



## Contribution to the Symposium: '5th International Otolith Symposium' Original Article

# Pitfalls in stock discrimination by shape analysis of otolith contours

Alf Harbitz\* and Ole Thomas Albert

Institute of Marine Research, PO Box 6404, 9294 Tromsø, Norway

\*Corresponding author: tel: +47 55 90 65 57; fax: +47 77 60 97 01; e-mail: [alf.harbitz@imr.no](mailto:alf.harbitz@imr.no)

Harbitz, A., and Albert, O. T. Pitfalls in stock discrimination by shape analysis of otolith contours. – ICES Journal of Marine Science, 72: 2090–2097.

Received 31 October 2014; revised 5 March 2015; accepted 8 March 2015; advance access publication 2 April 2015.

This paper focuses on artefacts that may corrupt stock discrimination by shape analysis of otolith contours, how one can examine if such artefacts are important, and how they can be avoided. The scope focuses on Fourier transforms of contour points, the linear Fisher discrimination technique, and success rates based on cross validation by the “leave one out at a time” technique. The “zero-score” technique is introduced as a tool to examine the importance of a possible artefact, based on the theoretical result that the probability of correct classification of any otolith from either of two identical groups is zero. If one of the identical groups is exposed to a possible influential factor, e.g. a different smoothing, a high classification rate will reveal that this factor is an important artefact. The concept of a “lasso contour” is introduced that drastically reduces the impact of smoothing and provides a non-concave shape that enables a one-dimensional representation of the contour without ambiguities. Results are illustrated by comparison between Greenland halibut (*Reinhardtius hippoglossoides*) otolith contours from southern Greenland and Northeast Arctic waters. The conclusion is that the probability of correct classification of locality based on the original contours is too optimistic (77–79%), while the scores based on lasso contours are insensitive to smoothing and still optimistically high (68–70%).

**Keywords:** Greenland halibut, lasso contour, stock separation, zero score.

## Introduction

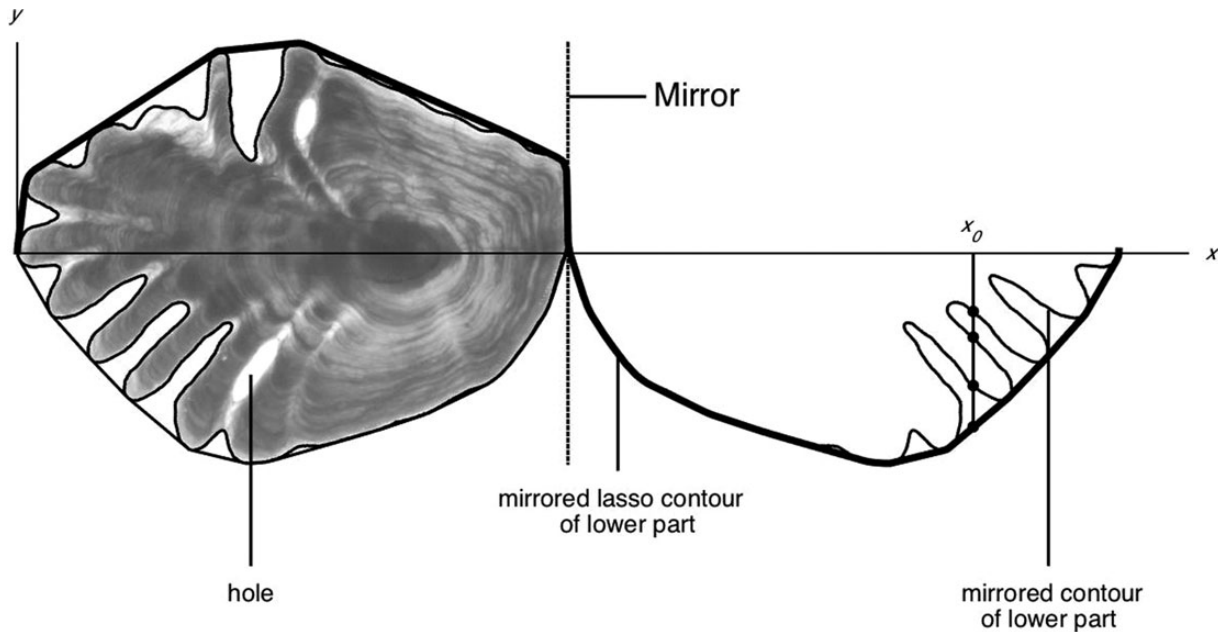
Discrimination of fish stocks by shape analysis of otolith contours has become a powerful and cheap tool with increasing popularity. A range of good results has been reported (e.g. *Micromesistius pou-tassou*, Keating *et al.*, 2014; *Gadus morhua*, Stransky *et al.*, 2007 and Jónsdóttir *et al.*, 2006; *Scomberomorus cavalla*, DeVries *et al.*, 2002), usually by applying classical elliptical Fourier analysis (EFA) introduced by Kuhl and Giardina (1982) and subsequently widely cited. EFA is applied to two-dimensional contour points ( $x$ ,  $y$ ) with straight lines between succeeding points. Briefly speaking, EFA approximates the contour by superimposing ellipses with increasing frequencies and needs four descriptors (coefficients) for each ellipse, two to generate the horizontal  $x$ -coordinates and two for the vertical  $y$ -coordinates.

An alternative and popular Fourier method was introduced by Haines and Crampton (2000), where the Fourier transform is taken of the tangent angle at contour points with the same distance between succeeding points, and the continuous contour is constructed by smooth interpolation between the sampled contour points ( $x$ ,  $y$ ). The contour variable thus is one-dimensional and

enables a one-dimensional fast Fourier transform (FFT) to be applied. Their paper also emphasizes the importance of smoothing and gives a rule for how extensively the sampled contour points should be smoothed before further analysis.

Reig-Bolaño *et al.* (2010) introduced a Fourier method that transforms the two-dimensional contour to a one-dimensional function. First, a best-fitting ellipse is found, and the contour is rotated so that the major axis of the ellipse is parallel with the horizontal  $x$ -axis. Then, the lower part of the (rotated) contour is mirrored around a vertical line at the rightmost point of the contour (Figure 1). Equidistant  $x$ -values are defined, and for each  $x$ -value, the corresponding contour-value  $y$  is found by interpolation between the nearest (sampled) contour points. An FFT is then applied to the  $y$ -values. To easily associate with the mirroring technique, we call this method MIRR (denoted partial reflection by the authors).

A great advantage of the methods where the one-dimensional FFT can be applied is that it is very fast, and there are only two descriptors for each frequency component. In addition, all descriptors are independent of each other, contrary to the EFA descriptors



**Figure 1.** Illustration of MIRR applied to a Greenland halibut otolith. With the original contour, ambiguities in the contour  $y$ -values occur for many  $x$ -values such as  $x_0$ . By the lasso contour, the ambiguities disappear.

(see Haines and Crampton, 2000, for a proof of the latter). Fourier methods, which are dedicated to periodic phenomena, are natural to use in contour analysis, because a closed contour can be considered a perfectly periodic signal with a smooth joint between successive loops when virtually moving around the contour many times. For shapes with very localized landmarks, however, wavelet analysis can be more powerful (Sadighzadeh *et al.*, 2014).

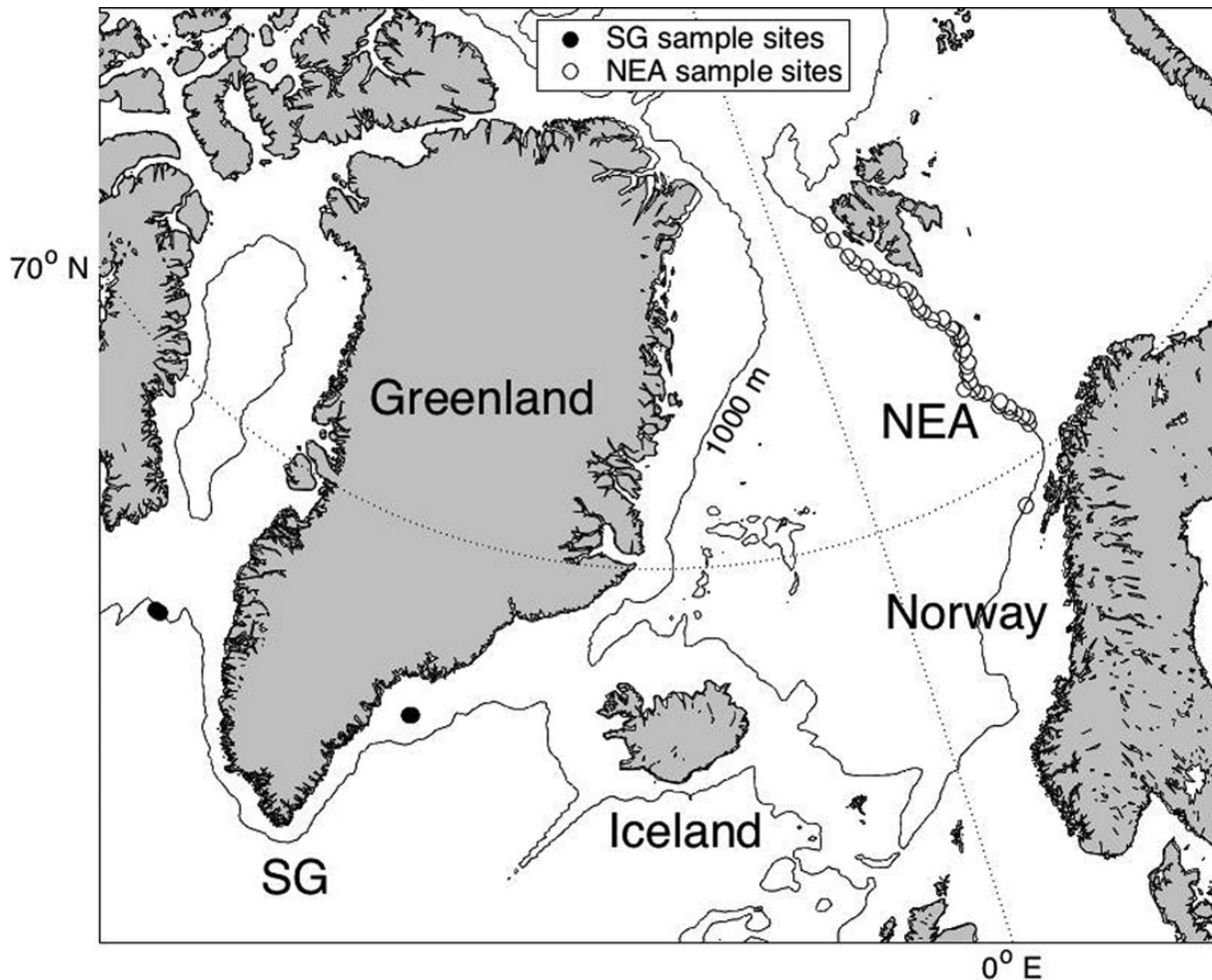
Our multivariate variable to discriminate between fish stocks now consists of a set with Fourier descriptors. These are standardized to avoid discrimination results influenced by differences other than shape between the groups to be compared (see the Material and methods section for further details). There is a range of different multivariate statistical discrimination methods (Johnson and Wichern, 2007) that can be applied, also handling more than two groups. In this paper, we limit the attention to the widely applied Fisher linear discrimination method (Johnson and Wichern, 2007) applied to only two groups to be distinguished with control samples where it is known to which group each individual belongs. In general, shape analysis is widely applied with a “black box” approach where outlines are automatically extracted from images by an image-analysis tool and the following discriminant analysis is performed by a computer program. There has been little focus, however, on the many pitfalls that can lead to biased results, how one can examine the impact of a suspected factor (such as the extent of smoothing), and how one can avoid such pitfalls.

The major objective of this paper is to present a tool for analysing whether a suspected (non-shape) factor can corrupt the discrimination analysis results and to introduce the “lasso contour” as an easy applicable large-scale feature tool with an inherent contour smoothing effect. In addition, the lasso contour has a non-concave shape enabling the application of MIRR without ambiguities (Figure 1). The tool for factor examination is based on a theoretical result that is proven in the Supplementary material. The probability of

correct classification of any otolith from either of two identical samples treated as two different groups is zero by applying linear Fisher discrimination combined with cross validation (leave one out at a time, see Lachenbruch and Mickey, 1968). If one of the two identical samples is manipulated, e.g. by applying a different smoothing, and the result is a large score rate, it is then revealed that smoothing might have a big impact on the discrimination result. If this exercise gives a different score rate for one sample from one group compared with the other sample from the other group, this indicates that the images are taken, and/or the contours extracted, with different settings for the two groups that can result in erroneously large score rates. In this way, a range of possible factors can be examined, e.g. otolith orientation in the image, pixel resolution, and contour extraction procedures including the criterion for thresholding. We define this approach the “zero-score technique”.

The objective of the paper is also to demonstrate how the suggested tool to examine for and mitigate pitfalls works in a real example. For this purpose, the suggested approach is applied to Greenland halibut (*Reinhardtius hippoglossoides*) otoliths from two different areas in the Northeast Atlantic (Figure 2). The use of otolith shape for discrimination purposes has not previously been applied to Greenland halibut. The hand-like shape is rather complex, with considerable individual variation regarding the number of fingers and their length (Figure 3), in contrast to much simpler outlines, such as illustrated in Figure 4 for the twoline eelpout (*Bothrocara brunneum*).

To the naked eye, it is hard to see any systematic difference between the shapes of the otoliths from southern Greenland waters (SG) in the upper panel of Figure 3 and the otoliths from the Northeast Arctic (NEA) in the lower panel. A big difference in otolith brightness is also seen between the two regions, which is due to different imaging setups, and not to opacity differences. This could be a typical situation when comparing images taken at



**Figure 2.** Map of the Northeast Arctic with the SG (83 otoliths) and NEA (828 otoliths) sample sites used in the discrimination analyses.

different labs. The Greenland halibut samples are, therefore, well suited to demonstrate the feasibility of the outlined method to examine possible pitfalls and to provide reliable score estimates useful for investigating stock structure.

The intention of the initial investigation was to compare otoliths from the two Northeast Atlantic Greenland halibut management units, the West Nordic (WN) stock of Southeast Greenland, Iceland, and the Faroe Islands, and the Northeast Arctic (NEA) stock of the Barents Sea and eastern Norwegian Sea. The sample from the West Nordic stock was from southeastern Greenland, and recent genetic (Roy *et al.*, 2014) and tagging (Albert and Vollen, 2015) studies indicate that fish from this area are similar to fish from West Greenland and do not mix with fish from the Northeast Arctic stock. The two locations are also located within separate large current systems with different environmental conditions (ICES, 1969; Blindheim and Malmberg, 2005). It is, therefore, expected that both genetic and environmental effects may create differences in otolith shape of Greenland halibut from the two localities, in similar ways as for some other species (Cardinale *et al.*, 2004).

### Material and methods

Most of the otolith samples from Southeast Greenland waters turned out to be broken and not useful for analyses of the contours.

Therefore, we had to add samples from Southwest Greenland to increase the sample size. The fish from both sides of southern Greenland were caught mostly in September and October 2007 (Figure 2). Of these, 34 unbroken right otoliths from the east coast (WN) and 49 from the west coast were analysed. These were lumped together to obtain a sufficiently large sample, and we applied SG (Southern Greenland waters) as an abbreviation for this combined sample. The sample of right otoliths from the NEA stock was provided from surveys in July and August 2007 (Figure 2). From this sample, 828 unbroken otoliths were included in the analyses. The length frequencies for each sex from the SG sample are shown in Figure 5. The much larger NEA sample contained at least as many fish in each length–sex group as the SG sample, so random samples of 83 otoliths from the NEA sample with the same length–sex distribution as the SG sample could be generated. In this way, differences in length–sex distributions between the two samples will not corrupt the discrimination analysis.

In general, smoothing of the raw, original “pixel” contour is strongly recommended before an analysis of the contour to avoid pixel noise effects. Haines and Crampton (2000) applied a 50% weight on each contour point and a 25% weight on each of the neighbours in each iteration and gave the following recommended minimum number  $n_{it}$  of smoothing iterations:





**Figure 3.** Examples of right sagittae Greenland halibut otoliths from southern SG waters 2007 (upper panel) and from the NEA 2007 (lower panel). Note the individual variation and complex finger-like shape of the contour in contrast to the twoline eelpout otolith contour in Figure 4.

$$n_{it} \geq 2 \left( \frac{N_{\text{samp}}}{N_{\text{FFT}}} \right)^2, \quad (1)$$

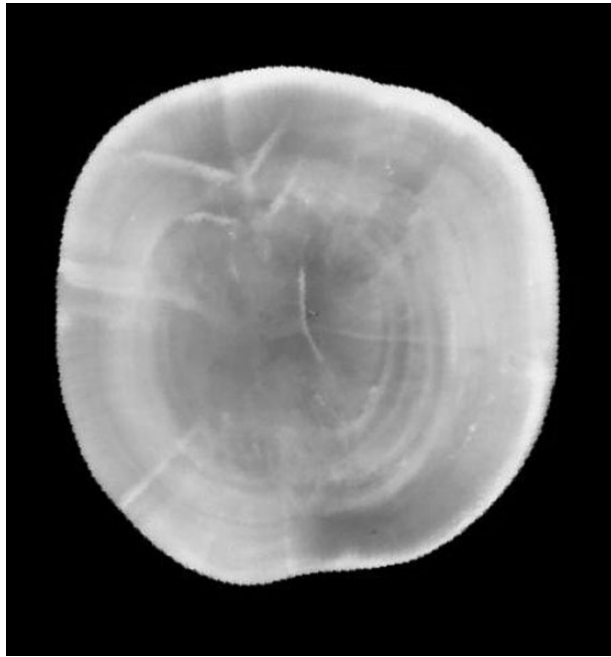
where  $N_{\text{samp}}$  is the number of sample points along the contour, and  $N_{\text{FFT}}$  the number of approximate contour points from the FFT. Too little smoothing may corrupt the discriminant analysis because of pixel noise, while too hard smoothing may change the shape excessively.

To reduce the effect of smoothing and to avoid ambiguities by MIRR, the concept of a lasso contour is introduced, where only the original points that provide a non-concave contour between succeeding points are remained (Figures 1 and 6, left and right panels). Between succeeding lasso contour points where there are skipped contour points in between, equidistant extra points are filled in with distances as close to one pixel as possible between succeeding points. This is not needed in EFA if the lasso contour is not further smoothed, but it is needed for the MIRR technique if non-linear interpolation is applied, and in the case of smoothing in general.

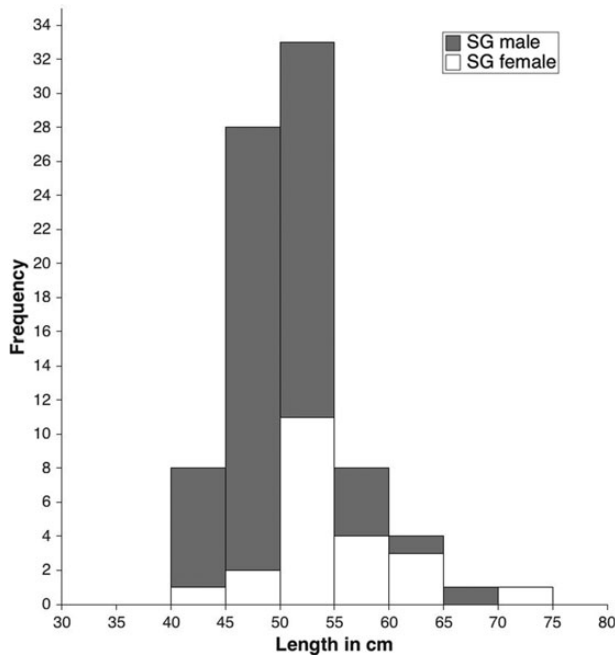
The introduced EFA and MIRR are the two Fourier methods applied in the discrimination analyses. By EFA, the predictors are

standardized with regard to orientation, scale, contour starting point, and tracing direction (clockwise or anticlockwise) (see, e.g. Kuhl and Giardina, 1982). With MIRR, the original contour variables are also standardized by orientation according to the best-fitting ellipse, and the coordinates are then scaled dividing by the square root of the otolith area.

When the standardized Fourier descriptors are established, the Fisher linear classification rule along with cross validation are run to calculate the correct classification score for each of the two regions, i.e. proportion of the SG otoliths correctly classified as SG otoliths, and proportion of the NEA otoliths correctly classified as NEA otoliths. This is done for each of the methods EFA and MIRR and with the number of frequency components equal to 1, 2, ..., 10, 15, and 20. For each set-up, 100 random samples consisting of 83 NEA otoliths with the same length–sex distribution as the SG sample are generated, producing 100 random classification results. A point estimate for the expected classification score (percentage correct classification of each group) is then calculated as the mean of the 100 simulations, and one-sided 95% bootstrap confidence intervals  $[L, 100\%]$  are calculated as well (Efron and Tibshirani, 1993). The interval limit  $L$  is the lower empirical 5th



**Figure 4.** Example of otolith image of twoline eelpout to illustrate a rather simple outline in contrast to the Greenland halibut otoliths in Figure 3.



**Figure 5.** Length–sex distributions for the 83 fish in the SG sample.

percentile of the simulated score percentages and can be interpreted as a conservative, lower estimate of the classification scores that would have been obtained with an infinitely large NEA sample. We denote  $L$  as the 95% significant score.

## Results

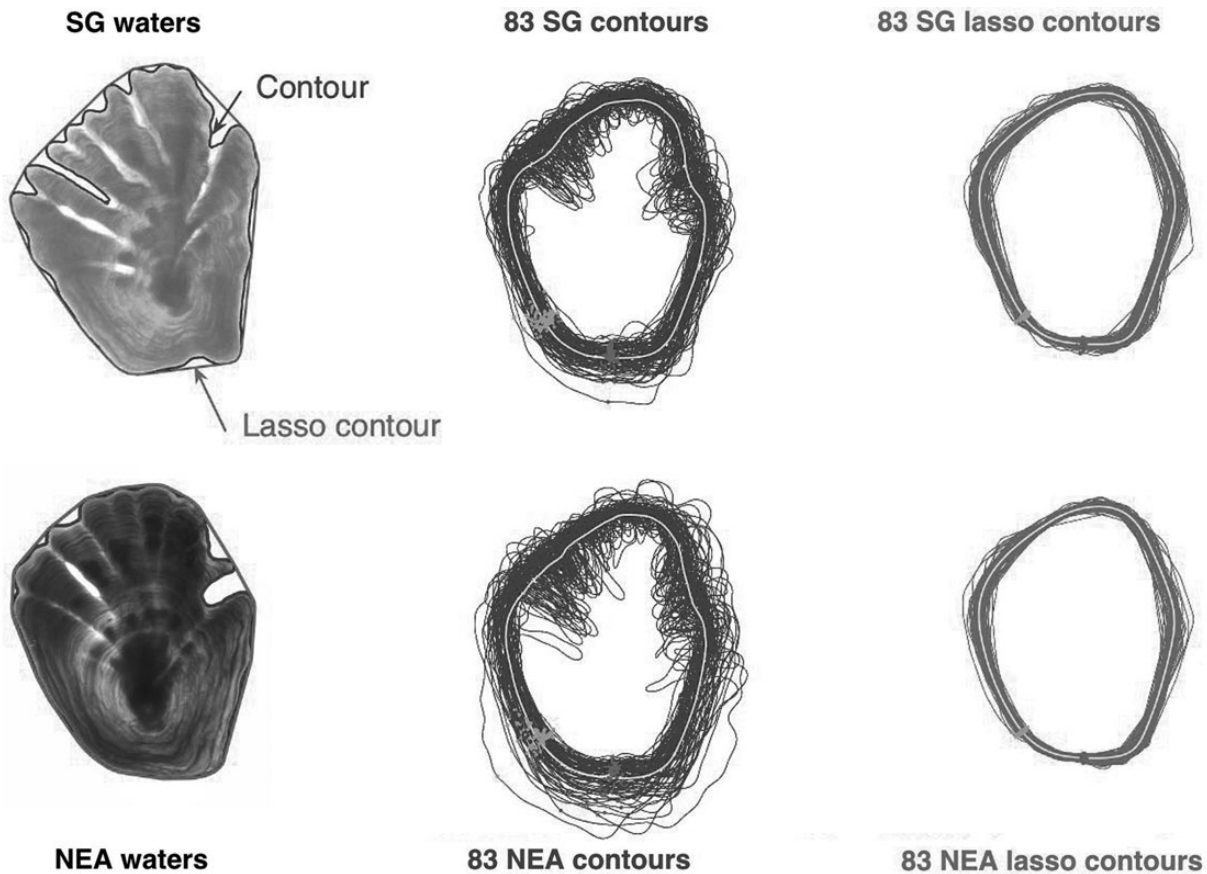
The 83 standardized SG contours created from EFA with 50 frequency components and a corresponding random sample of 83 contours from NEA with the same length–sex distribution as for NEA are shown in Figure 6 in the mid-upper and mid-lower panels, respectively. The mean contours are created from the mean EFA descriptors. Corresponding superimposed lasso contours are shown in the right panels. Note the much less variation between individual lasso contours compared with the variation between original contours.

A range of mean score rate results are shown in Figure 7, each as a function of the number of frequency components. Each result is an average over 100 random samples from the NEA otoliths, each with the same sex-at-length distribution (5-cm length intervals) as for the SG sample. Note that the score rates for the raw contours based on EFA are considerably larger than the other results and are drastically reduced even with moderate smoothing. We also see that MIRR and EFA applied to the lasso contour give about the same maximum scores, and between the EFA results based on raw and smoothed contours. The maximum rates are obtained for the number of frequency components  $k$  close to 5, and Table 1 shows a range of score rates with  $k = 5$ . The figures in parentheses are the 95% significant score values.

The first result in Table 1 (five frequency components) shows that the percentage score rates for SG and NEA based on EFA from the raw, unsmoothed contours are 79.2 (74.2) and 76.7 (70.3), respectively, with the significant 95% value within parentheses. The score rates by EFA applied to the raw lasso contours were considerably lower: 70.4 (66.2) and 68.1 (61.8) for SG and NEA, respectively. As a test example, the raw SG contours were both run as group one and two, and a zero score was obtained in accordance with the theoretical result. When the raw SG contours were compared with the (same) SG contours smoothed twice, the score rates were 68%, indicating a strong effect of smoothing. The same exercise run on non-smoothed and smoothed NEA contours provided score rates of 54 and 62%, respectively, i.e. lower than for the SG contours and indicating a stronger impact of smoothing for the SG contours than for the NEA contours. When comparing EFA applied to the raw NEA lasso contours with the smoothed NEA lasso contours (256 iterations), the low score rates of 10 and 18%, respectively, were obtained, i.e. quite close to 0 and clearly indicating that the smoothing effect is small for the lasso contours. When the smoothed SG and NEA contours were compared (16 iterations), the score rates were reduced rather drastically from 79.2 to 64.2% for SG and from 76.7 to 61.6% for NEA. EFA applied to the smoothed lasso contours (64 iterations) gave score rates of 69.4 and 67.6% for SG and NEA, respectively, very similar to the lasso contour result without smoothing (70.4 and 68.1%). For MIRR applied to the raw lasso contours, the SG and NEA scores were 69.9 and 67.8%, respectively, with a negligible effect of smoothing, and close to the results obtained with EFA applied to the lasso contours.

## Discussion

The paper has applied the zero-score technique to only two groups. An apparent next step would be to test if this result also holds for more than two groups. A very simple approach is to try this experimentally by using the same sample in all groups and see if the correct classification score for any otolith becomes zero in all groups.



**Figure 6.** Greenland halibut otoliths from the SG sample (upper panels) and NEA sample (lower panels). The left panels illustrate the concept of the lasso contour. The mid and right panels show 83 superimposed standardized contours generated from the EFA descriptors (50 frequency components), along with the average outline, based on the raw contours and the raw lasso contours, respectively.

In the Greenland halibut case study, we have seen that by applying the zero-score technique to the original contours, a score rate as high as 68% was obtained with EFD applied to raw and smoothed SG otolith contours, and circa 10% lower scores for the NEA otoliths. To the contrary, convincingly small scores (10 and 18%) were obtained based on raw and smoothed lasso contours, respectively, from NEA. So in these cases, the zero-score technique has clearly demonstrated its usefulness, showing a large and asymmetric effect of smoothing the original contours and a small effect of smoothing the lasso contours.

An apparent interpretation of the Greenland halibut discrimination results is that without smoothing, the score results based on EFA applied to the raw contours are too high. Even with 64 smoothing iterations, it is hard to see a contour difference with the naked eye, unless you zoom in so much that the pixels are clearly distinguished. The reason for the score reduction by smoothing can be a result of the different brightness of the images from SG and NEA. This interpretation is supported by the asymmetric results by the zero-score technique that revealed a larger sensitivity to smoothing for the SG contours than for the NEA contours.

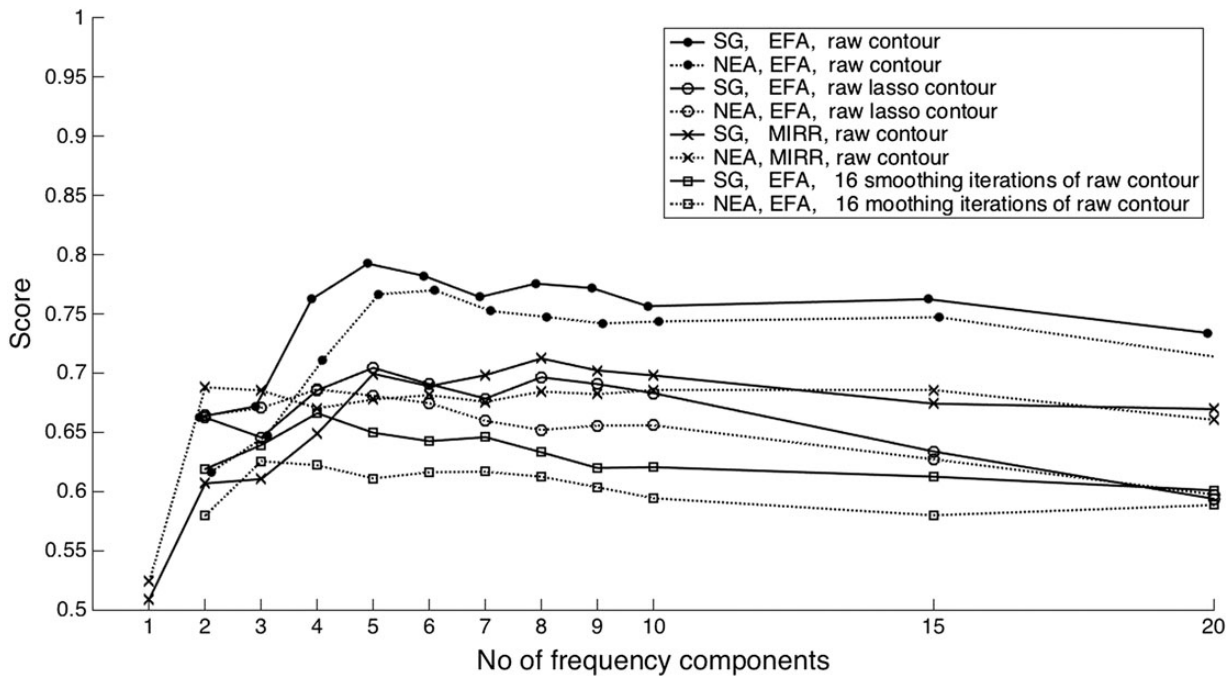
The score results close to 70% by EFA and MIRR applied to the lasso contour are probably the most realistic results so far, and they appear as robust methodological results that still indicate a stock separation potential by shape analysis of the otolith contour in this Greenland halibut case. The results are also in accordance

with recent findings, indicating a stock boundary at Iceland between a Northeast and a Northwest Atlantic gene pool, respectively (Roy *et al.*, 2014; Albert and Vollen, 2015). To uncover the finer structure and management implications, further studies should be conducted with samples from across the assumed boundary region from different seasons and years, and preferably combining genetics with otolith morphology.

The lasso contour has proven to be useful in the Greenland halibut case, but obviously has its limitations when small-scale features are separating two groups. For otoliths where the global large-scale shape is the most important feature, however, it might be a valuable contribution to the toolkit of shape analysis, not least due to the enabling of MIRR without ambiguities. While EFA intuitively is the most appropriate technique for elliptical-like large-scale features, MIRR might be more appropriate for eye-like large-scale shapes with sharp corners at the horizontal “turning” points. The latter shape is, for example, closer to the lasso contour of typical cod (*Gadus morhua*) otoliths than the lasso contour of Greenland halibut otoliths. Henriksen (2013) concludes that the MIRR technique applied to the lasso contour is superior to MIRR applied to the raw contour to discriminate between Arctic and coastal cod from Northeast Arctic waters.

All material in terms of original otolith contour coordinates and corresponding information needed to reproduce the results in the paper, along with the computer codes written in Matlab, are available free from the corresponding author.





**Figure 7.** Score results by MIRR and EFA, i.e. percentage correct classification of SG and NEA individuals, as functions of the number  $k$  of frequency components. Note the rather dramatic fall in classification score from the raw contours to the smoothed contours. The EFA applied to the lasso contours give very similar maximum scores as the MIRR of the lasso contours. The peaks are obtained close to five frequency components.

**Table 1.** Discrimination results with five frequency components.

Group 1	Group 2	Score 1%	Score 2%
SG EFA raw	NEA EFA raw	79.2 (74.2)	76.7 (70.3)
SG EFA raw lasso	NEA EFA raw lasso	70.4 (66.2)	68.1 (61.8)
SG EFA raw	SG EFA raw	0	0
SG EFA raw	SG EFA smooth	68	68
NEA EFA raw	NEA EFA smooth	54	62
NEA EFA raw lasso	NEA EFA smooth lasso	10	18
SG EFA smooth	NEA EFA smooth	64.2 (58.4)	61.6 (55.3)
SG EFA smooth lasso	NEA EFA smooth lasso	69.4 (63.9)	67.6 (61.7)
SG MIRR raw	NEA MIRR raw	69.9 (64.3)	67.8 (61.5)
SG MIRR smooth	NEA MIRR smooth	70.6 (65.1)	68.0 (62.5)

SG, Southern Greenland waters; NEA, Northeast Arctic. Decimal figures based on 100 simulations, integer values based on one replicate. Figures in parentheses are 95% significant scores. See text.

**Supplementary data**

Supplementary material is available at the ICES/JMS online version of the manuscript.

**Acknowledgements**

The Greenland halibut otoliths were provided in cooperation with Møreforsk AS (contact Inge Fossen, [inge@mfaa.no](mailto:inge@mfaa.no)). The original manuscript was improved thanks to the constructive effort by the editor and two referees.

**References**

Albert, O. T., and Vollen, T. 2015. A major nursery area around the Svalbard archipelago provides recruits for the stocks in both Greenland halibut management areas in the Northeast Atlantic. *ICES Journal of Marine Science*, 72: 872–879.

Blindheim, J., and Malmberg, S-Å. 2005. The mean sea level pressure gradient across the Denmark Strait as an indicator of conditions in the North Icelandic Irminger Current. *In The Nordic Seas: an Integrated Perspective*, pp. 65–71. Ed. by H. Drange, T. Dokken, T. Furevik, R. Gerdes, and W. Berger. Geophysical Monograph Series 158, American Geophysical Union, Washington, DC.

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

DeVries, D. A., Grimes, C. B., and Prager, M. H. 2002. Using otolith shape analysis to distinguish eastern Gulf of Mexico and Atlantic Ocean stocks of king mackerel. *Fisheries Research*, 57: 51–62.

Efron, B., and Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, London.

Haines, J., and Crampton, J. S. 2000. Improvements to the method of Fourier shape analysis as applied in morphometric studies. *Palaeontology*, 43: 765–783.

Henriksen, A. M. 2013. Discrimination of cod otolith shapes by two different Fourier methods. Master’s thesis in Industrial Mathematics. STA-3921. University of Tromsø, Norway.

ICES. 1969. Atlas of the hydrography of the northern North Atlantic Ocean. Conseil International pour l’Exploration de la Mer, Service Hydrographique, Charlottenlund Slot, Denmark.

Johnson, R. A., and Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*, 6th edn. Pearson Prentice Hall, Upper Saddle River, NJ. 800 pp.

Jónsdóttir, I. G., Campana, S. E., and Marteinsdóttir, G. 2006. Otolith shape and temporal stability of spawning groups of Icelandic cod (*Gadus morhua* L.). *ICES Journal of Marine Science*, 63: 1501–1512.

Keating, J. P., Brophy, D., Officer, R. A., and Mullins, E. 2014. Otolith shape analysis of blue whiting suggests a complex stock structure at their spawning grounds in the Northeast Atlantic. *Fisheries Research*, 157: 1–6.

- Kuhl, F. P., and Giardina, C. R. 1982. Elliptic features of a closed contour. *Computer Graphics and Image Processing*, 18: 237–258.
- Lachenbruch, P. A., and Mickey, M. R. 1968. Estimation of error rates in discriminant analysis. *Technometrics*, 10: 1–11.
- Reig-Bolaño, R., Martí-Puig, P., Lombarte, A., Soria, J. A., and Parisi-Baradad, V. 2010. A new otolith image contour descriptor based on partial reflection. *Environmental Biology of Fishes*, 89: 579–590.
- Roy, D., Hardie, D. C., Treble, M. A., Reist, J. D., and Ruzzante, D. E. 2014. Evidence supporting panmixia in Greenland halibut (*Reinhardtius hippoglossoides*) in the Northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, 71: 763–774.
- Sadighzadeh, Z., Valinassab, T., Vosugi, G., Motallebi, A. A., Fatemi, M. R., Lombarte, A., and Tuset, V. M. 2014. Use of otolith shape for stock identification of John's snapper, *Lutjanus johnii* (Pisces: Lutjanidae), from the Persian Gulf and the Oman Sea. *Fisheries Research*, 155: 59–63.
- Stransky, C., Baumann, H., Fevolden, S. E., Harbitz, A., Høie, H., Nedreaas, K., Salberg, A. B., *et al.* 2007. Separation of Norwegian coastal cod and Northeast Arctic cod by outer otolith shape analysis. *Fisheries Research*, 90: 26–35.

*Handling editor: Emory Anderson*