	49.2	57.0
fou20n	82.2	50.0	61.7	69.6	37.4	71.8	48.0	46.9	51.6	53.7
fou50n	81.5	54.7	60.1	68.0	47.5	63.2	42.9	49.0	54.2	53.3
Box	98.5	79.5	65.0	85.8	56.6	46.2	46.0	45.7	47.6	45.9
Axes	99.4	79.5	66.8	83.6	52.5	46.2	52.5	51.0	50.8	42.2
STAT	98.8	77.4	65.1	83.6	57.6	43.6	43.4	47.5	53.2	47.8
STAT$^+$	98.8	76.5	68.8	84.0	57.1	46.2	44.4	51.4	51.1	48.5
U-test (5%)	✓	✓	✓	✓						✓

Table 4.4: Table showing average classification results (averaged over algorithms) for each train/test set. Check-marks show age-sets where size-inclusive and size-exclusive methods were distinct from one another using U-tests (p>0.05). Transforms shown in bold are size-*inclusive*

However, whilst at age-5 size-exclusive methods perform better than size-inclusive ones, the results for size inclusive/exclusive methods are not significantly different, which is supported by the post-hoc testing in Figure 4.5, where inclusive/exclusive accuracies overlap. Figure 4.7 shows average classification accuracy by age for each of the image sets. Whilst at higher ages size-exclusive methods do appear to perform better (than size-inclusive methods) during ranking, the results are not as significant as those for younger image sets and perform only marginally better than random assignment of class.

4.2.3 Relative Performance by Otolith Age

Table 4.2 shows statistically different classification accuracies according to sample age (p = 0.00). Figure 4.8 illustrates post-hoc ANOVAN testing by age. It shows that

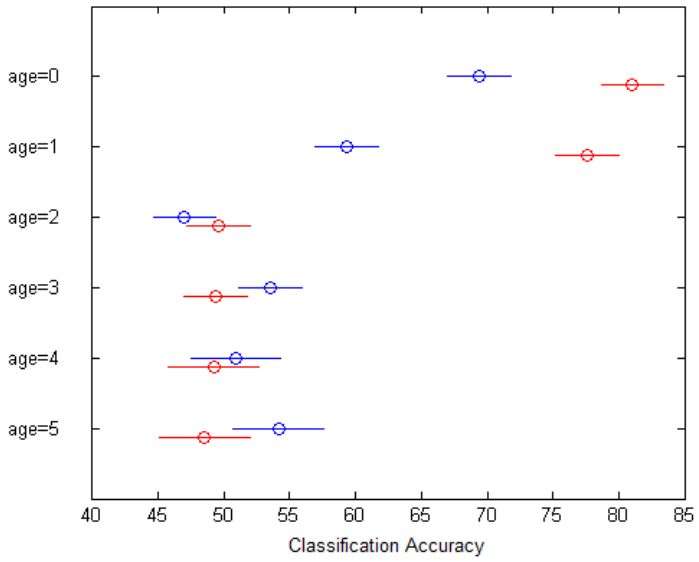


Figure 4.5: Results of post-hoc ANOVAN testing showing interaction between sample age and transform grouping (size-inclusive/exclusive). Modified/coloured for clarity, size-inclusive results (red), size-exclusive (blue).

classification of younger otoliths (ages 0/1) returns results statistically distinct from each other and from results using older samples (age 2+). Results of age 2+ test are indistinct (from each other).

Variance testing of results for individual datasets (BB, S and H) showed statistical difference between results for different sample ages for both BB, and S, where all ages produce significantly different results. However, results for separate ages of the H dataset are indistinct from one-another. It is clear from these tests that results for younger samples are significantly better regardless of the dataset tested.

4.3 Discussion

Our results show significant differences between size-inclusive and size-exclusive transform methods depending on the age-category of otoliths modelled, showing that type of transform must be considered when modelling otolith boundaries. We also show

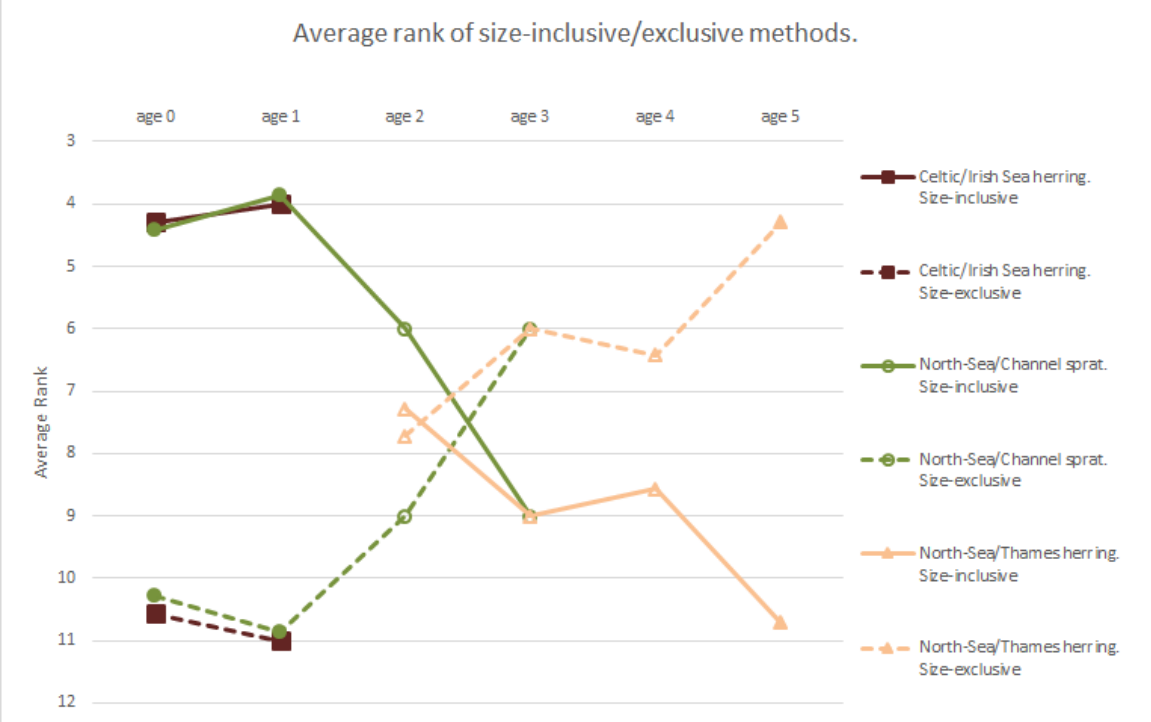


Figure 4.6: Average rank of size-inclusive and size-exclusive transform methods for each age-set. Tests where size-inclusive/exclusive results are statistically different (5%) are shown using solid markers.

that the age of the sample otoliths impacts results when single-age sets are used to construct and test classifiers regardless of the classification methods used in this study. Analysis of our learning algorithms was limited given the difference in sizes of our age-sets; however we demonstrate that no general difference exists between algorithms for these particular problem sets. Further analysis using consistent dataset sizes may yield better results in this area.

Results for the age-0 and age-1 sets show a significant difference between size-inclusive and size-exclusive transformation methods. Specifically, size based methods achieve up to 100% accuracy for age-0 herring (using individual classification algorithms), and average as high as 98.8% across all learning algorithms for size-inclusive transform methods. These results are far better than for size-exclusive transforms

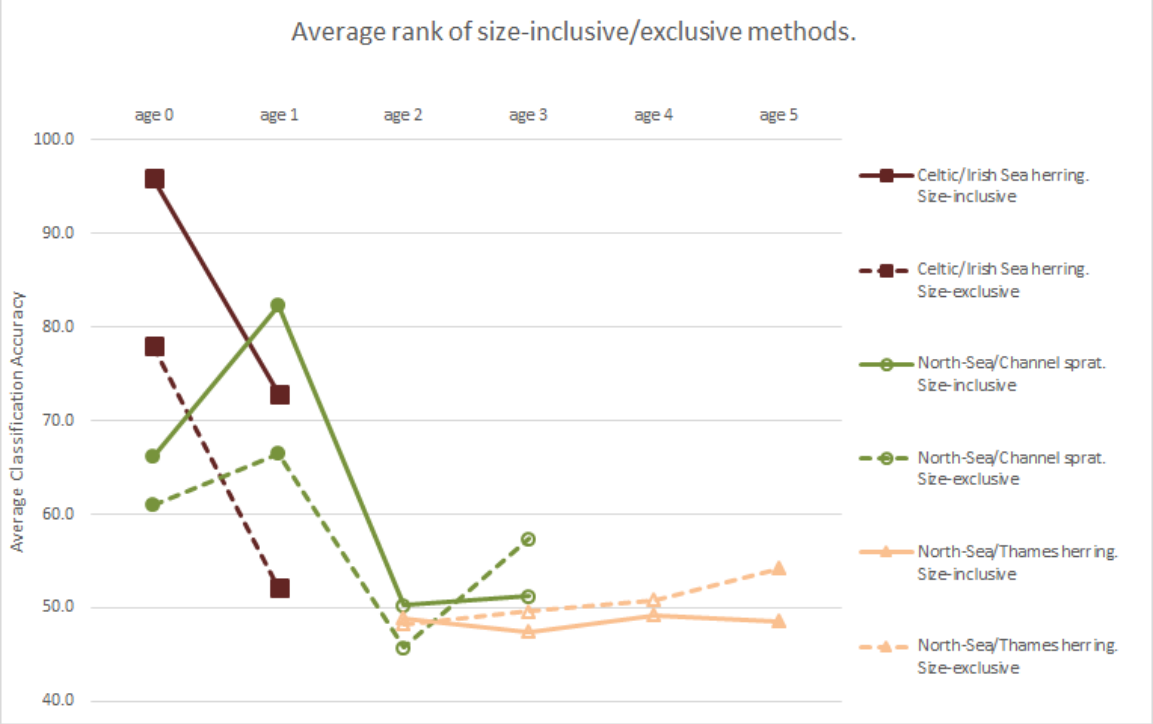


Figure 4.7: Average classification result of size-inclusive and size-exclusive transform methods for each age-set. Tests where size-inclusive/exclusive results are statistically different (5%) are shown using solid markers.

of the same sets which achieve up to 83.3%. This disparity in results between size-inclusive and size-exclusive methods can be seen across all age-0 and age-1 sets; indicating that for younger samples, size is perhaps a more useful metric for stock discrimination.

The use of size may be an obvious method when distinguishing between stocks that spawn at different times of the year, where we expect age-specific inter-class sizes. However, even when metrics recorded using the SPa transform method were normalised (by otolith length), we achieve similar result patterns and accuracies. For the Burke et al. [22] sets, the larger (otoliths) of the two classes (Celtic-Sea herring) are from fish that spawn later in the year than the smaller class (Irish-Sea) but have a much faster initial growth rate. Further investigation is therefore required to

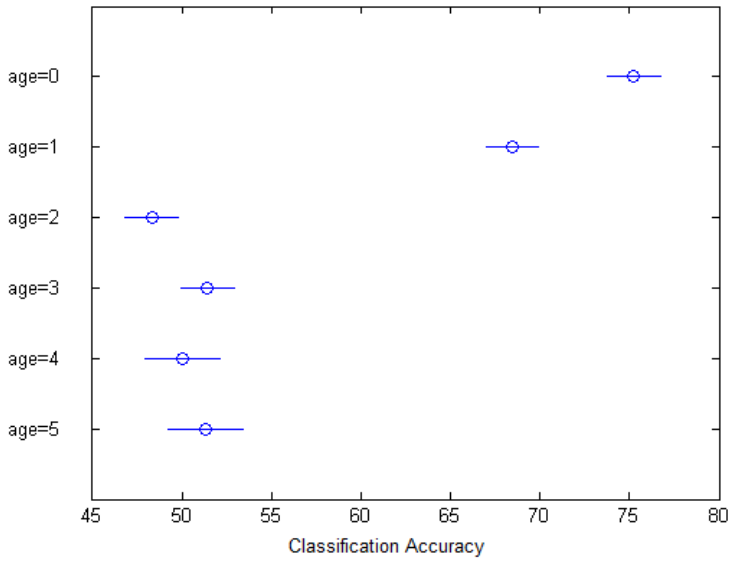


Figure 4.8: Results of post-hoc ANOVAN testing showing sample age impact on classification accuracies.

determine the age at which size-inclusive methods become incapable of separating spawning components for this particular discrimination problem.

Whilst for age-0 and age-1 sets we return favourable results, both in terms of classification accuracy and disparity between inclusive/exclusive methods; we find that as sets increase in age, classification accuracies reduce, becoming comparable to random assignment of class. However for the eldest of our age categories, size-inclusive and size-exclusive transform methods do perform differently (rejecting the Mann-Whitney U-test at $p<0.05$). Shape may therefore be the primary factor to consider for stock classification using otolith outlines for older fish. However more testing needs to be done in this area as once partitioned by age the datasets used in our study were of small size and inconsistent in number across ages-categories. Testing using larger sets may not show such a large drop in classification accuracies,

and may clearly credit or discredit the change from size-inclusive to size-exclusive methods, both for the stocks studied here, and for additional stocks and/or species.

In machine learning, parameter calculations should be performed using a separate training set [6, 26]. By using only static transform methods in this study, the need to calculate parameters was avoided. This was an important consideration as our age partitioned datasets are small. Further division required during validation of non-static methods would result in smaller sets with which to build and test classifiers, thus reducing robustness.

Results achieved using EFDs for the age-0 partition of the Irish/Celtic Sea herring set [22] did not match their published figure of 97%. However we feel that this is largely down to two factors. First we used only a portion of the original dataset, and we split the portion further into two single age-sets; whereas the results reported by Burke et al. [22] were based on the larger age-0 set, boosted by age-0 ring traces from the age-1 specimens. Second we remove test instances from the set prior to parameter selection (regression analysis) and to building classifiers; whilst the original work performs parameter selection on the full set (including test samples). The validation in this work therefore reduces accuracy, but goes some distance to reduce over fitting the classifiers to the training set, and would produce classifiers that are more reliable when classifying further otolith samples. Our results are consistent with those in the previous work however, showing that it is certainly possible to separate stocks of juvenile herring using Fourier analysis.

When compared with previous classifications of the same dataset [61] Curvature scale space methods performed to a similar degree when separated into single-age experiments. However the otolith sets used in each of our experiments each represented

a fraction of the sets used in previous work with no significant change in accuracy. Further experiments using larger training sets may yield increased classification accuracies, and may show curvature transform methods to be an important tool when classifying datasets containing otoliths from older fish.

When comparing classification results for our younger otolith experiments we attain accuracies in line with or higher than those in previous studies. Duarte-Neto et al. [31] reports classification results of dolphinfish using analysis of Fourier harmonics in the region 57-70% which is surpassed by our age 0-1 size-*inclusive* tests, while their lower result limit (57%) is comparable to our size-*exclusive* tests of older samples (age 2+). DeVries et al. [30] reports application accuracies of Atlantic mackerel classification at 71.5-77.5% which we match with our Fourier experiments. However the same report also gives feasibility results up to 86%, a result that we surpass with size-*inclusive* experiments using younger age-sets.

Results from the single age testing show a significant reduction in classification accuracies as the age of otoliths used to build and test classifiers increases, regardless of the methods used to transform the otolith data, indicating that otolith age may inpact classification accuracies regardless of the methods used to encode their shape. This suggestion is supported by statistical differences between age-1/2 and age-2+ tests, which are clearly visible in results and statistical testing regardless of the dataset tested.

However, the class sizes of a number of age sets used in this study were relatively small, and therefore classification accuracies may be impacted by insufficient training set sizes. Further testing with larger datasets is required to establish whether either

method is able to perform with adequate certainty to allow clear stock discrimination in older fish, or whether clear results are the reserve of younger otoliths.

Whilst most of our results appear to be correlated with tests on other sets; our age-0 sprat tests return much lower classification accuracies. This is despite the set being the same size as BB0 (40 instances per class). This may be due to one class of the set being comprised of age-0 ring traces from age-1 samples, in the absence of available age-0 samples for that class.

The drop in accuracies may be down to several reasons in this instance: poor accuracy of outlining; or visible rings being inconsistent or uncorrelated with age-0 growth. Furthermore, age-0 ring extraction has yet to be validated for this species [95]. Further studies on the impact of outlining methods are presented in Chapter 5 of this thesis.

4.3.1 Summary

We conclude that otolith age dictates which class of boundary transformation (including/excluding scale information) can be used to construct accurate classifiers. Whilst scale retaining methods show superior results for young (age 0/1) otolith samples, scale invariant transform methods perform better for older samples (age 2-5). However, we find that whilst scale invariant date performs *better* at higher ages, results are inferior to those obtained using younger samples (Objective 2).

As with Chapter 3, we find that curvature scale space methods offer no improvement on Fourier techniques (Objective 1), even given additional choice learning algorithms used to construct classifiers (Objective 4). However, we note that there was no

noticeable drop in classification accuracies when the datasets used (as single age sets) were considerably smaller than in previous study (Chapter 3).

Chapter 5

Susceptibility Of Fourier Based Classification To Outlining Methods

This chapter examines whether the method chosen for otolith boundary extraction affects stock discrimination accuracies using elliptical Fourier based classification. The work presented here focuses mainly on our third objective (Chapter 1, Section 1.2) and also pertains to Objective 4, whether learning algorithm affects extraction method choice. Boundaries extracted by two experts using two outlining methods were used to construct classifiers then used to assign further samples to their source populations.

Previous studies have used variable methods of boundary determination prior to Fourier based morphological classification. In the studies reported by Begg and Brown [11] and Begg et al. [12] otoliths were outlined by hand and transformed using Fourier methods before discriminant analysis was used to predict further samples, with classification accuracies in the range 56-81% dependent on whether single-age or multi-age tasks were undertaken. Results from DeVries et al. [30] also fell firmly within this accuracy range for Atlantic mackerel stock classification at 77.5% using elliptical Fourier descriptors and discriminant analysis.

Whilst the study by Burke et al. [22] reported accuracies far above this range at 97%, we were unable to replicate this result using similar methods described in the Chapter 4, with results consistently falling into the range above (56-81%) when using size normalised Fourier descriptors. However, Burke et al. [22] details hand-tracing of annuli in their study, whilst in the study presented in Chapter 4 samples are outlined automatically using thresholding methods.

Methods of automatically extracting the boundary from the digitized otolith have also been used with varying success. Atlantic cod was successfully separated using Fourier methods built on outlines automatically outlined using a threshold determined with analysis of the grey-scale histogram [25], reporting accuracies of 85-96% dependent on age class. In Stransky [90] classification of redfish returned mixed results, however accuracies for certain tasks were reported as high as 76% overall accuracy. However these results were for regional separation, and in the same study mixed accuracies were reported for stock separation, depending on the species being analysed.

We therefore compare two methods of boundary extraction in this study: Outlines derived by two expert readers, traced by hand; and outlines derived by intensity thresholding the otolith image using two different approaches. As in the chapter 4 our goal is not to achieve high classification rates, rather we assess how outlines created using each method differ when used to construct and test classifiers. To this end outlines from each method are transformed using elliptical Fourier methods creating a set of harmonics for each of the outlining methods. As in previous work we construct classifiers, each fully cross validated, using varying harmonic content and multiple learning algorithms available within the WEKA machine learning suite, allowing us to determine whether algorithm selection is dependent on outlining methods used.

We find that classifiers constructed using otolith boundaries traced by hand marginally outperform those that use boundaries extracted using thresholding methods, and that classifiers may be constructed by one expert and used to test further outlined samples (outlined by the alternative expert) with a statistically insignificant decrease in accuracy. The choice of learning algorithm used to construct classifiers is also an important consideration, with three algorithms performing well regardless of the outline methods used. However we also note that algorithm selection may significantly differ given additional classification tasks.

5.1 Materials and Methods

Whilst most outlining, transform and classification methods used in this chapter are detailed in Chapter 2, details of how the methods were used with regards to the study presented in this chapter are given here.

Two expert readers were employed to perform otolith outlining on a small image set (described in Section 5.1.1) using two different methods: hand outlining; and user defined thresholding. These methods are used to compile four sets of image outlines: Two for expert outlines (one set per expert) and two thresholded outlines (High and Low thresholds). The specific methods for these are detailed in Sections 5.1.2 and 5.1.3. With the respective sections also providing information on how each of the outline sets are used to build and test classifiers.

The choice of harmonic content, and the learning algorithms used for building and testing classifiers are discussed in Sections 5.1.4 and 5.1.5, followed by details of statistical tests used to analyse the results of tests in Section 5.1.6

It should be noted that when we refer to an 'expert' we refer to both the expert user, and that user's computing set-up. For example, both experts used the same software to perform the tasks, however, expert-1 used high precision hardware for the tasks (high precision mouse/mouse-mat etc.), whilst expert-2 used altogether more standard equipment to perform outlining tasks.

5.1.1 Image Set

We perform tests on a subset of the previously used Irish/Celtic Sea herring set (Chapter 4). The subset comprises ten Celtic-Sea otolith images, and ten Irish-Sea images. All images were selected from the age-0 portion of the original dataset and all images were chosen at random. The Irish and Celtic Sea selections can be seen in Figure 5.1 (Irish) and Figure 5.2 (Celtic).

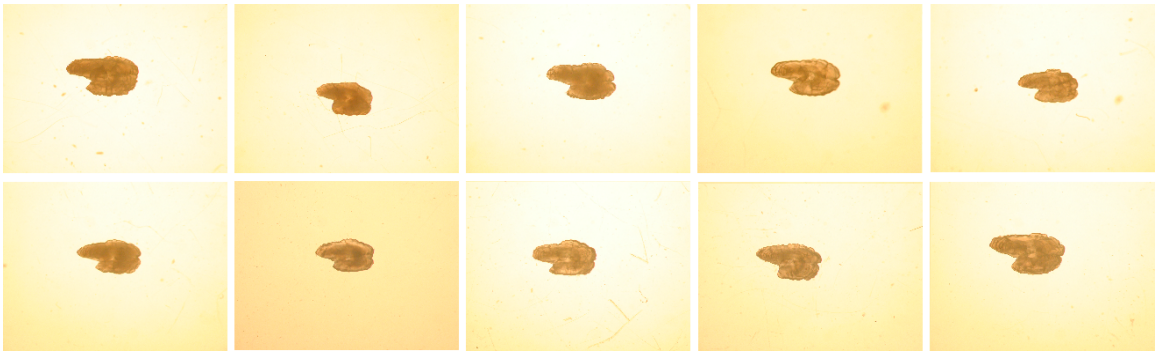


Figure 5.1: Irish-Sea herring otoliths used for the study.

5.1.2 Hand Outlining

Each expert used the hand outlining procedure described in Chapter 2 Section 2.1 to segment each otolith image four times. Each expert's hand-drawn outlines are then used to construct a mean outline for each otolith in the database, resulting in two mean outlines for each image: an outline for expert-1 and an outline for expert-2.

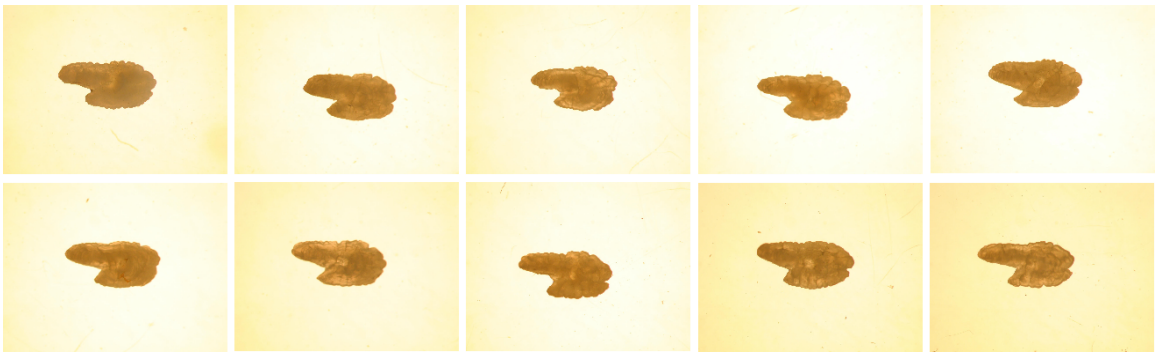


Figure 5.2: Celtic-Sea herring otoliths used for the study.

However, as the hand-drawn boundaries may be variable in length, and as coordinate extraction may not begin at the exact same point each time, a simple point-by-point average of pixels is not possible.

Each boundary outline traced by the expert is recorded as a binary image, the same height/width as the original otolith images. The pixels selected as otolith boundary during the hand tracing are set at value one, and background pixels set as zero. Figure 5.3-left shows a binary image for a simple example, for clarity the boundary drawn is shown as black.

For each of the four images we first invert the values (boundary becomes zero, background becomes one) then set each pixel value to its Euclidean distance to the nearest boundary pixel in the image (now zeros). Figure 5.3-center shows the result of this step for the simple example. The pixels at which a boundary point is present have distance zero, those orthogonally adjacent to a boundary point have a distance of one pixel etc. Pixels that fall within the boundary then have their sign inverted, so that those pixels have a negative distance, and those on the outside remain positive distance, seen on the right in Figure 5.3.

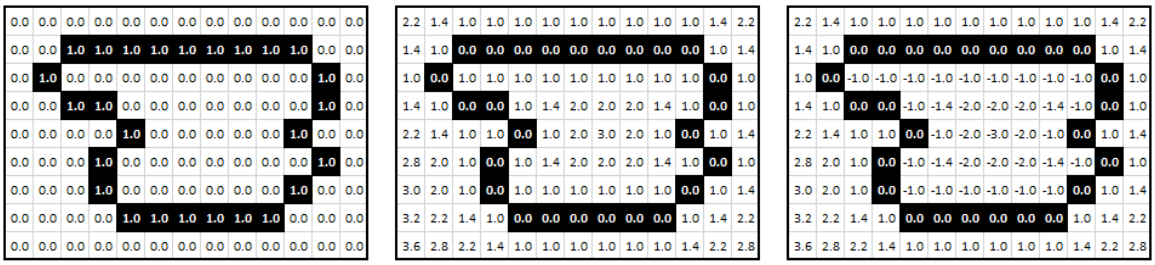


Figure 5.3: Simple example of boundary trace recording. Points set as boundary by the expert reader have value one in the left-hand image. The central image shows the calculated Euclidean distance to the nearest point on the boundary. Right-hand image shows the final distance image after sign inversion of encompassed pixels. All values show to 1 dcp.

Once all four of an experts hand-drawn outlines have been used to create a *distance image* in this manner we create a mean distance image. The mean image is the same height/width as the previous images and each pixel is set as the mean of the corresponding pixel values in each of the four distance images. Figure 5.4 shows four different (synthetic) traces of the same simple boundary (left and center columns). The top-right image shows the mean distance image calculated from the four boundaries (to 1 dcp).

New *mean otolith* masks can be created by thresholding the mean distance images at value zero. Any pixels with value less than or equal to zero are designated foreground (white in Figure 5.4) all other pixels are set as background. A mean boundary can be extracted from the mask using methods previously discussed in Chapter 2. The bottom right image in Figure 5.4 shows the resulting boundary (white) overlaying the mean distance image. The boundary coordinates are recorded as discussed in Chapter 2. The process is then repeated for the outlines from the second expert.

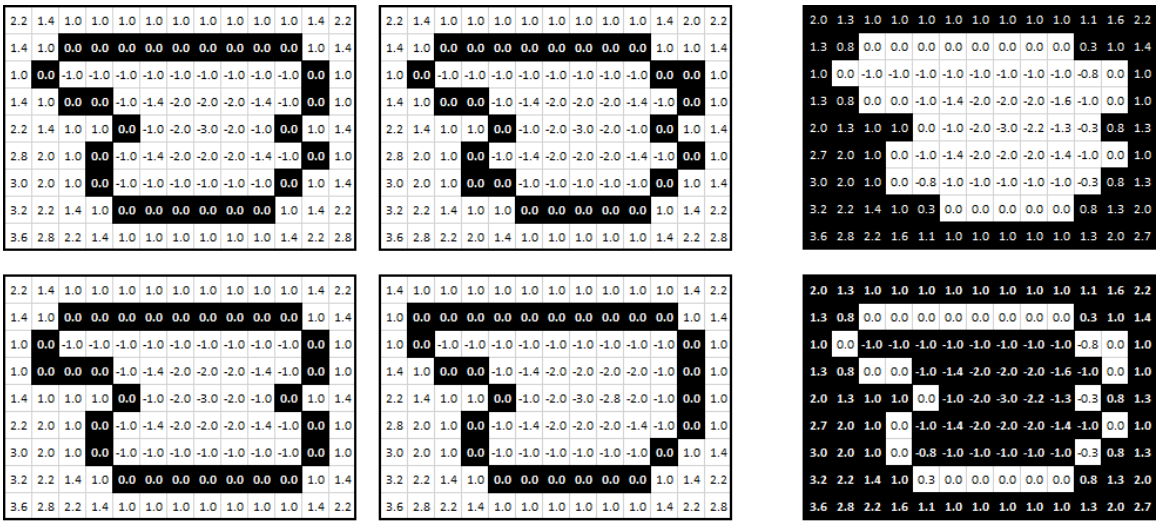


Figure 5.4: Example of mean boundary calculation. Left and center columns show distance images for four hand traced (synthesised) boundaries of a simple otolith example (with traced boundaries shown in black). Top-right image shows mean values for the four distance calculations and the resulting 'mean mask' shown in white. The extracted boundary is shown in white in the bottom-right image. All values to 1 decimal place.

Expert Selection

To assess whether the expert that outlined the images impacts classification results we construct classifiers using the Fourier harmonics of each expert, withholding each instance in turn, then using the constructed classifier to predict the class of the withheld instance (Leave one out cross validation). This returns two sets of classification accuracies: one for classifiers built and tested using the mean boundaries of expert-1 (*E1vE1*); and one set for those built and tested using expert-2 mean boundaries (*E2vE2*).

During the procedure, each constructed classifiers is also used to predict the class of the corresponding (withheld) otolith from the alternative expert's mean boundaries. By cross validating in this manner we ensure that the test instance is not included in the training data, even when that otolith was outlined by another expert. This

returns a further two sets of (cross-validated) accuracies: One for classifiers built using expert-1 boundaries and tested with those of expert-2 (*E1vE2*); and one vice versa (*E2vE1)*. We therefore end with four sets of accuracies, each of which contain accuracies of algorithm/harmonic content classification:

- E1vE1 → Cross validated accuracies for classifiers built using mean outlines of expert-1 when classifying mean outlines of expert-1.
- E1vE2 → Cross validated accuracies for classifiers built using mean outlines of expert-1 when classifying mean outlines of expert-2.
- E2vE1 → Cross validated accuracies for classifiers built using mean outlines of expert-2 when classifying mean outlines of expert-1.
- E2vE2 → Cross validated accuracies for classifiers built using mean outlines of expert-2 when classifying mean outlines of expert-2.

5.1.3 Thresholding

To outline the otoliths using thresholding, we use techniques similar to those described in the methods chapter (Chapter 2, Section 2.1). Each otolith was shown to the expert four times and the expert was asked to manually set an intensity threshold that adequately segmented the image. Whilst no corrections of the segmentation (by hand) were allowed, holes in the foreground were filled and background noise was removed so that only one hole-free mask was constructed. The range of pixel intensities in the grey-scale image to be thresholded was rescaled to range 0-100 for these tests.

Otoliths in the dataset are imaged using transient light, therefore the otolith appears (generally) darker in colour than the background. Otoliths are segmented by

determining pixels whose values fall below the threshold used. However, were images to be captured using incidental light, the otolith would likely appear lighter than the background, and segmentation would be performed by recording pixels which fall above each threshold.

Inside-Out (Low) Thresholding

The first two times each image was shown to each expert, the initial threshold was set at zero, so that no pixels in the image were shown as foreground (as no pixels have lower value than zero). The expert was asked to iteratively increase the threshold by one point at a time until the resulting mask and extracted boundary were deemed sufficiently accurate. At this point the resulting otolith mask and threshold were recorded, resulting in four otolith masks (two per expert) for the inside-out or *Low* Thresholding.

Due to otolith segments being darker in general, this method 'grows' the segmentation from within the otolith (as central pixels are darker than boundary pixels). Hence we call this method inside-out thresholding.

Outside-In (High) Thresholding

The third and fourth time through the dataset, each image was shown using an initial threshold of one-hundred, meaning that all pixels were set as foreground (as all pixels have a value lower than one-hundred). Each expert was asked to iteratively reduce the threshold by one point until the mask and boundary were sufficient, and the resulting mask and threshold were recorded. This again creates four otolith masks (two per expert) for the outside-in or *High* Thresholding. Due to otolith segments being darker in general, this method appears to 'shrink' the segmentation from outside the otolith.

Three example stages through the bottom-up method and two from the top-down method can be seen in Figure 5.5. It can be seen from this example that the threshold range deemed suitable for the shown otolith is 81-94. Thresholds below this range return masks that do not fully cover the otolith; Thresholds above the range cover the otolith but also classifies areas of the background as otolith.

Note that were the images captured while using incidental light, the otolith would be lighter than the background and would be segmented by detecting pixels higher in value than the threshold. Therefore the first method (where the threshold is gradually raised) would result as outside-in thresholding, and gradually reducing the threshold (second method) would result in inside-out thresholding.

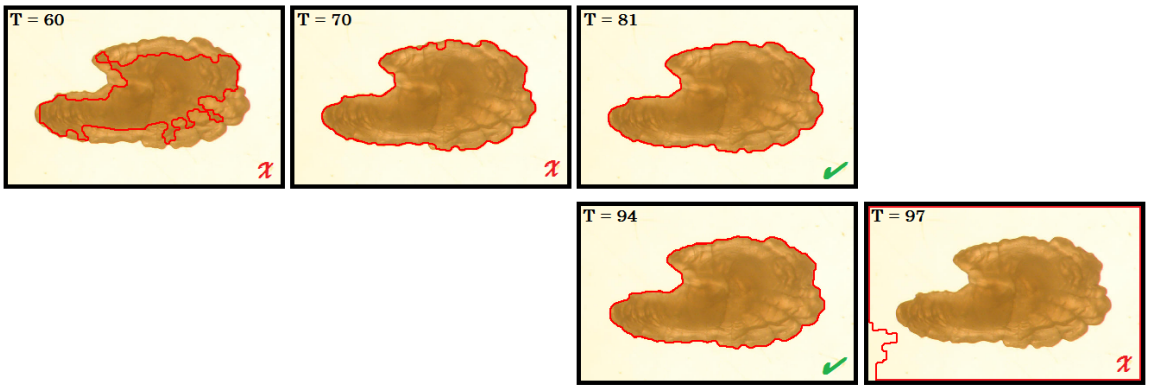


Figure 5.5: Top row: Three stages during the bottom-up thresholding method. T = 60 and T = 70 return a mask that does not contain the full otolith and are rejected. T = 97 also contains background pixels and is therefore rejected. T = 81 (during bottom-up) and T = 94 (top-down) return masks deemed sufficient by the expert.

Once each expert has completed this process we have eight sets of otolith masks and associated thresholds: Four *high* thresholded masks (two per expert), and four *low* thresholded masks (two per expert). From these we construct two sets of boundaries, a set using *High-thresholds*, and a *Low-threshold* set.

To construct the *High* set we visit each image in the dataset in turn and keep only the boundary constructed using the highest recorded acceptable threshold. The boundary coordinates are then extracted and recorded as described (Chapter 2). For the *Low* set we again visit each image in turn, keeping the boundary constructed with the lowest recorded threshold. Table 5.1 shows a partial threshold table. Shown are the eight segmentation thresholds deemed suitable by the experts as well as the selected high and low thresholds used to construct the boundary sets.

	Expert 1				Expert 2					
	High-1	High-2	Low-1	Low-2	High-1	High-2	Low-1	Low-2	High	Low
Img-1	96	96	91	91	96	93	90	88	96	88
Img-2	94	94	81	81	94	94	81	79	94	79
Img-3	93	96	94	92	96	95	86	90	96	86
Img-4	92	92	86	86	92	92	85	85	92	85
Img-5	94	94	89	86	93	94	84	85	94	84
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
Img-20	91	91	70	71	89	91	70	71	91	70

Table 5.1: Showing partial table for the segmentation thresholds. Four thresholds per expert are shown, two high and two low. High/Low columns show the selected threshold for each of the shown images which are used to construct the high/low boundary sets.

Threshold Selection

Threshold selection tests carried out to determine whether the thresholding method used to determine otolith boundary impacts classification accuracies. Tests are carried out in a similar manner to expert selection tests, however, rather than having two sets of mean boundaries created from expert outlines, we test two outline sets constructed using the peak **H**igh and **L**ow thresholds determined by the experts.

We construct and test classifiers using the high-threshold and low-threshold boundary sets, withholding each instance in turn to form the test 'set'. The constructed classifier is then used to classify both the withheld instance, and the corresponding instance from the set constructed using the other threshold method. Tests are carried out using each combination of algorithm/harmonic content to determine whether the method of threshold selection (top-down/bottom-up) shows any interaction therewith. When including these classification methods we end with another four sets of accuracies:

- HvH → Cross validated accuracies for classifiers built and tested using outlines determined with top-down (**H**igh) thresholding method.
- HvL → Cross validated accuracies for classifiers built using outlines determined with top-down thresholding method and tested using outlines from bottom-up (**L**ow) method.
- LvH → Cross validated accuracies for classifiers built using outlines determined with bottom-up thresholding method and tested using outlines from top-down method.
- LvL → Cross validated accuracies for classifiers built and tested using outlines determined with bottom-up thresholding method.

5.1.4 Fourier Transform

An in previous chapters we calculate the first fifty elliptical Fourier harmonics as reconstructions of the boundary using this number of harmonics shown a mean pixel distance between reconstructed and original boundaries less than one. It can be clearly seen from Figures 5.2 and 5.1 that the two classes of otolith are significantly different in scale.

We therefore remove the scale component by size-normalising the extracted Fourier harmonics using the scale normalisation methods described in Chapter 2. In addition all otoliths are normalised for translation (by removing the Fourier DC components) and are normalised for rotation by adjusting the image using the angle between the horizontal axis and the major axis of the otolith (calculated using MATLAB regionprops function).

Harmonic Content

Chapter 3 found that when classifying herring otolith boundaries, best results were obtained when using the first twenty elliptical Fourier harmonics. However, to assess whether classification is dependent on harmonic content for different outlining methods, we conduct classification using differing harmonic content. For this study we construct and test classifiers using the first 10, 20, 30, 40 and 50 Fourier harmonics.

5.1.5 Classification

As in Chapter 4 we perform multiple cross-validated classifications of the transformed boundary sets to test whether algorithm selection, harmonic content or outlining method has greatest impact on classification accuracies.

Algorithm Selection

Whilst the study conducted for Chapter 4 shows no overall difference between classifications using different learning algorithms, tests did show minor decrease in accuracy when using increased harmonic content. We therefore again conduct tests using multiple classification algorithms to discern whether, for age-0 Celtic/Irish Sea herring, choice of learning algorithm interacts with harmonic content. The algorithms used in this study are as follows:

- NB → Naïve Bayes
- BN → Bayesian Networks
- Log → Logistic
- HP → HyperPipes
- J48 → J48/C4.5
- RaF → Random Forest
- IBk → k-Nearest Neighbours
- SMO → Support Vector machine

When paired with harmonic content selection this produces forty classification results, one for each combination of harmonic content and classification algorithm. Each of our eight experiments (four for expert selection and four for threshold) is carried out using each of the forty combinations of algorithm and harmonic content.

5.1.6 Statistical Testing of Results

N-Way analysis of variance tests (ANOVAN) [7] are used to determine which factors have a significant effect on classification accuracies. Our null Hypotheses are that mean results are the same for each group selection for: algorithm selection; differing harmonic content; training data outline method; and test data outline method.

We also test for interaction between selected groups to determine whether algorithm selection, or harmonic content, differs in importance dependent on which outlining methods are used for classifier construction/testing. Post-hoc testing of variance test results are used to show (visually) differences between results.

5.2 Results

The susceptibility to outlining methods results are shown in this section. Section 5.2.1 presents results of expert selection results, where otolith boundary contours were determined by hand-tracing by expert readers (expert-1 and expert-2). Statistical tests are performed on the results to determine whether the expert that traces the boundary impact classification accuracies. Results for thresholding method tests can be seen in Section 5.2.2, where we show classification accuracies obtained when otolith boundaries are determined using the different thresholding methods described in Section 5.1.3. Again, we perform analysis of variance tests on the results to determine impact of thresholding methods on classification accuracies. Finally, Section 5.2.3 shows statistical testing of results to determine if overall outlining method (hand-tracing or thresholding) impacts classification results.

5.2.1 Expert Selection Results

Full results of the expert selection tests are shown in Table 5.2.

Expert Selection Statistical Testing

Analysis of variance testing of hand outlined classification results can be seen in Table 5.3, showing results of testing for learning algorithm, harmonic content, and choice of outlines (by expert) used to construct classifiers. Testing showed that classification accuracy was significantly ($p < 0.05$) impacted by which experts outlines were used to construct the classifiers (training expert). From the table it can be seen that algorithm selection and harmonic content are also shown to impact on classification results, however there is significant interaction between algorithm selection and training data expert. The choice of expert outlines used for testing (=test expert) showed as insignificant during tests ($p=0.76$).

		Learning Algorithm							
	HC	NB	BN	Log	HP	J48	RaF	IBk	SMO
E1vE1	10	75%	65%	65%	65%	60%	70%	80%	90%
	20	85%	55%	90%	65%	45%	85%	75%	85%
	30	85%	55%	80%	60%	75%	70%	80%	85%
	40	80%	75%	85%	60%	75%	80%	80%	85%
	50	80%	70%	80%	60%	75%	65%	65%	70%
E1vE2	10	70%	70%	65%	60%	65%	65%	75%	80%
	20	80%	70%	90%	65%	60%	85%	75%	85%
	30	85%	70%	90%	80%	55%	75%	80%	90%
	40	90%	75%	85%	85%	55%	65%	80%	90%
	50	90%	75%	90%	75%	55%	75%	60%	90%
E2vE1	10	80%	70%	75%	65%	90%	80%	75%	80%
	20	80%	85%	85%	60%	85%	85%	80%	85%
	30	85%	90%	90%	70%	90%	75%	80%	80%
	40	85%	90%	95%	75%	90%	70%	70%	80%
	50	75%	90%	85%	65%	90%	80%	60%	85%
E2vE2	10	75%	65%	70%	70%	90%	85%	75%	85%
	20	85%	90%	90%	80%	85%	90%	85%	90%
	30	85%	85%	90%	75%	80%	70%	80%	95%
	40	85%	95%	95%	70%	80%	75%	75%	95%
	50	85%	90%	90%	75%	80%	80%	50%	80%

Table 5.2: Classification accuracies for expert selection tests. Showing cross-validated accuracies for combinations of algorithm/harmonic content (HC) when: classifiers are built and tested using mean outlines of expert-1 (E1vE1); classifiers are built using expert-1 and tested with expert-2 outlines (E1vE2); classifiers are built using expert-2 and tested with expert-1 outlines (E2vE1); classifiers are built and tested using mean outlines of expert-2 (E2vE2).

	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
algorithm	4499.84	7.00	642.83	15.80	0.00
content	1029.69	4.00	257.42	6.33	0.00
training expert	1722.66	1.00	1722.66	42.34	0.00
=test expert	3.91	1.00	3.91	0.10	0.76
algorithm*content	3247.81	28.00	115.99	2.85	0.00
algorithm*training expert	2921.09	7.00	417.30	10.26	0.00
algorithm*=test expert	139.84	7.00	19.98	0.49	0.84
content*training expert	82.81	4.00	20.70	0.51	0.73
content*=test expert	132.81	4.00	33.20	0.82	0.52
training expert*=test expert	150.16	1.00	150.16	3.69	0.06
Error	3865.47	95.00	40.69		
Total	17796.09	159.00			

Table 5.3: Results of N-way analysis of variance (ANOVAN) of expert selection results. All figures to 2 dcp.

Figure 5.6 shows marginal means diagram for all combinations of expert used to train and expert used to test. It highlights classifiers trained using outlines of expert-2 which (generally) return greater accuracies, regardless of the expert outlines that are used for testing. Further, it reinforces the finding that training and testing experts need not be the same as training experts in order to perform better.

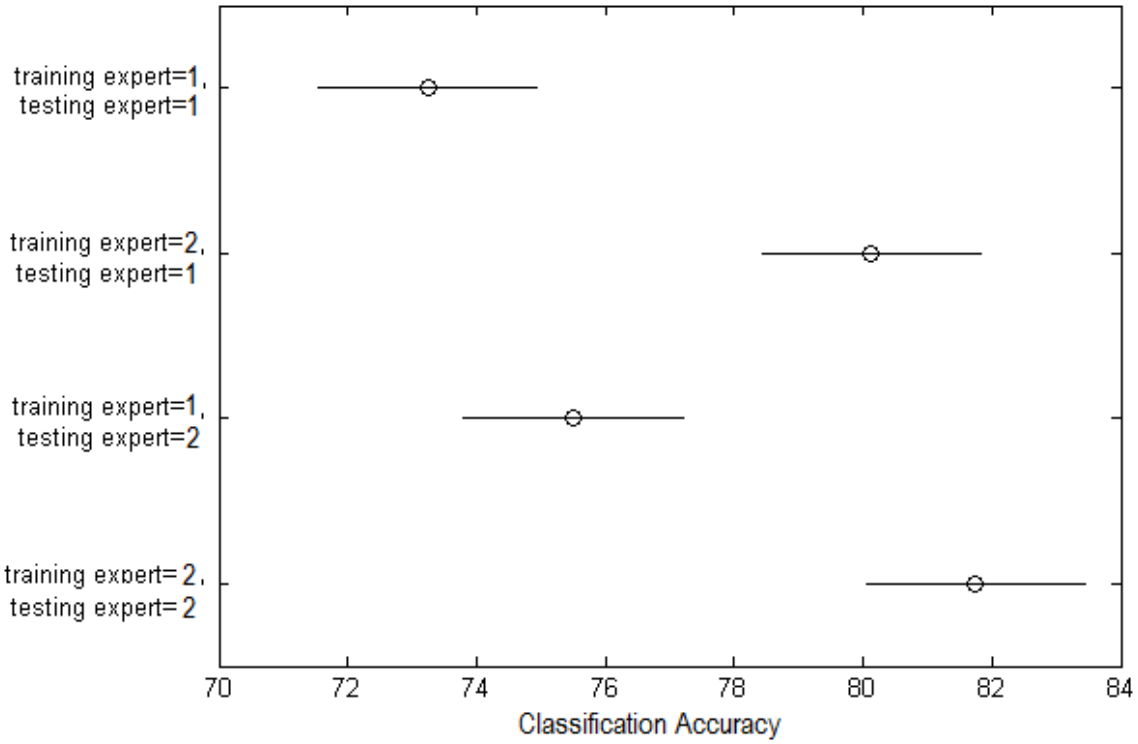


Figure 5.6: Results of post-hoc ANOVAN testing showing classification accuracies for combinations of training/testing experts. Expert-2 outlines used for training return higher accuracies regardless of testing outlines, or whether training outlines were produced by the same expert.

5.2.2 Threshold Selection Results

Full results of the threshold selection tests are shown in Table 5.4.

		Learning Algorithm							
	HC	NB	BN	Log	HP	J48	RaF	IBk	SMO
HvH	10	80%	65%	85%	75%	60%	75%	80%	80%
	20	85%	70%	85%	75%	50%	85%	90%	75%
	30	85%	65%	90%	75%	45%	60%	85%	90%
	40	90%	70%	90%	75%	45%	75%	90%	90%
	50	85%	80%	90%	70%	45%	65%	70%	85%
HvL	10	80%	70%	85%	70%	65%	80%	75%	90%
	20	90%	75%	85%	75%	55%	80%	75%	75%
	30	80%	70%	85%	80%	50%	70%	80%	80%
	40	80%	70%	85%	85%	50%	65%	80%	80%
	50	85%	70%	85%	75%	50%	55%	70%	85%
LvH	10	90%	70%	90%	70%	60%	75%	80%	75%
	20	90%	75%	85%	70%	65%	80%	80%	80%
	30	85%	75%	85%	65%	65%	70%	75%	80%
	40	80%	80%	80%	70%	65%	65%	70%	80%
	50	80%	80%	75%	65%	65%	70%	75%	75%
LvL	10	90%	75%	90%	70%	55%	85%	85%	85%
	20	85%	90%	85%	75%	70%	80%	85%	85%
	30	85%	90%	90%	75%	70%	65%	80%	80%
	40	80%	90%	90%	70%	70%	60%	70%	75%
	50	80%	85%	85%	55%	70%	65%	70%	80%

Table 5.4: Classification accuracies for threshold selection tests. Showing cross-validated accuracies for combinations of algorithm/harmonic content (HC) when: classifiers are built and tested using outlined segmented with high thresholds (HvH); classifiers are built using high threshold outlines and tested with low threshold outlines (HvL); classifiers are built using low threshold outlines and tested with high threshold outlines (LvH); classifiers are built and tested using low threshold outlines (LvL).

Threshold Selection Statistical Testing

Analysis of variance testing for the threshold selection tests can be seen in Table 5.5.

From the table it can be seen that algorithm selection and harmonic content (the

number of Fourier harmonics used for classification) again significantly impact classification accuracies. However, in contrast to expert selection tests, the threshold used to segment otoliths used for training classifiers appears to be far less significant, although not entirely insignificant ($p < 0.10$). Also, expert selection results show that the choice of expert used for training data need not be the same as the training expert, while threshold selection tests show that accuracies for tests where the training and test threshold were the same are statistically distinct from those where the thresholds differ.

	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
algorithm	10909.84	7.00	1558.55	57.36	0.00
content	464.69	4.00	116.17	4.28	0.00
training threshold	82.66	1.00	82.66	3.04	0.08
=test threshold	131.41	1.00	131.41	4.84	0.03
algorithm*content	1450.31	28.00	51.80	1.91	0.01
algorithm*training threshold	1793.59	7.00	256.23	9.43	0.00
algorithm*=test threshold	194.84	7.00	27.83	1.02	0.42
content*training threshold	125.94	4.00	31.48	1.16	0.33
content*=test threshold	24.06	4.00	6.02	0.22	0.93
training threshold*=test threshold	35.16	1.00	35.16	1.29	0.26
Error	2581.09	95.00	27.17		
Total	17793.59	159.00			

Table 5.5: Results of N-way analysis of variance (ANOVAN) of threshold selection results for algorithm, harmonic content and threshold method selection. Including whether the thresholding method used for testing is the same as for the training data (= test threshold), and factor interactions ($*$ factors). All figures to 2 decimal places.

Figure 5.7 shows marginal means diagram for all combinations of threshold used to train and test classifiers. Whilst accuracies are not significantly different for the different training thresholds, mean accuracies are clearly different depending on whether the training/test thresholds are the same.

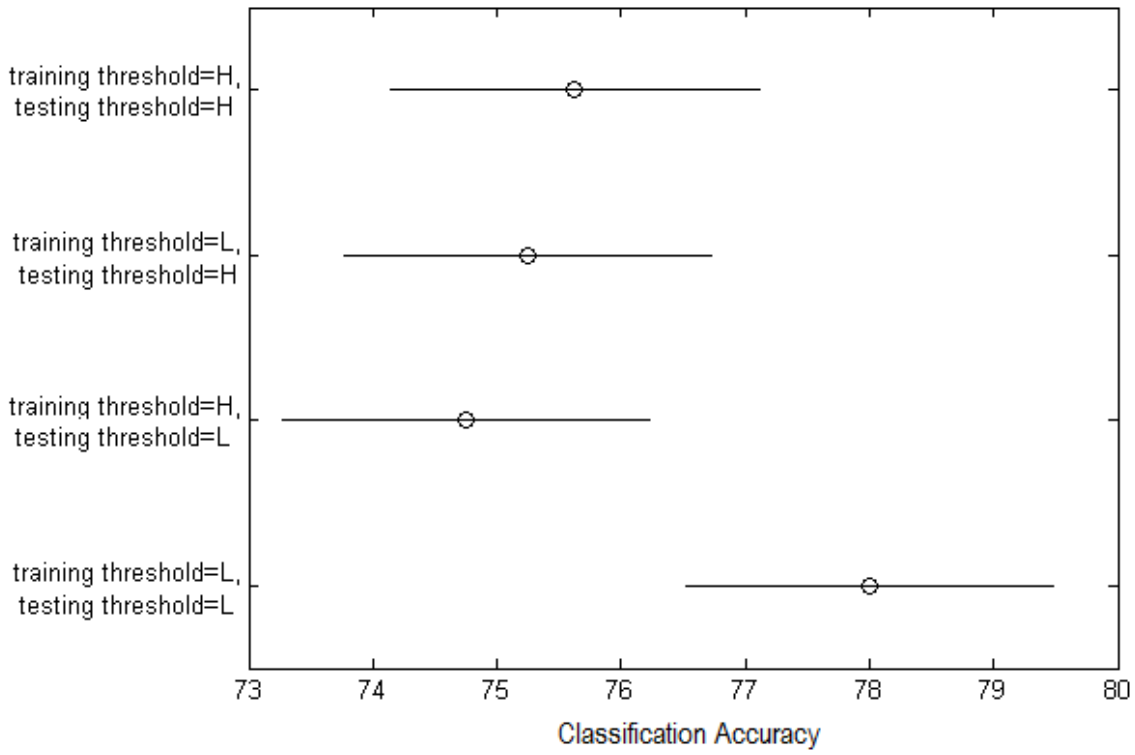


Figure 5.7: Results of post-hoc ANOVAN testing showing classification accuracies for combinations of training/testing thresholding methods. Best results are returned when low (inside-out) thresholding methods are used to outline both training and testing instances.

5.2.3 Method Selection Results Statistical Testing

Results from variance testing of otolith outlining methods are shown in Figure 5.6. The results show that while the selection of classification algorithm impacts classification accuracies most, the choice of outlining method used for otolith segmentation

is also significant (p=0.04), with mean accuracies of hand segmented outlines around 2% higher than those of threshold segmentation.

Whilst method itself is insignificant, interaction tests do show that algorithm choice does interact with the segmentation methodology. Two post-hoc (multi-comparison) tests were performed and the results are shown in figure 5.8. Mean accuracies for the threshold tests are shown in red, expert tests are shown in blue, both with group means denoted as a circle symbol and group interval as lines.

	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
algorithm	12484.7	7.00	1783.53	31.61	0.00
content	830.8	4.00	207.7	3.68	0.01
method	245	1.00	245	4.34	0.04
algorithm*content	3563	28.00	127.25	2.26	0.00
algorithm*method	2925	7.00	417.86	7.41	0.00
content*method	663.6	4.00	165.9	2.94	0.02
Error	15122.7	268.00	56.43		
Total	35834.7	319.00			

Table 5.6: Results ov N-way analysis of variance (ANOVAN) of outlining method selection results, for learning algorithm, harmonic content, and outlining method used (expert trace or thresholding). Including interaction testing ($*$ factors). All figures to 2 decimal places.

The J48/C4.5 algorithm does show interaction with outline method used. However all other algorithms do not return statistically different results dependent on outline method. A number of other algorithms are distinct from one another however, depending on methods. For example: when expert outlines are used, Naïve Bayes and Bayesian Networks are indistinct (margin lines overlap); when threshold methods are used, the same two algorithms return distinct results.

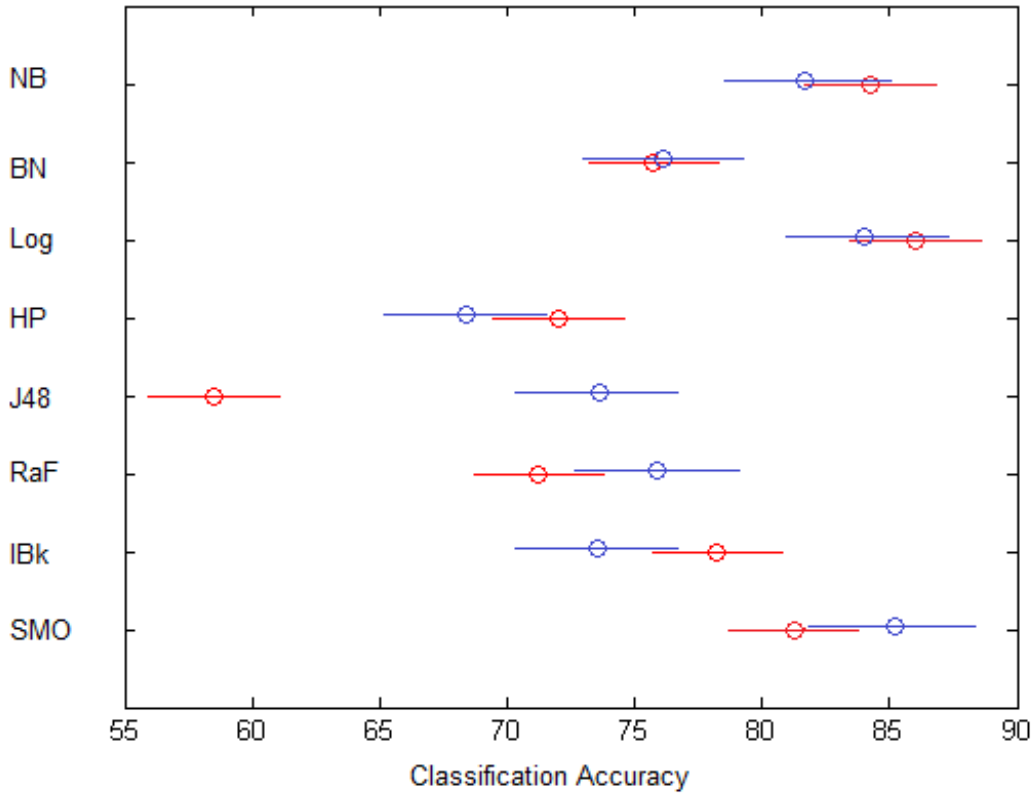


Figure 5.8: Results of post-hoc ANOVAN testing showing classification accuracies for different classification algorithms. Accuracies for threshold tests are shown in red, while expert tests are shown in blue.

Post-hoc test results are show in Figure 5.9 for different amounts of harmonic content for both expert and threshold tests. As before, threshold tests are shown in red, expert tests are shown in blue. It can be seen in the results that expert tests perform marginally better given larger harmonic content, whilst threshold tests show peak accuracies at twenty Fourier harmonics used.

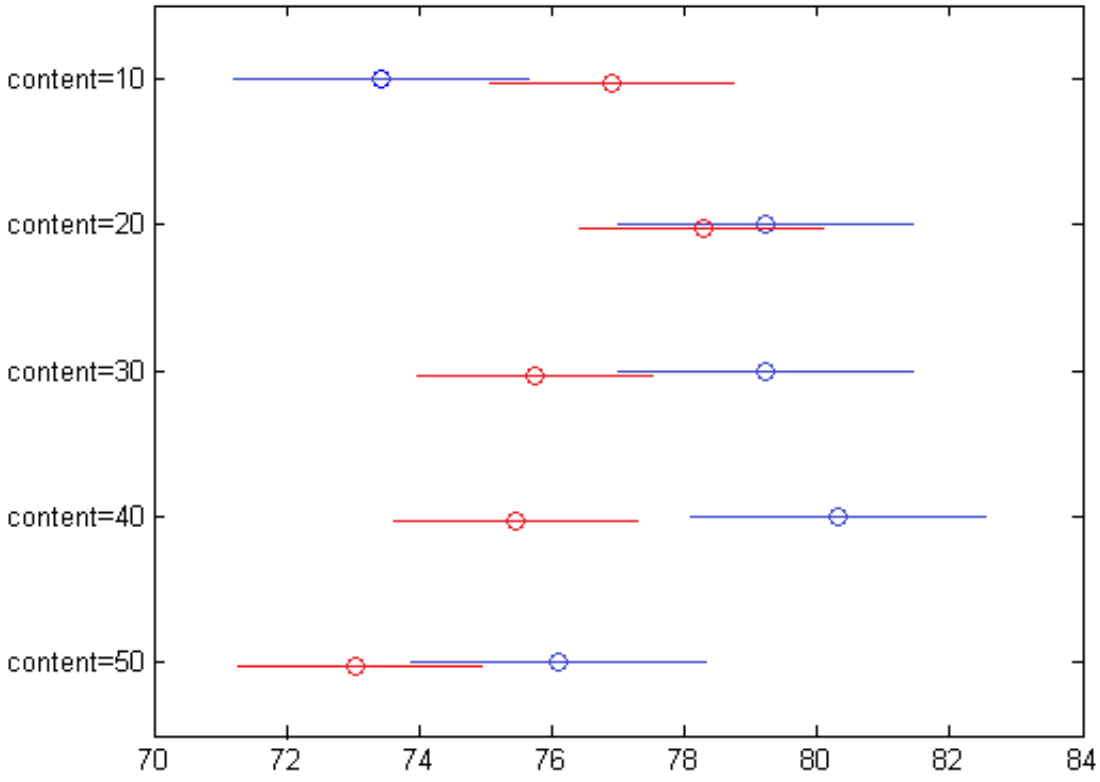


Figure 5.9: Results of post-hoc ANOVAN testing showing classification accuracies for different amounts of harmonic content. Accuracies for threshold tests are shown in red, while expert tests are shown in blue.

When variance testing is performed on a restricted set of results we see similar impact from the outlining method used. In this test we only compare the best performing combination of expert train/test sets to the best performing combination of thresholds. Classifiers built and tested using expert-2 were compared to classifiers built and tested using inside-out (low) thresholding methods. The results of the restricted variance testing are seen in Table 5.7

	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
algorithm	2613.75	7.00	373.39	10.02	0.00
content	601.88	4.00	150.47	4.04	0.01
method	281.25	1.00	281.25	7.55	0.01
algorithm*content	1908.13	28.00	68.15	1.83	0.06
algorithm*method	853.75	7.00	121.96	3.27	0.01
content*method	246.87	4.00	61.72	1.66	0.19
Error	1043.13	28.00	37.25		
Total	7548.75	79.00			

Table 5.7: Restricted testing of method selection. Only classifiers built/tested using the *Low* threshold, and those built and tested using expert-2 outlines are compared ('method' factor). All figures to 2 decimal places.

5.3 Discussion

Overall results of stock classification using the two outline methods, hand-outlines and intensity thresholded, are similar to results obtained from scale normalised Fourier based classification in previous work [11, 12, 22, 25, 63, 90], despite training classifiers with small numbers of samples.

Testing of expert selection shows a substantial change in accuracies over the classification tests. Post-hoc testing shows that classifiers trained using the outlines of expert-2 perform substantially better than classifiers trained using the outlines of

expert-1, despite the use of standard, rather than high-precision computing hardware. Mean accuracy for expert-2 falls at 81.94% whilst the mean accuracy for expert-1 is roughly 7.57% lower, falling at 74.38%.

Whilst the choice of expert outlines used for testing data returns marginally lower results overall, the drop in accuracies does not show as significant. This suggests that the accuracies of classifiers pre-built using expert outlined data should not be significantly impacted when used for further otolith classification, where the outlining expert has changed. Further, additional otoliths may be classified using outlines of multiple experts, enabling outline or Fourier data from multiple locations to be combined and then classified using a single system.

Whilst variance testing of thresholding methods shows minor significance ($p = 0.08$), post-hoc testing perhaps gives a clearer indication of the difference in results. Showing only minor difference in accuracies, classifiers trained using inside-out (Low) thresholding methods perform better overall, with best results returned when testing data is also segmented using the inside-out method. This is supported by the variance testing which shows statistical difference in result depending on whether the test threshold is the same as the training threshold. Therefore further classification tests using thresholded segmentation of otoliths should be performed using conservative segmentation (inside-out methods).

Results suggest, however, that when training classifiers using thresholded outlines, the choice of classification algorithm is the overriding consideration. Logistic (regression) and Naïve Bayes algorithms performed best in threshold based classification tests, with both performing significantly better than most other algorithms. Support vector machines also performed well in threshold selection tests.

In both full and restricted method testing results we see that the selection of method used for otolith segmentation can significantly impact results, although in practice the difference is only minor at around 1.76% in favour of hand outlined boundaries over all results, and 3.75% when comparing the best from each method (Low threshold Vs expert-2).

Whilst there is some interaction between algorithm selection and outlining method, the post-hoc testing shows that the top three performing algorithms (Naïve Bayes, Logistic and Support Vector Machines) remain the top three performing algorithms regardless of outline methods used. Of these three algorithms, the best performing (given both outline methods) is the logistic algorithm. However, the support vector machine algorithm is notable as it also performed well in the study presented in Chapter 3.

The choice of harmonic content is far less clear, and interaction between outlining method and harmonic content apparent. Thresholded methods perform best when given restricted content (20 harmonics) whilst hand outlining classification favours higher content, however, there is little difference between 30/40 harmonics for hand lined tests. Therefore in further studies, where the method of outlining is unclear, or mixed method outlines are used, the first 20-30 Fourier harmonics should be used for best results.

With hand outlining and inside-out thresholding methods performing best from the two methods, this suggests that mid resolution detail of the boundary is an important feature for classification. Such fine concavities of the boundary contour are retained during hand outlining, and are more likely to be retained during inside-out thresholding methods.

5.3.1 Summary

We find that, with regards to stock classifier construction, expertly hand traced boundaries have the potential to return significantly higher accuracies than intensity thresholded images, depending on the expert that performs the task (Objective 3).

Additionally, tests show that classifiers constructed using hand-traced images are more robust to boundaries determined by a different segmentation method. While trace-build classifiers show no marked drop in accuracy when testing another experts' outlines, or thresholded boundaries, classifiers constructed using thresholded boundaries returned significantly reduced accuracies when separating samples segmented using other methods, even testing other thresholding methods (Objective 3).

Whilst not significantly different from other front-runners (Naïve Bayes and Logistic), the Support Vector Machine learning algorithm is of note as it was among the top performers in previous research discussed in Chaper 3 (Objective 4).

Chapter 6

Three Dimensional Otolith Reconstruction And Virtual Slicing Using Synchrotron Tomography

This chapter presents a three dimensional reconstruction of a plaice (*Pleuronectes platessa*) otolith using data acquired at the Diamond Light synchrotron, beamline I12 X-ray source. We assess whether these methods of scanning can be used to produce three-dimensional models of internal otolith structure, or to virtually slice through otoliths without invasive procedures. The research presented relates to research Objective 5, on whether internal otolith scans are possible, where previous attempts to do so have failed. We also investigate whether complex slices along estimated growth planes offer improvement on the flat-slicing methods widely used in the industry (Objective 6).

Whilst factors concerning otolith growth are the subject of many ongoing studies, until recently, experiments were, for the most part, undertaken using whole or sectioned otolith samples and examined using standard optical techniques. Previously, it has only been possible to approximate three-dimensional otolith structure by stacking series of flat-plane slices through the otolith [9]. However, flat plane slices are not able to follow the complex growth patterns that otoliths exhibit. As otiliths grow

they often curve to follow the shape of the ear chamber (towards or away from the viewing angle of otolith images in this thesis). Therefore, a flat plane slice may not give clear indication of where the growth rings lie. Additionally, such slicing methods are destructive in nature, and once performed the otolith cannot be sectioned along other planes.

Therefore x-ray tomography presents the possibility of recovering a complete three-dimensional model of otolith shape, density or composition using non-invasive/non-destructive procedures. Previous attempts to achieve this using conventional absorption x-ray micro-CT were unsuccessful [75] as only the outer surface of the otolith was recovered from scans, despite the authors being hopeful that phase-contrast micro-CT would be successful. We have recently repeated this experiment using a SkyScan 1172 desktop microCT with 5 micron spot size source operating at 60 kV. The detector used for the experiment was a 12-bit CCD camera filled with Gallium Oxide scintillator and fibre-optic connection. Results were consistent with previous studies in that we were unable to resolve sufficient absorption contrast to determine annual growth marks within the sample.

We investigate here whether higher energy x-rays are capable of recovering internal growth features from a single otolith sample, and whether resulting scans can be used to produce a complete three-dimensional model of the otolith. Further, we investigate whether it is possible to construct virtual slices through the otolith along the plane of growth, and compare results to a synthetic flat-plane slice through the same sample.

6.1 Image Capture

Otolith images were captured at Diamond Light [52], the UK's national synchrotron science facility and one of only 23 such facilities worldwide. In this process electromagnetic radiation produced by a linear accelerator is harnessed as a light source and split into multiple 'beamlines' which are used for different types of experiments. The radiation used is vastly more intense than that produced by conventional x-ray systems, often with a final energy in the order of Giga-electron-volts.

The facility's physicists recommended beamline I12 for these experiments, the high energy white-beam synchrotron beamline, known as the Joint Engineering, Environmental and Processing (JEEP) beamline. This beamline has previously been used to investigate a number of high density materials at high rates of precision/resolution previously and as such is a good candidate for internal scans of otoliths.

A Single otolith was used for this proof of concept study and is shown in Figure 6.1. The area 'low-lighted' in the image denotes the portion of the otolith that was scanned by the facility for use in this study. The area was imaged using monochromatic x-rays of wavelength $\lambda = 0.0234nm$ (53 keV), and propagation phase contrast, inherent in this type of monochromatic x-ray images [28, 88], was used to observe variations in the sample. This process has been used with success when studying other specimens exhibiting weak variations in contrast; such as in palaeontological studies, where non-invasive studies were required [94].

The detector used in the I12 beamline for this study was a Cadmium Tungstate (CdWO4) scintillator viewed through bespoke radiation-hard microscope optics (SILL, Germany) by a PCO.EDGE camera (PCO, Germany) with a 2560x2150 pixel sensor (roughly 5 Megapixel). To achieve the phase contrast effect the camera was positioned

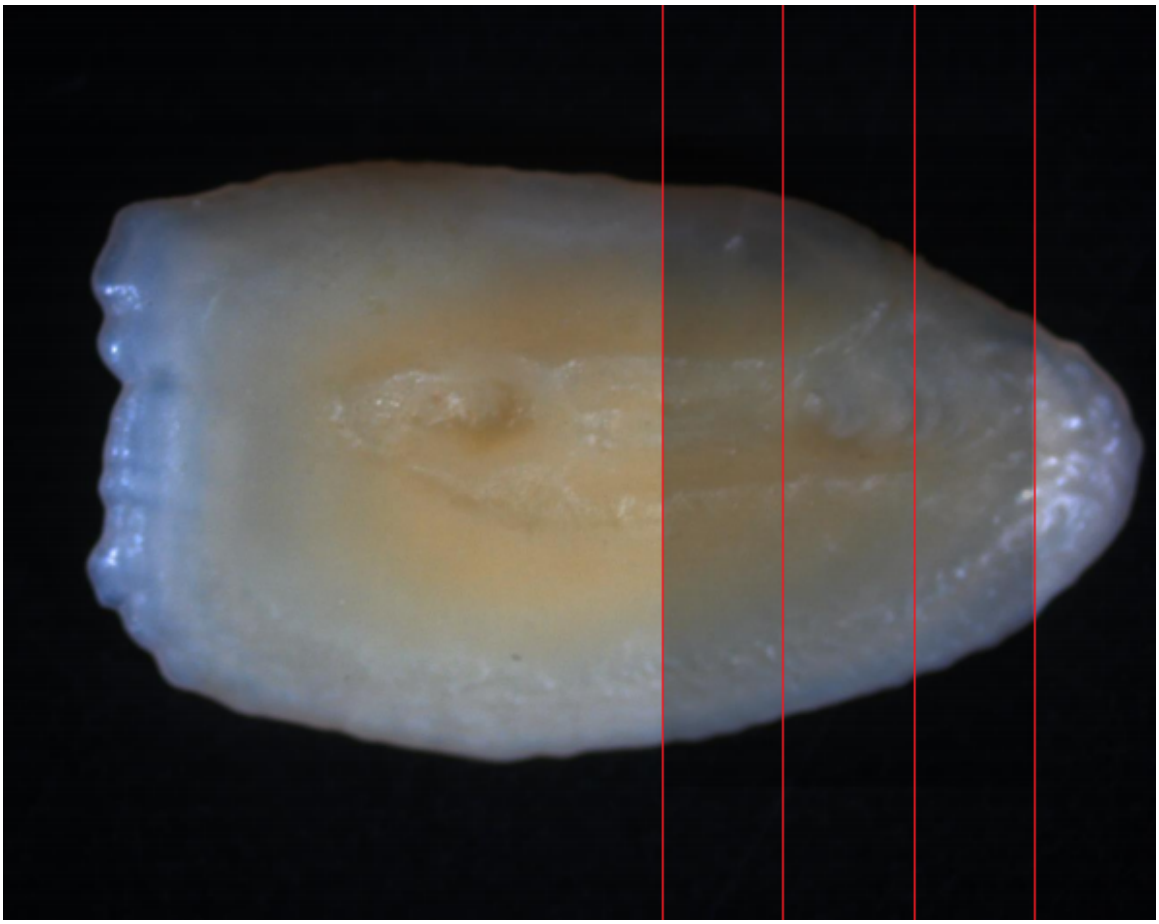


Figure 6.1: The plaice otolith used in this study. The area low-lighted shows the approximate area scanned by the facility and reconstructed in further tests. Red lines show approximate position of the scan images shown in Figure 6.2.

1,000mm beyond the sample. The resulting images have a resolution of approximately 5um per pixel.

Scan images were processed by the facility using methods described by Paganin et al. [72] and reconstructed using filtered back-projection methods [45] creating 1,554 images with 153x160 pixel dimensions, each a virtual slice through the otolith. Examples of these can be seen in figure 6.2. The red arrow on the far left example shows the approximate viewing angle of the camera for the image shown in figure 6.1. A flick-book style animation of the sequential slices can be seen on-line [57].

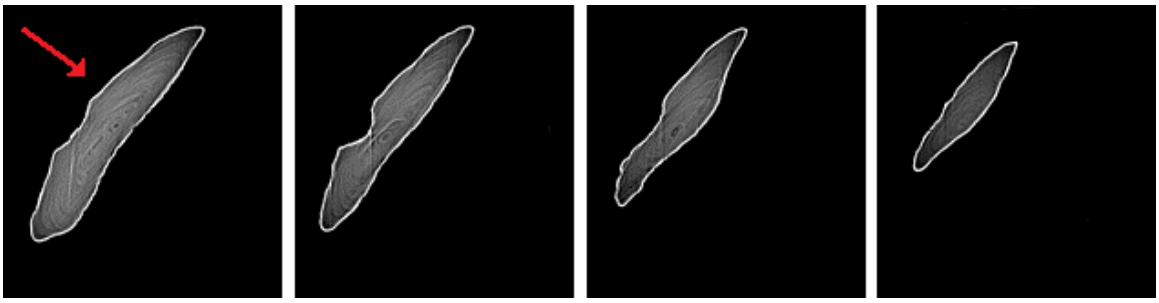


Figure 6.2: Four example images created by the process. Red arrow in the leftmost image shows approximate viewing angle of the camera for the image in figure 6.1.

6.2 Tomographic Reconstruction

Each of the scan images is processed in turn to remove background detail. For each slice we calculate the Otsu threshold [71] (where the threshold is determined from the intensity histogram) which is then used to binarise the image using a simple thresholding technique, so that the foreground (otolith) is shown as ones, and background as zeros. The resulting mask array is multiplied with the original image array, resulting in background pixels being forced to zero whilst otolith pixels remain their original intensity. An example of the Otsu segmentation can be seen in figure 6.3. Whilst background pixels in the original image are almost uniformly zero before processing, this process removes any noise that may be present.

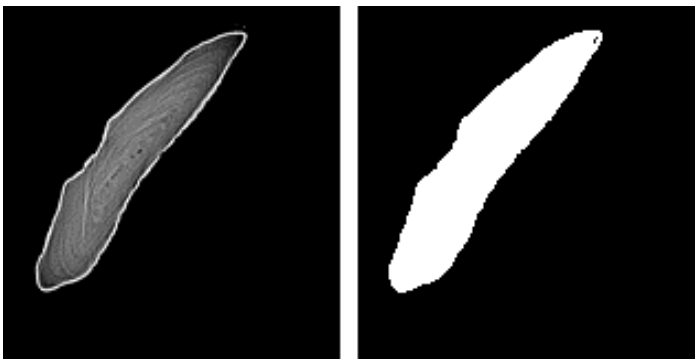


Figure 6.3: Example 'slice' created by the scanning process (left) and after Otsu based segmentation (right).

Once background Noise has been removed, the slice images are 'stacked' to create a 153x160x1554 matrix. Matlab's 'isosurface' function can be used to create a surface around the non-zero portion of the matrix (the otolith). It is important to note that the surface model created by this method does not show surface intensities of the otolith, and so appears uniformly grey, and hollow, by default. In order to show the otolith surface we create a virtual light source which casts shadow and highlights areas of the surface, allowing ridge details etc to be seen. In addition, Figure 6.4 shows the reconstruction with the upper and bottom-most slices added to the ends of the reconstruction to aid visualisation (only the uppermost visible in the figure).

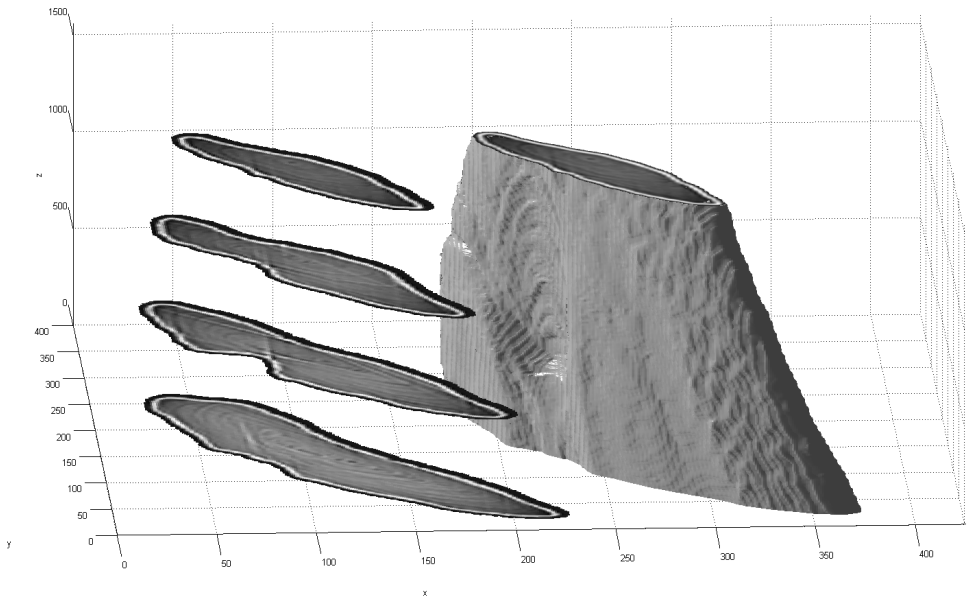


Figure 6.4: Showing 3d reconstruction of the scanned segment of the plaice otolith. The reconstruction does not show otolith image intensities (on the reconstructed surface) any shading is due to a virtual light source highlighting/low-lighting the surface due to texture.

6.3 Virtual Slicing

Physical sections of the otolith are limited to flat-plane slices through the otolith and are destructive in nature. Once an otolith has been sliced through one plane, it is impractical to perform a further slice through a different plane at differing angles. Creating a three dimensional model of the otolith from synchrotron scans allows virtual slices to be taken through the otolith at any angle, and allows slicing the model multiple times.

Additionally, we are able to slice through the otolith along complex, non-flat planes; where we reconstruct a surface along a 'spline-plane', fitted to the approximated maximum seasonal growth.

In the simplified example shown in this chapter(Figures 6.5–6.10), only four scan images were used to calculate the plane by which to slice the otolith, those seen previously in this chapter (Figure 6.4). Points of maximum growth (ring peaks) are selected by an expert reader for each image using visible ring peaks in each image. A bicubic spline is then fitted to these selected points, of equal length in each image. Examples of these selections are seen in Figure 6.5 where the selected ring peaks are shown in red, and the spline fitted is shown in green (two hundred interpolated points per spline).

However, to ensure that all points are equally spaced along the spline, rather than equally distributed between selected points, further processing is required before fitting the spline. Experts are not required to mark the same growth rings in each image, neither is there any requirement to mark the same number of points per image. Where points are not equally spaced along each two-dimensional spline, the spline-plane may not be constructed correctly. Figure 6.6 shows the first of our four

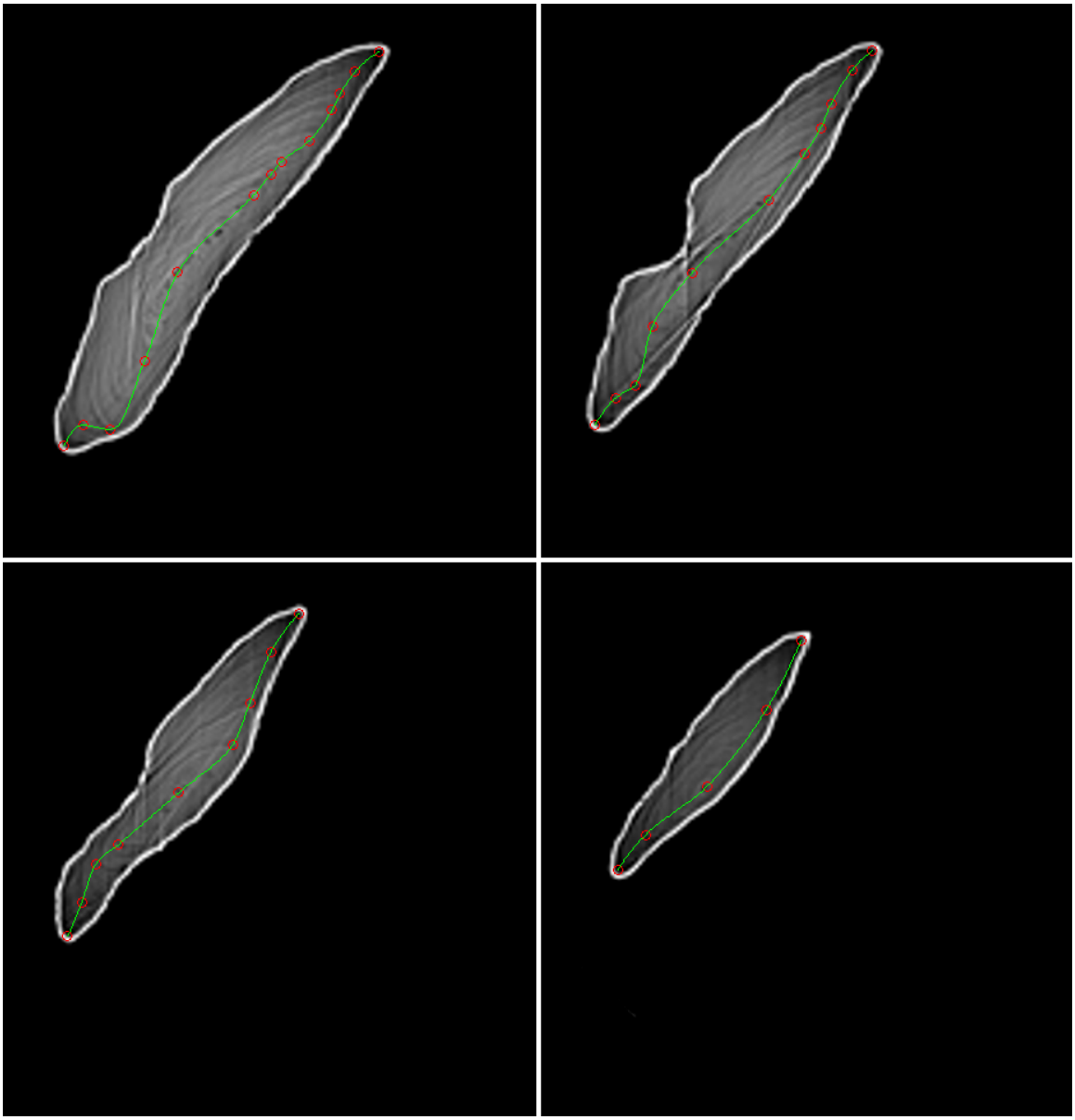


Figure 6.5: The four previously seen scan images overlayed with calculated spline (green) and expert designated growth peaks (red).

scan images used for the example reconstruction, with selected points marked as red circles and points joined by straight lines.

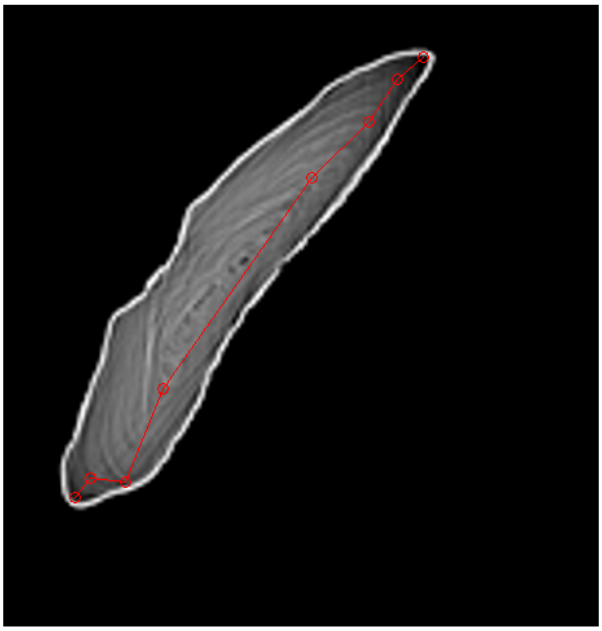


Figure 6.6: Expert selected points of maximum growth (red circles) joined by straight lines.

When a bicubic spline is fit to the selected points using standard Matlab methods, the interpolated points are distributed equally between line segments. Figure 6.7 shows an example of this issue, with a total of fifteen interpolated points (low number for clarity) these are distributed with two points along each spline section, one on each expert selected point, and one between each selected point, rather than spaced equally along the whole spline.

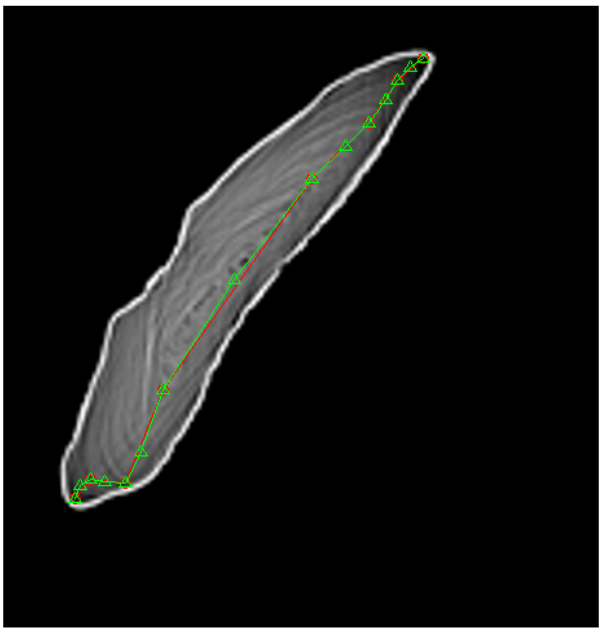


Figure 6.7: Bicubic spline shown in green given twenty interpolated points (green triangles). Interpolated points are (incorrectly) distributed equally between expert selected points, rather than distributed evenly along the length.

To overcome this issue each of the expert selected points is marked as its Euclidean distance along the linear 'curve' for interpolation (shown as a red line in figure 6.6), rather than as its point in the sequence of expert selections. Figure 6.8 shows the selected points as sequence number along the curve (red), and as distance along the curve (blue). Note that for clarity the distance along curve is rescaled to 1-8 in this example, so that first an last points have the same distance along the curve as their sequence number (distance/sequence shown on the x-axis). In practice the distance is rescaled 0-n, where n is the number of interpolation points required. The rescaled distances along the curve are then used in place of sequence numbers as the interpolation landmarks.

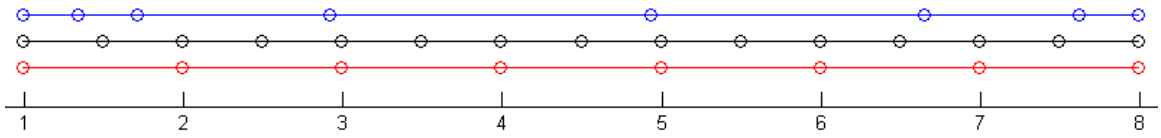


Figure 6.8: Showing landmark values for the spline interpolation. Black shows the interpolation points to be created to ensure even distribution; red shows landmarks as sequence numbers; blue shows landmarks as distance along curve (rescaled).

It can be seen from Figure 6.8 that interpolated points are evenly distributed between sequence number points (red) but not distributed between distance points (blue). Indeed, when using this small number of interpolation points the first and last line segments have no corresponding interpolated points (other than the selected end points).

Interpolation is carried out in this manner for both x and y coordinate arrays for the expert selected landmarks. This creates x/y coordinates for all interpolated points. Figure 6.9 shows our simple example with 15 interpolated points. Whilst these points were evenly distributed between line segments when using sequence number landmarks, using distance landmarks the points are distributed (approximately) evenly along the length of the spline. Whilst this method does not space interpolated points exactly evenly, due to difference between distance along the linear curve and bicubic curve, the method is sufficient for this proof of concept.

Once each of the four scan images have been processed in this way, the resulting splines are used to create a spline-plane. As the distance landmark method creates an equal number of evenly spaced points, the corresponding points from each processed scan image can be used to interpolate points on all other scan images. Figure 6.10 shows our example. Each of the four splines created from the four processed scan

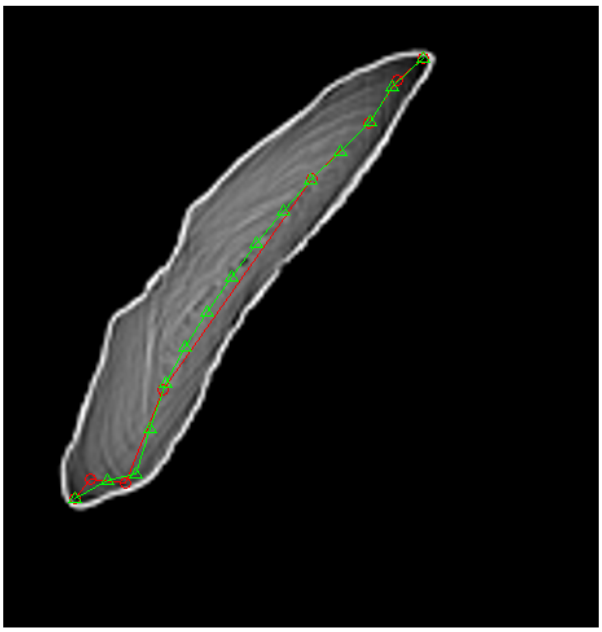


Figure 6.9: Correct bicubic spline shown in green given twenty interpolated points (green triangles). Points are distributed (approximately) equally evenly along the length of the spline.

images are shown in green, in our example fifteen interpolated points have been calculated along each spline. Each of these interpolated points is then used to interpolate fifteen points between all 1,554 scan images. For example, the four spline points marked with green circles are used to interpolate the points along the 'vertical' spline shown in blue.

The calculated spline-plane is stored as three n-by-m coordinate grids, where n is the number of interpolated points per spline, and m is the number of vertically interpolated points. The three grids hold the x, y and z coordinates for each of the spline-plane points. These grids can then be used to create a virtual slice through the spline, and seen as the otolith scan intensity at those coordinates in the scan image stack created during 3d reconstruction. A resulting model can be seen in figure 6.11

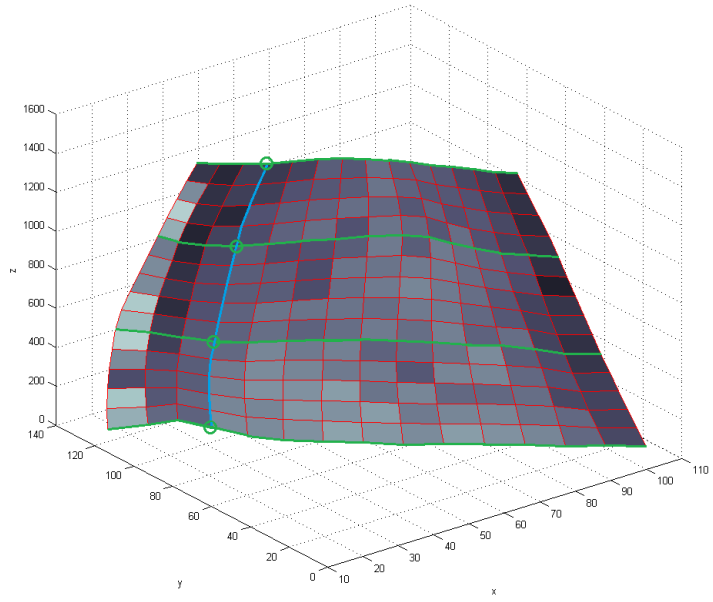


Figure 6.10: Showing spline-plane interpolation points for our simplified example. Expert marked and interpolated splines are shown in green. Green circles are corresponding points on 2d splines that are used to interpolate points on the spline plane along that 'vertical' blue line.

where n=m=200 created from ten expert landmarked scan images. An animation of the spline-plane slice can be seen on-line [57]

6.3.1 Comparison of Virtual Slices to Flat-plane Slices

As the interpolated spline-plane is a collection of coordinates for an n-by-m slice, the plane is easily flattened and shown as a square grid. Figure 6.11 shows a 200-by-200 point slice interpolated from ten scan images, landmarked by an expert reader. Shown in figure 6.12 is the same 200x200 plane, flattened and stretched into a square grid, whilst keeping the intensities calculated using the complex plane. Figure 6.13 shows a flat virtual slice through approximately the same angle as the spline-plane.

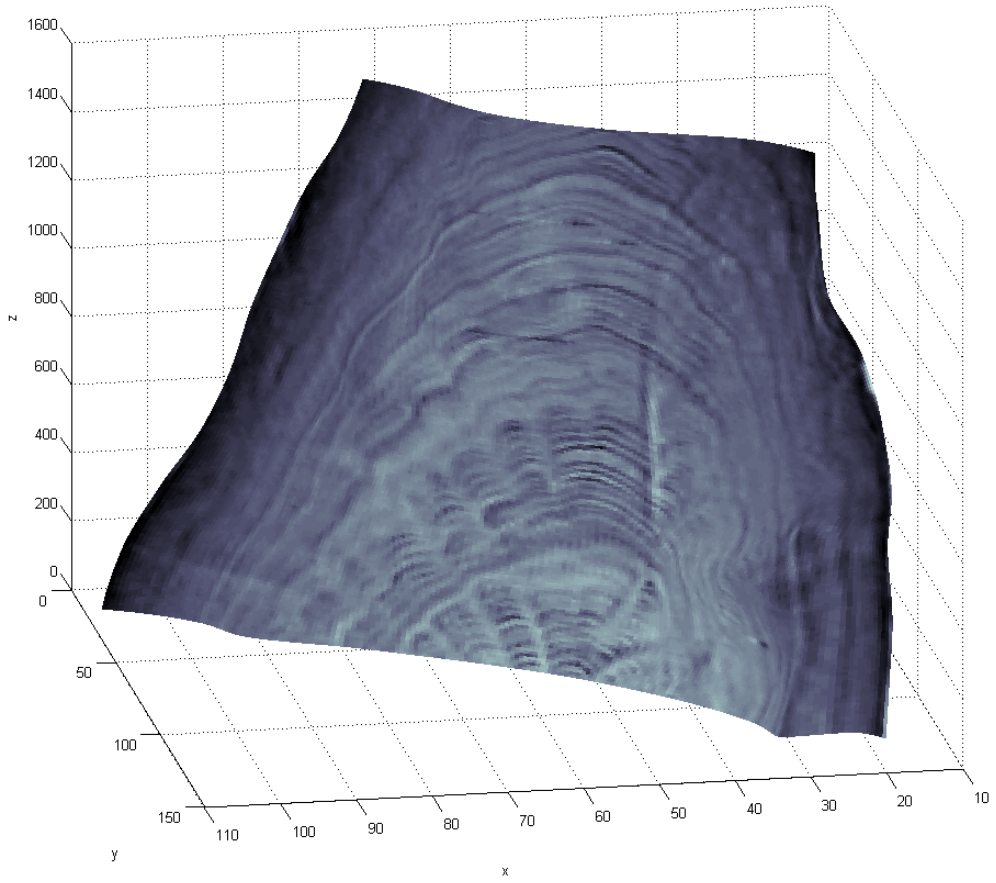


Figure 6.11: Complex virtual slice through the plane of maximum seasonal growth of a plaice otolith segment.

Clear differences can be seen between the flattened spline-plane and the flat slice when compared in this manner. Not only are the growth rings much clearer in the flattened spline-plane, but ring shape is also remarkably different.

6.4 Discussion

Previous synchrotron studies of otoliths have used Synchrotron Rapid Scanning X-ray Fluorescence (SRS-XRF) to investigate trace elements (e.g. (Doubleday et al. 2014, Limburg et al. 2007)). This technique is inherently limited to an examination of the

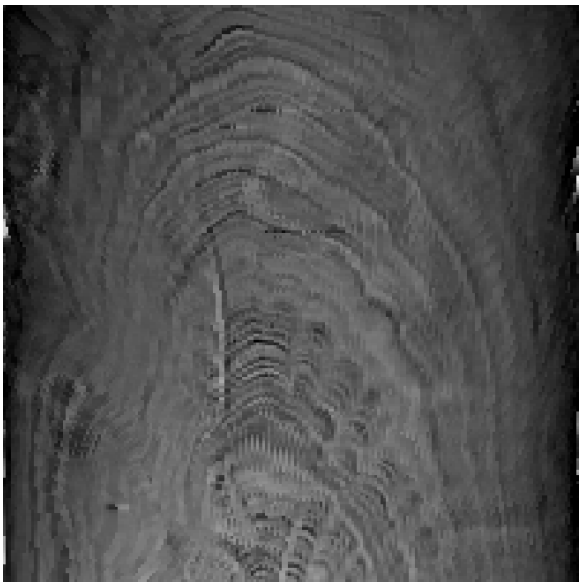


Figure 6.12: Spline-plane after flattening/stretching to a square grid, retaining intensities at spline-plane coordinates.

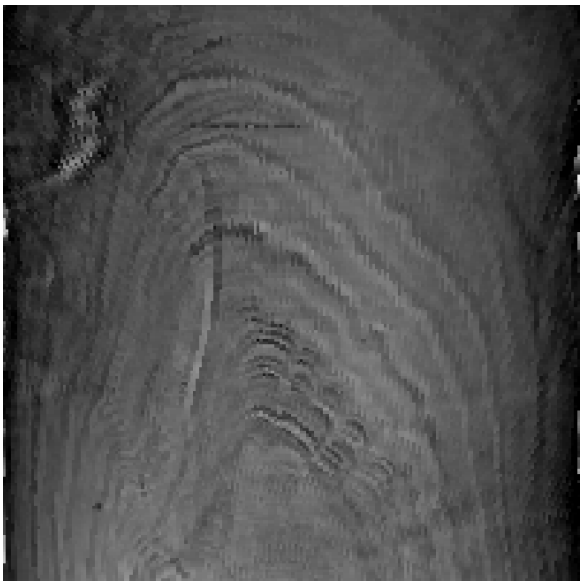


Figure 6.13: Image created by stretching a flat-plane slice through the otolith at angle approximately equal to the spline-plane.

otolith surface and so may require thin sections of material to be prepared, resulting in damage to the specimen. The preliminary data from this virtual slicing study using the JEEP beamline are encouraging because they illustrate the potential for more accurate measurement of total seasonal accreted volume (as opposed to an estimate taken from a 2-D section). 3-D analysis is also potentially more robust to anomalous secondary growth signatures that do not correspond to seasonal deposits. Such artefacts continue to challenge even the most experienced readers of 2-D otolith sections thereby contributing to uncertainties in age estimates, and consequently stock assessments [29]. Currently, synchrotron studies are very expensive but as the technology becomes more accessible virtual otolith studies using computer graphics could provide a historical perspective for each individual within their environmental context.

We hope to build on this preliminary study through funded access to the beamline at DiamondLight to allow us to render whole otolith specimens, investigate species

specific growth traits, and factors that affect the direction of maximal accretion, which seems to change with age (Fig. 3). This may help in understanding why certain species are particularly problematic for human otolith readers and provide a valuable insight into how and why the accreted biomineralisation is related to anatomy, physiology and life-cycle. Preliminary observations of the specimen suggest the study was totally non-invasive and that the otolith in this study was undamaged by the experimental process. This feature of tomographic analysis could be particularly valuable, for example where the otoliths of individual fish have been recovered accompanied by archival data storage tags which may have recorded ambient experience of the same fish for periods over seasons or sometimes years (Sturrock et al. 2012). Such otoliths with accompanying 'ground-truth' data represent a rich resource of information concerning individual lifetime movements.

6.4.1 Summary

We successfully show for the first time that high energy scanning methods presented in this chapter are able to render both the external *and internal* otoliths structures using non-invasive methods (Objective 5).

Preliminary testing has shown that virtual plane slicing through otoliths rendered using these scanning techniques offer increased clarity of ring structure. Additionally we show discrepancies between actual growth ring position at point of maximal yearly growth, and the visible location of the growth ring during flat plane slicing (Objective 6).

Chapter 7

Thesis Conclusions

This chapter summarises the work presented in this thesis. We give a summary of the findings from Chapters 3 to 6, followed by our main findings in relation to our stated research objectives. We close with some notes on potential future work, including suggestions for methods used in the wider otolith sciences field.

7.1 Summary of findings

Whilst Shapelet based classification is showing to be a useful tool in other fields, our studies show that classification of herring otoliths using shapelets methods returned marginally inferior results to classification using the unprocessed boundary. Further, no individual candidates extracted using the shapelet method showed to be particularly discriminatory of otolith class, as was the case with prior tests performed on the SHAPE dataset.

The (double) cross-validation procedure used when extracting and testing shapelet based methods reduced the size of training sets by approximately one-fifth compared to tests conducted with other transform methods, while returning results similar to other tests on the same dataset. Increasing the dataset may mitigate this reduction

and return higher classification accuracies. However, given no candidates show as particularly discriminatory, we find this eventuality unlikely.

Similarly, Curvature Scale Space methods (when used to separate herring stocks) do not provide any significant improvement on Fourier based methods already in widespread use, despite being an industry standard specifically established to encode object shape. The improvements seen when pre-indexing methods are used based on gross shape metrics (circularity, eccentricity and CSS aspect-ratio) implies that otoliths of different stock overlap in scale-space. The difference in CSS pre-indexing thresholds returning peak results during early tests suggests, however, those gross shape differences between classes may vary depending on the task at hand. For example, eccentricity of herring otoliths from different stocks may show high inter-class variance, whilst variance may be high for convex area between sprat stocks.

Whilst we show that classification of juvenile otoliths is certainly possible using complex transformations such as Fourier, Shapelets or Curvature Scale Space, the addition of simple measurements such as otolith length, height or aspect-ratio, or refraining from performing scale normalisation of Fourier harmonics, significantly increases classification accuracies. The importance of early growth for classification therefore supports methods whereby early incremental growth microstructure measurements are used for stock separation.

The age of otoliths used for classification studies is likely to impact classification accuracies. Experiments using mixed age otoliths may suffer due to the differing defining feature (shape or size) depending on otolith age composition of the dataset. This impact may be overcome by separating samples (by age) and applying separate

classification methods, or potentially by performing size-inclusive tests using inner ring traces for the older otoliths, where ring extraction has been validated.

The method of otolith segmentation has shown to be an important consideration when boundaries are used to construct Fourier based classifiers. Where hand-outlining is used for segmentation, the expert reader employed to trace the boundaries used for training may significantly effect the accuracy of the resulting classifier. In contrast, the range of thresholds that can be employed for automatic extraction is large, providing larger margin of error when segmenting training samples.

Whilst hand-trace based classifiers are less robust with regard to training set composition, the constructed classifiers are (statistically) more accurate than those built using automatically extracted outlines. Additionally, classifiers constructed using hand-traces show little sensitivity to the expert employed to outline test data. Classifiers constructed using automatic boundary extract return lower accuracies when test data is segmented using alternate thresholding methods, with best results returned using inside-out thresholding for both training and testing classifiers.

A number of factors indicate that low to mid resolution otolith detail may hold the best potential for stock discrimination of younger specimens. Retaining scale information during Fourier analysis increases classification accuracies, whilst during transform studies the best performing classifiers were constructed using size-inclusive data. Peak Fourier based classification results (both scale-normalised and non-normalised) are returned when classifiers are constructed using restricted harmonic content (around 20-30 Fourier harmonics). In many cases inclusion of higher level harmonics, or exclusion of the mid level harmonics, reduces classification accuracy.

Differences in accuracy between classifiers constructed by experts may also be explained by variation in the mid level detail captured by each expert, where one expert may retain more mid-level boundary detail than another. Inside-out thresholding methods may also retain more detail, particularly fine concavities present along the ventral edge of many herring otoliths.

Choice of learning algorithm used for constructing classifiers appears to be a much more complex issue. Discrete age tests show little pattern in best performing algorithms between ages, however when building classifiers for age-0 herring (from either of the datasets used in these studies) choice of algorithm is seen as significant ($p < 0.05$) with Support Vector Machines being among the top performing, regardless of outline method used.

We have shown that three-dimensional reconstruction of internal otolith features is indeed possible given correct scanning procedure and high energy x-ray sources, and confirmed the feasibility of non-invasive tomographic analysis of internal features. Such reconstructions suggest exciting new possibilities for otolith growth analysis, and virtual slicing of otoliths reveals many potential possibilities for future analysis.

Differences seen between a flattened growth-plane spline and that of a virtual flat slice may also go some way to explain the sprat anomaly encountered in chapter 4 whereby classification accuracies were significantly reduced when growth rings were used as an approximation of age-0 growth. Growth vectors can be seen to be non-linear as the scanned sample aged, and the true age-0 growth ring would clearly have been occluded when examining otolith surface growth rings.

7.2 Main Findings

The main findings of this thesis are given here. They are structured according to our initial research objectives laid out in Chapter 1

1. To critically compare recent techniques taken from the fields of computer vision and time-series analysis to methods traditionally used for otolith classification.

Using comparison testing we have shown that, with regards to otolith based stock classification, curvature scale space and time-series shapelet methods *do not* perform as well as methods widely used in otolith shape analysis (Fourier analysis).

2. To establish whether otolith age impacts classification accuracies dependent on scale-invariance of transformed boundaries.

We clearly show that age composition of otolith datasets significantly impacts classification accuracies depending on boundary encoding methods used. Size inclusive methods, whether non-scale-normalised Fourier harmonics or simple morphometric indices return statistically higher accuracies for younger samples, whilst size-exclusive encoding methods are superior when building/testing classifiers using older otolith samples.

3. To determine whether the methods used for otolith boundary determination impact classification accuracies.

Classifiers constructed using hand traced boundaries return significantly higher classification accuracies when used to separate fisheries stocks. Whilst automatically segmented boundaries remove potential human error/bias, the resulting classifiers are less robust to changes in segmentation method. Therefore until threshold segmentation methods are standardised, hand-tracing outlines may generate more robust

and distributable classifiers for stock-separation, regardless of segmentation methods and/or learning algorithms.

4. To establish whether the choice of machine learning algorithm affects classification accuracies for otolith stock separation.

Best Performing learning algorithms appear to be specific to the classification task at hand. However Support Vector Machines do appear to be more robust to boundary segmentation and encoding methods. Whilst this algorithm did not always return peak results during tests, it was commonly found in the top results throughout this work.

5. To establish whether three dimensional modelling of otoliths is possible using non-invasive methods, and whether complex plane slices give clearer indication of internal otolith structure.

High energy x-ray synchrotron scans have shown to be able to penetrate beyond the otolith surface and model internal structures that are easily reconstructed as a three-dimensional virtual otolith. Although some imaging artefacts are evident in the reconstruction (e.g. the corona around the air/otolith interface), the proof of principle study provides a glimpse of the possibilities for otolith imaging using phase contrast synchrotron radiation and also demonstrates the potential for further 3-D rendered tomographic reconstructions using the JEEP beamline.

6. To establish whether complex plane slices give clearer indication of internal otolith structure than traditional flat plane slicing.

Synchrotron scans enable us to perform complex slices through the otolith along whichever plane we choose. Slicing through the plane of estimated growth has shown

distinct differences between actual growth, and that seen when the otolith is sliced using industry methods.

7.3 Future work

The study presented in this thesis on whether time-series shapelets can improve on industry methods (Fourier based analysis) focus on a sole classification task. Further research should be carried out to determine whether poor performance of shapelet encoding methods is unanimous, or whether additional stock-separation tasks return increased accuracies.

Our results suggest that mid resolution boundary detail may hold more potential for otolith shape classification. Future work may be undertaken to determine which elliptical Fourier harmonics offer higher stock separation capability. Our suggestion in the absence of such research is to construct classifier using the first 20–30 harmonics.

Further classification tasks should take into consideration cohort age before selecting boundary transform methods used to encode boundaries prior to classification. Whilst our research has suggested that younger samples hold more potential for stock-separation, further research should investigate whether this is the case with all species, and whether there may be an age at which shape based classification becomes futile.

In future studies, choice of outlining method used for classifier construction should take into consideration whether further classification tasks are to be undertaken with the constructed classifier. Those constructed using expertly hand-traced otoliths appear to be more robust to difference in outlining method, perhaps allowing tests to be carried out on data outlined using unknown or unpredicted methods.

Chapter 6 presents reconstruction and complex-plane slicing of a single Plaice otolith. However the work suggests that high-energy scanning could be a valuable method for future research. When the affect of age on classification is considered, scans of this type may facilitate the 'stripping' of otolith growth rings in three dimensions, essentially returning the otolith to an earlier stage of fish development, allowing the sample to be classified at a younger age.

The work presented here has only performed classification tasks using single classifiers. Ensemble methods, where samples are classified using a range of techniques before using the results to determine the predicted class, are a potential avenue to investigate.

It is our suggestion, that in future research where testing of individual learning algorithms, or use of ensemble methods may not be possible, the Support Vector Machine learning algorithm should be considered for use when training classifiers.

Abbreviations

AFBI	Agri-Food Biosciences Institute
ANOVA	Analysis Of Variance
ANOVAN	Analysis of Variance, N-way
BL	Boundary Length
BN	Bayesian Networks
Cefas	Center for Environment, Fisheries and Aquaculture Science
CSS	Curvature Scale Space
DC	Direct Current
EFD	Elliptical Fourier Descriptor
EFT	Eliptical Fourier Transform
FFT	Fast Fourier Transform
GMIT	Galway-Mayo Institute of Technology
HC	Harmonic Content
HP	HyperPipes
IBk	k-Nearest Neighbour
ICES	International Center for the Exploration of the Seas
IOS	International Otolith Symposium
JEEP	Joint Engineering, Environment and Processing
Log	Logistic (learning algorithm)
LOOCV	Leave One Out Cross Validation
MPEG7	Moving Picture Expert Group version 7
NB	Naïve Bayes
NN	Nearest-Neighbour
NNDTW	Nearest Neighbour Dynamic Time Warping
PCA	Principle Componant Analysis
RaF	Random Forest
RoF	Rotation Forest
SMO	Support Vector Machines
SPa	Shape Parameter transformation
UV-B	Univariate Boundary
WEKA	Waikato Environment for Knowledge Analysis

Bibliography

[1] S. Abbasi, F. Mokhtarian, and J. Kittler. Curvature scale space image in shape similarity retrieval. *Multimedia systems*, 7(6):467–476, 1999.

[2] S. Abbasi, F. Mokhtarian, and J. Kittler. Enhancing css-based shape retrieval for objects with shallow concavities. *Image and Vision Computing*, 18(3):199–211, 2000.

[3] D. Aha and D. Kibler. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.

[4] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2004.

[5] A. Amanatiadis, VG Kaburlasos, A. Gasteratos, and SE Papadakis. Evaluation of shape descriptors for shape-based image retrieval. *Image Processing, IET*, 5 (5):493–499, 2011.

[6] Christophe Ambroise and Geoffrey J McLachlan. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the national academy of sciences*, 99(10):6562–6566, 2002.

[7] F. J. Anscombe. The validity of comparative experiments. *Journal of the Royal Statistical Society. Series A (General)*, 111(3):181–211, 1948.

[8] Anthony Bagnall, Eamonn Keogh, Stefano Lonardi, Gareth Janacek, et al. A bit level representation for time series data mining with shape based similarity. *Data Mining and Knowledge Discovery*, 13(1):11–40, 2006.

[9] KM Bailey, AL Brown, A Nishimura, and MT Reilly. Three-dimensional imaging of walleye pollock otoliths: reconstruction from serial sections and fluorescent laser cytometry. *Journal of fish biology*, 47(4):671–678, 1995.

[10] A Bani, S Poursaeid, and Víctor M Tuset. Comparative morphology of the sagittal otolith in three species of south caspian gobies. *Journal of fish biology*, 82(4):1321–1332, 2013.

[11] G.A. Begg and R.W. Brown. Stock identification of haddock melanogrammus aeglefinus on georges bank based on otolith shape analysis. *Transactions of the American Fisheries Society*, 129(4):935–945, 2000.

[12] GA Begg, WJ Overholtz, NJ Munroe, et al. The use of internal otolith morphometrics for identification of haddock (melanogrammus aeglefinus) stocks on georges bank. *Fishery Bulletin*, 99(1), 2001.

[13] G.A. Begg, S.E. Campana, A.J. Fowler, and I.M. Suthers. Otolith research and application: current directions in innovation and implementation. *Marine and Freshwater Research*, 56(5):477–483, 2005.

[14] S. Bermejo, B. Monegal, and J. Cabestany. Fish age categorization from otolith images using multi-class support vector machines. *Fisheries research*, 84(2): 247–253, 2007.

[15] M. Bober. Mpeg-7 visual shape descriptors. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6):716–719, 2001.

[16] L.J. Bolle, A.D. Rijnsdorp, W. van Neer, R.S. Millner, P.I. van Leeuwen, A. Ervynck, R. Ayers, and E. Ongenae. Growth changes in plaice, cod, haddock and saithe in the north sea: a comparison of (post-) medieval and present-day growth rates based on otolith measurements. *Journal of sea research*, 51(3): 313–328, 2004.

[17] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[18] Deirdre Brophey, Paula Haynes, Haritz Arrizabalaga, Igaratza Fraile, Jean Fromentin, Ivan Garibaldi, Fulvio andKatavic, David Tinti, Fausto andMacias, Dheeraj Saadet Karakulak, F andBusawon, Ai Hanke, Alex andKimoto, Osamu Sakai, Simeon Deguarra, Abid Nouredinne, and Miguel Santos. Otolith shape variation in bluefin tuna (thunnus thynnus) from different regions of the north atlantic: a potential marker for stock origin. page N/A In final review, 2015.

[19] D. Brophy and B.S. Danilowicz. Tracing populations of atlantic herring (clupea harengus l.) in the irish and celtic seas using otolith microstructure. *ICES Journal of Marine Science: Journal du Conseil*, 59(6):1305–1313, 2002.

[20] D. Brophy, B.S. Danilowicz, and P.A. King. Spawning season fidelity in sympatric populations of atlantic herring (clupea harengus). *Canadian Journal of Fisheries and Aquatic Sciences*, 63(3):607–616, 2006.

[21] N. Burke, D. Brophy, and P.A. King. Otolith shape analysis: its application for discriminating between stocks of irish sea and celtic sea herring (clupea harengus) in the irish sea. *ICES Journal of Marine Science: Journal du Conseil*, 65(9):1670–1675, 2008.

[22] N. Burke, D. Brophy, and PA King. Shape analysis of otolith annuli in atlantic herring (clupea harengus); a new method for tracking fish populations. *Fisheries Research*, 91(2):133–143, 2008.

[23] N. Burke, D. Brophy, P.J. Schön, and P.A. King. Temporal trends in stock origin and abundance of juvenile herring (clupea harengus) in the irish sea. *ICES Journal of Marine Science: Journal du Conseil*, 66(8):1749–1753, 2009.

[24] S.E. Campana and J.M. Casselman. Stock discrimination using otolith shape analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 50(5):1062–1083, 1993.

[25] M. Cardinale, P. Doering-Arjes, M. Kastowsky, and H. Mosegaard. Effects of sex, stock, and environment on the shape of known-age atlantic cod (gadus morhua) otoliths. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(2): 158–167, 2004.

[26] Gavin C Cawley and Nicola LC Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, 11:2079–2107, 2010.

[27] Media Cybernetics. *OPTIMAS, version 6.2, eighth ed. Media Cybernetics*. Silver Spring., Maryland, 1996.

[28] TJ Davis, D Gao, TE Gureyev, AW Stevenson, and SW Wilkins. Phase-contrast imaging of weakly absorbing materials using hard x-rays. *Nature*, 373(6515): 595–598, 1995.

[29] Helene De Pontual, Anne Laure Groison, Carmen Piñeiro, and Michel Bertignac. Evidence of underestimation of european hake growth in the bay of biscay, and its relationship with bias in the agreed method of age estimation. *ICES Journal of Marine Science: Journal du Conseil*, 63(9):1674–1681, 2006.

[30] D.A. DeVries, C.B. Grimes, and M.H. Prager. Using otolith shape analysis to distinguish eastern gulf of mexico and atlantic ocean stocks of king mackerel. *Fisheries Research*, 57(1):51–62, 2002.

[31] P. Duarte-Neto, R. Lessa, B. Stosic, and E. Morize. The use of sagittal otoliths in discriminating stocks of common dolphinfish (coryphaena hippurus) off northeastern brazil using multishape descriptors. *ICES Journal of Marine Science: Journal du Conseil*, 65(7):1144–1152, 2008.

[32] Herbert Freeman. Computer processing of line-drawing images. *ACM Computing Surveys (CSUR)*, 6(1):57–97, 1974.

[33] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

[34] M. Gagliano and M.I. McCormick. Feeding history influences otolith shape in tropical fish. *Marine Ecology Progress Series*, 278:291–296, 2004.

[35] E.A. Galley, P.J. Wright, and F.M. Gibb. Combined methods of otolith shape analysis improve identification of spawning areas of atlantic cod. *ICES Journal of Marine Science: Journal du Conseil*, 63(9):1710–1717, 2006.

[36] RW Gauldie and JS Crampton. An eco-morphological explanation of individual variability in the shape of the fish otolith: comparison of the otolith of hoplostethus atlanticus with other species by depth. *Journal of fish biology*, 60 (5):1204–1221, 2002.

[37] R Gonzalez, R Woods, and S Eddins. *Digital image processing using MATLAB*. Pearson Education India, 2004.

[38] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[39] Robert M Haralock and Linda G Shapiro. *Computer and robot vision*. Addison-Wesley Longman Publishing Co., Inc., 1991.

[40] J Hills, J Lines, E Baranauskas, J Mapp, and A Bagnall. Classification of time series by shapelet transformation. *Data Mining and Knowledge Discovery*, 28 (4):851–881, 2014.

[41] ICES. Report of the international bottom trawl survey working group (IBTSWG). *ICES CM 2013/SSGESST*, 10:272, 2013.

[42] ICES. Report of the working group of international pelagic surveys (WGIPS). *ICES CM 2014/SSGESST*, 1:360, 2014.

[43] Ramesh Jain, Rangachar Kasturi, and Brian G Schunck. *Machine vision*, volume 5. McGraw-Hill New York, 1995.

[44] Sharif Jemaa, Mahmoud Bacha, Gaby Khalaf, David Dessailly, Khalef Rabhi, and Rachid Amara. What can otolith shape analysis tell us about population structure of the european sardine, sardina pilchardus, from atlantic and mediterranean waters? *Journal of Sea Research*, 96:11–17, 2015.

[45] Avinash C.. Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. Society for Industrial and Applied Mathematics, 2001.

[46] James P Keating, Deirdre Brophy, Rick A Officer, and Eugene Mullins. Otolith shape analysis of blue whiting suggests a complex stock structure at their spawning grounds in the northeast atlantic. *Fisheries Research*, 157:1–6, 2014.

[47] W.H Kruskal. A nonparametric test for the several sample problem. *The Annals of Mathematical Statistics*, 23(4):525–540, 1952.

[48] F.P. Kuhl and C.R. Giardina. Elliptic fourier features of a closed contour. *Computer Graphics and Image Processing*, 18(3):236–258, 1982. ISSN 0146-664X.

[49] L.J. Latecki, R. Lakamper, and T. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 424–429. IEEE, 2000.

[50] S. le Cessie and J.C. van Houwelingen. Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191–201, 1992.

[51] D. Lewis. Naive (bayes) at forty: The independence assumption in information retrieval. *Machine Learning: European Conference on Machine Learning - 98*, pages 4–15, 1998.

[52] Diamond light. Diamondlight synchrotron facility., 2015. URL `http://www.diamond.ac.uk/`.

[53] Jason Lines and Anthony Bagnall. Alternative quality measures for time series shapelets. In *Intelligent Data Engineering and Automated Learning-IDEAL 2012*, pages 475–483. Springer, 2012.

[54] Jason Lines, Luke M Davis, Jon Hills, and Anthony Bagnall. A shapelet transform for time series classification. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 289–297. ACM, 2012.

[55] H.B Mann and D.R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 18(1):50–60, 1947.

[56] A Manurung. Elliptic fourier for shape analysis., 2011. URL `http://www.mathworks.com/matlabcentral/fileexchange/32800`.

[57] James Mapp. Uea otolith reconstruction demonstration videos., 2015. URL `https://www.youtube.com/results?search_query=ueaotolith`.

[58] James Mapp, Mark Fisher, Richard Atwood, Duncan Bell, Mark Greco, Sally Songer, and Ewan Hunter. Three-dimensional analysis of otolith growth using phase contrast synchrotron tomography. In *Journal of Fish Biology*. Wiley Online Library, 2016.

[59] James Mapp, Mark Fisher, and Ewan Hunter. Boundary based stock classification: Expert otolith readers outperform automated outlining methods. In *ICES Annual Science Conference (submitted)*, 2016.

[60] James Mapp, Ewan Hunter, Sally Songer, Jeroen Van Der Kooij, and Mark Fisher. Operational viability of stock-separation using shape indices derived from the otolith morphometric outline. an example using sprat and herring. In *In preparation*, 2016.

[61] James Mapp, Mark Fisher, Anthony Bagnall, Jason Lines, Sally Warne, and Joe Scutt Phillips. Clupea harengus: Intraspecies distinction using curvature

scale space and shapelets. In *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, pages 138–143, SciTePress, 2013.

[62] MATLAB. *version 7.11.0 (R2010b)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[63] B. Mérigot, Y. Letourneur, and R. Lecomte-Finiger. Characterization of local populations of the common sole solea solea (pisces, soleidae) in the nw mediterranean through otolith morphometrics and shape analysis. *Marine Biology*, 151 (3):997–1008, 2007.

[64] F. Mokhtarian. Silhouette-based isolated object recognition through curvature scale space. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(5):539–544, 1995.

[65] F. Mokhtarian and A.K. Mackworth. A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):789–805, 1992.

[66] F. Mokhtarian, S. Abbasi, and J. Kittler. Robust and e cient shape indexing through curvature scale space. In *Proceedings of the 1996 British Machine and Vision Conference BMVC*, volume 96, 1996.

[67] F. Mokhtarian, S. Abbasi, and J. Kittler. Efficient and robust retrieval by shape content through curvature scale space. *Series on Software Engineering and Knowledge Engineering*, 8:51–58, 1997.

[68] A.M Mood. *Introduction to the Theory of Statistics.* McGraw-hill, 1950.

[69] MPEG7. Ce shape-1 part b, 1400 binary shape images., 1999. URL `http://www.imageprocessingplace.com/downloads_V3/root_downloads/image_databases/MPEG7_CE-Shape-1_Part_B.zip`.

[70] Abdullah Mueen, Eamonn Keogh, and Neal Young. Logical-shapelets: an expressive primitive for time series classification. In *Proceedings of the 17th ACM*

SIGKDD international conference on Knowledge discovery and data mining, pages 1154–1162. ACM, 2011.

[71] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[72] David Paganin, SC Mayo, Tim E Gureyev, Peter R Miller, and Steve W Wilkins. Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *Journal of microscopy*, 206(1):33–40, 2002.

[73] V. Parisi-Baradad, A. Lombarte, E. García-Ladona, J. Cabestany, J. Piera, and O. Chic. Otolith shape contour analysis using affine transformation invariant wavelet transforms and curvature scale space representation. *Marine and freshwater research*, 56(5):795–804, 2005.

[74] V. Parisi-Baradad, A. Manjabacas, A. Lombarte, R. Olivella, Ó. Chic, J. Piera, and E. García-Ladona. Automated taxon identification of teleost fishes using an otolith online database - aforo. *Fisheries Research*, 105(1):13–20, 2010.

[75] A Patwardhan, A Sasovii, M Scnarzii, and DVAN Dyc. A new method for three-dimensional otolith analysis. *Journal of fish biology*, 54:223–225, 1999.

[76] K Paul, R Oeberst, and C Hammer. Evaluation of otolith shape analysis as a tool for discriminating adults of baltic cod stocks. *Journal of Applied Ichthyology*, 29(4):743–750, 2013.

[77] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[78] G. Petursdottir, G.A. Begg, and G. Marteinsdottir. Discrimination between icelandic cod (*Gadus morhua* l.) populations from adjacent spawning areas based on otolith growth and shape. *Fisheries research*, 80(2):182–189, 2006.

[79] G. Pierre. Pattern extraction for time series classification. In *Principles of Data Mining and Knowledge Discovery*, pages 115–127. Springer, 2001.

[80] J. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1998. URL `http://research.microsoft.com/~jplatt/smo.html`.

[81] D. Ponton. Is geometric morphometrics efficient for comparing otolith shape of different fish species? *Journal of Morphology*, 267(6):750–757, 2006.

[82] J.R. Quinlan. *C4.5: programs for machine learning*. Morgan Kaufmann, 1993.

[83] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, and E. Keogh. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 262–270. ACM, 2012.

[84] C. Ratanamahatana and E. Keogh. Three myths about dynamic time warping data mining. In *Proc. 5th SDM*, 2005.

[85] S. Richter, G. Kühne, and O. Schuster. Contour-based classification of video objects. In *Proceedings of SPIE*, volume 4315, page 608, 2001.

[86] J.J. Rodriguez, L.I. Kuncheva, and C.J. Alonso. Rotation forest: A new classifier ensemble method. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(10):1619–1630, 2006.

[87] C.E Shannon and W Weaver. The mathematical theory of communication (urbana, il. *University of Illinois Press*, 19(7):1, 1949.

[88] A Snigirev, I Snigireva, V Kohn, S Kuznetsov, and I Schelokov. On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation. *Review of Scientific Instruments*, 66(12):5486–5492, 1995.

[89] C. Spampinato, D. Giordano, R. Di Salvo, Y.H.J. Chen-Burger, R.B. Fisher, and G. Nadarajan. Automatic fish classification for underwater species behavior understanding. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams*, pages 45–50. ACM, 2010.

[90] C. Stransky. Geographic variation of golden redfish (sebastes marinus) and deep-sea redfish (s. mentella) in the north atlantic based on otolith shape analysis. *ICES Journal of Marine Science: Journal du Conseil*, 62(8):1691–1698, 2005.

[91] C. Stransky and S.E. MacLellan. Species separation and zoogeography of redfish and rockfish (genus sebastes) by otolith shape analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 62(10):2265–2276, 2005.

[92] C. Stransky, H. Baumann, S.E. Fevolden, A. Harbitz, H. Høie, K.H. Nedreaas, A.B. Salberg, and T.H. Skarstein. Separation of norwegian coastal cod and northeast arctic cod by outer otolith shape analysis. *Fisheries Research*, 90 (1-3):26–35, 2008.

[93] Michael J Swain and Dana H Ballard. Color indexing. *International journal of computer vision*, 7(1):11–32, 1991.

[94] P Tafforeau, R Boistel, E Boller, A Bravin, M Brunet, Y Chaimanee, P Cloetens, M Feist, J Hoszowska, J-J Jaeger, et al. Applications of x-ray synchrotron microtomography for non-destructive 3d studies of paleontological specimens. *Applied Physics A*, 83(2):195–202, 2006.

[95] E Torstensen, ATGW Eltink, M Casini, WJ McCurdy, and LW Clausen. Report of the work shop on age estimation of sprat. *Institut of Marine Research, Fl devigen, Arendal, Norway*, pages 14–17, 2004.

[96] V.M. Tuset, P.L. Rosin, and A. Lombarte. Sagittal otolith shape used in the identification of fishes of the genus serranus. *Fisheries research*, 81(2):316–325, 2006.

[97] WEKA. WEKA hyperpipes classifier., 2002. URL `http://www.cs.tufts.edu/~ablumer/weka/doc/weka.classifiers.HyperPipes.html`.

[98] M.H.F. Wilkinson, J.B.T.M. Roerdink, S. Droop, and M. Bayer. Diatom contour analysis using morphological curvature scale spaces. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 3, pages 652–655. IEEE, 2000.

[99] L Ye and E Keogh. Time series shapelets: a new primitive for data mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 947–956. ACM, 2009.

[100] D. Zhang and G. Lu. Content-based shape retrieval using different shape descriptors: A comparative study. In *IEEE international conference on multimedia and expo*, pages 1139–1142, 2001.

[101] D. Zhang and G. Lu. A comparative study of curvature scale space and fourier descriptors for shape-based image retrieval. *Journal of Visual Communication and Image Representation*, 14(1):39–57, 2003.

[102] D. Zhang and G. Lu. Evaluation of mpeg-7 shape descriptors against other shape descriptors. *Multimedia Systems*, 9(1):15–30, 2003.

Appendices

Appendix A

Full Classification Results From Chapter 4

	NB	BN	Log	HP	J48	RaF	IBk	SMO	RoF
bou	60.00	65.00	65.00	61.25	73.75	58.75	57.50	67.50	70.00
bouG	83.75	87.50	78.75	75.00	87.50	92.50	75.00	80.00	93.75
evo	68.75	61.25	70.00	61.25	70.00	67.50	75.00	72.50	72.50
evoG	90.00	86.25	73.75	67.50	86.25	90.00	86.25	78.75	91.25
fou10	97.50	97.50	95.00	93.75	98.75	100.00	98.75	96.25	98.75
fou20	97.50	97.50	95.00	91.25	98.75	96.25	98.75	95.00	100.00
fou50	96.25	96.50	96.25	73.75	98.75	97.50	78.75	97.50	98.75
fou10n	65.00	78.75	82.50	85.00	81.25	87.50	83.75	83.75	88.75
fou20n	71.25	85.00	73.75	83.75	81.25	87.50	81.25	85.00	91.25
fou50n	72.50	88.75	78.75	73.75	87.50	81.25	80.00	78.75	92.50
Box	98.75	98.75	98.75	98.75	97.50	96.25	98.75	100.00	98.75
Axes	98.75	100.00	100.00	98.75	98.75	100.00	100.00	100.00	98.75
STAT	98.75	100.00	98.75	98.75	97.50	98.75	97.50	100.00	98.75
STAT+	98.75	100.00	97.50	98.75	97.50	100.00	97.50	100.00	98.75

Table A.1: Age-0 Celtic/Irish Sea herring: Full results table for transform method/learning algorithm tests.

	NB	BN	Log	HP	J48	RaF	IBk	SMO	RoF
bou	34.62	50.00	46.15	46.15	50.00	46.15	53.85	53.85	34.62
bouG	34.62	46.15	42.31	57.69	53.85	46.15	61.54	46.15	34.62
evo	69.23	42.31	53.85	57.69	50.00	53.85	53.85	57.69	69.23
evoG	69.23	42.31	50.00	65.38	50.00	65.38	50.00	57.69	65.38
fou10	65.38	73.08	61.54	57.69	84.62	84.62	69.23	65.38	69.23
fou20	69.23	69.23	61.54	69.23	84.62	61.54	46.15	61.54	73.08
fou50	53.85	65.38	57.69	42.31	84.62	57.69	57.69	57.69	73.08
fou10n	50.00	38.46	50.00	53.85	57.69	61.54	57.69	53.85	57.69
fou20n	53.85	38.46	50.00	57.69	53.85	42.31	50.00	46.15	57.69
fou50n	61.54	57.69	46.15	61.54	46.15	65.38	50.00	50.00	53.85
Box	88.46	73.08	84.62	84.62	76.92	73.08	73.08	80.77	80.77
Axes	84.62	84.62	84.62	80.77	84.62	80.77	73.08	84.62	57.69
STAT	84.62	80.77	69.23	76.92	73.08	73.08	80.77	80.77	76.92
STAT+	84.62	80.77	65.38	76.92	73.08	73.08	76.92	76.92	80.77

Table A.2: Age-1 Celtic/Irish Sea herring: Full results table for transform method/learning algorithm tests.

	NB	BN	Log	HP	J48	RaF	IBk	SMO	RoF
bou	62.50	47.50	63.75	45.00	56.25	62.50	50.00	67.50	57.50
bouG	62.50	47.50	62.50	45.00	52.50	55.00	50.00	63.75	57.50
evo	65.00	71.25	60.00	60.00	65.00	63.75	66.25	62.50	66.25
evoG	67.50	71.25	57.50	58.75	68.75	66.25	65.00	61.25	68.75
fou10	66.25	71.25	57.50	66.25	61.25	65.00	71.25	63.75	73.75
fou20	72.50	68.75	66.25	65.00	72.50	62.50	73.75	73.75	72.50
fou50	61.25	61.25	62.50	61.25	68.75	57.50	57.50	62.50	62.50
fou10n	57.50	71.25	60.00	50.00	58.75	57.50	71.25	73.75	67.50
fou20n	63.75	71.25	51.25	45.00	63.75	60.00	67.50	62.50	70.00
fou50n	56.25	70.00	51.25	57.50	71.25	57.50	60.00	58.75	58.75
Box	67.50	72.50	66.25	46.25	67.50	60.00	68.75	63.75	72.50
Axes	71.25	72.50	68.75	47.50	73.75	61.25	71.25	65.00	70.00
STAT	68.75	73.75	66.25	47.50	66.25	65.00	68.75	66.25	63.75
STAT+	71.25	73.75	71.25	50.00	75.00	68.75	70.00	72.50	66.25

Table A.3: Age-0 North-Sea/Channel sprat: Full results table for transform method/learning algorithm tests.

	NB	BN	Log	HP	J48	RaF	IBk	SMO	RoF
bou	58.00	50.00	56.00	44.00	66.00	66.00	64.00	58.00	54.00
bouG	60.00	72.00	56.00	48.00	78.00	78.00	50.00	66.00	68.00
evo	60.00	72.00	76.00	46.00	74.00	72.00	48.00	70.00	60.00
evoG	72.00	80.00	70.00	50.00	64.00	76.00	58.00	74.00	68.00
fou10	86.00	84.00	82.00	76.00	78.00	88.00	82.00	88.00	86.00
fou20	82.00	82.00	86.00	62.00	72.00	82.00	78.00	84.00	84.00
fou50	86.00	82.00	84.00	58.00	70.00	72.00	64.00	86.00	88.00
fou10n	68.00	70.00	80.00	74.00	62.00	78.00	78.00	82.00	78.00
fou20n	70.00	66.00	70.00	70.00	70.00	60.00	68.00	82.00	70.00
fou50n	70.00	62.00	64.00	68.00	58.00	72.00	64.00	70.00	84.00
Box	86.00	90.00	88.00	76.00	86.00	82.00	86.00	90.00	88.00
Axes	82.00	86.00	84.00	76.00	82.00	80.00	90.00	86.00	86.00
STAT	86.00	86.00	78.00	80.00	82.00	82.00	84.00	88.00	86.00
STAT+	84.00	86.00	80.00	80.00	82.00	86.00	82.00	88.00	88.00

Table A.4: Age-1 North-Sea/Channel sprat: Full results table for transform method/learning algorithm tests.

	NB	BN	Log	HP	J48	RaF	IBk	SMO	RoF
bou	27.27	50.00	50.00	40.91	63.64	50.00	54.55	45.45	63.64
bouG	31.82	50.00	40.91	40.91	63.64	50.00	54.55	45.45	50.00
evo	54.55	45.45	45.45	40.91	27.27	50.00	77.27	45.45	63.64
evoG	50.00	45.45	40.91	45.45	27.27	54.55	45.45	45.45	59.09
fou10	50.00	36.36	36.36	59.09	54.55	59.09	50.00	45.45	45.45
fou20	45.45	36.36	31.82	45.45	45.45	45.45	68.18	40.91	45.45
fou50	27.27	27.27	31.82	40.91	36.36	54.55	36.36	27.27	31.82
fou10n	40.91	50.00	36.36	45.45	27.27	54.55	54.55	40.91	31.82
fou20n	31.82	45.45	36.36	54.55	18.18	36.36	59.09	31.82	22.73
fou50n	45.45	45.45	36.36	54.55	63.64	45.45	54.55	36.36	45.45
Box	54.55	45.45	45.45	45.45	54.55	68.18	54.55	72.73	68.18
Axes	50.00	40.91	45.45	50.00	59.09	40.91	63.64	63.64	59.09
STAT	59.09	40.91	54.55	50.00	54.55	63.64	59.09	68.18	68.18
STAT+	54.55	40.91	50.00	50.00	59.09	77.27	50.00	68.18	63.64

Table A.5: Age-2 North-Sea/Channel sprat: Full results table for transform method/learning algorithm tests.

	NB	BN	Log	HP	J48	RaF	IBk	SMO	RoF
bou	34.62	46.15	69.23	46.15	34.62	53.85	42.31	61.54	42.31
bouG	30.77	46.15	73.08	50.00	34.62	65.38	42.31	65.38	50.00
evo	42.31	46.15	57.69	61.54	53.85	57.69	46.15	53.85	57.69
evoG	42.31	46.15	53.85	57.69	53.85	53.85	46.15	53.85	50.00
fou10	53.85	69.23	76.92	42.31	61.54	73.08	53.85	76.92	69.23
fou20	61.54	57.69	69.23	50.00	53.85	46.15	53.85	73.08	65.38
fou50	53.85	30.77	61.54	53.85	50.00	53.85	46.15	73.08	57.69
fou10n	53.85	46.15	76.92	46.15	65.38	76.92	50.00	84.62	76.92
fou20n	69.23	57.69	73.08	53.85	84.62	84.62	61.54	84.62	76.92
fou50n	61.54	57.69	53.85	53.85	84.62	57.69	50.00	69.23	80.77
Box	42.31	50.00	50.00	46.15	50.00	38.46	42.31	53.85	42.31
Axes	38.46	50.00	46.15	46.15	42.31	53.85	46.15	42.31	50.00
STAT	30.77	50.00	57.69	34.62	38.46	34.62	38.46	50.00	57.69
STAT+	34.62	50.00	61.54	38.46	42.31	50.00	30.77	53.85	53.85

Table A.6: Age-3 North-Sea/Channel sprat: Full results table for transform method/learning algorithm tests.

	NB	BN	Log	HP	J48	RaF	IBk	SMO	RoF
bou	50.00	81.82	36.36	59.09	59.09	36.36	31.82	36.36	50.00
bouG	50.00	81.82	27.27	50.00	59.09	59.09	27.27	27.27	36.36
evo	59.09	40.91	68.18	54.55	40.91	50.00	54.55	59.09	45.45
evoG	59.09	40.91	59.09	54.55	40.91	59.09	50.00	54.55	50.00
fou10	54.55	45.45	54.55	72.73	36.36	54.55	36.36	63.64	63.64
fou20	59.09	40.91	59.09	63.64	36.36	45.45	40.91	68.18	54.55
fou50	50.00	50.00	68.18	59.09	36.36	50.00	59.09	40.91	40.91
fou10n	50.00	40.91	40.91	45.45	45.45	54.55	36.36	54.55	54.55
fou20n	40.91	40.91	45.45	59.09	45.45	50.00	50.00	50.00	50.00
fou50n	31.82	59.09	31.82	40.91	50.00	45.45	45.45	36.36	45.45
Box	59.09	50.00	50.00	45.45	45.45	27.27	31.82	54.55	50.00
Axes	54.55	50.00	63.64	45.45	50.00	45.45	54.55	63.64	45.45
STAT	54.55	50.00	45.45	27.27	40.91	31.82	40.91	59.09	40.91
STAT+	59.09	50.00	50.00	31.82	40.91	36.36	40.91	45.45	45.45

Table A.7: Age-2 North-Sea/Thames herring: Full results table for transform method/learning algorithm tests.

	NB	BN	Log	HP	J48	RaF	IBk	SMO	RoF
bou	53.70	46.30	53.70	44.44	38.89	46.30	50.00	48.15	59.26
bouG	51.85	46.30	57.41	44.44	33.33	44.44	50.00	51.85	48.15
evo	57.41	70.37	42.59	51.85	55.56	50.00	46.30	44.44	51.85
evoG	57.41	70.37	44.44	57.41	59.26	55.56	42.59	48.15	55.56
fou10	50.00	44.44	42.59	61.11	48.15	61.11	33.33	42.59	48.15
fou20	50.00	44.44	37.04	57.41	42.59	37.04	40.74	38.89	37.04
fou50	46.30	35.19	46.30	53.70	40.74	44.44	53.70	44.44	44.44
fou10n	62.96	46.30	53.70	46.30	40.74	44.44	35.19	53.70	51.85
fou20n	62.96	46.30	48.15	57.41	29.63	40.74	44.44	44.44	48.15
fou50n	59.26	51.85	50.00	50.00	42.59	55.56	38.89	46.30	46.30
Box	44.44	50.00	44.44	46.30	50.00	46.30	29.63	50.00	50.00
Axes	48.15	50.00	46.30	48.15	50.00	59.26	61.11	46.30	50.00
STAT	50.00	50.00	57.41	46.30	50.00	53.70	35.19	33.33	51.85
STAT+	48.15	50.00	53.70	48.15	61.11	53.70	50.00	46.30	51.85

Table A.8: Age-3 North-Sea/Thames herring: Full results table for transform method/learning algorithm tests.

	NB	BN	Log	HP	J48	RaF	IBk	SMO	RoF
bou	47.62	40.48	61.90	50.00	38.10	71.43	61.90	54.76	61.90
bouG	42.86	40.48	59.52	52.38	35.71	47.62	66.67	61.90	64.29
evo	42.86	50.00	50.00	52.38	47.62	47.62	54.76	42.86	30.95
evoG	40.48	50.00	45.24	54.76	47.62	45.24	61.90	42.86	42.86
fou10	52.38	50.00	54.76	57.14	73.81	54.76	35.71	45.24	52.38
fou20	47.62	47.62	38.10	47.62	54.76	54.76	30.95	42.86	40.48
fou50	50.00	52.38	33.33	50.00	52.38	50.00	28.57	40.48	38.10
fou10n	57.14	45.24	42.86	59.52	33.33	57.14	57.14	47.62	42.86
fou20n	57.14	52.38	35.71	59.52	66.67	52.38	52.38	47.62	40.48
fou50n	64.29	52.38	47.62	52.38	59.52	64.29	40.48	42.86	64.29
Box	50.00	50.00	50.00	52.38	42.86	50.00	47.62	38.10	47.62
Axes	54.76	50.00	57.14	57.14	45.24	52.38	40.48	50.00	50.00
STAT	50.00	50.00	47.62	61.90	40.48	54.76	66.67	47.62	59.52
STAT+	47.62	50.00	45.24	57.14	40.48	52.38	71.43	40.48	54.76

Table A.9: Age-4 North-Sea/Thames herring: Full results table for transform method/learning algorithm tests.

	NB	BN	Log	HP	J48	RaF	IBk	SMO	RoF
bou	63.33	50.00	60.00	56.67	56.67	46.67	56.67	56.67	43.33
bouG	63.33	50.00	56.67	53.33	56.67	56.67	46.67	53.33	50.00
evo	63.33	50.00	43.33	50.00	66.67	70.00	53.33	40.00	50.00
evoG	63.33	50.00	40.00	50.00	66.67	43.33	60.00	43.33	53.33
fou10	50.00	46.67	73.33	53.33	43.33	53.33	46.67	56.67	50.00
fou20	50.00	73.33	53.33	53.33	43.33	53.33	46.67	43.33	33.33
fou50	50.00	66.67	46.67	46.67	46.67	70.00	46.67	46.67	53.33
fou10n	40.00	50.00	73.33	53.33	80.00	43.33	53.33	73.33	46.67
fou20n	46.67	46.67	60.00	46.67	56.67	50.00	53.33	66.67	56.67
fou50n	30.00	60.00	53.33	46.67	60.00	50.00	53.33	66.67	60.00
Box	43.33	50.00	40.00	53.33	50.00	36.67	50.00	43.33	46.67
Axes	33.33	50.00	20.00	53.33	50.00	36.67	53.33	36.67	46.67
STAT	43.33	50.00	43.33	50.00	43.33	43.33	50.00	53.33	53.33
STAT+	46.67	50.00	43.33	50.00	43.33	46.67	56.67	50.00	50.00

Table A.10: Age-5 North-Sea/Thames herring: Full results table for transform method/learning algorithm tests.