UiT THE ARCTIC UNIVERSITY OF NORWAY

Department of Mathematics and Statistics

Discrimination of cod otolith shapes by two different Fourier methods

This thesis compares two Fourier methods performance with regards to cod discrimination by otolith shape analysis.

—

Are Murberg Henriksen

Master's thesis in Industrial Mathematics - STA-3921, December 2013

Preface

This thesis is written during the fall of 2013 at the Department of Mathematics and Statistics at UIT, the Arctic University of Norway. This is the final thesis for a degree in Industrial Mathematics and it accounts for 30 credits in a semester. The course code for the thesis is Sta-3921.

I would like to thank my supervisor Sigrunn Holbek Sørbye for aiding me in writing the thesis and I would also like to thank my external supervisor, Alf Harbitz and the Institute of Marine Research for giving me the assignment and providing help to all aspects of the thesis.

Are Murberg Henriksen,
Tromsø, December 15, 2013

Abstract

In this thesis we have applied Fisher's discrimination method to cod samples to discriminate between cod originating from the coastal cod stock and cod originating from the arctic cod stock. The discrimination is based on Fourier coefficients extracted from cod otolith contours. We have optimized and compared two different Fourier methods of extracting these coefficients. These methods are evaluated and compared in terms of discrimination results and robustness.

Contents

Chapter 1

Introduction

Along the coast and offshore regions of Northern Norway we find two different stocks of the Northeast Atlantic cod population (*Gadus morhua*), the Northeast arctic cod and the Norwegian coastal cod, hereafter referred to as arctic- and coastal cod. The arctic cod has its nursery and feeding area in the Barents Sea and is the largest cod stock within the Northeast Atlantic cod population. Each year arctic cod migrates to the coast of Northern Norway for feeding and spawning. The coastal cod are on the other hand located along the coastline of Northern Norway all year and has been red-listed as a near-threatened fish stock (Kålås et al., 2006). Unfortunately, the arctic and coastal cod mix during the spawning season, making fishermen unable to catch arctic cod exclusively. To effectively stop all catch of coastal cod along the coastline of Norway would mean stopping all catch of arctic cod as well. This is not possible due to the huge impact it would have on the economy of the Norwegian fisheries. The fishery management have come up with other fishery regulations to try to reduce the pressure on the coastal cod stock. To be able to detect if such measures are effective, they need a way to identify if a cod caught along the coast of Norway belongs to the arctic cod stock or the coastal cod stock, and thus be able to give good assessments of the development and status of the two stocks.

To differentiate between different stocks within a population, otolith shape analysis is an often used method, not only for cod, but also for other fish populations (Stransky et al. (2008), Cañas et al. (2012) and Benzinou et al. (2013)). Otoliths, also known as ear-stones, are calcium carbonate structures located behind the brain of the fish. Each fish has three pairs of otoliths which function as balancing organs as well as organs to convert sound into electrical signals. The otoliths continue to grow throughout the life of the fish and their shapes are highly dependent on age, sex, heritage and environment. Because of this, otolith shape analysis is an important tool in fish studies.

In this thesis we are trying to separate between coastal- and arctic cod with the use of otolith shape analysis. Our sample contains images of cod otoliths provided by the Institute of Marine Research. We can see three of the images contained within the sample in Figure 1.1. The cod as to which the images originate from were caught in the Barents Sea and along the coast of Norway and Svalbard between 1999 and 2001. This sample is the same sample as was studied in Stransky et al. (2008). More details regarding this sample can be found in Section 3.1 of this thesis and in the aforementioned article.

The use of Fourier shape analysis to investigate variations in the Atlantic cod population was first

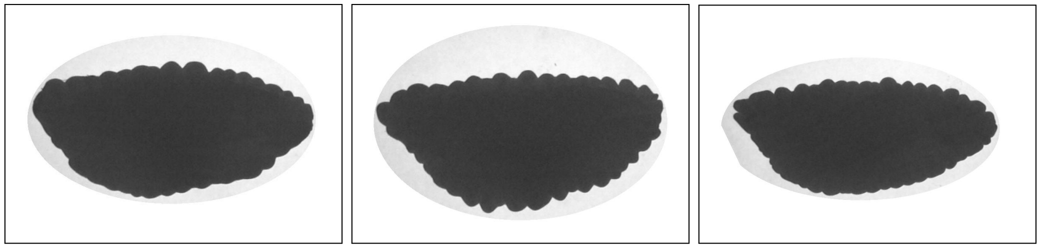

Figure 1.1: Three of the images contained within the cod sample analyzed in this thesis.

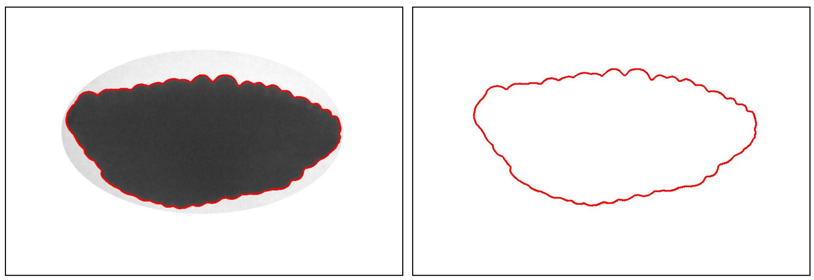

Figure 1.2: The highlighted, red, otolith outline can be seen along the perimeter of the otolith. This outline will also be referred to as the otolith contour.

introduced in Campana and Casselman (1993) and has later been used in Stransky et al. (2008). In this thesis we are looking at two different methods of using Fourier shape analysis to discriminate between the coastal- and arctic cod stocks. The first being the use of elliptic Fourier descriptors (EFDs) introduced by Kuhl and Giardina (1982) and improved by Haines and Crampton (2000) and the second being the use of one-dimensional Fourier analysis(1DF) introduced by Reig-Bolaño et al. (2010).

Going through the steps of the analysis, we start off by extracting the otolith outlines from the images, forming otolith contours. This is illustrated in Figure 1.2. We then approximate the contours and extract the Fourier coefficients with the use of the EFD- and the 1DF method. In Henriksen (2013), the different Fourier methods ability to approximate and reconstruct the original contours of halibut otoliths were studied. The focus in this thesis is to determine which of the two methods gives the best discrimination results. We are also highlighting how the smoothing of the otolith contours before extracting the Fourier coefficients effects the discrimination scores for both methods, and we are studying how orientation of the ototlihs can effect the scores.

We are using the Fourier coefficients provided by the EFD- and the 1DF method to discriminate between coastal and arctic cod by using Fisher's discrimination method and a leave-one-out algorithm both described in Johnson and Wichern (2007). We are doing the discrimination analysis for different groups based on age, sex and length of the cod the otoliths originate from. The information as to what stock the otolith images are collected from is provided by the Institute of Marine Research and with this knowledge we are able to see how well the discrimination methods perform. We will measure this performance in the percentage of correctly discriminated cod otoliths. With the use of bootstrapping we are able to repeat the discriminations for 1000 samples, giving robust discrimination scores accompanied with confidence intervals.

Ultimately, the goal of this thesis is to see which Fourier method provides the highest discrimination scores. We have built up the results of the thesis in two steps. The first step optimizes the use of the different Fourier methods and the last step compares the discrimination results based on the two methods, leading to a conclusion as to which method is preferable.

Chapter 2 contains the theory of the methods applied in this thesis. The first two sections describe the Fourier methods used in the analysis, followed by theory of the smoothing methods applied to the otolith contours to improve the discrimination. The last two sections of Chapter 2 is dedicated to the bootstrapping method and the discrimination method used in the thesis. In chapter 3 we present the data set in the first section. In the following sections we do the optimization of the two Fourier methods and in the final section we compare their ability to discriminate between coastal and arctic cod groups. In Chapter 4 we give some concluding remarks, discuss some challenges and give thoughts about future work on the subject. For all programming purposes we have used matlab and a link to the code can be found in Appendix A.

Chapter 2

Methodology

In this chapter we present the theory of the methods applied in this thesis. Sections 2.1 and 2.2 describe the EFD- and the 1DF methods used to extract Fourier coefficients from the cod otolith contours. In Section 2.3 we describe a weighted-moving-average method used to smooth the otolith contours before extracting the Fourier coefficients. In Section 2.4 we are presenting an alternative smoothing method which turns the otolith contours into concave contours. We are using bootstrapping as a tool in the discrimination analysis and we are presenting the details and usage of this bootstrapping in Section 2.5. For the discrimination between the arctic and coastal stocks with the use of Fourier coefficients, we are applying Fisher's discrimination method. This method is presented in Section 2.6. Section 2.7 is the chapter appendix.

2.1 Elliptic Fourier Descriptors (EFD)

A well known technique for analyzing otolith contours is the use of EFDs which was first introduced by Kuhl and Giardina (1982) and was later studied in Haines and Crampton (2000) and Harbitz (2007). This is a method based on approximating a closed contour with a sum of elliptic Fourier functions. From these sums of functions we extract EFD coefficients that can unveil differences between otoliths. These differences may be hard to spot for the human eye. The coefficients we get from the otolith contours will enable us to create allocation rules and apply discrimination to cod samples containing different cod stocks, e.g. coastal cod and arctic cod. That is, the differences in coefficients between the arctic- and the coastal cod otoliths should enable us to determine which otoliths that originate from which stocks. The EFD method is a popular choice for otolith discrimination studies as can be seen in Stransky et al. (2008), Cañas et al. (2012) and Benzinou et al. (2013).

The theory behind the EFD method is presented in this section, but the details surrounding the implementation of this method are not described in this thesis. The programs were provided by Alf Harbitz at the Institute of Marine Research and the code can be found in Appendix A.

Now looking at the theory, let $\mathbf{x} = [x_1, x_2, ..., x_K]$ and $\mathbf{y} = [y_1, y_2, ..., y_K]$ denote the coordinates of a 2D contour, in our case the outline of a cod otolith, where K is the number of otolith contour points. Now picture us tracing the contour in a counter-clockwise direction at a constant speed. The time spent traversing the first m contour points is given by:

$$t_m = \sum_{i=1}^{m} \Delta t_i \quad \text{with} \quad \Delta t_i = t_i - t_{i-1}$$

and the traversed x and y segments of the contour are equal to:

$$\begin{aligned} x_m &= \sum_{i=1}^{m} \Delta x_i \quad \text{with} \quad \Delta x_i = x_i - x_{i-1} \\ y_m &= \sum_{i=1}^{m} \Delta y_i \quad \text{with} \quad \Delta y_i = y_i - y_{i-1}. \end{aligned}$$

The total time it takes to cover a complete cycle of the contour is denoted as $t_K = T$.

The Fourier expansion of the contour $\mathbf{x}$-vector is then given as:

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos\frac{2\pi nt}{T} + b_n \sin\frac{2\pi nt}{T} \quad , \quad 0 < t < T \tag{2.1}$$

and the EFD coefficients are

$$\begin{aligned} a_0 &= \tfrac{1}{T}\int_0^T x(t)dt, \\ a_n &= \tfrac{2}{T}\int_0^T x(t)\cos\tfrac{2\pi nt}{T}dt, \\ b_n &= \tfrac{2}{T}\int_0^T x(t)\sin\tfrac{2\pi nt}{T}dt. \end{aligned} \tag{2.2}$$

As we are studying a closed otolith contour made out of pixel points we know that $x(t)$ is piecewise linear and continuous for all time values. Because of this, the time derivative, $\dot{x}(t)$, consists of a sequence of piecewise constant derivatives, $\Delta x_m/\Delta t_m$ where:

$$\Delta x_m = x_m - x_{m-1} \quad \text{and} \quad \Delta t_m = t_m - t_{m-1}.$$

This time derivative is periodic with period T and can be represented by it's own Fourier series:

$$\dot{x}(t) = \sum_{n=1}^{\infty} \alpha_n \cos\frac{2\pi nt}{T} + \beta_n \sin\frac{2\pi nt}{T} \tag{2.3}$$

where

$$\alpha_n = \frac{2}{n}\int_0^T \dot{x}(t)\cos\frac{2\pi nt}{T}dt \quad \text{and} \tag{2.4}$$

$$\beta_n = \frac{2}{n}\int_0^T \dot{x}(t)\sin\frac{2\pi nt}{T}dt. \tag{2.5}$$

We can write $\dot{x}(t) = \sum_{m=1}^{K} \Delta x_m/\Delta t_m$. Inserting this into (2.4) and (2.5) yields:

$$\begin{aligned}\alpha_n &= \frac{2}{T}\sum_{m=1}^{K}\frac{\Delta x_m}{\Delta t_m}\int_{t_{m-1}}^{t_m} \cos\frac{2\pi nt}{T}dt \\ &= \frac{2}{T}\sum_{m=1}^{K}\frac{\Delta x_m}{\Delta t_m}(\sin\frac{2\pi nt_m}{T} - \sin\frac{2\pi nt_{m-1}}{T})\frac{T}{2\pi n}\end{aligned}$$

and

$$\begin{aligned}\beta_n &= \frac{2}{T}\sum_{m=1}^{K}\frac{\Delta x_m}{\Delta t_m}\int_{t_{m-1}}^{t_m} \sin\frac{2\pi nt}{T}dt \\ &= \frac{2}{T}\sum_{m=1}^{K}\frac{\Delta x_m}{\Delta t_m}(\cos\frac{2\pi nt_m}{T} + \cos\frac{2\pi nt_{m-1}}{T})\frac{T}{2\pi n}.\end{aligned}$$

Based on Equation (2.1), we also notice that:

$$\dot{x}(t) = \frac{d}{dt}x(t) = \sum_{n=1}^{\infty} -\frac{2\pi n}{T}a_n\sin\frac{2\pi nt}{T} + \frac{2\pi n}{T}b_n\cos\frac{2\pi nt}{T}.$$

Now we have two expressions for $\dot{x}(t)$ and can combine these to find expressions for the elliptic Fourier coefficients:

$$\begin{aligned}a_n &= \frac{T}{2n^2\pi^2}\sum_{m=1}^{K}\frac{\Delta x_m}{\Delta t_m}(\cos\frac{2\pi nt_m}{T} - \cos\frac{2\pi nt_{m-1}}{T}) \quad \text{and} \\ b_n &= \frac{T}{2n^2\pi^2}\sum_{m=1}^{K}\frac{\Delta x_m}{\Delta t_m}(\sin\frac{2\pi nt_m}{T} - \sin\frac{2\pi nt_{m-1}}{T}).\end{aligned}$$

Following the same steps for

$$y(t) = c_0 + \sum_{n=1}^{\infty} c_n\cos\frac{2\pi nt}{T} + d_n\sin\frac{2\pi nt}{T} \quad , \quad 0 < t < T \tag{2.6}$$

we find the Fourier coefficients:

$$
\begin{aligned}
c_n &= \frac{T}{2n^2\pi^2}\sum_{m=1}^{K}\frac{\Delta y_m}{\Delta t_m}\left(\cos\frac{2n\pi t_m}{T} - \cos\frac{2n\pi t_{m-1}}{T}\right) \quad \text{and} \\
d_n &= \frac{T}{2n^2\pi^2}\sum_{m=1}^{K}\frac{\Delta y_m}{\Delta t_m}\left(\sin\frac{2n\pi t_m}{T} - \sin\frac{2n\pi t_{m-1}}{T}\right).
\end{aligned}
$$

Now we can find the expression for a_0 and c_0 by doing a piecewise integration. Integrating over all the line segments x_m and y_m yields:

$$
\begin{aligned}
a_0 &= \frac{1}{T}\sum_{m=1}^{K}\frac{x_m + x_{m-1}}{2}\Delta t_m \quad \text{and} \\
c_0 &= \frac{1}{T}\sum_{m=1}^{K}\frac{y_m + y_{m-1}}{2}\Delta t_m.
\end{aligned}
$$

The point (a_0, c_0) can be interpreted as the weighted sum of the mid points of the line segments Δx_m and Δy_m.

Normalization

For the elliptic Fourier descriptors to be used in discriminant analysis, a normalization of the otolith contours with regards to starting point, rotation and size, is needed. Without normalization we will be unable to detect whether the discrimination results we get originates as a result of otoliths from different stocks or otoliths oriented in different ways while being scanned.

We start off by creating a best fitted ellipse, E, to the otolith contour. We define the normalization parameters: R and r as the length of the major- and minor axis of E, respectively. If we look at Figure 2.1, U and V form a coordinate system and are parallel with the best fitted ellipses major- and minor axis. We can now define the normalization rotation parameter, ψ, as the rotation angle between the XY-coordinate system and the UV-coordinate system. We can also define the translation parameter θ as a function of the time it takes to traverse the contour from the starting point, x_1, y_1 to the point where the contour crosses the U-axis, (x_u, y_u). As shown in Kuhl and Giardina (1982), these four normalization parameters can be calculated based on the elliptic Fourier coefficients a_1, b_1, c_1 and d_1 as follows:

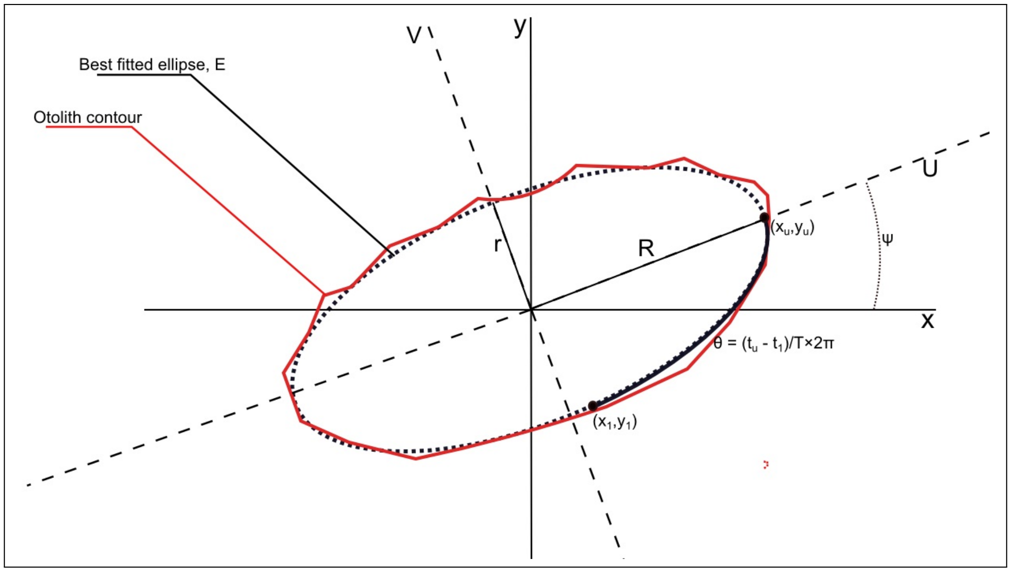

Figure 2.1: The red otolith contour is not an original contour image, but an illustration.

$$
\begin{aligned}
\theta &= \frac{1}{2}\tan^{-1}\frac{2(a_1b_1+c_1d_1)}{(a_1^2+c_1^2-b_1^2-d_1^2)}, \quad 0 \le \theta \le \pi, \\
a_\theta &= a_1\cos\theta + b_1\sin\theta,\ c_\theta = c_1\cos\theta + d_1\sin\theta, \\
b_\theta &= a_1\sin\theta + b_1\cos\theta,\ d_\theta = -c_1\sin\theta + d_1\cos\theta, \\
R &= \sqrt{a_\theta^2+c_\theta^2}, \\
r &= \sqrt{b_\theta^2+d_\theta^2} \quad \text{and} \\
\psi &= \tan^{-1}(c_\theta/a_\theta).
\end{aligned}
$$

Now let a_{0n}, b_{0n}, c_{0n} and d_{0n} be the normalized elliptic Fourier coefficients. These are given by:

$$
\begin{bmatrix} a_{0n} & b_{0n} \\ c_{0n} & d_{0n} \end{bmatrix} = \frac{1}{R}\begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}\begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix}\begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix}.
$$

By using this normalization $a_{01} = 1$, $b_{01} = c_{01} = 0$ and the best fitted ellipse, E, will have a semi major axis with length 1 and semi minor axis with length < 1.

The contour points (x_i, y_i) are transformed to normalized contour points u_{0i}, v_{0i}, as follows:

$$
\begin{bmatrix} u_{0i} \\ v_{0i} \end{bmatrix} = \frac{1}{R}\begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}\begin{bmatrix} x_i - a_0 \\ y_i - c_0 \end{bmatrix}.
$$

The implementation of this method was provided by Alf Harbitz and the Institute of Marine Research. We have not focused on further details regarding the normalization method in this thesis. However, more details can be found in Kuhl and Giardina (1982) where it was introduced the first time.

2.2 One Dimensional Fourier Method (1DF)

This new method was first introduced by Reig-Bolaño et al. (2010) where they studied its ability to reconstruct otolith contours from two different species of fish, *Liza aurata* and *Liza ramada*. They got significant results in their studies, showing that the 1DF method was able to reconstruct contours at the same accuracy as the EFD method using a smaller number of coefficients. In Henriksen (2013) we did the same study for halibut otoliths, but were unable to find significant results showing that the 1DF method gives any better accuracy. In this thesis, our focus is on discrimination and we want to see if we are able to get better discrimination scores using 1DF coefficients instead of EFD coefficients.

The main idea of this method is to convert a closed 2D contour, in our case an otolith contour, into a one dimensional function and then into a one dimensional Fourier series. We have continued the work on the programs developed in Henriksen (2013), adapting them to discrimination purposes instead of contour approximations. These programs can be found in Appendix A. As in the case of the EFD method, using the coefficients for discrimination purposes requires normalization of the otolith contours. We normalize the contours in the same way for both the EFD- and the 1DF method, as described in Section 2.1.

The following sections describes the theory and implementation of the 1DF method.

2.2.1 Discrete Fourier Transform(DFT)

Given a sequence of L numbers, $y(1), y(2), ..., y(L-1)$, representing the discrete function $y(k)$, the goal with the DFT is to decompose this function into a sum of complex sinusoids on the form $e^{i\omega k}$. These sinusoids are to have a period of L. To be able to do this we need that $e^{i\omega k} = e^{i\omega(k+L)}$ for all values of ω. This implies that $e^{i\omega L} = 1$, which again implies that ω has to be one of the numbers:

$$\omega_n = \frac{2\pi n}{L}, \quad n = 0, 1, ..., L-1.$$

The discrete Fourier series expansion of $y(k)$ takes the form:

$$y(k) = \frac{1}{L}\sum_{n=0}^{L-1} z_n e^{ik\omega_n} = \frac{1}{L}\sum_{n=0}^{L-1} z_n e^{\frac{i2\pi nk}{L}}, \quad z_n \in \mathbb{C}.$$

Now we can find the complex values $z_n = a'_n + ib'_n$ by using the fast Fourier algorithm (FFT). This gives the 2L Fourier coefficients: $a'_0, b'_0, a'_1, b'_1, a'_2, ..., a'_{L-1}, b'_{L-1}$. These are the 1DF coefficients used in the discriminant analysis.

How we turn the otoliths contours into discrete functions $y(k)$, enabling the use of the discrete Fourier transform and extraction of the 1DF coefficients, is described in the following section. We use the ' -notation to separate these coefficients from the EFD coefficients extracted from the contours.

2.2.2 The Implementation Steps of The 1DF Method.

The following steps describe the implementation of the 1DF method in Matlab. The code can be found in Appendix A.

1st step

Load the contour pixel values from the scanned cod otolith image. If wanted, apply smoothing to the contour. Standardize the contour by fitting a single ellipse to the contour. Rotate the contour so the ellipses major axis become parallel with the x-axis and reshape it so the length of the major half-axis of the ellipse is equal to 1.

2nd step

Find the maximum and minimum values of **x** and **y**, for the otolith contour. Position a rectangular box that encloses the entire contour and mirror the contour around the maximum x-value (See Figure 2.2).

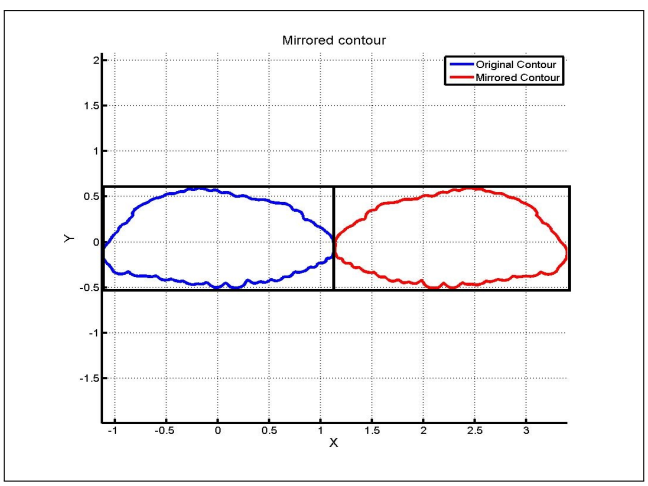

Figure 2.2: The mirrored contour

3rd step

Combine the two contours (see Figure 2.2) at the mirroring point. Start at the minimum x-value of the left-most contour. Iterate clockwise until you reach the mirror-point and continue by iterating in a counter clockwise direction of the right-most contour to create a one dimensional contour(see Figure 2.3).

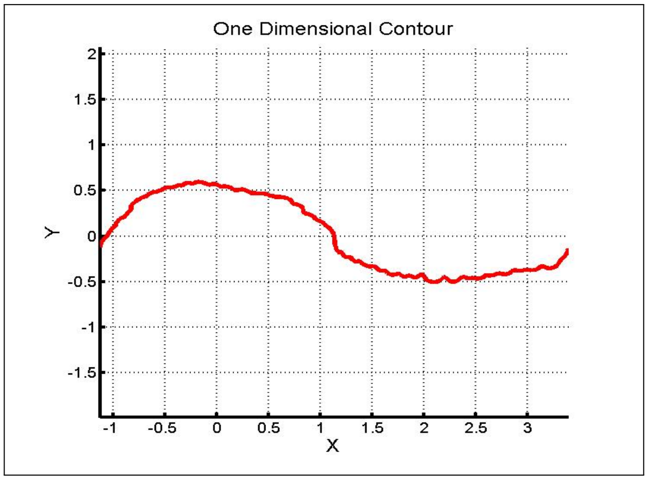

Figure 2.3: The 1D contour.

4th step

Remove all the ambiguities of the function. This is done by comparing the y-values of the points where there are matching x-values. Replace the areas containing ambiguities with a line, containing the same amount of points as the ones removed, going from the last point that wasn't an ambiguity to the next point that isn't an ambiguity. (see Figure 2.4).

5th step

Interpolate along the points of the curve and create new x-values with equal spacing. We use linear interpolation with the matlab function *interp*.

6th step

We have that $x(0) = 0$ and after the interpolation we have equal spacing between the x-values. By setting $x = k$ we can represent the contour as a discrete function y(k), just as described in Section 2.2.1. We can now apply the Matlab FFT function to the contour and extract the one dimensional Fourier coefficients we use in the discrimination analysis.

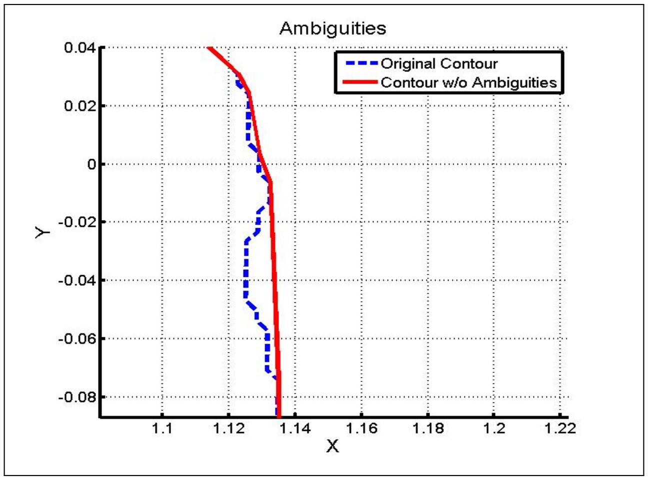

Figure 2.4: Handling of ambiguities

2.3 Weighted-Moving-Average Smoothing (WMA)

When scanning a cod otolith, the original contour outline which is a continuous smooth curve, inevitably, is turned into a finite number of straight line segments, decided by the scanned image's resolution. This will in turn create a lot of high frequency noise that may disrupt or distort the analysis. To remove this unwanted noise we have applied different smoothing methods to the otolith contours. These smoothing methods are applied to the original otolith contours and the smoothing is done before the 1DF- and EFD methods are applied to extract the Fourier coefficients. We will compare the discrimination scores for differently smoothed contours and end up with a smoothing method best suited for the EFD method and the 1DF method. The smoothing analysis is presented in Chapter 3. In this chapter we focus on describing how the methods are applied.

In this section, we describe a smoothing method based on on weighted moving average and in Section 2.4 we describe another smoothing method based on making the contours concave. The following box describes the implementation of the WMA-smoothing:

Given the otolith contour points $\mathbf{x} = (x_1, x_2, ..., x_n)$ and $\mathbf{y} = (y_1, y_2, ..., y_n)$, we apply the WMA-smoothing:

$$\begin{aligned} (x_{i,new}, y_{i,new}) &= 0.25 * (x_{i-1}, y_{i-1}) + 0.5 * (x_{i,old}, y_{i,old}) + 0.25 * (x_{i+1}, y_{i+1}). \\ \text{Set} \quad (x_i, y_i) &= (x_{i,new}, y_{i,new}). \end{aligned}$$

Iterate for $i = 1, 2, ..., n$.

As we are smoothing a closed contour, the special cases for $i = 1$ and $i = n$ will be:

$$\begin{aligned} (x_{1,new}, y_{1,new}) &= 0.25 * (x_n, y_n) + 0.5 * (x_{1,old}, y_{1,old}) + 0.25 * (x_2, y_2) \\ (x_{n,new}, y_{n,new}) &= 0.25 * (x_{n-1}, y_{n-1}) + 0.5 * (x_{n,old}, y_{n,old}) + 0.25 * (x_1, y_i) \end{aligned}$$

In the results and analysis in this thesis we are talking about WMA-smoothing iterations. One iteration is one full traverse of the contour, smoothing every single point.

In Figure 2.5 we can see the results of applying this smoothing to a contour with different numbers of WMA-smoothing iterations. We are showing only a small segment of the contour to emphasize the effect of the smoothing.

The reasoning behind our choice of using this WMA-smoothing method is that it is fast, easy to implement and gives good control over smoothing levels. By testing different numbers of WMA-smoothing iterations we can get a good look at how different smoothing levels effect the discrimination results. A link to the implementation of this method can be found in Appendix A.

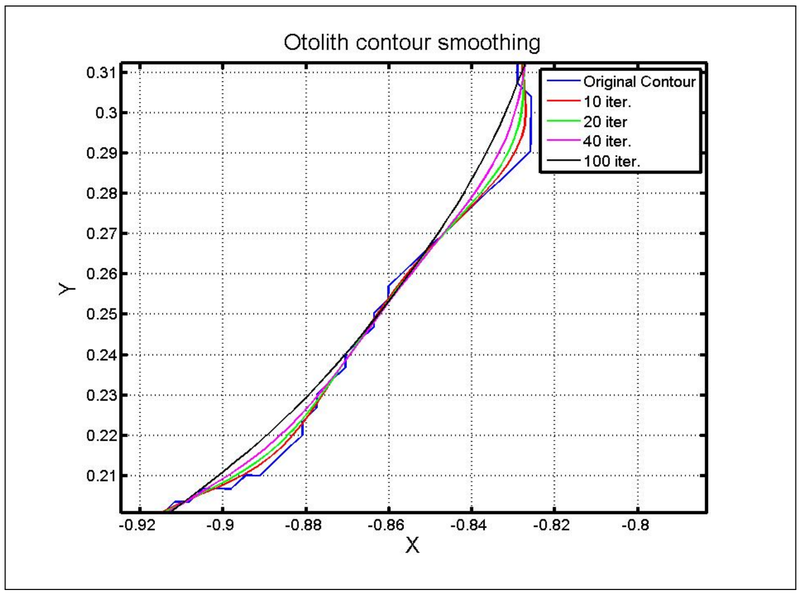

Figure 2.5: Close up image of a segment of a cod otolith contour smoothed with 5 different number of WMA-smoothing iterations.

2.4 Concavisation

The concavisation method is the second contour smoothing method we are looking at in this thesis. This method was proposed by Alf Harbitz at the Institute of Marine Research. The method hasn't been published yet, but in preliminary testing it gave promising results, encouraging the study of it in this thesis.

In Section 2.2 we described how we handled the ambiguities of the contours in the implementation of the 1DF method. When we have applied the concavisation to the otolith contours, this is unnecessary as the concavisation effectively removes all ambiguities from the contours. This is an effect of turning the contours concave, which is illustrated in Figure 2.6. In this thesis we are comparing this smoothing method with the WMA-smoothing described in Section 2.3. The comparison is based on the discrimination results the differently smoothed contours present. Our hope is that the concavisation method will be able to remove unneeded attributes of the contours, leaving contour shapes that give better discrimination results than differently smoothed contours. It may, however, remove too much of the contours, reducing the ability to discriminate between different cod stocks. This is analyzed in Chapter 3.

Now we look at the implementation of this method, presented as pseudo-code. The program used in this thesis was provided by Alf Harbitz and are linked in Appendix A.

Start off with the $(n \times 1)$ otolith contour vectors $\mathbf{x} = [x_1,\ x_2,\ ...,\ x_n]$ and $\mathbf{y} = [y_1,\ y_2,\ ...,\ y_n]$. Define m such that

$$y_m = max(\mathbf{y})$$

and shift the contour vectors in a counter clockwise direction such that

$$x_1 = x_m \text{ and } y_1 = y_m$$

Set i = 2 and find:

$$\mathbf{v}_i = (x_i, y_i) - (x_{(i-1)}, y_{(i-1)})$$

for j = (i+1) to n, find:

$$\mathbf{u}_j = (x_j, y_j) - (x_i, y_i)$$

Find the angle, θ_j, between $\mathbf{v}_i$ and $\mathbf{u}_j$.

Now find the j such that:

$$\theta_j = \min(\theta_{i+1},\ \theta_{i+2}, ..., \theta_n)$$

Replace all contour points between $(x_i,\ y_i)$ and (x_j, y_j) with a straigth line.

Set $i = j$ and repeat.

Continue until $i = n$ and the new $\mathbf{x}$ and $\mathbf{y}$ vectors will form a concave contour. When replacing the contour points with straight lines, we are using the matlab, linspace function.

Illustration of the method can be seen in Figure 2.7 where the upper part of the figure shows how the method is applied to the contours and the lower half shows the finished concave contour.

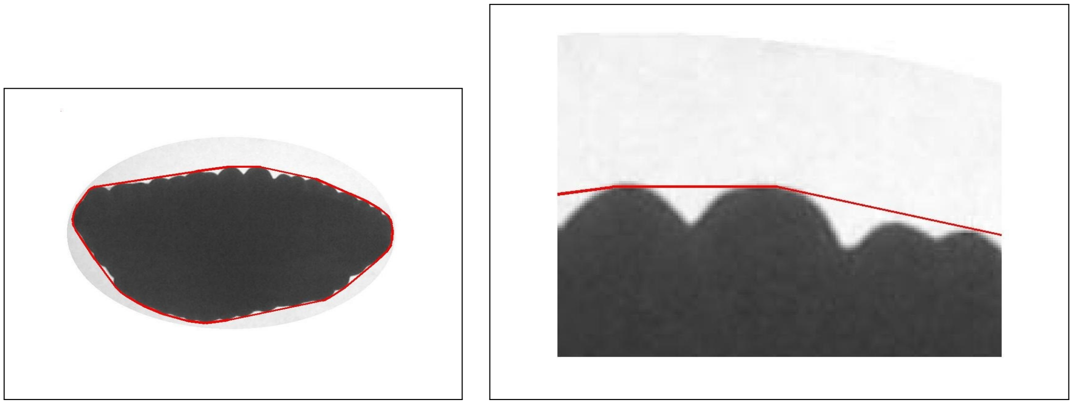

Figure 2.6: Images showing how the concave contours are compared to the cod otoliths.

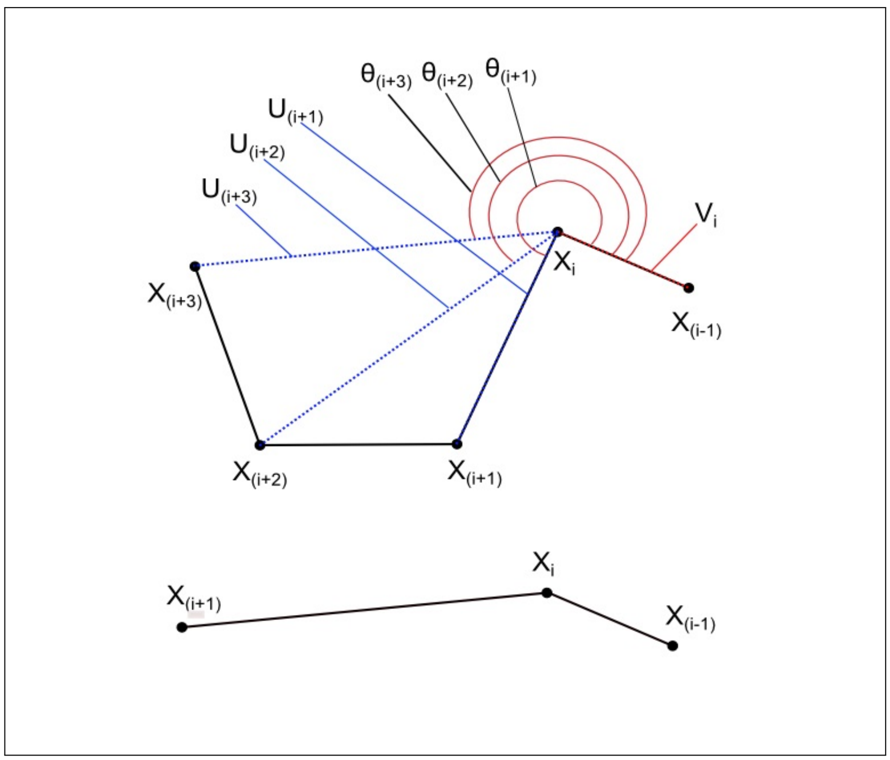

Figure 2.7: In the upper contour we see the V and U vectors and the angles between them. The smallest angle is θ_{i+3}. This will then be θ_j and all points between x_i and $x_j = x_{i+3}$ are replaced with points forming a straight line. In the lower contour we see how this segment looks after the concavisation.

2.5 Bootstrap Method

Bootstrapping are computational intensive methods used to measure the uncertainty in statistical estimators. Different bootstrapping methods are used when the distribution of an estimator is unknown or complicated. These methods can be used to find variance, bias and confidence intervals of estimators as well as for hypothesis testing. In our case, the boostrapping is based on drawing with replacement from the original sample. The reason for us to apply bootstrapping is the idea that the arctic- and coastal cod samples are the best estimates of the arctic and coastal cod populations. Because of this, drawing with replacement from the original sample can give a number of new bootstrap samples. By doing the discrimination analysis for all these samples, we can find a mean discrimination score together with confidence intervals for the discrimination results.

In Section 2.1 we introduced the coefficient vectors, $\mathbf{w}$, the EFD coefficients and, $\mathbf{v}$, the 1DF coefficients of the otolith contours. The method is the same for both $\mathbf{v}$ and $\mathbf{w}$ so for simplicity we are looking at $\mathbf{w}$ with k coefficients.

Given a coastal cod group with n_c otoliths and an arctic cod group with n_a otoliths, the EFD coefficient matrices are given as:

$$\mathbf{W}_c = \begin{bmatrix} \mathbf{w}_{c1} \\ \mathbf{w}_{c2} \\ \dots \\ \mathbf{w}_{cn_c} \end{bmatrix}_{(n_c \times k)} \quad \text{and} \quad \mathbf{W}_a = \begin{bmatrix} \mathbf{w}_{a1} \\ \mathbf{w}_{a2} \\ \dots \\ \mathbf{w}_{an_a} \end{bmatrix}_{(n_a \times k)}$$

with $\mathbf{W}_c$ being the coefficient matrix of the coastal group and $\mathbf{W}_a$ being the coefficient matrix of the arctic group (the different groups in the studies can be found in Section 3.1).

Now we find the minimum group size, n, defined to be:

$$n = \min(n_c, n_a).$$

We draw, with replacement, n otoliths from each cod group, giving the first bootstrap samples:

$$\mathbf{W}_c^{*1} = \begin{bmatrix} \mathbf{w}_{c1}^{*1} \\ \mathbf{w}_{c2}^{*1} \\ \dots \\ \mathbf{w}_{cn}^{*1} \end{bmatrix}_{(n \times k)} \quad \text{and} \quad \mathbf{W}_a^{*1} = \begin{bmatrix} \mathbf{w}_{a1}^{*1} \\ \mathbf{w}_{a2}^{*1} \\ \dots \\ \mathbf{w}_{an}^{*1} \end{bmatrix}_{(n \times k)}.$$

Now let θ_c^{*1} be the number of correctly discriminated coastal cod and θ_a^{*1} be the number of correctly discriminated arctic cod. We find these scores by discriminating between $\mathbf{W}_c^{*1}$ and $\mathbf{W}_a^{*1}$. How the discrimination is performed can be seen in Section 2.6. For now, we just define the discrimination function as

$$(\theta_c^{*1},\ \theta_a^{*1}) = disc(\mathbf{W}_c^{*1}, \mathbf{W}_a^{*1}).$$

We repeat the drawing with replacement and discrimination 1000 times. This leaves 1000 discrimination scores for each of the two stocks, θ_c^{*1}, θ_c^{*2}, ..., θ_c^{*1000} and θ_a^{*1}, θ_a^{*2}, ..., θ_a^{*1000}. The two discrimination score estimators are now defined as:

$$\hat{\theta}_c = \frac{1}{1000}\sum_{i=1}^{1000}\theta_c^{*i},\ \hat{\theta}_a = \frac{1}{1000}\sum_{i=1}^{1000}\theta_a^{*i},$$

A link to the implementation of this method can be found in Appendix A.

2.6 Fisher's Discrimination Method

For the discrimination in this thesis we have used the EFD and 1DF coefficients extracted from the cod otolith contours. In the discrimination, each otolith is represented as a vector of Fourier coefficients instead of a contour. Because of the normalization described in Section 2.1, the first three EFD coefficients are the same for all otoliths, the same goes for the first two 1DF coefficients. Because of this, they are irrelevant for discrimination purposes and are removed from the analysis.

We define the EFD coefficient vector with 4(n+1) coefficients to be:

$$\mathbf{w} = [a_0,\ b_0,\ c_0,\ d_0,\ a_1,\ b_1,\ c_1,\ ...,\ a_n,\ b_n,\ c_n,\ d_n]$$

The following are the EFD coefficients used in the discrimination. We are using ten as an example as this is the max number used in the discrimination analysis.

$$\mathbf{w} = [d_0,\ a_1,\ b_1,\ c_1,\ d_1,\ a_2,\ b_2,\ c_2,\ d_2,\ a_3]$$

Now lets define the $2(n+1)$ 1DF coefficients as:

$$\mathbf{v} = [a'_0,\ b'_0,\ a'_1,\ b'_1,\ ...,\ a'_n,\ b'_n]$$

and the following are the ten 1DF coefficients used in the discrimination

$$\mathbf{v} = [a'_1,\ b'_1,\ a'_2,\ b'_2,\ a'_3,\ b'_3,\ a'_4,\ b'_4,\ a'_5,\ b'_5]$$

For discriminating between the arctic cod stock and the coastal cod stock, we have used Fisher's approach for classification with populations described in Johnson and Wichern (2007). The discrimination is equal for both the EFD and the 1DF coefficients so for simplicity we are looking at the discrimination method using the EFD coefficients, $\mathbf{w}$.

The idea behind Fisher's discrimination method is to transform the multivariate observations $\mathbf{w}$, in our case the otolith contour Fourier coefficients, into univariate observations u such that the different groups are separated as much as possible. The reasoning for creating these univariate observations being that they are easy to handle, simple functions of the coefficient vectors $\mathbf{w}$.

2.6.1 Theory

For the discrimination we are always presented with two groups of Fourier coefficient vectors. One group originating from the coastal cod stock and one group originating from the arctic cod stock. Each of the coefficient vectors contain k coefficients and each group contains n vectors. The coastal- and arctic groups will always contain the same number of vectors as a result of the bootstrapping we described in Section 2.5. We can present the groups of coefficient vectors as $(n \times k)$ matrices

$$\mathbf{W}_c = \begin{bmatrix} \mathbf{w}_{c1} \\ \mathbf{w}_{c2} \\ \cdots \\ \mathbf{w}_{cn} \end{bmatrix}_{(n \times k)} , \quad \mathbf{W}_a = \begin{bmatrix} \mathbf{w}_{a1} \\ \mathbf{w}_{a2} \\ \cdots \\ \mathbf{w}_{an} \end{bmatrix}_{(n \times k)}$$

for the coastal- and arctic groups, respectively. Now we define:

$$u_{ci} = \hat{\boldsymbol{\lambda}}' \boldsymbol{w}_{ci} \quad \text{and} \quad u_{ai} = \hat{\boldsymbol{\lambda}}' \mathbf{w}_{ai}$$

as some linear combination of the coefficient vectors, which gives the following values:

$$u_{c1}, u_{c2}, \ ..., \ u_{cn}, \quad \text{and} \quad u_{a1}, u_{a2}, \ ..., \ u_{an}$$

for the two groups. We give the separation between these two sets, expressed in standard deviation units, as:

$$separation = \frac{|\bar{u}_{c.} - \bar{u}_{a.}|}{s_u},$$

where

$$\bar{u}_{c.} = \frac{1}{n}\sum_{i=1}^{n} u_{ci}, \quad \bar{u}_{a.} = \frac{1}{n}\sum_{i=1}^{n} u_{ai} \quad \text{and} \quad s_u^2 = \frac{\sum_{j=1}^{n}(u_{cj} - \bar{u}_{c}.)^2 + \sum_{j=1}^{n}(u_{aj} - \bar{u}_{a}.)^2}{2n-2}.$$

We define the following:

$$\bar{\mathbf{w}}_{c.} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_{ci}, \quad \bar{\mathbf{w}}_{a.} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_{ai} \quad \text{and} \quad \mathbf{S}_p = \frac{(n-1)(\mathbf{S}_c + \mathbf{S}_a)}{2n-2}$$

with $\mathbf{S}_c$ and $\mathbf{S}_a$ as the $(k \times k)$ estimated sample covariance matrices of the coastal- and arctic coefficient matrices, $\mathbf{W}_c$ and $\mathbf{W}_a$, respectively. To get the best possible separation between the coastal and the arctic group we use the linear combination that creates the maximum separation between the sample means. The linear combination

$$\hat{u} = \hat{\boldsymbol{\lambda}}' \mathbf{w} = (\bar{\mathbf{w}}_{c.} - \bar{\mathbf{w}}_{a.})' \mathbf{S}_p^{-1} \mathbf{w} \tag{2.7}$$

maximizes the ratio:

$$\frac{(\bar{u}_{c.} - \bar{u}_{a.})^2}{S_u^2} = \frac{(\hat{\boldsymbol{\lambda}}' \bar{\mathbf{w}}_{c.} - \hat{\boldsymbol{\lambda}}' \bar{\mathbf{w}}_{a.})^2}{\hat{\boldsymbol{\lambda}}' \mathbf{S}_p \hat{\boldsymbol{\lambda}}} \tag{2.8}$$

$$= \frac{(\hat{\boldsymbol{\lambda}}' \mathbf{d})^2}{\hat{\boldsymbol{\lambda}}' \mathbf{S}_p \hat{\boldsymbol{\lambda}}} \tag{2.9}$$

for all possible vectors $\hat{\boldsymbol{\lambda}}$ where $\mathbf{d} = (\bar{\mathbf{w}}_{c.} - \bar{\mathbf{w}}_{a.})$. The maximum of the ratio (2.9) is $D^2 = (\bar{\mathbf{w}}_c - \bar{\mathbf{w}}_a)' \mathbf{S}_p^{-1} (\bar{\mathbf{w}}_c - \bar{\mathbf{w}}_a)$. The proof of this maximum is given in Section 2.7

2.6.2 The Allocation Rule and Discrimination Scores

Based on the maximization of ratio (2.9). we can create an allocation rule.

> Given the coastal and the arctic groups, and a contour coefficient vector $\mathbf{w}_0$, allocate $\mathbf{w}_0$ to the coastal group if
>
> $$\hat{u}_0 = (\bar{\mathbf{w}}_{\mathbf{c}\cdot} - \bar{\mathbf{w}}_{\mathbf{a}\cdot})'\mathbf{S}_p^{-1}\mathbf{w}_0 \geq \frac{1}{2}(\bar{\mathbf{w}}_{\mathbf{c}\cdot} - \bar{\mathbf{w}}_{\mathbf{a}\cdot})\mathbf{S}_p^{-1}(\bar{\mathbf{w}}_{\mathbf{c}\cdot} + \bar{\mathbf{w}}_{\mathbf{a}\cdot}) = \frac{1}{2}(\bar{u}_{1\cdot} + \bar{u}_{2\cdot}) = \hat{m}$$
>
> or
>
> $$\hat{u}_0 - \hat{m} \geq 0$$
>
> Allocate $\mathbf{w}_0$ to the arctic group if
>
> $$\hat{u}_0 < \hat{m}$$
>
> or
>
> $$\hat{u}_0 - \hat{m} < 0$$

We have discussed two different approaches to the discrimination process. The first is getting discrimination scores using a leave-one-out algorithm, presented here in pseudo code.

> Given the two group's coefficient matrices, $\mathbf{W}_c$ and $\mathbf{W}_a$,
>
> iterate from $i = 1$ to n
>
> remove $\mathbf{w}_{ci}$ from $\mathbf{W}_c$ and create an allocation rule based on $\mathbf{W}_a$ and the new $\mathbf{W}_c$.
>
> allocate $\mathbf{w}_{ci}$ to one of the groups based on the allocation rule
>
> return $\mathbf{w}_{ci}$ to $\mathbf{W}_c$.
>
> After the n iterations you have a percentage of correctly allocated coastal coefficient vectors. This is the discrimination score, θ_c, for the coastal group.
>
> Repeat the iterations, this time removing $\mathbf{w}_{ai}$ from $\mathbf{W}_a$ to get the discrimination score, θ_a, for the arctic group.

The second method is less computational demanding. It is based on randomly dividing the coastal and the arctic groups in half, using one half to create allocation rules and find discrimination scores by allocating the other half of the groups. This could be a valid choice, but the otolith groups contain around 30 otoliths. By dividing these groups in half, we get groups containing around 15 otoliths. When bootstrapping 1000 times from these groups we often ended up with covariance matrices being singular or close to singular, making any discrimination scores unreliable. Because of this, we decided to use the leave-one-out algorithm. Our main concern with using the leave-one-out algorithm is that outliers may have a bigger impact on the scores. As a precaution for this, we have checked the data set for outliers in Section 3.1.

In Figure 2.8 we are presenting an illustration of the discrimination method using two groups $\mathbf{W}_c^*$ and $\mathbf{W}_a^*$ to create allocation rules and then applying them to the groups $\mathbf{W}_c$ and $\mathbf{W}_a$. This is a 2D figure,

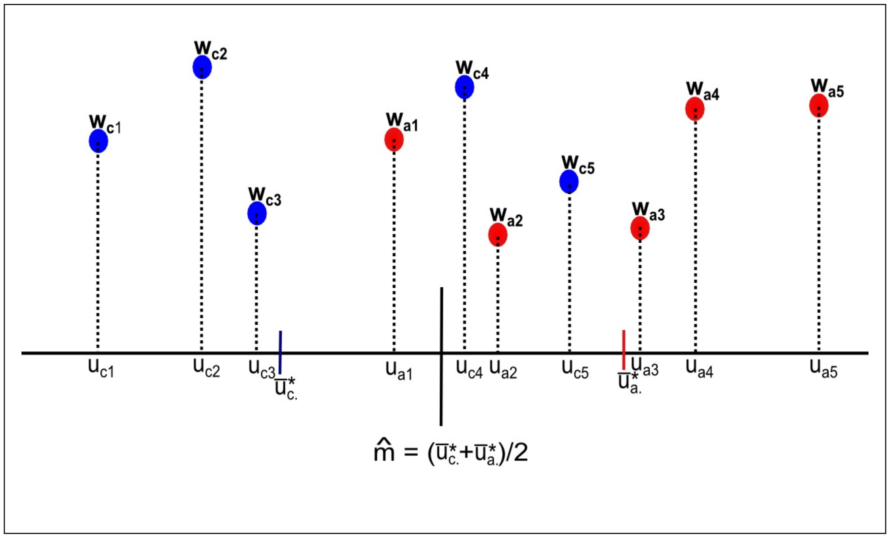

Figure 2.8: Illustration of Fisher discrimination, applied to contour vectors $\mathbf{w}_{ij}$ from the i groups. $\bar{u}^*_{1.}$, $\bar{u}^*_{2.}$ and $\hat{m}_0$ are calculated based on a different sample of contour vectors $\mathbf{w}^*_{in}$, originating from the same i groups.

but we can imagine the vectors being k-dimensional. We see how the coefficient vectors are transformed into scalars and are allocated to groups dependent on which side of the mid point $\hat{m}_0$ they are. From the figure, we see that the coastal coefficient vectors $\mathbf{w}_{c4}$ and $\mathbf{w}_{c5}$ will be allocated to the arctic group, while the arctic coefficient vector $\mathbf{w}_{a1}$ will be allocated to the coastal group. This will give a coastal discrimination score of $\theta_c = 60\%$ and an arctic discrimination score of $\theta_a = 80\%$. We have decided to illustrate this second method, instead of the leave-one-out method, to give a good overview of how the discrimination scores are found.

Recall from Section 2.5 that we are repeating the discrimination for 1000 bootstrap samples, giving 1000 discrimination scores for each group. All discrimination scores presented in Chapter 3 are the mean values of 1000 such scores. We have done the implementation of the discrimination method and the leave-one-out algorithm and the code can be found in Appendix A.

In this discrimination method there is no assumption of normality, however, it is implicitly assumed that the covariance matrices for the coefficients are equal. This is because of the use of a pooled estimate of the common variance matrix. Introducing $\mathbf{\Sigma}_c$ as the covariance matrix of the coastal cod stock Fourier coefficients and $\mathbf{\Sigma}_a$ as the covariance matrix of the arctic contour Fourier coefficients , we have tested the hypothesis $H_0 : \mathbf{\Sigma}_c = \mathbf{\Sigma}_a$ vs. $H_1 : \mathbf{\Sigma}_c \neq \mathbf{\Sigma}_a$ with a Box's F-test and significance level of $\alpha = 0.05$. We keep H_0 for both the EFD coefficients and the 1DF coefficients and continue with the assumption of equal covariance matrices.

Box's F-test is built on the assumption of normality, an assumption which we haven't made. However,we acknowledge that it is robust to non-normality in cases where the sample sizes are large. Also, the main issue with the test with regards to non-normality is that H_0 is too easily rejected. Because of the size of the samples and the fact that we keep H_0, we find the test trustworthy.

2.7 Chapter Appendix

Proof Eq.(2.9)

The maximization lemma from matrix algebra states that for a positive definite matrix $\mathbf{B}_{pxp}$ and a given vector $\mathbf{d}_{px1}$ with an arbitrary non-zero vector $\mathbf{x}_{(px1)}$

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{d}'\mathbf{B}^{-1}\mathbf{d} \tag{2.10}$$

with maximum achieved when $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{d}$ for any constant $c \neq 0$.

The maximum of the ratio in (2.9) is given by applying (2.10) directly. By setting $\mathbf{d} = (\bar{\mathbf{w}}_1 - \bar{\mathbf{w}}_2)$, we have

$$\max_{\hat{\boldsymbol{\lambda}}} \frac{(\hat{\boldsymbol{\lambda}}'\mathbf{d})^2}{\hat{\boldsymbol{\lambda}}'\mathbf{S}_p\hat{\boldsymbol{\lambda}}} = \mathbf{d}'\mathbf{S}_p^{-1}\mathbf{d} = (\bar{\mathbf{w}}_1 - \bar{\mathbf{w}}_2)'\mathbf{S}_u^{-1}(\bar{\mathbf{w}}_1 - \bar{\mathbf{w}}_2) = D^2 \tag{2.11}$$

where D^2 is the sample squared distance between the two means.

Chapter 3

Results

In this chapter we use the EFD method (Section 2.1) and the 1DF method (Section 2.2) to extract Fourier coefficients from contours extracted from 604 otolith images. These coefficients are used in Fisher's discrimination method (Section 2.5) to discriminate between different groups of coastal- and arctic cod. The goal is to see which of these Fourier methods are best suited for discrimination (e.g. which method provides us with the highest discrimination scores).

We introduce the data set in Section 3.1. Here we check the data for outliers and discuss some choices with regards to cod groups used in the discrimination. In Section 3.2 and 3.3 we aim to optimize the performance of the 1DF method and EFD method, respectively. We are studying how they perform with regards to differently smoothed otolith contours, how different numbers of Fourier coefficients affect the scores, and we are studying how robust the methods are with regards to the scanned otoliths orientation. After finding the optimized usage of each method, we are comparing their discrimination abilities in Section 3.4 and we conclude as to which method is preferable for discrimination between arctic- and coastal cod.

3.1 Data Set

In this section we give a description of the data set together with some of its attributes (Section 3.1.1). Furthermore, we are dividing the data set into discrimination groups based on different covariates (Section 3.1.2), before we check for outliers (Section 3.1.3) in all the groups.

3.1.1 Description of the Data Set

The dataset consists of 1177 scanned otolith images with 915 left otoliths and 262 right otoliths. The cod were caught between the years 1999 and 2001 in the Barents Sea and along the coast of Norway and Svalbard. In Figure 3.1 we can see the different distributions within the sample. Some more details surrounding the acquisition of the cod otoliths and cod otolith images can be found in Stransky et al. (2008). The images used in the analysis were provided by the Institute of Marine Research.

The otolith images are scanned from left otoliths from cods with length between 30 and 70 cm and ages between 2 and 6 years that originate from either the arctic- or the coastal stock, without uncertainty. These are the 604 otolith images mentioned in the introduction to this chapter. The reasonings behind our choice of selection can be seen in Section 3.1.2.

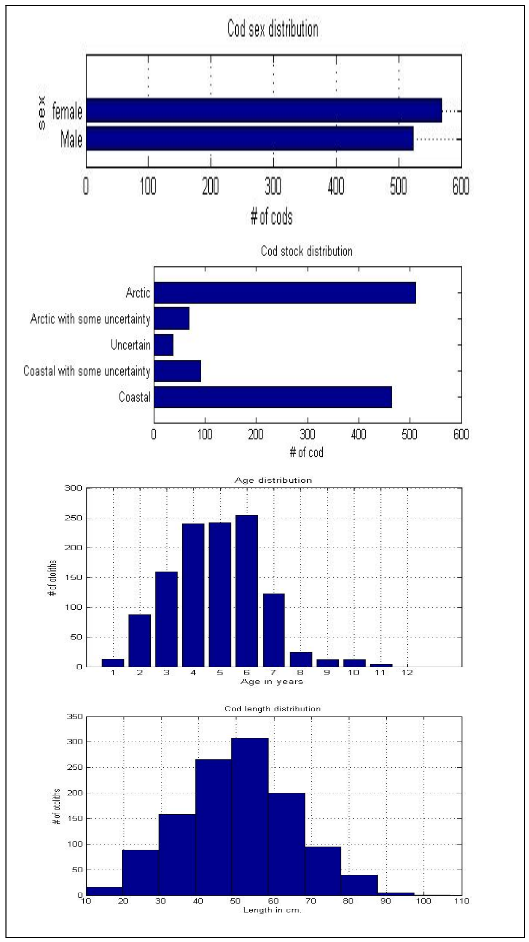

Figure 3.1: Cod's length, age, sex and stock distribution.

3.1.2 Discrimination Groups

To be able to effectively discriminate between coastal- and arctic cod, we need to ensure that the discrimination scores are a result of different fish stocks, and not some other covariate. In table 3.1 the average length, weight and age together with the sex distribution of the coastal- and arctic cods used in the analysis are found. The average length of the arctic cod sample is 49,94 cm, while the average length of the coastal cod sample is 49,14 cm. The average coastal cod weight is 1179g, while the average arctic cod

Sample attributes:

	Coastal cod	Arctic cod
Sex:	39,5% males	48,5 % males
Age:	5,27 years	4,64 years
Length:	49,14 cm	49,94 cm
Weight:	1179 g	1257 g

Table 3.1: Details surrounding the coastal and arctic samples used in the analysis. Average age, length and weight together with percentage of males. All data based on cods with length between 30 and 70 cm.

Discrimination groups:

	Male		Female	
	Arctic cod	Coastal cod	Arctic cod	Coastal cod
Age: 2-4 years Length: 30-40 cm	Gr. 1 : 18	Gr. 2 : 28	Gr. 3 : 32	Gr. 4 : 20
Age: 3-5 years Length: 40-50 cm	Gr. 5 : 27	Gr. 6 : 25	Gr. 7 : 37	Gr. 8 : 32
Age: 4-6 years Length: 50-60 cm	Gr 9 : 34	Gr. 10 : 29	Gr. 11 : 26	Gr. 12 : 27

Table 3.2: Groups used in the discriminant analysis. In the cells, group numbers and group sizes are shown.

weight is 1257g. Because of these differences, the discrimination scores may not be a product of otoliths from different fish stocks, but may be caused by otoliths originating from different weight classes of fish, or different lengths of fish. To avoid this, we try to create arctic and coastal groups which are close in length, age and sex so that we are certain that the discrimination scores are based on different stocks. We divide the otoliths into the 12 groups that are shown in Table 3.2. The groups have overlapping ages, but they are distinct. No otoliths are found in more than one group. In Sections 3.2, 3.3 and 3.4 we plot discrimination results and all the plots have a number in the upper left corner. These numbers refer to the groups described in Table 3.2.

3.1.3 Checking for Outliers

Mahalanobis Distance

The Mahalanobis Distance (Mahalanobis, 1936) is a measure of the distances between a point and the common group point (e.g. the group mean). It is a convenient way of measuring distances as it does take into account the correlation between observations. These distances are often used for discrimination purposes, but in our case, we are measuring the distance between otolith Fourier coefficients and the mean Fourier coefficients of the group of otoliths to identify potential outliers in the data set. In our case, which is the multivariate case, we can find the Mahalanobis distance as follows.

Given a vector with Fourier coefficients $\mathbf{w}_i = [w_{i1}, w_{i2}, w_{i3}, ..., w_{ik}]^T$ from a group of vectors $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N$ with mean $\boldsymbol{\mu} = [\bar{w}_{.1}, \bar{w}_{.2}, ..., \bar{w}_{.k}]^T$ the Mahalanobis distance is given as:

$$D_M(\mathbf{w}_i) = \sqrt{(\mathbf{w}_i - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\mathbf{w}_i - \boldsymbol{\mu})}$$

where $\mathbf{S}$ is the $(k \times k)$, sample covariance matrix for the given group.

Group Dot Plots

To check the data set, we have made a study of the Mahalanobis distances and plotted them in dot plots to check for outliers. We have done this for the overall arctic group and the overall coastal group. The results can be seen in Figure 3.2. We don't see any immediate outliers in these sets. There are some high values, but that is to be expected with the size of these data sets. We have done the same analysis for all the 12 different groups in Table 3.2 and we found no outliers.

3.2 1DF Optimization

We would like to find the best smoothing to apply to the contours before extracting the 1DF coefficients. Our decision will be based on what gives the highest discrimination scores when using the coefficients in Fisher's method. To do this, we plotted the discrimination scores for the groups described in Section 3.1.2. To find the best smoothing we have tested the WMA-smoothing method (Section 2.3) with three different levels of smoothing: 20, 60 and 100 smoothing iterations. We have compared these results with the results obtained when applying the concavisation method (Section 2.4) and when not applying any smoothing to the contours at all. This analysis is done both for the original otolith images and for rotated otolith

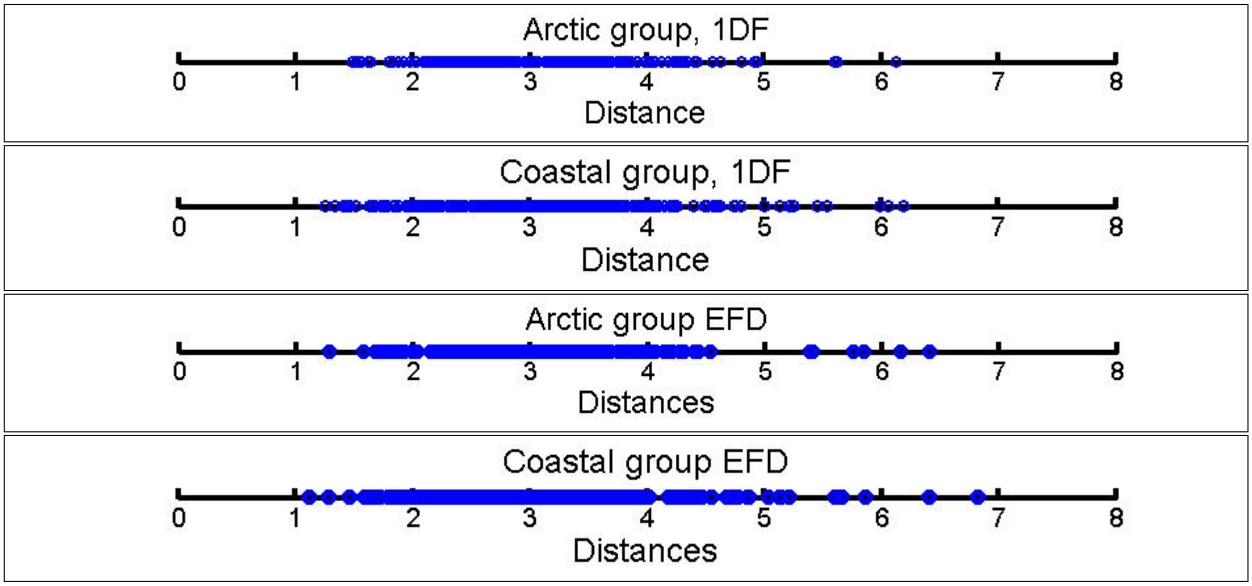

Figure 3.2: Dot plots of the Mahalanobis distances of the arctic and coastal group with EFD coefficients and 1DF coefficients.

images. The reasons to use smoothing on the contours are described in Section 2.3 where we introduced the WMA-smoothing. In the plots, the bars along each graph represent the bootstrap confidence intervals. These intervals are not Bonferroni-corrected as our focus lies in scores for specific coefficient numbers and not for all coefficients numbers. Later on we want to find what number of coefficients provide us with the highest results. Because of this, we have to find the smoothing methods that performs best for these numbers. As an example, if we look at Figure 3.5 and the plot for group 1, we see that using nine and ten coefficients give the best results for all the smoothing methods. We have decent results for coefficients numbers between four and eight as well, but not as good. For less then four coefficients we see that the scores drop drastically. As we won't be using less then four coefficients in a future discriminations, we don't really focus on how the methods perform for these coefficient numbers. In most cases using six or more coefficients will be preferable. What is interesting is to see how these smoothing methods perform for nine and ten coefficients and then between six and eight coefficients. Least important how they perform for less than six coefficients.

The orientation of the otoliths when scanning them may have an influence on the discrimination. We want to see how robust the discrimination method using 1DF coefficients is with regards to this. Detecting this effect requires us to find the best overall smoothing for both the original and the rotated otoliths. To get the rotated otoliths we have rotated the images instead of rescanning the otoliths. This rotation can be seen in Figure 3.3. Notice that the images have different sizes, though the otoliths have the same size. When rotating the image you either have to resize the image or get a smaller otolith. To be able to detect the rotation effect we decided to resize the image. The reason for testing the rotation is that the contours can get different levels of pixel noise based on different orientations. A straight line may turn into a jagged line if it's rotated. You can see this effect in Figure 3.4. We have looked at a rotation of 45 degrees as we expect it to gives the largest differences in pixel noise. The reasoning behind this is that a straight line will turn into a total jagged line when rotated 45 degrees. An illustration of this can be seen in Figure 3.4. We will look at how the orientation gives different scores. Keep in mind though, that whether we are looking at rotated or non-rotated images, all images within one discrimination are oriented in roughly the same way.

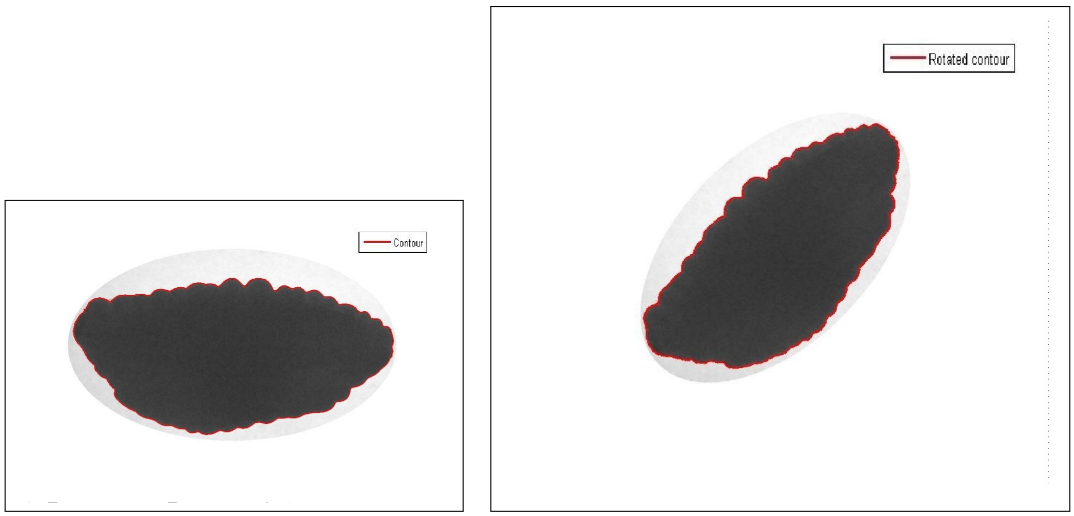

Figure 3.3: Original otolith image and the 45 degree rotated otolith image. The red outline represents the closed contours extracted from the otolith images.

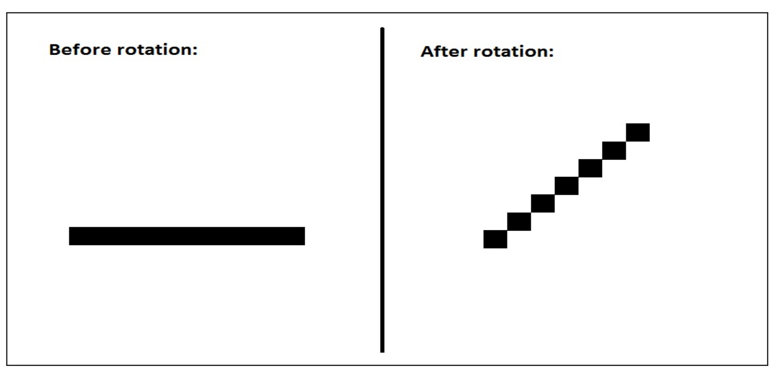

Figure 3.4: How rotation effects a straight line. To the left we see an unrotated straight line, and to the right we see it after a 45 degree rotation. This is a line represented with 11 pixels.

3.2.1 Smoothing Analysis Using Non Rotated Otolith Images

We start off by looking at Figure 3.5. As mentioned earlier, we don't focus on the results for low coefficient numbers as we see they perform worse for all smoothing methods. Starting at the plot for group 1, we see that the concave contours provide slightly worse results for 4, 5 and 6 coefficients while it performs better for 7 and 8 coefficients. The WMA-smoothed contours are giving worse results for all number of coefficients and it also seems like a higher number of WMA-smoothing iterations gives worse results. For 9 and 10 coefficients the non-smoothed and the concave contours give the highest scores and close to equal scores. Looking at groups 2, 3 and 4 we find the same patterns. The concave contours give worse results for fewer number of coefficients and almost equal scores for 9 and 10 coefficients. The plot for group 2 gives slightly better scores for 100 iteration WMA-smoothed contours, but the differences aren't significant. For groups 3 and 4 the non smoothed contours give the best results. Overall, we would prefer to use the non smoothed contours based on these results, with the concave contours being our second choice.

Moving on to Figure 3.6 we see the same tendencies for all four plots. The non smoothed contours perform better than the WMA-smoothed contours. More WMA-smoothing iterations gives worse results in all four cases. What is interesting to see, though, is that the concave contours gives much better performances for groups 5, 7 and 8. It is slightly worse for group 6, but not by that much and overall, based on these four plots, the concave contours are by far best choice.

In Figure 3.7 we see that the differences between the methods are much less in these groups. There aren't really any significant differences between the non-smoothed contours and the WMA-smoothed contours, but the concave contours seems to give slightly worse results for all but 9 and 10 coefficients.

For all 12 groups I recommend using the concave contours as they perform good in all cases and significantly better for some groups. A second choice would be the non smoothed contours. It seems that applying WMA-smoothing worsens the results in almost all cases.

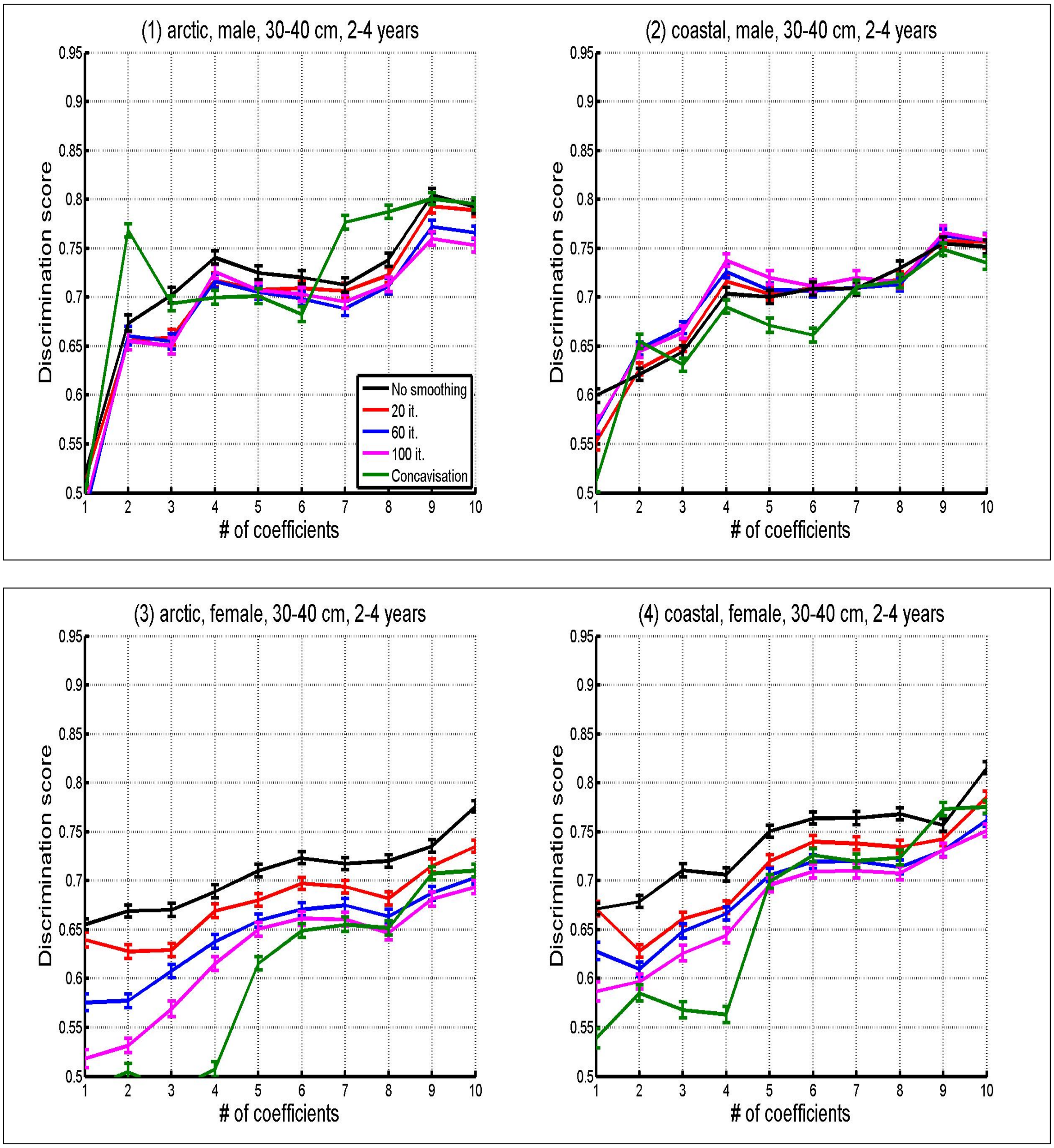

Figure 3.5: Scores with different smoothed contours for arctic and coastal cod, 30-40cm and 2-4 years. Average discrimination scores based on 1000 bootstrap samples are plotted against the number of coefficients used in the discrimination.

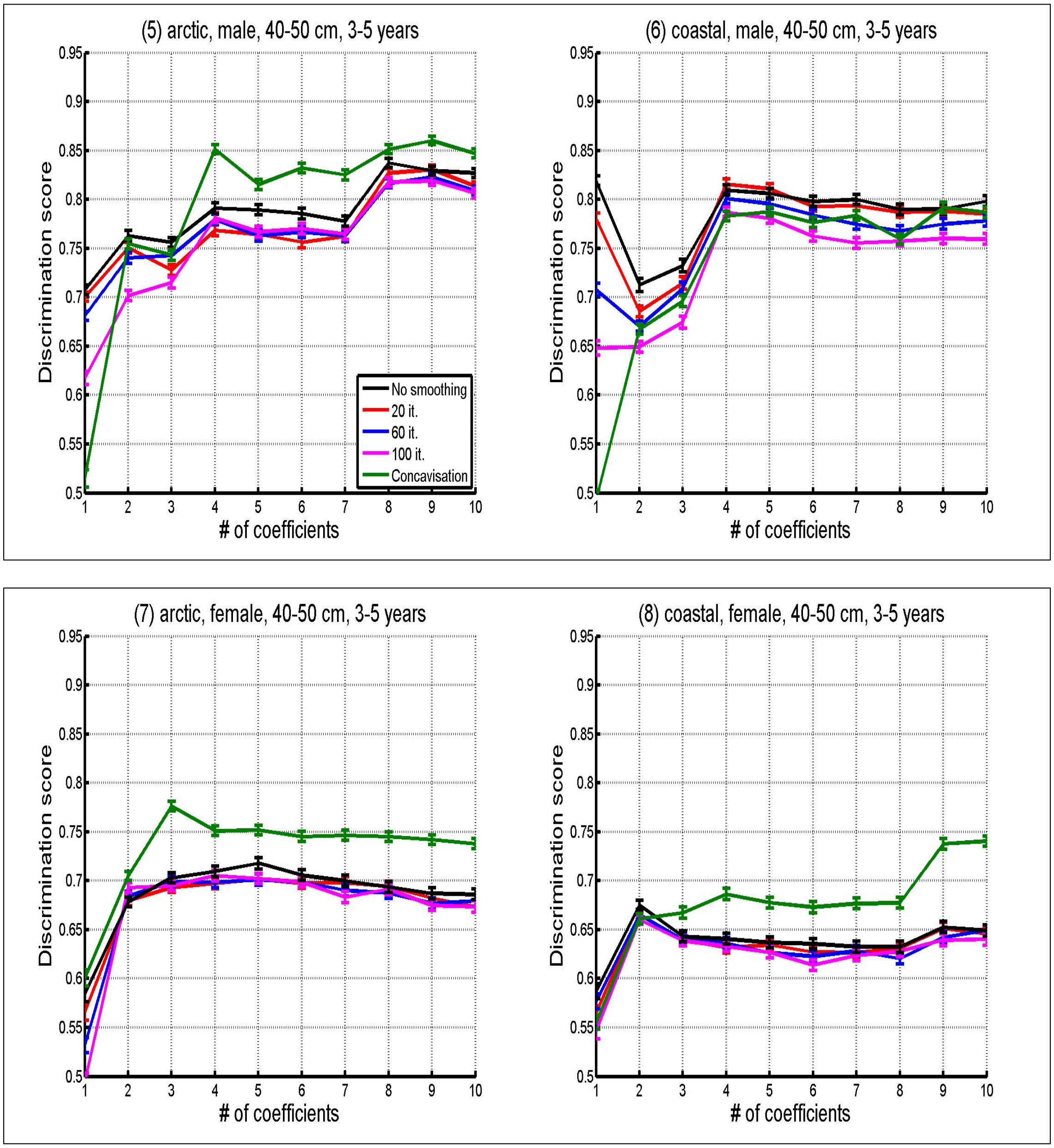

Figure 3.6: Scores with different smoothed contours for arctic and coastal cod, 40-50cm and 3-5 years. Average discrimination scores based on 1000 bootstrap samples are plotted against number of coefficients used in the discrimination.

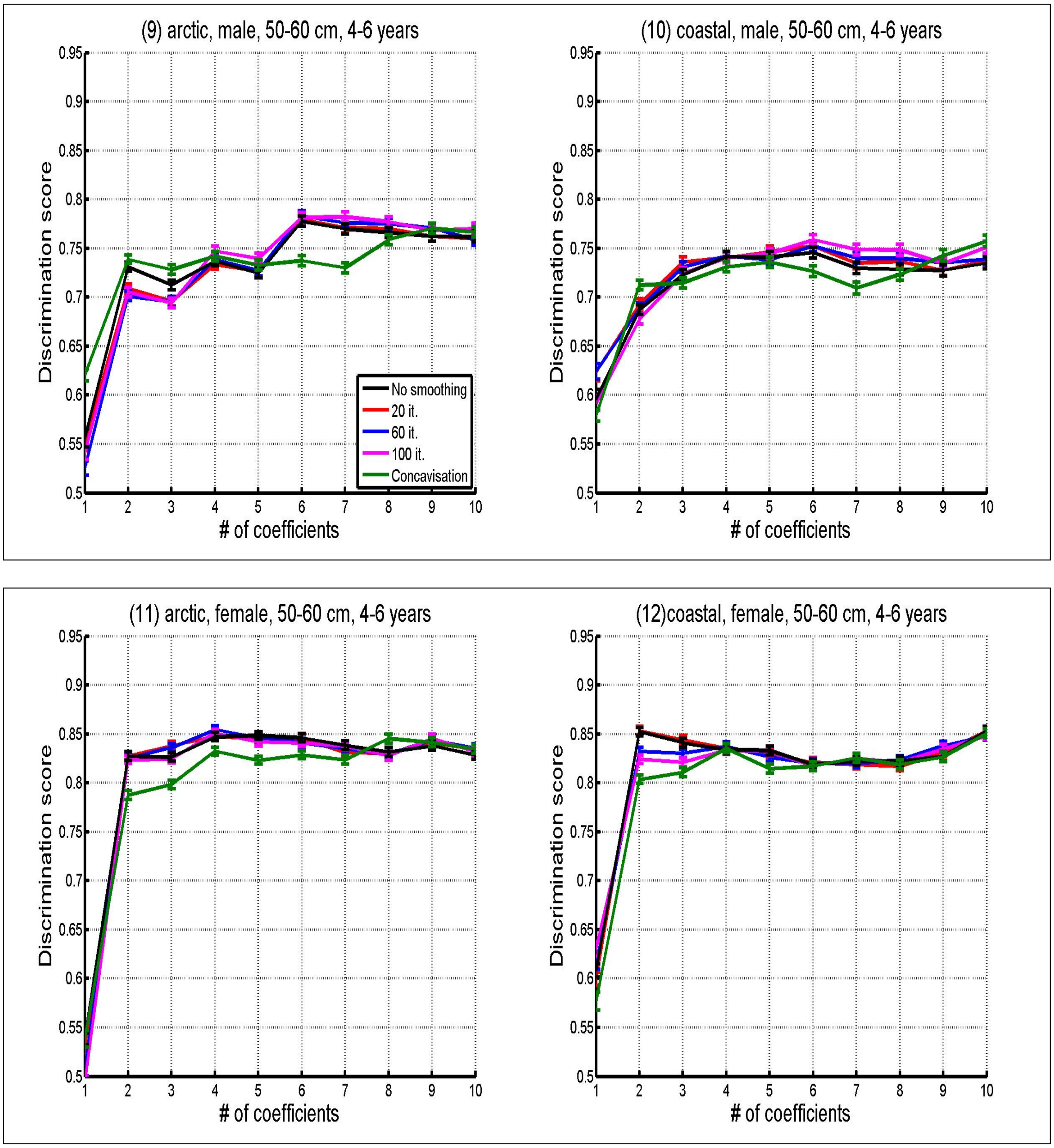

Figure 3.7: Scores with different smoothed contours for arctic and coastal cod, 50-60cm and 4-6 years. Average discrimination scores based on 1000 bootstrap samples are plotted against number of coefficients used in the discrimination.

3.2.2 Smoothing Analysis Using Rotated Images

We now repeat the analysis in the previous section, but this time for results based on rotated otolith images.

In Figure 3.8 we see that the concave contours again give worse results for lower number of coefficients. For group 1 however, they give significantly higher results for 6 or more coefficients. For groups 3 and 4, the non smoothed contours gives best results while for group 1 and 2 the concave and the 100 iteration WMA-smoothed contours provides the highest scores. I would say using 20 iteration, WMA-smoothing, would be benefitial based on these four plots. The concavisation may be a good second choice because of the high scores obtained for group 1.

In Figure 3.9 we see the same as we did for the non-rotated images. The concave contours give much higher results for groups 5, 7 and 8. Overall there isn't much difference between the different iteration, WMA-smoothed contours. An exception for group 6 where 20 iterations gives highest scores. The concave contours are natural choices based on these plots. 20-iteration WMA-smoothing would be a good second choice.

For Figure 3.10 we again see that the differences are minimal, however the non smoothed contours seems to perform worse in all cases. The concave contours give slightly more varying results for all groups so using WMA-smoothing may be best based on these four groups. What number of iterations to use in the WMA-smoothing is difficult to see from these plots, as they perform equally.

All in all, the concave contours does seem to give high performances for all groups and they give much better results in some cases. So for the rotated images, the concave contours would again be the wisest choice. However, for these rotated contours, 20 iteration WMA-smoothing seems like a second best choice instead of the non rotated contours.

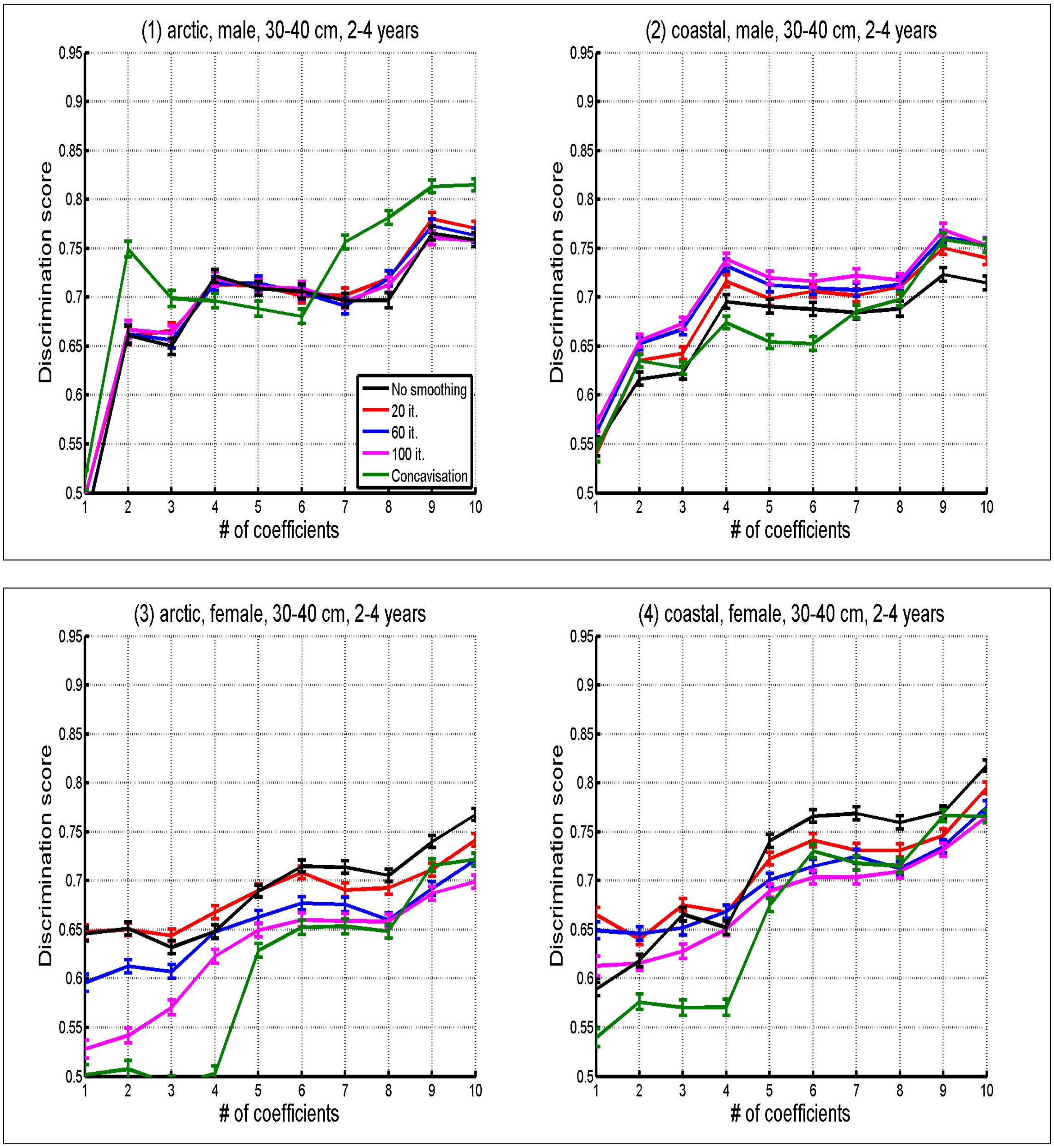

Figure 3.8: Scores with different smoothed contours for arctic and coastal cod, 30-40cm and 2-4 years. Average discrimination scores based on 1000 bootstrap samples are plotted against the number of coefficients used in the discrimination.

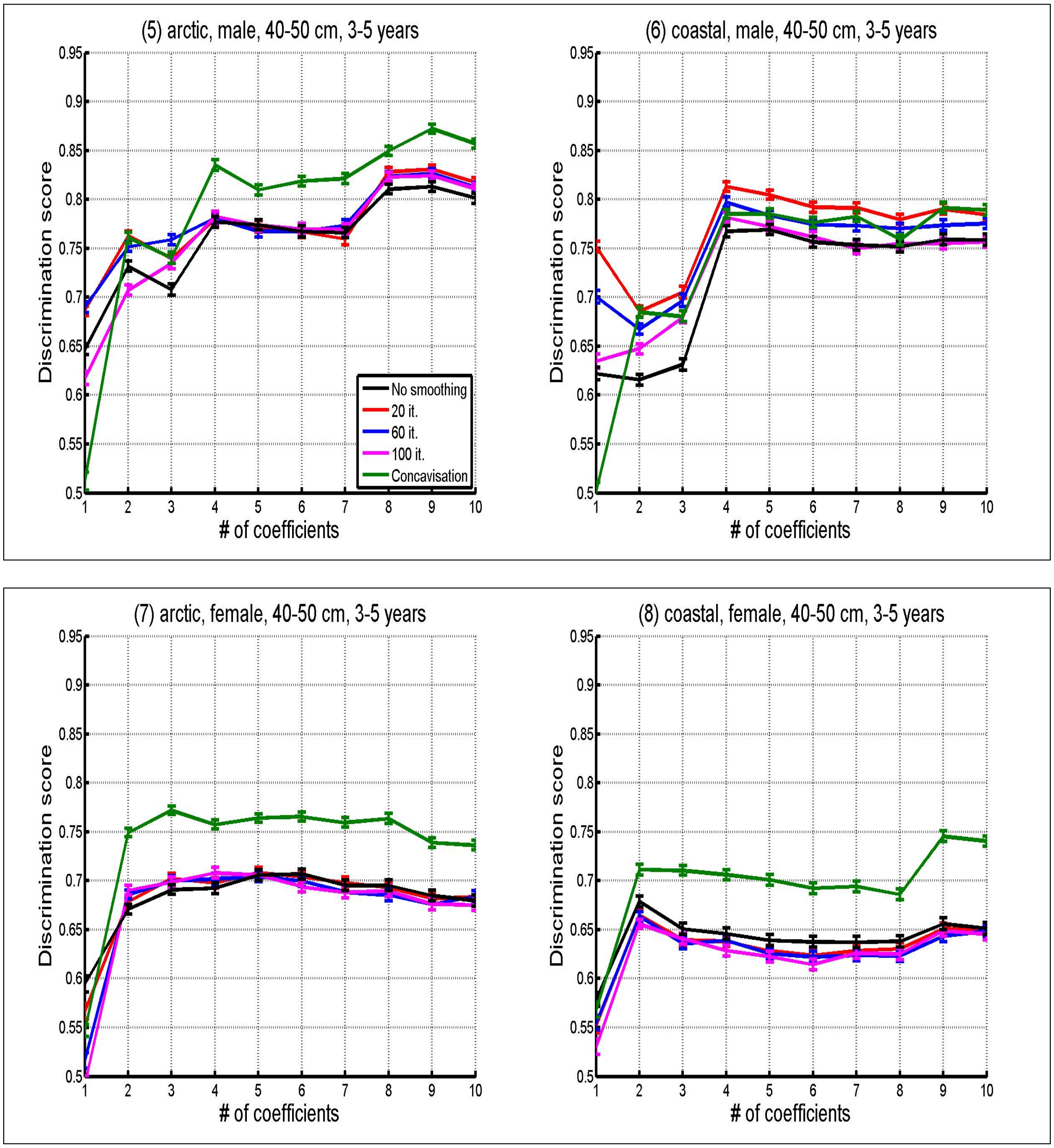

Figure 3.9: Scores with different smoothed contours for arctic and coastal cod, 40-50cm and 3-5 years. Average discrimination scores based on 1000 bootstrap samples are plotted against the number of coefficients used in the discrimination.

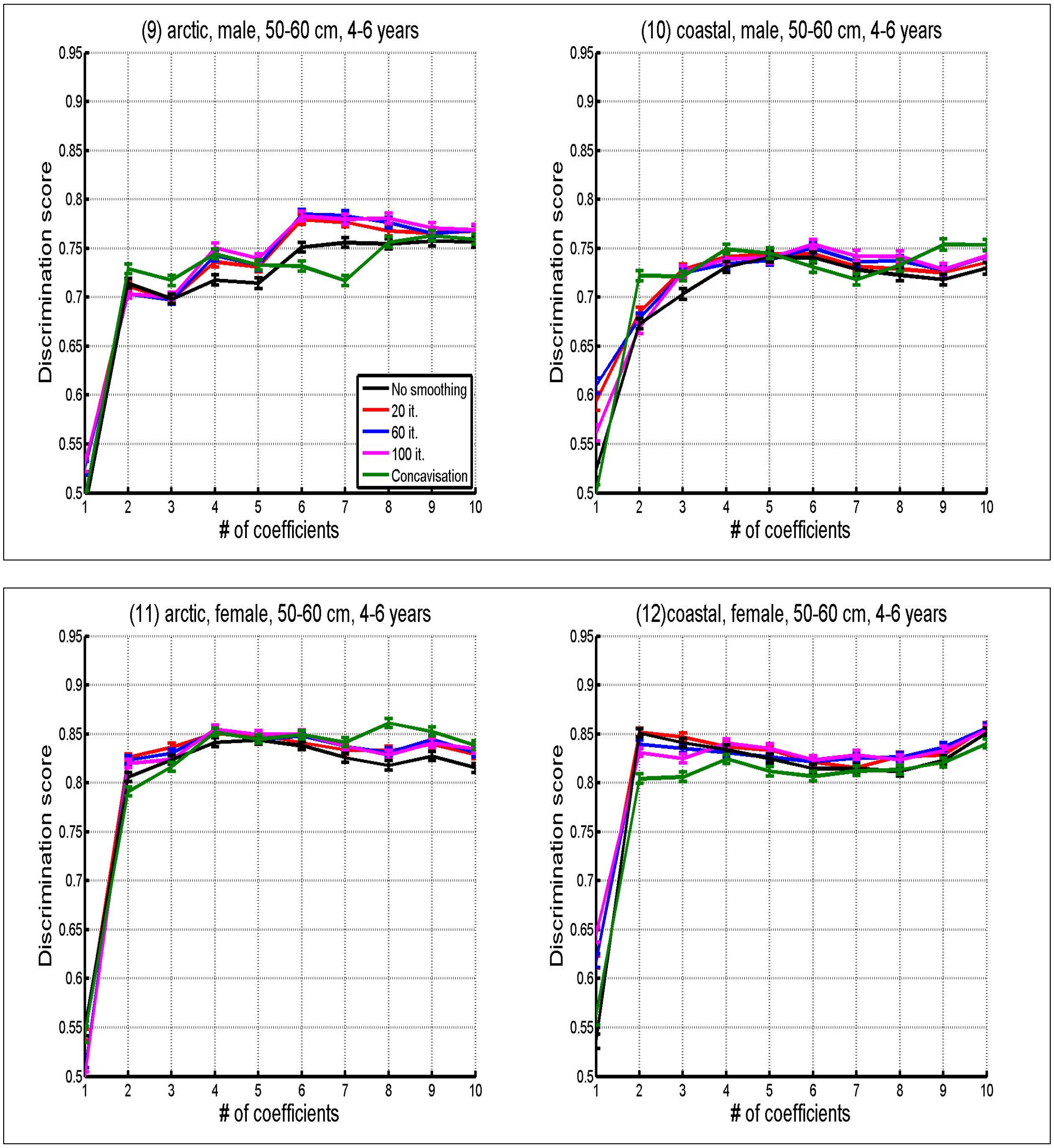

Figure 3.10: Scores with different smoothed contours for arctic and coastal cod, 50-60cm and 4-6 years. Average discrimination scores based on 1000 bootstrap samples are plotted against the number of coefficients used in the discrimination.

3.2.3 Summary on the Different Smoothing Techniques

From all the plots in this chapter we see that concavisation is a valid method to smooth the contours. It doesn't provide the highest scores for all groups, but it performs well in all cases and it gives the highest scores for some of the groups. In general, WMA-smoothing performs better for low numbers of coefficients, but as low numbers of coefficients generally give varying results, how the different smoothing methods effect the scores for higher numbers of coefficients are more interesting. As the aim is to find an overall smoothing method that will work well for both the original images and the rotated images we have decided to find the average discrimination scores obtained with the different smoothings over all 12 groups with original oriented images and the same 12 groups with rotated images. The results can be found in Table 3.3.

Based on the results in this chapter, we recommend to use at least 6 coefficients in a discrimination analysis. Looking at the table we see that the concavisation performs better if we average over all the cod groups and 6 to 10 coefficients. For exact numbers of coefficients the concave contours give best results for 7, 8, 9 and 10 coefficients and stands out as an obvious choice when using the 1DF method. To ensure that the scores aren't products of one or two groups which gives brilliant discriminations scores with the concave contours, we have decided to look at the smoothing methods with the use of rankings. When looking at Table 3.3 we see that 10 coefficients give the best results with almost all smoothing methods. As a result we look at the results for every group when using 10 coefficients. We are ranking the different smoothing methods with numbers from 1 to 5, giving the smoothing method providing the highest score, 5, and the method providing the lowest score, 1. We average the ranks for each method across the 24 groups (12 cod groups $\times$ rotated- and non rotated images). We now have an average rank telling us how well the method performs on average compared to the others. The rankings can be seen in Table 3.4. It is clear that the concave contours provides the highest average scores.

For the non rotated images, the contours without smoothing gave the second best results while for the rotated otolith images the second best results were obtained using 20, WMA-smoothing iterations. This may imply that rotating the otoliths creates some more pixel noise in the contours, leaving us with having to use smoothing on the contours. If we compare the results for the rotated and non rotated images, we see small variations. Based on the results, the orientation when scanning the otoliths does not have a huge impact on the score. Remember, this is based on all otoliths from each discrimination being oriented the same way.

1DF Smoothing Results:

Coefficients:	No smoothing	20 iter. WMA	60 iter. WMA	100 iter. WMA	Concavisation
6	74,65 %	74,41 %	73,91 %	73,61 %	73,84 %
7	74,22 %	73,94 %	73,59 %	73,37 %	74,35 %
8	74,63 %	74,55 %	73,91 %	73,82 %	75,36 %
9	75,84 %	75,93 %	75,47 %	75,16 %	78,15 %
10	76,56 %	76,43 %	75,91 %	75,42 %	78,03 %
Mean:	75,18 %	75,05 %	74,56 %	74,28 %	75,95 %

Table 3.3: Performance of different smoothing methods averaged across 24 groups with rotated and non-rotated otolith images.

1DF Smoothing Ranks:

Coefficients:	No smoothing	20 iter. WMA	60 iter. WMA	100 iter. WMA	Concavisation
10	2,88	3,08	2,75	2,33	3,96

Table 3.4: Mean rank of the different smoothing methods based on ranking for different groups.

3.3 EFD Optimization

In this section we are repeating the study in Section 3.2, but this time using the EFD method to extract the coefficients from the contours (Section 2.1). In the section 3.4 we want to compare the discrimination scores we get from the EFD coefficients with the scores from the 1DF coefficients, but first we need to find the best suitable smoothing for the contours before extracting the EFD coefficients.

We repeat the same steps as we did for the 1DF method in Section 3.2.

3.3.1 Smoothing Analysis Using Non Rotated Contours

Figures 3.11, 3.12 and 3.13 show how the EFD method performs given contours with no smoothing, with three different levels of WMA-smoothing and with concave contours.

Looking at Figure 3.11 we see that the concave contours provide the best results for groups 1 and 2, while it performs worse for groups 3 and 4. There isn't much of a difference between scores based on WMA-smoothed contours and non smoothed contours, but it seems like scores based on contours with no smoothing and 20 WMA-smoothing iterations work best. Looking at these groups, we find the concave contours to give the best scores, with 20 iteration WMA-smoothing being a second choice.

Moving on to Figure 3.12 we see that for groups 5 and 6 the concave contours provide much lower scores for all numbers of coefficients above 3. For groups 7 and 8 the concave contours give good results for

8, 9 and 10 coefficients, but give much worse results for 6 and 7 coefficients. Based on these groups the non smoothed contours or the 20 iteration WMA-smoothed contours are the best choices. The concave contours give good results in some cases, but overall we find the scores to be too varying.

In Figure 3.13 it seems that the non smoothed contours give worse scores than the WMA-smoothed contours. Again we find the concave contours to give too varying results. Based on these groups we would say the 100 iteration, WMA-smoothed contours are preferable. However, the differences between the results given different levels of WMA-smoothing aren't huge, so any choice would be valid.

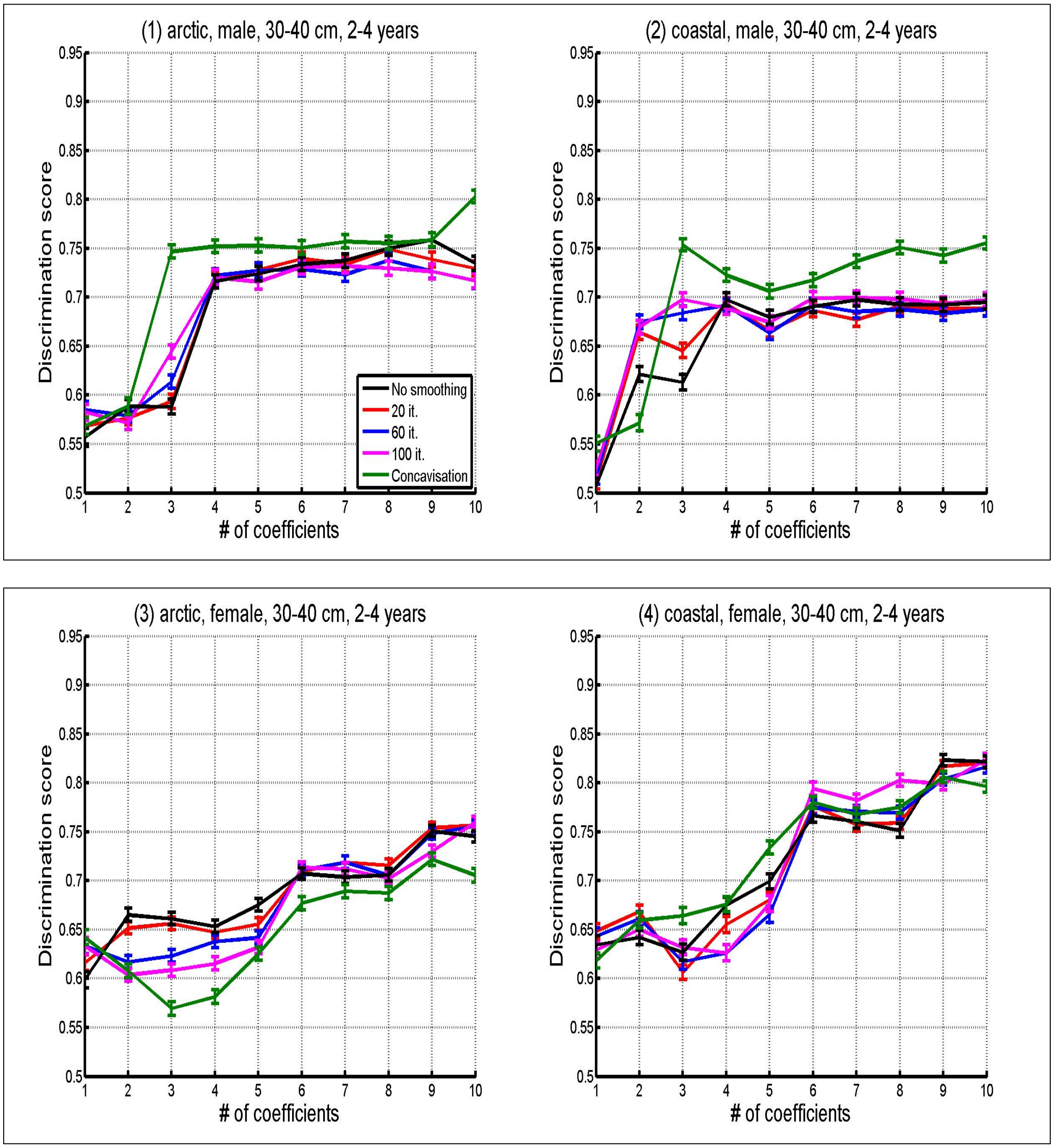

Figure 3.11: Scores with different smoothed contours for arctic and coastal cod, 30-40cm and 2-4 years. Average discrimination scores based on 1000 bootstrap samples are plotted against the number of coefficients used in the discrimination.

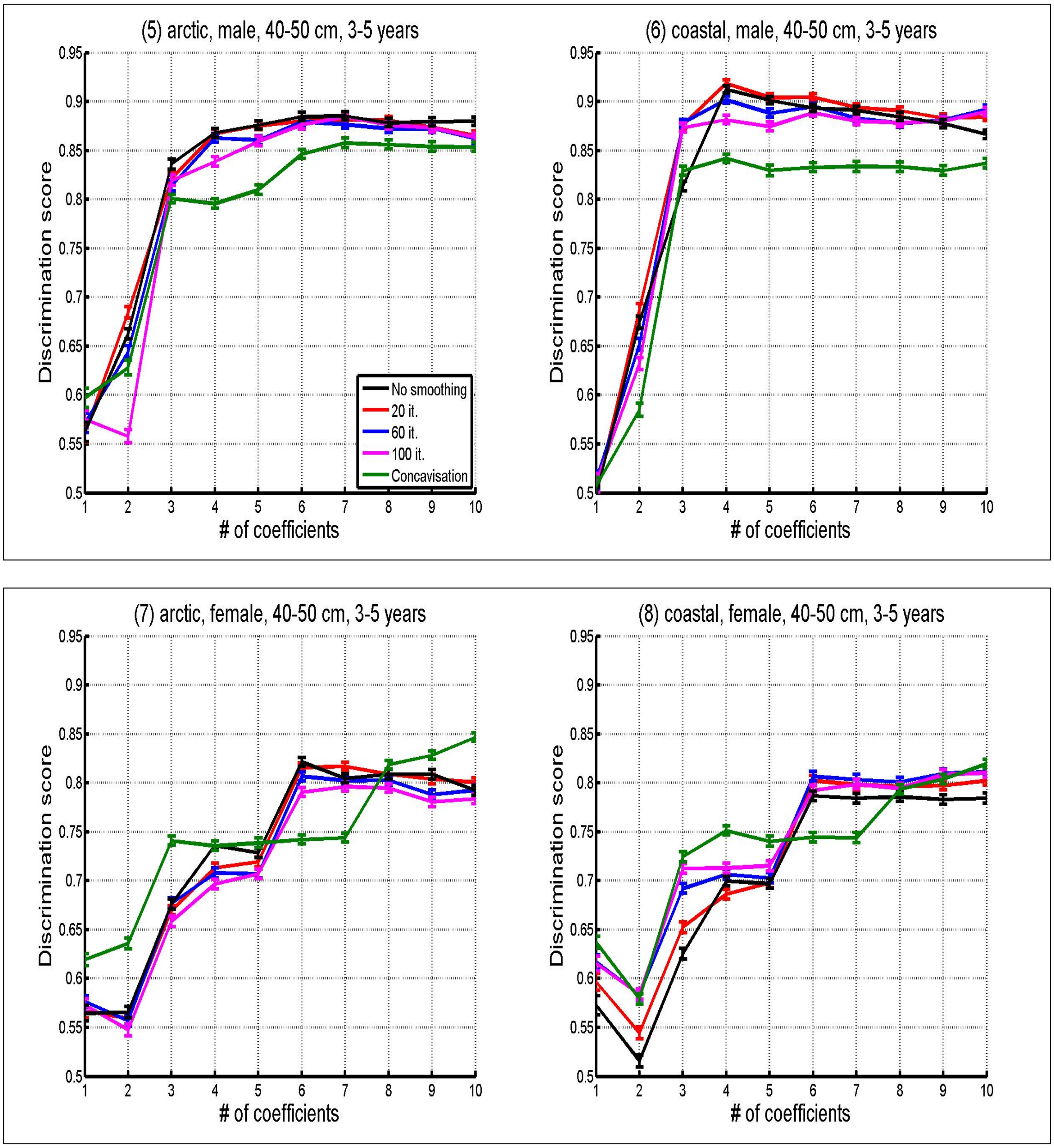

Figure 3.12: Scores with different smoothed contours for arctic and coastal cod, 40-50cm and 3-5 years. Average discrimination scores based on 1000 bootstrap samples are plotted against the number of coefficients used in the discrimination

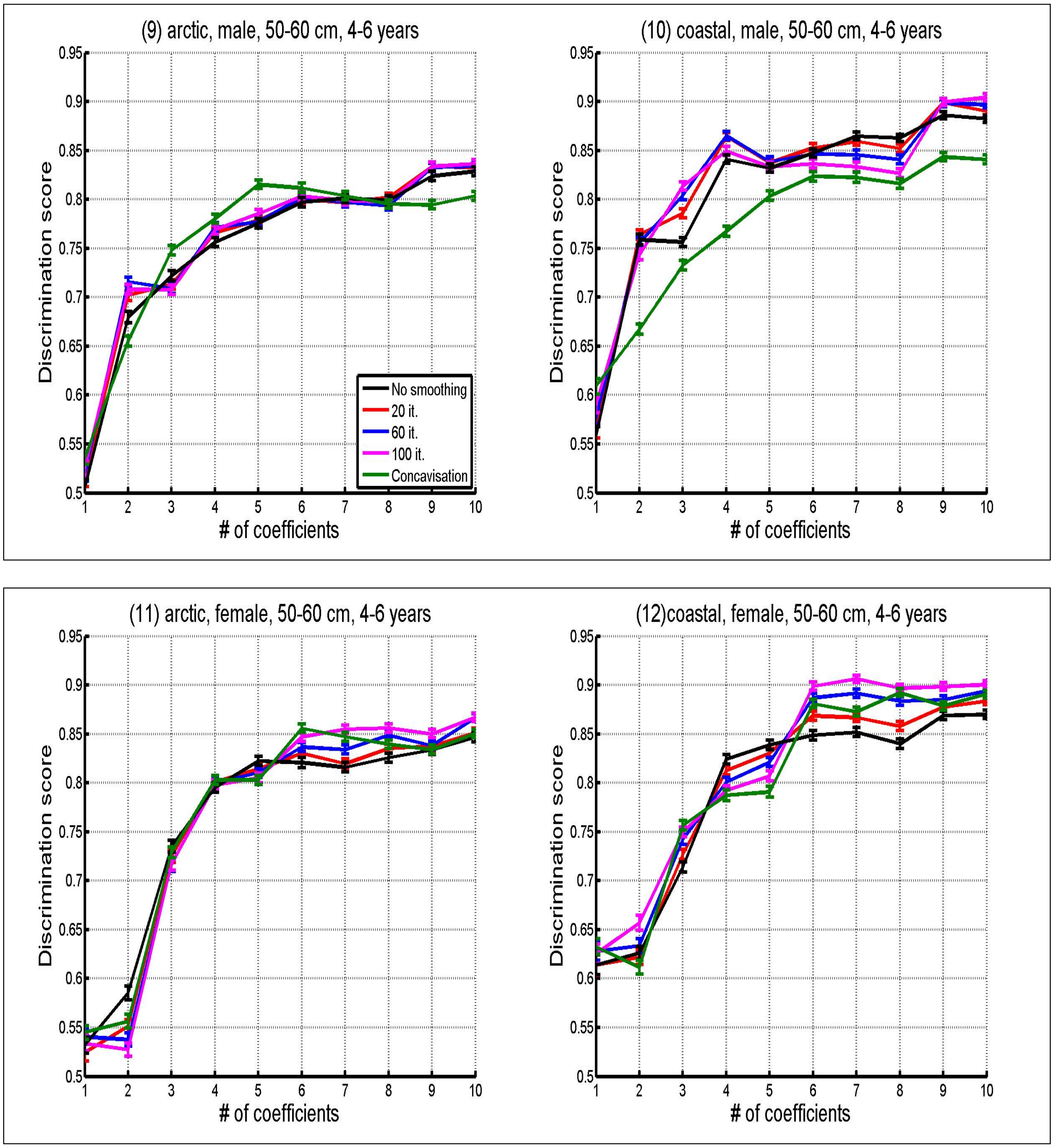

Figure 3.13: Scores with different smoothed contours for arctic and coastal cod, 50-60cm and 4-6 years. Average discrimination scores based on 1000 bootstrap samples are plotted against the number of coefficients used in the discrimination

3.3.2 Smoothing Analysis Using Rotated Contours

Now we repeat the same analysis, but this time for the rotated contours. The results can be found in Figures 3.14, 3.15 and 3.16. All details and reasons for looking at rotated contours can be found in Section 3.2.

Starting off with Figure 3.14, we see that the concave contours give much higher scores for groups 1 and 2, same as we found for the non rotated contours. The non smoothed contours give significantly lower scores than the WMA-smoothed contours for groups 3 and 4. The concave contours give good results in all four plots, and would be a natural choice. There aren't much difference between the different WMA-smoothing levels.

Looking at Figure 3.15 we see the concave contours giving significantly better results for groups 7 and 8. For groups 5 and 6, the concave contours give significantly worse results for small numbers of coefficients, while at higher numbers the scores are more even. It seems like using WMA-smoothing with 60 iterations is the best method. In general for all groups, 60 or 100, WMA-smoothing iterations give most stable, good results.

The results displayed in Figure 3.16 are a bit more difficult to interpret. The concave contours does seem to give worse results in almost all cases except for group 11, while using WMA-smoothing with 60 iterations gives the best overall results. The differences aren't significant between the different levels of WMA-smoothing, but the concave contours tend to give more varying results based on coefficient numbers and cod groups.

Overall for these rotated contours, the non-smoothed contours give much more varying results than the different smoothed contours. This is an effect that wasn't as clear for the non-rotated contours, so using some smoothing is a good idea, as the pixel noise does seem to have an effect. By looking only at the results for the rotated contours, the 100-iteration, WMA-smoothed contours are the best choice, while using 60 iterations as a second choice.

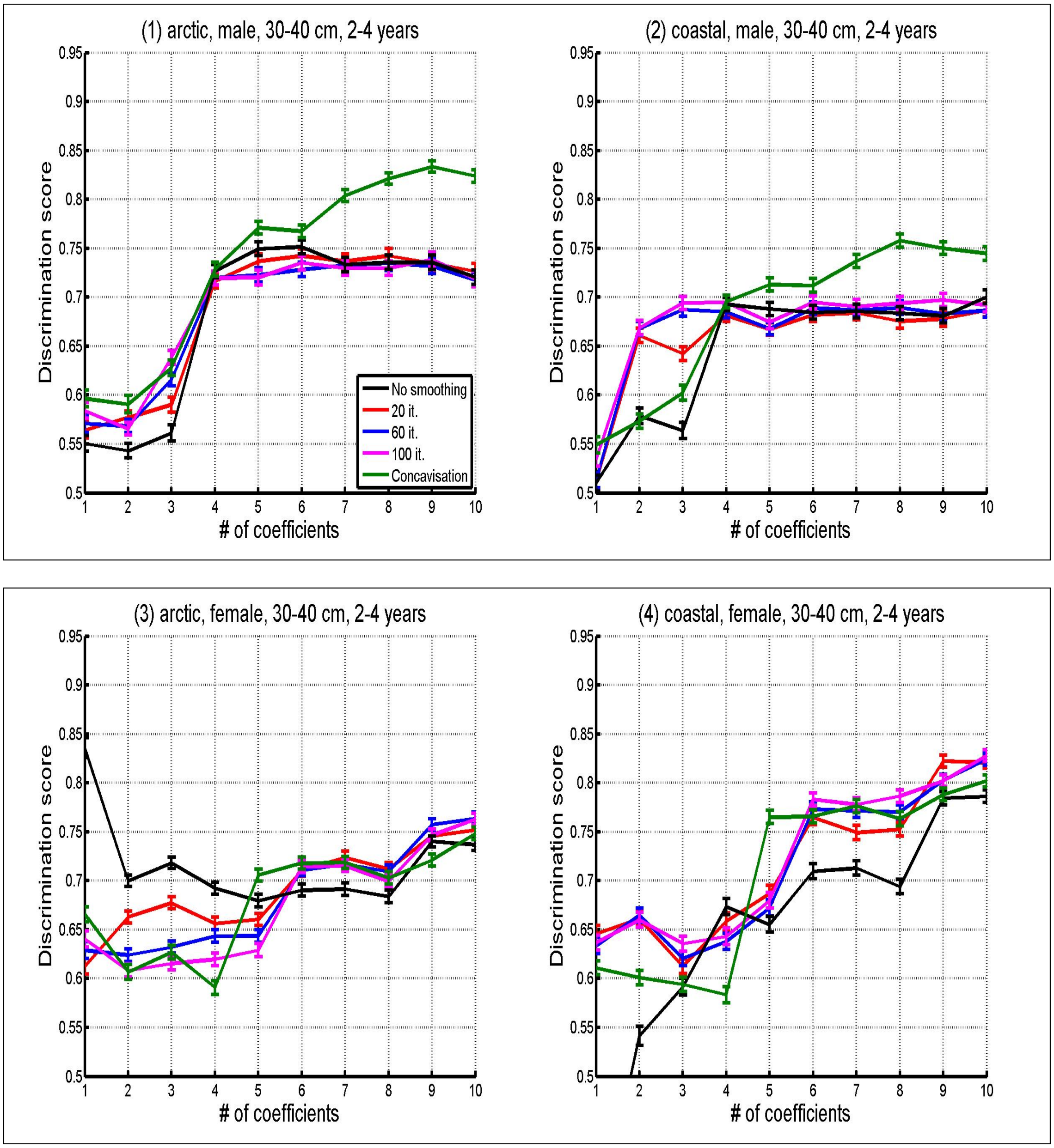

Figure 3.14: Scores with different smoothing iteration numbers for arctic and coastal cod, 40-50cm and 3-5 years. In plot (3) we have males and in plot (4) we have female. Average discrimination scores based on 1000 bootstrap samples are plotted against number of coefficients used. The otlith contours analyzed are extracted from rotated images.

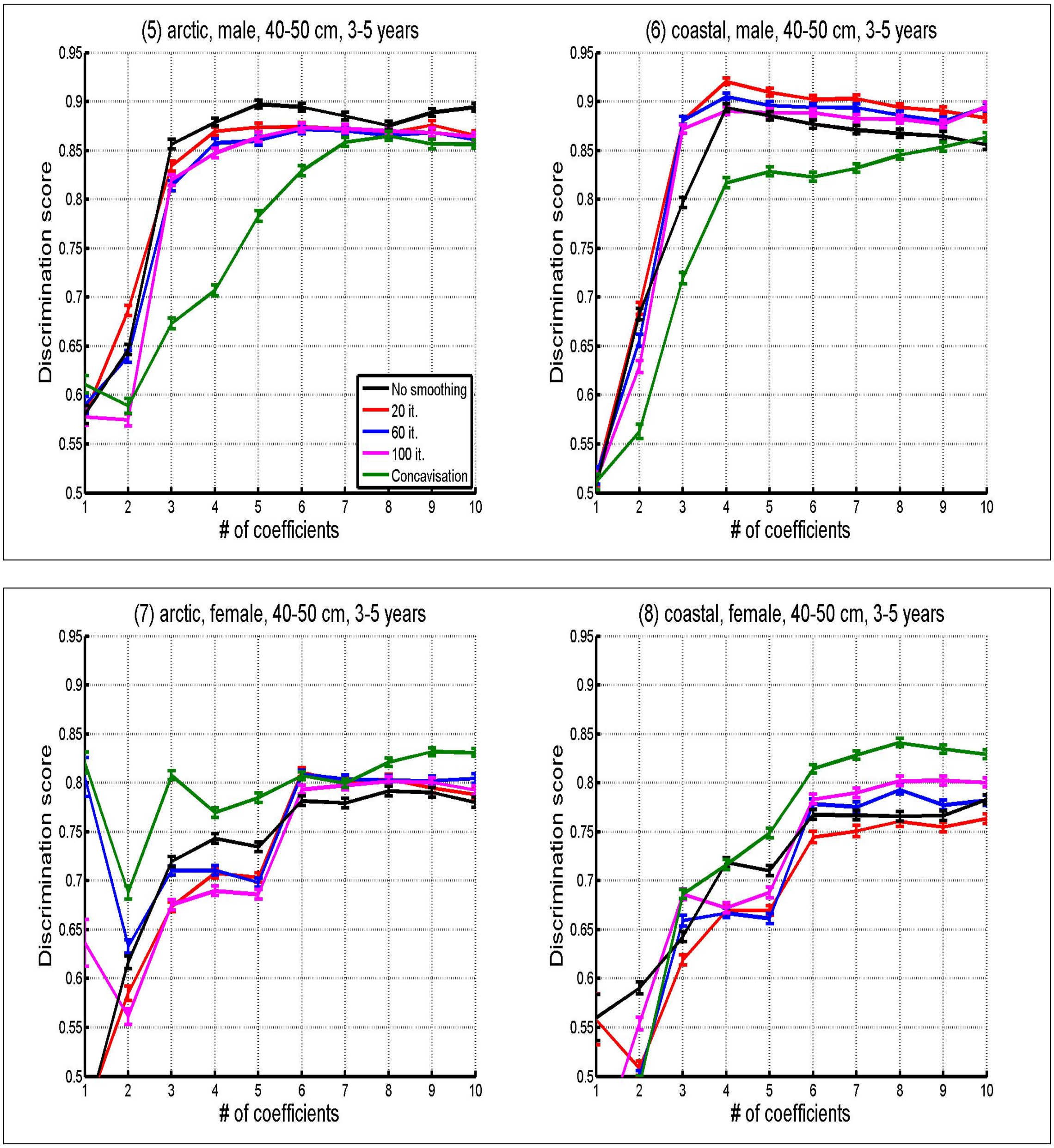

Figure 3.15: Scores with different smoothing iteration numbers for arctic and coastal cod, 40-50cm and 3-5 years. In plot (3) we have males and in plot (4) we have female. Average discrimination scores based on 1000 bootstrap samples are plotted against number of coefficients used. The otlith contours analyzed are extracted from rotated images.

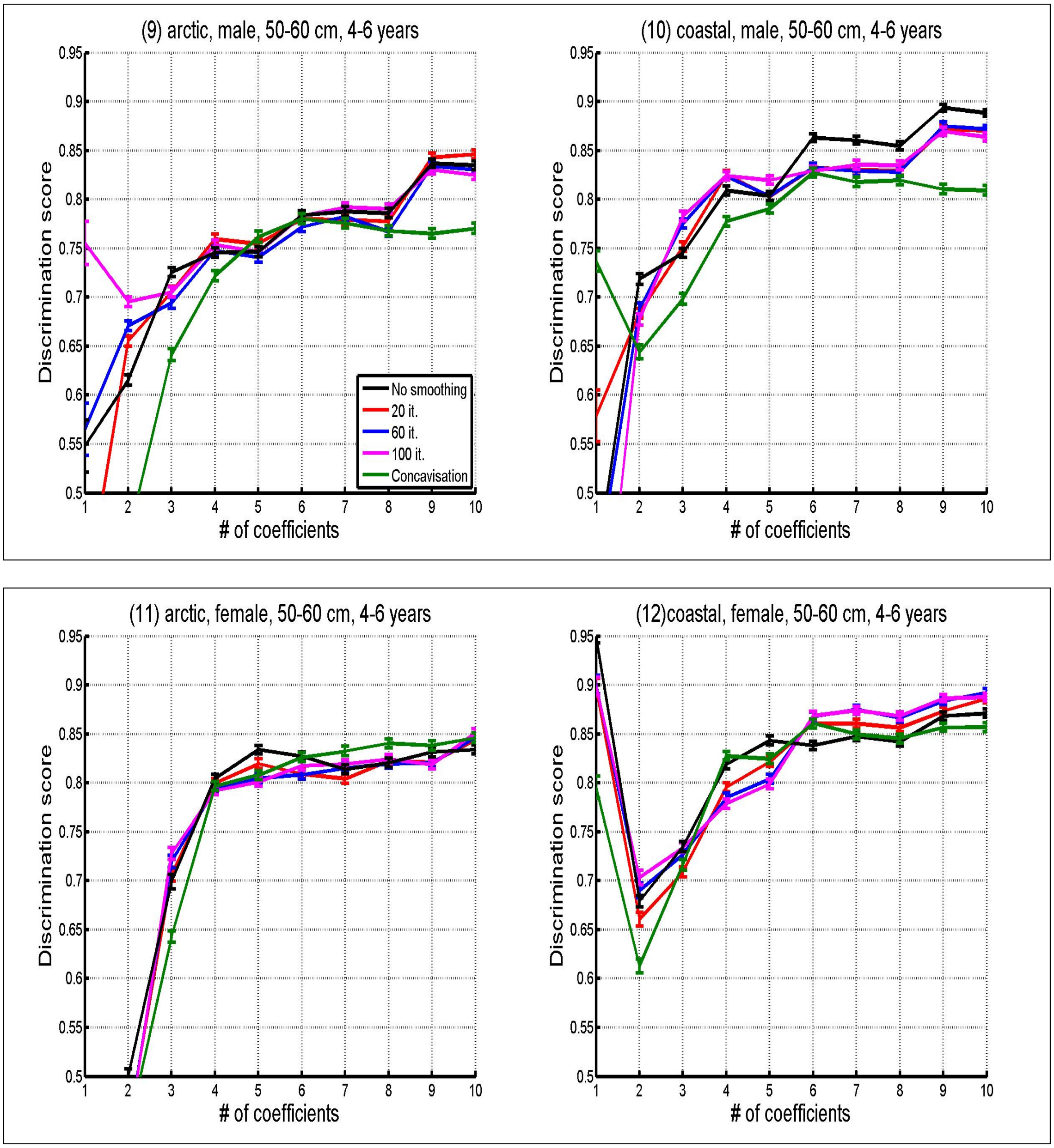

Figure 3.16: Scores with different smoothing iteration numbers for arctic and coastal cod, 40-50cm and 3-5 years. In plot (3) we have males and in plot (4) we have female. Average discrimination scores based on 1000 bootstrap samples are plotted against number of coefficients used. The otlith contours analyzed are extracted from rotated images.

3.3.3 Summary on the Different Smoothing Techniques

With regards to the concave contours, they gave best results for some groups, but also the worst results for other groups. All in all, they vary a bit too much for the different groups for us to trust the concavisation method to provide stable, good results. Table 3.5 shows that the concave contours give the highest average score. However, looking at the ranks for the different methods (Table 3.6, we see that the concavisation has the second worse score. This supports the conclusion that the concavisation method gives too varied results.

We found that using contours without any smoothing was not an option, and we can see this in the tables as well. The non smoothed contours give the worst average rank and they give the worst average discrimination scores.

We are left with the three different iteration numbers for the WMA-smoothing method. For the non-rotated contours we found that 20 iterations were sufficient and gave the best results. For the rotated contours it was apparent that more iterations gave better results. Our original thought was that a choice of 60 iterations would be best, as it was a middle road between 20 and 100 iterations. However, the differences in scores between the smoothing levels are quite small for all groups, and by looking at Table 3.5 and 3.6, we see that 100 iterations give the best overall scores and the best average rank. Based on this, we conclude with the 100 iteration WMA-smoothing method being the best choice when using the EFD method.

EFD Smoothing Results:

Coefficients:	No smoothing	20 iter. WMA	60 iter. WMA	100 iter. WMA	Concavisation
6	79,44 %	79,91 %	80,00 %	80,14 %	80,09 %
7	79,31 %	79,64 %	79,93 %	80,23 %	80,48 %
8	79,11 %	79,72 %	79,80 %	80,15 %	81,22 %
9	81,11 %	81,29 %	81,18 %	81,29 %	81,66 %
10	80,97 %	81,41 %	81,67 %	81,81 %	81,70 %
Mean:	79,99 %	80,39 %	80,52 %	80,72 %	81,03 %

Table 3.5: Performance of different smoothing methods averaged across 24 groups with rotated and non-rotated otolith images.

EFD Smoothing Ranks:

Coefficients:	No smoothing	20 iter. WMA	60 iter. WMA	100 iter. WMA	Concavisation
10	2,58	2,88	3,29	3,46	2,67

Table 3.6: Mean rank of the different smoothing methods based on ranking for different groups.

3.4 EFD and 1DF Comparison

In this section we are doing the comparison between the discrimination scores obtained when using EFD coefficients and when using 1DF coefficients. In Sections 3.2 and 3.3 we found the optimal smoothing to apply to the contours before extracting these coefficients. For extracting the EFD-coefficients we are using 100 iteration, WMA-smoothed contours as this was the choice providing the most stable, good results for all groups and for both rotated and non rotated otolith images. For the 1DF-coefficients we are using the concave contours for the same reasons. All the contours used for the comparison in this section are based on the non rotated images. All the results are presented with graphs in Figures 3.18, 3.19 and 3.20 with numbers in Table 3.7.

We have done the analysis with the rotated contours as well, but these results aren't shown in plots. However, we have presented the differences in discrimination scores between rotated and non rotated images in Section 3.4.2 where we will give further comments.

3.4.1 Score Comparison

We are comparing the discrimination scores when using the EFD coefficients and when using 1DF coefficients. We have done the comparison the same way as when we compared the different smoothing methods in Sections 3.2 and 3.3. All results plotted in this section can be found in Table 3.7.

Starting with Figure 3.18, for the female groups 3 and 4, the EFD coefficients give the best results. Looking at groups 1 and 2, the EFD coefficients give best scores as well, but only for less than 6 coefficients. When using more than 6, the 1DF coefficients give much higher scores. Moving on to Figure 3.19, in all four groups, the EFD coefficients give higher discrimination scores. For group 5 we see that the 1DF coefficients give similar results for the highest numbers of coefficients. For the final four groups in Figure 3.20, we again see that the EFD coefficients are the best to base the discrimination upon. Group 11 gives somewhat similar results, but based on the other three groups, there is no doubt that the EFD coefficients are preferable. Overall we see the 1DF method performing better for groups 1 and 2 and for groups 7, 11 and 12 for low number of coefficients. However, we need at least 4 coefficients to get stable, good results. In many cases 6 and even 9 or 10 is needed. It is clear that the best way of using these methods is with more than 6 coefficients, and above 6 coefficients, the EFD method is better for 10 out of 12 groups.

Let us analyze the graphs more in depth and see how the methods perform for different groups. It is clear that groups 1, 2, 3 and 4, which contain the youngest/smallest cod, have the lowest discrimination scores. If the differences between coastal- and arctic cod otoliths are caused by the fish being in different environments it seems plausible that the differences will get larger as the fish gets older. Of course, if the differences are caused by different genetics, it is plausible that these differences are as well enlarged when the fish gets older. Either way, it is natural that the scores for the youngest/smallest groups are lower. If we look at the other groups, the results aren't obviously better for the oldest groups. The otoliths grow throughout the life of the cod, but maybe the otolith, after 3-4 years, harbors enough information of the stock to give high discrimination scores, the differences not getting bigger when the cod get even older.

Now we are trying to detect any tendencies for the different sexes. In Figure 3.18 we see that the female groups have higher scores. In Figure 3.19 the male groups have much higher scores, and in Figure 3.20 there aren't any big differences. We can't detect any clear tendencies.

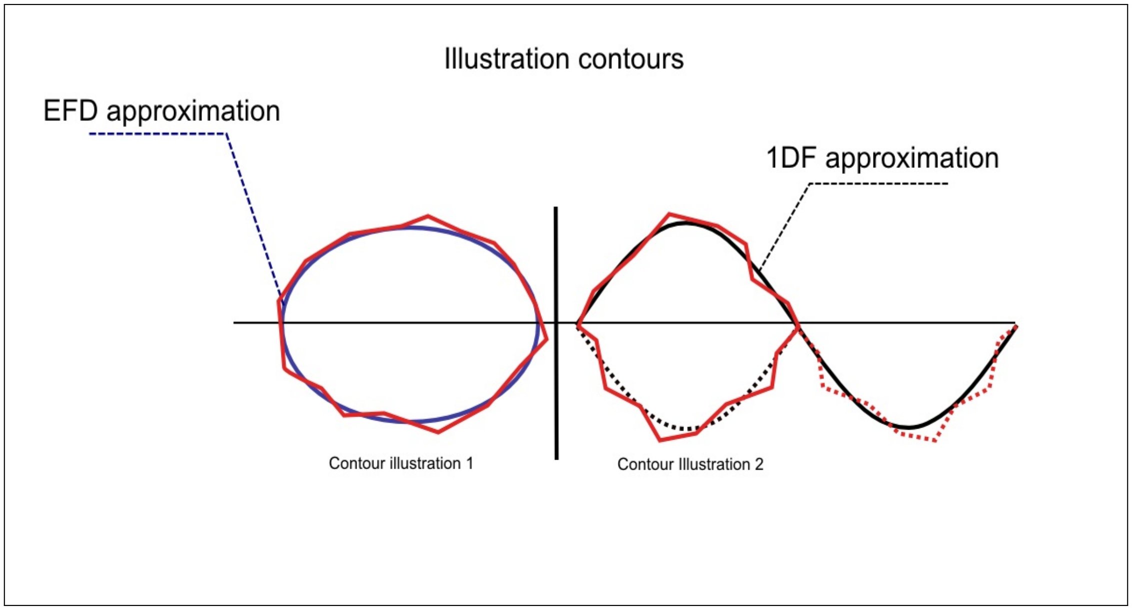

Figure 3.17: An illustration of how the EFD method best approximates contours close to ellipses, while the 1DF method best approximates contours close to sinus curves.

If we look at the differences between the discrimination scores for the coastal groups and the discrimination scores for the arctic groups, it looks like the coastal scores are generally higher than the arctic scores. What is causing these differences is hard to predict, something that will be discussed in Chapter 4.

Based on these results the EFD method applied to 100 iteration WMA smoothed contours, is the best choice for discrimination between coastal- and arctic cod groups. The focus in this thesis hasn't been how good the best Fourier method performs, but how it performs against the other method. However, it is interesting to see how good discrimination we get, using the optimal method with regard to contour smoothing, Fourier method and number of coefficients. All the results are presented in Table 3.7. We see stable good results for all groups. Some groups get almost 90% score and even for the youngest/shortest groups, we have scores around 70%.

A possible explanation for the EFD method to outperform the 1DF method may lie within the general shapes of the contours. The EFD method approximates the contours with the use of harmonic ellipses, while the 1DF method approximates the otolith contours with harmonic functions. A sine function is quite far from an ellipse and if the cod otolith contours are close to ellipses it may explain why the EFD method gives better results. This is illustrated in Figure 3.17. We see that the elliptic shape of the contour illustration 1 suits best for the EFD method and we see, from contour illustration 2, how the shape that is closer to a sinus, when the lower half is flipped, is better suited for the 1DF method. This may also be a reason for why the concavisation method is best together with the 1DF method, but not for the EFD method. Again, finding definitive reasons for such results isn't possible at this time, for now, we are content documenting them.

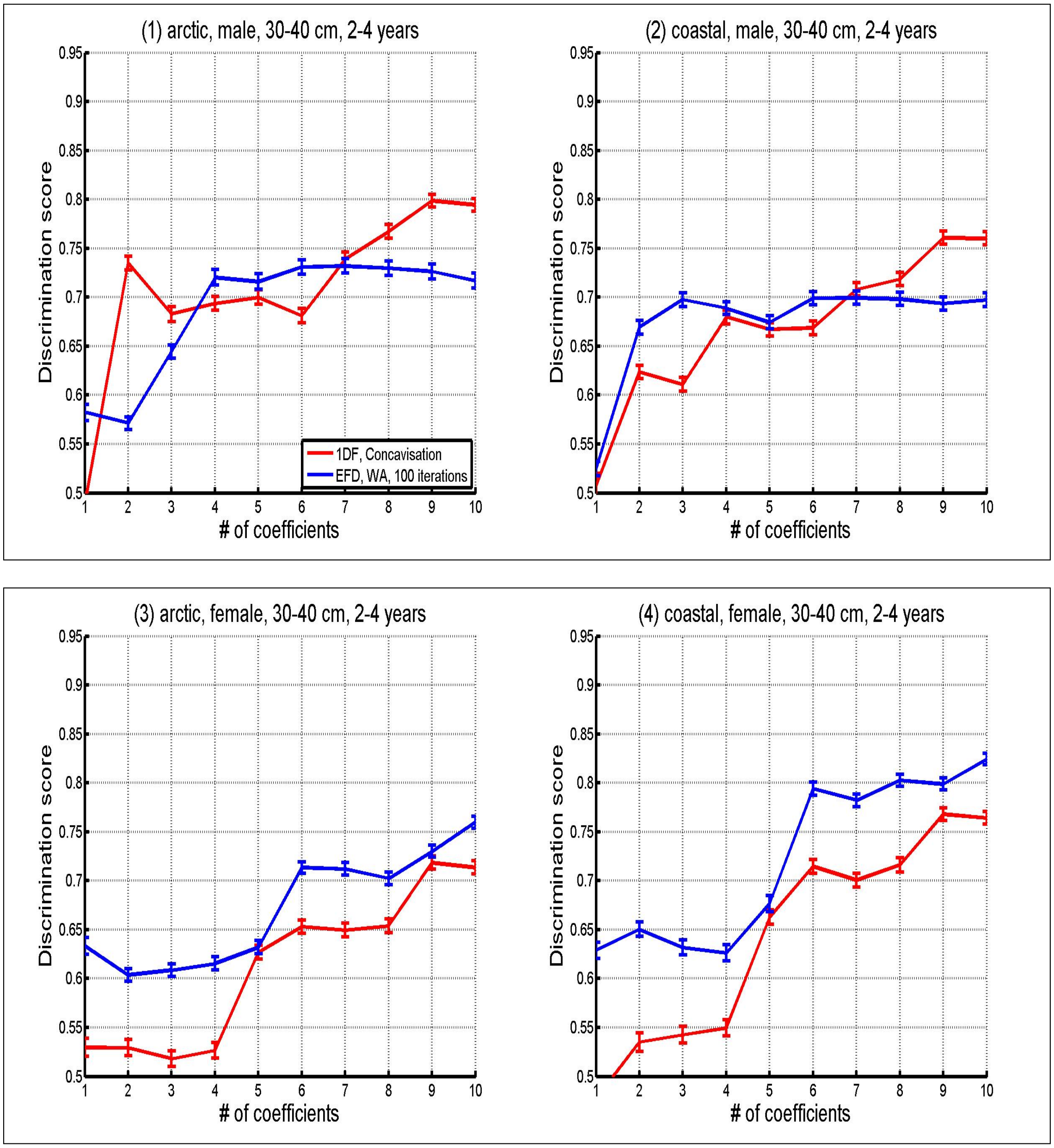

Figure 3.18: Discrimination scores for 10 different number of coefficients. 1DF extracted coefficients from concave contours and EFD extracted coefficients from 100 iteration WMA smoothed contours. Non rotated contours in both cases. Arctic and coastal groups between 30-40cm and 2-4 years.

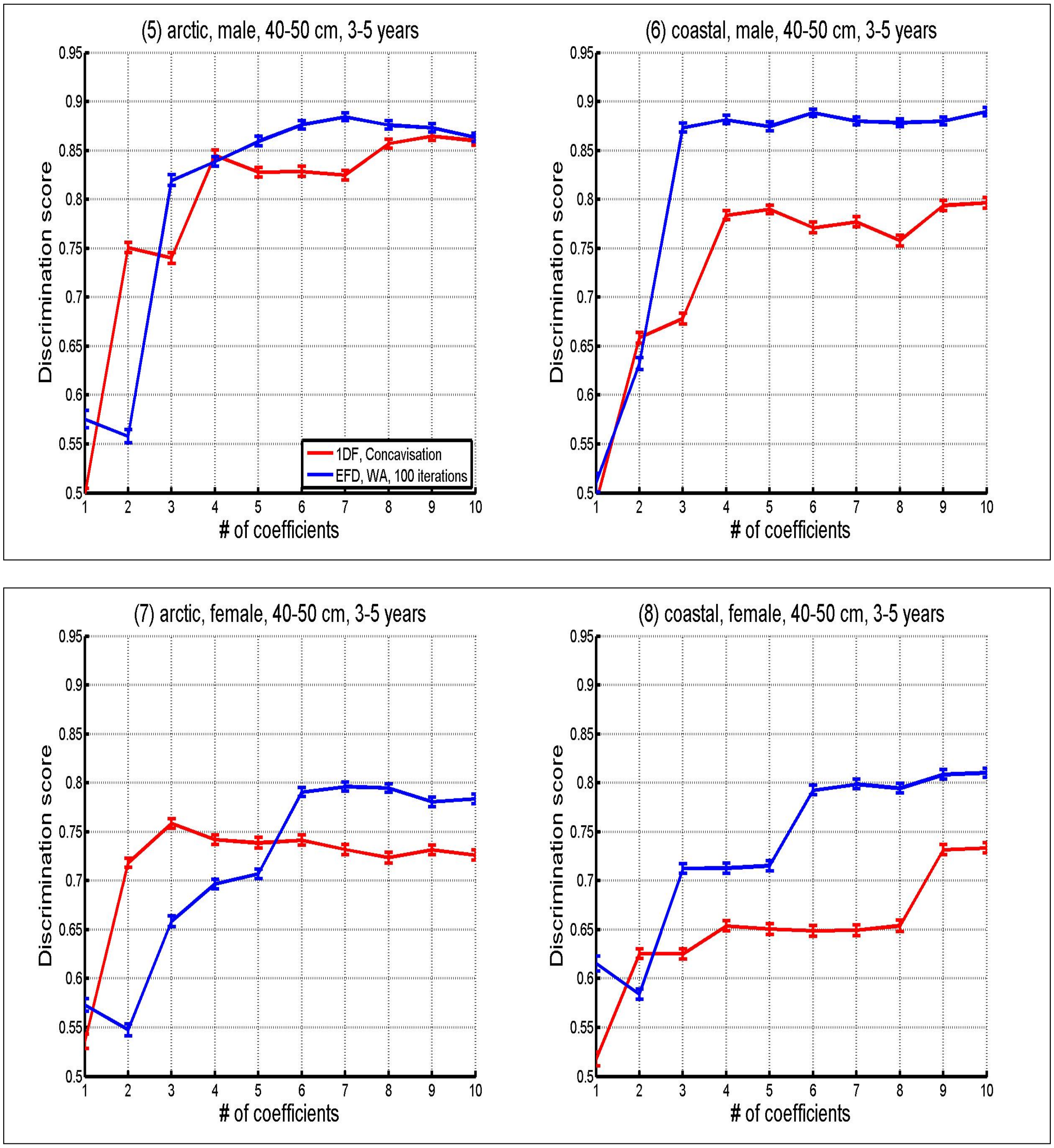

Figure 3.19: Discrimination scores for 10 different number of coefficients. 1DF extracted coefficients from concave contours and EFD extracted coefficients from 100 iteration WMA smoothed contours. Non rotated contours in both cases. Arctic and coastal groups between 40-50cm and 3-5 years.

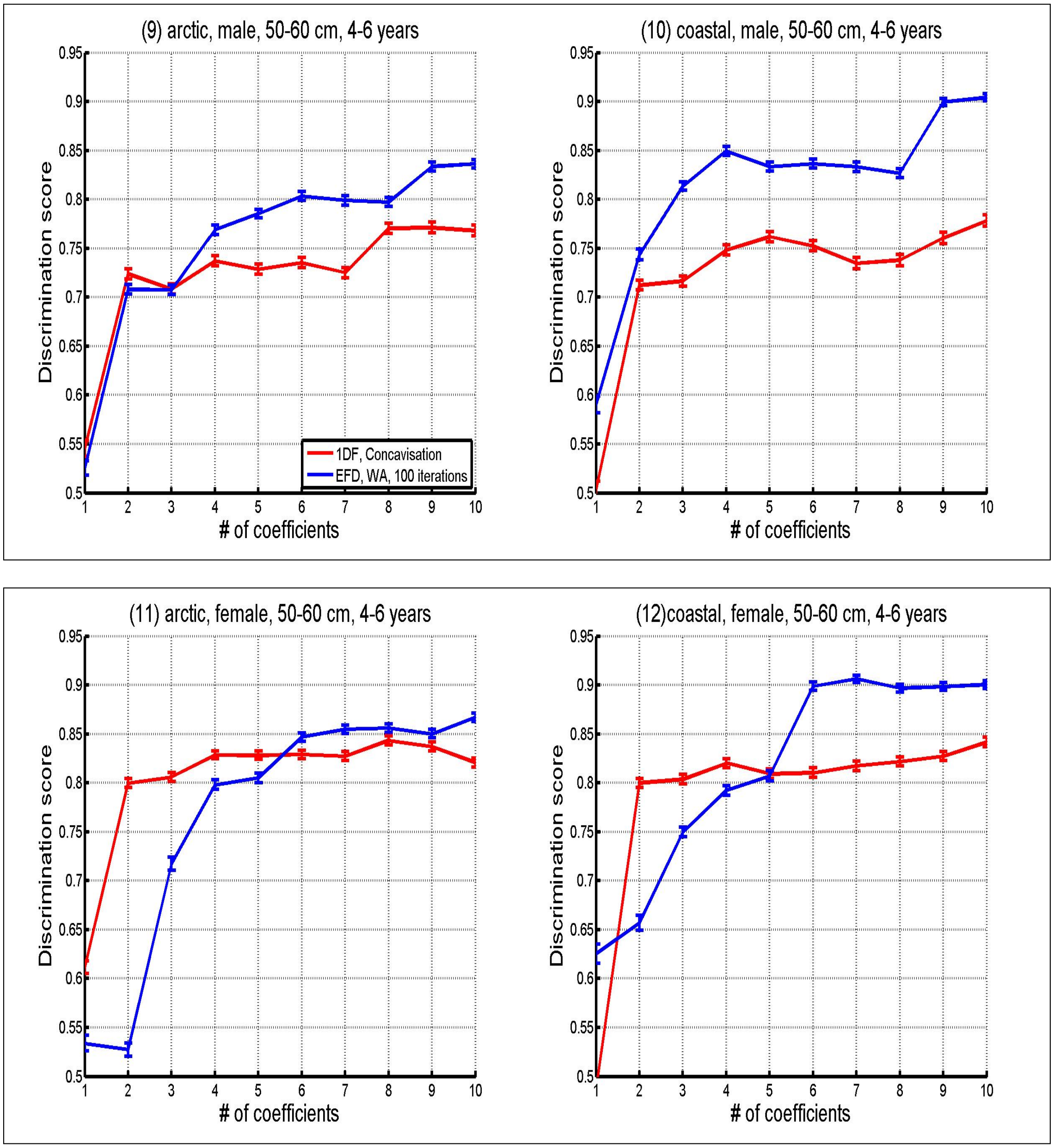

Figure 3.20: Discrimination scores for 10 different number of coefficients. 1DF extracted coefficients from concave contours and EFD extracted coefficients from 100 iteration WMA smoothed contours. Non rotated contours in both cases. Arctic and coastal groups between 50-60cm and 4-6 years.

Discrimination Scores:

Coefficients:

Groups:		2	3	4	5	6	7	8	9	10	Max	Coeff max:
(1)	1DF	73,50 %	68,29 %	69,37 %	70,00 %	68,11 %	73,88 %	76,73 %	79,87 %	79,44 %	79,87 %	9
	EFD	57,12 %	64,44 %	72,01 %	71,51 %	73,09 %	73,19 %	72,96 %	72,63 %	71,69 %	73,19 %	7
(2)	1DF	62,37 %	61,08 %	67,98 %	66,71 %	66,86 %	70,81 %	71,54 %	76,07 %	76,02 %	76,07 %	9
	EFD	66,94 %	69,76 %	68,88 %	67,42 %	69,91 %	69,95 %	69,81 %	69,34 %	69,73 %	69,95 %	7
(3)	1DF	52,94 %	51,79 %	52,67 %	62,68 %	65,30 %	64,96 %	65,37 %	71,85 %	71,35 %	71,85 %	9
	EFD	60,37 %	60,86 %	61,55 %	63,19 %	71,34 %	71,19 %	70,23 %	73,00 %	75,99 %	75,99 %	10
(4)	1DF	53,50 %	54,26 %	54,96 %	66,27 %	71,47 %	70,07 %	71,61 %	76,80 %	76,42 %	76,80 %	9
	EFD	65,04 %	63,20 %	62,62 %	67,68 %	79,39 %	78,21 %	80,27 %	79,87 %	82,43 %	82,43 %	10
(5)	1DF	75,07 %	74,00 %	84,53 %	82,77 %	82,85 %	82,48 %	85,70 %	86,46 %	86,00 %	86,46 %	9
	EFD	55,79 %	81,96 %	83,87 %	85,95 %	87,62 %	88,44 %	87,61 %	87,31 %	86,34 %	88,44 %	7
(6)	1DF	65,85 %	67,83 %	78,38 %	78,98 %	77,13 %	77,71 %	75,79 %	79,38 %	79,66 %	79,66 %	10
	EFD	63,19 %	87,33 %	88,14 %	87,47 %	88,85 %	88,00 %	87,83 %	87,98 %	88,96 %	88,96 %	10
(7)	1DF	71,81 %	75,84 %	74,20 %	73,88 %	74,14 %	73,20 %	72,35 %	73,14 %	72,61 %	75,84 %	3
	EFD	54,75 %	65,85 %	69,65 %	70,70 %	79,05 %	79,64 %	79,46 %	78,06 %	78,37 %	79,64 %	7
(8)	1DF	62,52 %	65,52 %	68,39 %	65,06 %	64,88 %	64,93 %	65,40 %	73,17 %	73,36 %	73,36 %	10
	EFD	58,39 %	71,25 %	71,29 %	71,54 %	79,25 %	79,88 %	79,43 %	80,85 %	81,02 %	81,02 %	10
(9)	1DF	72,40 %	70,82 %	73,72 %	72,85 %	73,53 %	72,52 %	77,06 %	77,13 %	76,81 %	77,13 %	9
	EFD	70,81 %	70,77 %	76,89 %	78,54 %	80,33 %	79,89 %	79,73 %	83,35 %	83,65 %	83,65 %	10
(10)	1DF	71,23 %	71,66 %	74,83 %	76,18 %	75,26 %	73,46 %	73,79 %	76,06 %	77,85 %	77,85 %	10
	EFD	74,35 %	81,36 %	84,95 %	83,34 %	83,64 %	83,33 %	82,68 %	89,96 %	90,42 %	90,42 %	10
(11)	1DF	80,00 %	80,60 %	82,85 %	82,82 %	82,89 %	82,73 %	84,34 %	83,70 %	82,07 %	84,34 %	8
	EFD	52,74 %	71,73 %	79,82 %	80,51 %	84,67 %	85,47 %	85,62 %	85,01 %	86,70 %	86,70 %	10
(12)	1DF	80,00 %	80,36 %	82,01 %	80,92 %	81,03 %	81,73 %	82,18 %	82,72 %	84,17 %	84,17 %	10
	EFD	65,67 %	74,97 %	79,56 %	80,70 %	89,88 %	90,64 %	89,69 %	89,84 %	90,04 %	90,64 %	7
Means:	1DF	68,43 %	68,50 %	71,99 %	73,26 %	73,62 %	74,04 %	75,16 %	78,03 %	77,98 %	78,62 %	
	EFD	62,10 %	71,96 %	74,94 %	75,71 %	80,59 %	80,65 %	80,44 %	81,43 %	82,11 %	82,59 %	

Table 3.7: All discrimination data from using the 1DF method with concavisation and the EFD method using a 100 iteration WMA-smoothing.

3.4.2 Method Orientation Influence

In this section we are taking a look at how big the difference in discrimination scores is for otoliths scanned with different orientations. We are applying the optimal methods found in Sections 3.2 and 3.3 to the original otolith images and to the rotated otolith images. We are finding the absolute difference between the two orientations with regards to discrimination scores for each group and each number of coefficients. All the results can be seen in Table 3.8. In the lower right of the table, the average difference between the two results for both the EFD and the 1DF method are presented. The differences are on average 1.38% and 1.25%. We find this to be relatively small deviations and not something we need to think about when scanning the otoliths.

Just to clarify, the main point is that the orientation you have when scanning the otoliths doesn't have a huge impact, as long as you scan them all with approximately the same orientation. We have only studied how the score changes when you scan everyone with one orientation against scanning everyone with another orientation. It doesn't tell anything about the effects of scanning all the otoliths with different orientations. More details can be found in the discussion in Chapter 4.

Rotation Difference:

Coefficients:

Groups:		2	3	4	5	6	7	8	9	10	Means:
(1)	1DF	1,42 %	1,58 %	0,27 %	1,18 %	0,06 %	1,75 %	1,41 %	1,44 %	2,04 %	1,24 %
	EFD	0,56 %	0,60 %	0,05 %	0,51 %	0,44 %	0,21 %	0,01 %	1,20 %	0,17 %	0,42 %
(2)	1DF	1,12 %	1,69 %	0,55 %	1,27 %	1,60 %	2,27 %	1,76 %	0,14 %	0,76 %	1,24 %
	EFD	0,07 %	0,34 %	0,59 %	0,02 %	0,44 %	0,91 %	0,41 %	0,39 %	0,52 %	0,41 %
(3)	1DF	2,19 %	2,55 %	2,39 %	0,19 %	0,06 %	0,38 %	0,54 %	0,28 %	0,83 %	1,05 %
	EFD	0,44 %	0,69 %	0,41 %	0,29 %	0,10 %	0,35 %	0,28 %	1,66 %	0,31 %	0,50 %
(4)	1DF	4,12 %	2,77 %	2,10 %	1,24 %	1,57 %	1,72 %	0,07 %	0,15 %	0,16 %	1,54 %
	EFD	0,96 %	0,38 %	1,68 %	0,30 %	1,08 %	0,43 %	1,62 %	0,35 %	0,38 %	0,80 %
(5)	1DF	1,03 %	0,01 %	1,03 %	1,81 %	0,97 %	0,34 %	0,75 %	0,76 %	0,30 %	0,78 %
	EFD	1,67 %	0,02 %	0,85 %	0,42 %	0,30 %	1,20 %	0,62 %	0,47 %	0,06 %	0,62 %
(6)	1DF	2,61 %	0,22 %	0,13 %	0,46 %	0,50 %	0,56 %	0,09 %	0,24 %	0,75 %	0,62 %
	EFD	0,29 %	0,13 %	0,90 %	1,37 %	0,01 %	0,25 %	0,38 %	0,29 %	0,58 %	0,47 %
(7)	1DF	3,09 %	1,34 %	1,55 %	2,51 %	2,41 %	2,76 %	4,01 %	0,75 %	1,01 %	2,16 %
	EFD	1,36 %	1,70 %	0,69 %	2,10 %	0,26 %	0,10 %	0,68 %	2,00 %	0,88 %	1,09 %
(8)	1DF	8,61 %	5,53 %	2,21 %	5,02 %	4,35 %	4,49 %	3,20 %	1,38 %	0,69 %	3,94 %
	EFD	2,99 %	2,64 %	4,08 %	2,74 %	0,91 %	0,91 %	0,76 %	0,61 %	1,00 %	1,85 %
(9)	1DF	0,51 %	0,90 %	0,71 %	0,43 %	0,34 %	0,83 %	1,41 %	0,84 %	0,89 %	0,76 %
	EFD	1,26 %	0,24 %	1,24 %	3,92 %	2,03 %	0,69 %	0,69 %	0,33 %	1,16 %	1,28 %
(10)	1DF	0,97 %	0,48 %	0,08 %	1,68 %	2,16 %	1,64 %	0,52 %	0,68 %	2,49 %	1,19 %
	EFD	6,64 %	3,09 %	2,55 %	1,38 %	0,71 %	0,21 %	0,82 %	3,02 %	4,05 %	2,50 %
(11)	1DF	0,89 %	1,04 %	2,33 %	1,66 %	2,03 %	1,41 %	1,78 %	1,53 %	1,76 %	1,60 %
	EFD	8,83 %	1,05 %	0,61 %	0,42 %	2,92 %	3,56 %	3,20 %	3,14 %	1,58 %	2,81 %
(12)	1DF	0,43 %	0,26 %	0,43 %	0,26 %	0,35 %	0,50 %	0,85 %	0,64 %	0,14 %	0,43 %
	EFD	4,68 %	1,60 %	1,68 %	0,82 %	3,02 %	3,24 %	2,86 %	1,19 %	1,29 %	2,26 %
Means:	1DF	2,25 %	1,53 %	1,15 %	1,48 %	1,37 %	1,55 %	1,37 %	0,74 %	0,98 %	1,38 %
	EFD	2,48 %	1,04 %	1,28 %	1,19 %	1,02 %	1,01 %	1,03 %	1,22 %	1,00 %	1,25 %

Table 3.8: Absolute difference between the discrimination scores using non rotated and rotated otolith images. The EFD method extracts coefficients from 100 iteration, WMA-smoothed contours. The 1DF method extracts coefficients from concave contours.

Chapter 4

Discussion and Concluding Remarks

In this thesis we discovered that for discriminating between coastal- and arctic cod, the EFD-method applied to 100 iteration, WMA-smoothing contours provides the best scores. Compared to the 1DF method, the EFD method gave scores averaging close to 4% better for 10 coefficients and the EFD method gave better results for 10 out of 12 cod groups. When applying the 1DF method, the concave contours gave the best results. The 1DF method was best for only the youngest/shortest male groups. With regards to number of coefficients, using 10 coefficients gave the best results, but we also saw that 6 coefficients were enough to give stable, good results. Looking at otolith orientation, both Fourier methods were robust and there were small differences between scores using differently oriented images.

As we detected different trends and tendencies in the results, we tried to find causes for methods being better, reasons for getting different scores based on different number of coefficients and possibilities as to why different cod groups gave different results. There are some ideas and possible explanations mentioned in Chapter 3, but generally it is hard to explain the results based on otolith shapes or group attributes. The main idea behind the use of Fourier coefficients in a discrimination analysis is that it will enable us to detect differences in the otoliths that are hard to detect with the human eye. There are some trends that may seem logical, like the fact that the youngest/smallest groups gave worse discrimination scores. If the otolith differences between the two stocks are caused by different environments it is natural that the differences will grow with time. If the differences are genetic it may be the same case, however the differences may be masked when the otoliths grow as environmental effects distorts the original otolith shapes. As the discrimination proved to be best for the middle groups, it's hard to say anything exact. We did detect some discrimination differences between the coastal- and arctic groups. Our first guess would be the stocks having different variances, but we tested this hypothesis for equal variances in Section 3.1 and found no significant difference in variance.

With regards to the use of coefficients we decided to do all the analysis for 1 to 10 coefficients. Using 1 coefficient doesn't give any good results so the analysis didn't take those results into account. The reason for us to try different number of coefficients is that we would like to find the number that gives the highest scores. Using as many coefficients as possible may seem like the better choice, as it gives a more detailed description of the otlith contour, however, we may end up with the coefficients being to descriptive. If the differences between the otoliths are mostly found within the main shape of the otolith, introducing more coefficients could mask the original shapes, lowering the differences between the stocks. We did find that overall, using 10, the maximum number of coefficients, did give the best results. However, there were several cases where you got the best results using 7, 8 and 9 coefficients. In many of the plots, the graphs indicated that we might get better results for more than 10 coefficients. The reason not to do the analysis for more than 10 coefficients was that using more than 10 coefficients ended in us getting sample covariance matrices that were close to singular. We still got results, but as some of the entries in the covariance matrices was far below computer accuracy, these results could not be trusted. With the given data set, 10 was the maximum number of coefficients we were able to consistently get results for.

The motivation for the study of these two Fourier methods were based on the results in Reig-Bolaño et al. (2010), where they were able to get significantly better accuracy when reconstructing contours using the 1DF method instead of the EFD method. Their study was based on otoliths from Mullets (*Liza aurata* and *Liza Ramada*). Henriksen (2013) showed that there were no significant differences between the two methods when trying to approximate halibut (*Hippoglossus hippoglossus*) otolith contours. We were hoping that we would be able to get better discrimination between the cod stocks with the use of lesser coefficients, when applying the 1DF method. This was not the case. The information regarding the implementation of the 1DF method in Reig-Bolaño et al. (2010) is sparse, which led us to have to make many decisions regarding different implementation details. In this thesis we have continued the work done in Henriksen (2013) where we did the implementation and the discussion. The method has been tweaked slightly for compatibility reasons. The EFD method used in this thesis is the same implementation we used in Henriksen (2013), and was provided by Alf Harbitz.

The programs used in the analysis are linked to in Appendix A. This does not include all programs used to present results. The implementation of all the different methods used in this thesis has been a major part of the work, still the programs need more work. At the moment the programs are made for our specific analysis. We would like to improve the implementations and create an easy to use program, which can accept a folder of scanned otolith images as input and give out discrimination results. Getting an operational, user-friendly program was not the main concern for this thesis, though, and will have to be done at a later time. The bottleneck with regards to computation time in the programs is the calculation of the inverse of the covariance matrices, which is done for every single bootstrap sample. If we have a predetermined set of allocation rules and are to apply them to a new set of otoliths, the computation time is insignificant. As time isn't a major issue, the need for as few as possible coefficients isn't that high. There aren't any reasons for giving away accuracy to gain computation speed, not for the coefficient numbers we have studied.

For future work there are a lot of things we would like to do. Doing this study on bigger samples, enabling us to use more coefficients in the analysis would be a good place to start. This would give more robust results for different groups and for different numbers of coefficients. This could enable us to read more from the result and be able to understand the underlying reasons for different scores. In this thesis we did one study on how the orientation did effect the scores, however, we just saw the effects when all otoliths were scanned in the same way. It would be interesting to test and see how the different smoothing methods and different Fourier methods performed on randomly oriented otoliths. That is, different rotations for all the otoliths within a sample. This would give some more knowledge about the robustness of the methods.

A major interest for us is to discover how the discrimination results varies based on other sources than the different stocks. We mentioned in Chapter 3 when we described the groups that we ordered them to avoid getting discrimination based on sex, age or length of the cod. There are a lot of other pitfalls in the discrimination analysis that we need to be vary of as well. First off, we could study how orientation of the otoliths may be a cause for discrimination. Imagine using the same groups as in this thesis, but instead of discriminating between coastal- and arctic cod, we discriminate between a coastal group with the original otolith orientations and the same coastal group, but this time with the otolith images rotated. If we would get high discrimination scores when doing this, it would indicate that when scanning the otoliths, the orientations have to be the same. Another idea was to study how image resolution effected the discrimination scores. Would scores be inflated if the coastal- and arctic cod otoliths were scanned with different resolutions. This is something we could test the same way we proposed to test the orientation influence. By discriminating between a group of otoliths based on high resolution images and the same group of otoliths based on low resolution images. The images provided to us in this thesis had all similar resolutions so for us to do this study would require a lot more work on the original cod otolith images. This was not a priority. Other things like background lighting, contrast and general details regarding the image processing could all be tested to see if they have effect on discrimination scores. It would also be interesting to analyze these things to find an optimal method of scanning the cod otoliths. A closer study of how cod age, length, weight and sex effect the discrimination scores would also be interesting. Trying to discriminate between young arctic cod and old arctic cod is just one example of a study that could be done.

We have mentioned that the different otolith shapes of the coastal- and arctic cod may be caused by the environment they live in, or they may be caused by genetics. This is without doubt an interesting object of study, however, reasons as to why we get discrimination isn't that important when the goal is to find the coastal- and arctic cod distribution withing a cod population. If we manage to discriminate between the cod stocks based on them having different motion pattern, different diets or any other environmental effect, or if we discriminate between them because of some genetic reason, it doesn't matter, as long as we get discrimination. Discrimination based on age, length or other physical attributes may be less robust as these variables changes within the populations.

One last note. We have found the EFD method to be best suited. However, with the sparse information regarding the implementation of the 1DF method, we think it is wise to continue studying it, especially in light of the results obtained by Reig-Bolaño et al. (2010). The concavisation method was also introduced in this thesis, and we got good discrimination results while using it. It's definitely a method worth doing more research on. As described in this section, there is plenty of future work to be done in this interesting field.

Bibliography

Benzinou, A., S. Carbini, K. Nasreddine, R. Elleboode, and K. Mahe (2013). Discriminating stock of striped red mullet (*Mullus surmuletus*) in the northwest european seas using three automatic shape classification methods. *Fisheries Research 143*, 153–160.

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

Cañas, L., C. Stransky, J. Schlickeisen, M. P. Sampedro, and A. C. Fariña (2012). Use of the otolith shape analysis in stock identification of anglerfish (*Lophius piscatorius*) in the northeast atlantic. *ICES Journal of Marine Science 69(2)*, 250–256.

Haines, A. J. and J. S. Crampton (2000). Improvements to the method of fourier shape analysis applied in morpometric studies. *Palaeontology 43(4)*, 765–783.

Harbitz, A. (2007). Improved elliptic fourier descriptors in shape analysis of a closed contour. *Internal note, The Institute of Marine Research.*

Henriksen, A. M. (2013). The use of fourier descriptors in otolith shape analysis. *Statistics project, The Arctic University of Norway 2.*

Johnson, R. A. and D. W. Wichern (2007). *Applied Multivariate Statistical Analysis.* Pearson Education.

Kålås, J. A., Å. Viken, and T. Bakken (2006). *Norsk rødliste - Norwegian Red List.* Artsdatabanken, Norway.

Kuhl, F. P. and C. R. Giardina (1982). Elliptic fourier features of a closed contour. *Computer Graphics and Image Processing 18*, 236–258.

Mahalanobis, P. C. (1936). On the general distance in statistics. *Proceedings of the National Institute of Sciences of India 2*, 49–55.

Reig-Bolaño, R., P. Marti-Puig, A. Lombarte, J. A. Soria, and V. Parisi-Baradad (2010). A new otolith image contour descriptor based on partial reflection. *Environmental Biology of Fishes 89*, 579–590.

Stransky, C., H. Baumann, S.-E. Fevolden, A. Harbitz, H. HÃÿie, K. H. Nedraas, A.-B. Salberg, and T. H. Skarstein (2008). Seperation of norwegian coastal cod and northeast arctic cod by outer otolith shape analysis. *Fisheries Research 90*, 26–35.

Appendix A

Matlab Code

Some of the matlab code used in the thesis can be downloaded here:
`https://www.dropbox.com/sh/rx9h8ruftnq77n2/q29mzhqo2j`
Check program descriptions for author names.