Assessing the performance of statistical classifiers to discriminate fish stocks using Fourier analysis of otolith shape

Szymon Smoliński [1,2,*], Franziska Maria Schade[3], Florian Berg[1,4]

[1] Institute of Marine Research, P.O. Box 1870 Nordnes, 5817 Bergen, Norway

[2] Department of Fisheries Resources, National Marine Fisheries Research Institute, Kołłątaja 1, 81-332 Gdynia, Poland

[3] Thuenen Institute of Baltic Sea Fisheries, Alter Hafen Süd 2, 18069 Rostock, Germany

[4] University of Bergen, Department of Biological Sciences, P.O. Box 7803, 5020 Bergen, Norway

[*] E-mail: szymon.smolinski@hi.no

Abstract

The assignment of individual fish to its stock of origin is important for reliable stock assessment and fisheries management. Otolith shape is commonly used as the marker of distinct stocks in discrimination studies. Our literature review showed that the application and comparison of alternative statistical classifiers to discriminate fish stocks based on otolith shape is limited. Therefore, we compared the performance of two traditional and four machine learning classifiers based on Fourier analysis of otolith shape using selected stocks of Atlantic cod (*Gadus morhua*) in the southern Baltic and Atlantic herring (*Clupea harengus*) in the western Norwegian Sea, Skagerrak and the southern Baltic Sea. Our results showed that the stocks can be successfully discriminated based on their otolith shapes. We observed significant differences in the accuracy obtained by the tested classifiers. For both species, support vector machines (SVM) resulted in the highest classification accuracy. These findings suggest that modern machine learning algorithms, like SVM, can help to improve the accuracy of fish stock discrimination systems based on the otolith shape.

Key words: Atlantic cod *Gadus morhua*, Atlantic herring *Clupea harengus*, fish stock discrimination, machine learning, support vector machines

1. Introduction

Discrimination of fish stocks is essential for reliable fisheries resource management and is currently an integral part of modern fish stock assessments (Begg et al. 1999). Many commercially exploited fish stocks show strong habitat overlaps, resulting in a temporal mixing. A disregard of stock mixing, particularly when stocks differ in productivity, may lead to the overexploitation of unique spawning components (Kell et al. 2004; Kerr et al. 2017). Therefore, individuals from mixed-stock catches need to be assigned to their stock of origin using reliable stock discrimination methods with high classification accuracy (Cadrin et al. 2014).

One widely applied stock discrimination technique involves otoliths; calcium carbonate structures located in the inner ear of fishes (Campana and Casselman 1993). Otolith shape is mostly driven by a combination of environmental and genetic factors and contains stock-specific features, which are usable as a relevant marker of distinct stocks (Vieira et al. 2014; Berg et al. 2018). In recent years, diverse methods enabling the description of the otolith shape were developed and tested, such as curvature-based descriptors, wavelets, shape geodesics or mirroring techniques (Parisi-Baradad et al. 2005; Nasreddine et al. 2009; Harbitz and Albert 2015). However, otolith outlines are still most frequently investigated with a mathematical scheme of Fourier decomposition, namely fast Fourier transform or elliptical Fourier analysis (Stransky 2014). Both fast Fourier transform and elliptical Fourier techniques decompose shape, which is a polygon of two-dimensional coordinates, into a spectrum of harmonically related trigonometric curves and calculate coefficients describing each of these curves (for details see Haines and Crampton 2000; Kuhl and Giardina 1982). Calculated coefficients may be then used as predictors for the discrimination of fish stocks in multivariate statistical analysis (Stransky 2014).

However, once shape coefficients are extracted, little attention has been paid to apply and compare performances of alternative statistical systems to assign fish individuals to known groups (stocks or species) based on their otolith shape. Available classifiers arise from different fields, like statistics (e.g., linear discriminant analysis), artificial intelligence and data mining (e.g., decision-trees) or connectionist approaches (e.g., neural networks) (Fernández-Delgado et al. 2014). Most machine learning (ML) algorithms are not yet part of the traditional statistical modeling, hence their application in ecology is still scarce (Olden et al. 2008). However, modern ML algorithms have a high potential to outperform traditional parametric classifiers in solving real-world classification problems (Fernández-Delgado et al. 2014). They are much more flexible than conventional models and are able to handle the non-linear relationships and interacting elements that often characterize biological data (Guisan and Zimmermann 2000). Current computational capabilities and freely available statistical software allow relatively easy implementation of these modern algorithms and they may be valuable in the development of fish stock discrimination routines. The advantages of ML applications have been already considered in other stock discrimination approaches, like in otolith chemistry (e.g., Mercier et al. 2011) or analysis of parasitological markers (e.g., Perdiguero-Alonso et al. 2008). These studies strongly suggest that current ML classifiers are already well suited to assign fish to stocks and that classification abilities are improved compared to traditional discriminant analysis.

Few studies used ML algorithms and Fourier analysis of otolith shape to discriminate fish stocks (e.g., Zhang et al. 2016; Mapp et al. 2017). However, these studies did not compare the ML performance with traditional classifiers like linear discriminant analysis. Only recently Jones and Checkley (2017) compared random forest with discriminant analysis to identify otoliths found in sediment cores and showed that the ML approach outperformed the traditional classifier. However, they applied these algorithms to distinguish between species, i.e. between higher taxonomic groups that naturally show stronger otolith shape differences than between fish stocks. To the best of our knowledge, no comprehensive comparison of traditional and modern ML classifiers to assign individuals to fish stocks has been conducted. Here, we apply six statistical classifiers (two traditional: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and four machine learning classifiers: K-nearest neighbors (KNN), classification and regression trees (CART), random forest (RF) and support vector machines (SVM)) to discriminate stocks of two commercially exploited fish species, where Fourier analysis of otolith shape is required for accurate estimation of mixing ratios for a proper stock assessment: Atlantic cod (*Gadus morhua*) in the southern Baltic and Atlantic herring (*Clupea harengus*) in the northeastern Atlantic.

This paper aims to i) conduct a systematic review of the available scientific literature focusing on statistical classifiers associated with Fourier analysis of otolith shape for discrimination purposes; ii) investigate the otolith shape variability of cod and herring stocks by applying elliptical Fourier analysis; and iii) assess the performance of traditional and recent ML classifiers to assign fish individuals to their group of origin based on their otolith shape.

2. Materials and methods

2.1. Literature review of the use of statistical classifiers

Peer-reviewed literature was searched in the Web of Science Core Collection database using the keywords: “otolith$” and “Fourier”. Only English-language studies on otolith shape that applied Fourier analysis to discriminate fish groups at different biosystematics levels (ecotype, stock, population, species) were chosen for further investigation. Selected literature was reviewed to analyze which statistical classification algorithm was applied to discriminate different fish groups. Different types of algorithms based on the framework of Fisher discriminant analysis (Fisher 1936), including parametric and nonparametric extensions, were aggregated as one group (”discriminant analysis”). The list of 106 publications used in the review process is given in the supplementary materials (Table S1).

2.2. Study species and datasets

2.2.1. Atlantic cod *Gadus morhua*

Atlantic cod is one of the most important commercially exploited fish species across the North Atlantic Ocean, inhabiting also the brackish waters of the Baltic Sea. Here, Baltic cod is managed as two separate stocks: one western stock (ICES subdivisions (SDs) 22-24) and one eastern stock (SDs 24-32, ICES 2019a). The genetically distinct cod stocks coexist in the Arkona Basin (SD 24, Hemmer-Hansen et al. 2018, Weist et al. 2019), resulting in uncertainties in the stock assessment. Since the ICES benchmark in 2015, otoliths of cod from commercial samples from the mixing area are assigned to their respective stock of origin using Elliptic Fourier descriptors and LDA (ICES 2015, 2019b; Hüssy et al. 2016). For this study, we used otolith images of genetically validated Baltic cod samples (N=507, Weist et al. 2019) from the mixing area (SD24, Fig.1) and from adjacent areas (Belt Sea (SD 22), Øresund (SD 23) and Bornholm Basin (SD 25)). The dataset consists of 52% western Baltic cod (WBC) and 48% eastern Baltic cod (EBC) (Table 1). For further details refer to Schade et al. (2019).

2.2.2. Atlantic herring *Clupea harengus*

Atlantic herring is a commercially exploited fish species in the northeastern Atlantic that has been a key species for stock discrimination studies (Geffen 2009). Herring stocks in this region consist of multiple spawning components. In this study, we analyzed otoliths from four distinct spawning components (Table 1): Norwegian spring spawners (NSS, 27% of herring data), coastal Skagerrak spring spawners (CSS, 20%), Greifswald Bay herring (GB, 31%) and central Baltic northern component (CBNC, 22%) (ICES 2018a, 2018b). While NSS is clearly a separate stock, CSS and GB are managed within the stock of western Baltic spring spawners (WBSS), whereas CBNC is part of the central Baltic herring (CBH) stock. To ensure that distinct components were sampled, we only used herring sampled in spawning condition. Further, only herring of age 5-6 were selected to reduce age effects on shape variability (Libungan et al. 2015). Herring were mainly collected during scientific surveys, except for GB and some samples of CSS that were caught by local fishers using gillnets.

2.3. Otolith shape analysis

For cod and herring, shape images of clean and unbroken sagittal otoliths were used. Images of the right otolith were preferred; otherwise, the image of the left otolith was flipped. There are no differences between left and right otoliths for cod (Campana and Casselman 1993; Cardinale et al. 2004) and herring (Libungan et al. 2015). High-resolution images were binarized using the threshold function of the GNU Image Manipulation program (Natterer and Neumann 2008).

For the shape analysis, outlines were automatically obtained from converted images using the *Momocs* package (Bonhomme et al. 2014) in the R environment (R Core Team 2018). Elliptical Fourier analysis proposed by Kuhl and Giardina (1982) was used to quantify otolith outlines. This technique decomposes two-dimensional shape with a sum of harmonics, where each harmonic is described by four coefficients (two for x-axis and two for y-axis coordinates). Precision of approximate reconstruction of shapes increases with the number of harmonics used, but it is recommended to reduce the number of harmonics for multivariate analysis. To define the appropriate number of harmonics, 100 otoliths were randomly sampled from the whole set and the Fourier power (PF_n) spectrum and cumulated Fourier power (PF_c) was calculated with the following formulas:

$$PF_n = \frac{A_n^2 + B_n^2 + C_n^2 + D_n^2}{2}$$

$$PF_c = \sum_{1}^{n} PF_n$$

where A_n, B_n, C_n, D_n are the coefficients of n^{th} harmonic (Lord et al. 2012). The number of harmonics that reaches 99% of cumulated Fourier power of 30 harmonics were chosen to summarize shapes of otoliths (Stransky et al. 2008b; Vieira et al. 2014). The first three coefficients were taken as fixed values (A_1=1; B_1=C_1=0) to normalize otoliths for size, orientation and starting point (Tracey et al. 2006). Mean otolith shapes of different stock components were calculated by invert transformation of Fourier coefficients. Overall variance in the shape of otoliths was assessed with principal component analysis (PCA) integrated with morphospaces (theoretical shapes were reconstructed based on the PCA scores) (Bonhomme et al. 2014).

2.4. Statistical classifiers

Among the six selected algorithms, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) were chosen as one of the most popular classifiers, widely implemented in otolith-based fish stock and species discrimination (e.g., Paul et al. 2013; Zhang et al. 2013). They are applied to predict the affiliation of observations from two or more known classes. Both classifiers use the best combination of several characters that provide the strongest separation of classes by maximizing the standard deviation between obtained groups and minimizing them within groups (Fisher 1936).

K-nearest neighbors (KNN) algorithm is one of the simplest ML classifier that can be applied both to binary and multiclass problems (Hall et al. 2008). In the first step, it selects the nearest neighbors and then determines the class of observation using these selected neighbors. One of the KNN advantages is its higher tolerance of the data structure (Hastie et al. 2009).

Similarly, classification and regression trees (CART), a nonparametric procedure, requires no assumptions about the distribution of the data. These models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition. As a result, the partitioning can be represented graphically as a decision tree (Loh 2011). Random forest (RF) is an ensemble technique, based on a set of CARTs, where a bootstrap approach is implemented to select a random set of observations and variables used to construct each tree in ensemble. Finally, decisions of all trees on object allocation are aggregated and the majority is used in order to provide final class prediction (Breiman 2001). Support vector machines (SVM) was selected among the broad range of ML approaches, because of its ability to deal with high-dimensional datasets and its flexibility in modeling diverse data sources (Ben-Hur et al. 2008). This technique uses kernel functions to project the predictive variables into feature space with more dimensions than the initial space of the input data, allowing the construction of linear models (Cortes and Vapnik 1995).

2.5. Statistical analysis

All predictors (Fourier coefficients) were examined for normality with graphical tools (Zuur et al. 2010). None of the variables showed significant deviation from normal distribution. For each fish species, differences in total fish length between stock components were tested and found to be significant using one-way ANOVA (TukeyHSD, $p<0.001$). To test allometric effects of fish length on shape coefficients, analyses of covariance (ANCOVAs) were conducted. Information on stock components origin was included in the model as fixed factors and fish length as covariate. If the interaction between fixed factor and covariate was significant, the variable was excluded from the dataset, otherwise, shape coefficients with significant fish length effect were standardized using the common slope for all stock components (Zhuang et al. 2014).

Classification and Regression Training package *caret* (Kuhn 2008) for R was used to compare performances of selected classifiers. The package allows for different algorithms to be trained in a consistent environment and to conduct the tuning of the machine learning parameters. All predictor variables were scaled and centered in a preprocessing stage. Optimal hyperparameters of KNN (k), CART (cp), RF ($mtry$) and SVM (σ and C) were defined during preliminary tuning (Fig. S1 and S2). Following Mercier et al. (2011) and Zhang et al. (2016), a 4-fold cross-validation resampling method was used to provide the data for the assessment of the performance of each classifier. This validation method is advised as a reasonably stable and low biased measure of model performance (Hastie et al. 2009), but typically indicates lower accuracy of the evaluated algorithms than most often applied leave-one-out cross-validation. Datasets were randomly split into four equal subsets with preservation of class ratios, where three subsets (75% of observations) were used as training data to classify the remaining subset (25% of observations). Validation was repeated for each of the four splits. Additionally, 100 repetitions of the whole process were conducted using a bootstrap approach with independent resampling (Hastie et al. 2009). Confusion (error) matrices (e.g. Kuhn 2008; Perdiguero-Alonso et al. 2008) were generated and classification accuracy (the percentage of fish correctly assigned to their actual class) was calculated as a measure of classifier quality. In order to assess the influence of the number of Fourier harmonics used for the shape representation on classification accuracy, each cross-validation procedure (400 repetitions) was conducted on datasets produced with between 2 to n harmonics, where n is the number of harmonics that reach 99% of cumulated Fourier power. When number of variables was lower than the specified optimal hyperparameter *mtry* for RF, the default *mtry* was applied, which equals the square root of the number of variables. Moreover, in order to assess the influence of the number of classes on the performance of classifiers, herring dataset was split into two-class subsets and similar cross-validation was run for each pair of spawning components. The algorithms were developed in parallel, using the same training and test sets. Therefore, paired t-tests with adjusted p-values to control false discovery rates (Benjamini and Hochberg 1995) were used to test differences in accuracies of classifiers in relation to the dataset with the *n* number of Fourier harmonics. The importance of Fourier descriptors was calculated with the *varImp* function of the *caret* package and was visualized in decreasing order using mean importance for all models. All of the models were built using following the R packages: LDA and QDA with *MASS* (Brian et al. 2015), KNN with *caret* (Kuhn 2008), CART with *rpart*, RF with *randomForest* (Liaw and Wiener 2002) and SVM based on the radial basis function (RBF) kernel with *kernlab* (Karatzoglou et al. 2015).

3. Results

3.1. Literature review of the use of statistical classifiers

Among 106 selected papers published in the period from 1990 to 2018 that incorporate Fourier analysis as the method for otolith shape description, the framework of Fisher discriminant analysis (DA) was the most popular statistical approach. Studies that applied only DA constituted ~92%, while one study (<1%) used DA and RF in parallel (Jones and Checkley 2017). The remaining ~7% of the publications applied classifiers other than DA to assign samples to their respective class, e.g., support vector machines or K-nearest neighbors classifier (Reig-Bolaño et al. 2010b; Benzinou et al. 2013), boundary-based shape classification (Nasreddine et al. 2009), between-class correspondence analysis (Ponton 2006), or random forest (e.g., Zhang et al. 2016).

Application of more than one classifier in the same analysis was scarce (~8% of papers). Comprehensive comparison of accuracy of nine ML algorithms was done by Mapp et al. (2017), including naive Bayes, Bayesian networks, logistic regression, HyperPipes, C4.5, RF, KNN, SVM, and rotation forest. Jones and Checkley (2017) showed that RF algorithms outperformed DA in terms of accuracy. Torres et al. (2000) presented that QDA was superior to LDA, while Finn et al. (1997) found no differences between LDA and QDA models. SVM performed better than KNN in terms of correct classification rate, but the second classifier resulted in more stable performances across the classes and has been chosen for discrimination of fish based on otolith shape in Benzinou et al. (2013).

3.2. Otolith shape variability

Precision of approximate reconstruction of shapes increased with the number of harmonics used (Fig. 2). For both species, 13 harmonics were needed to achieve 99% of cumulative Fourier power summarizing the otolith shapes. Consequently, the first 13 harmonics were used in further analyses. Due to the significant interaction between stock components and fish total length in the ANCOVA models ($p<0.001$), six and 12 Fourier descriptors were excluded from cod and herring data, respectively. A further 23 (cod) and 29 (herring) descriptors were corrected for the fish length effect using a common slope.

Visual inspection of mean otolith shape identified differences between cod stocks and herring components (Fig. S3). Among cod stocks, WBC had wider otoliths than EBC. Otoliths of NSS and CBNC herring were generally wider than those of CSS and GB herring, which mean otolith shapes were very similar.

For cod, the first two PCA axes explained 72.6% of the overall variance in the shape of otoliths (Fig. 3a). The two cod stocks were mainly separated along the first axis, even though a strong overlap was observed. For herring, 66.3% of the overall variance was explained by the first two axes (Fig. 3b).

3.3. Classification accuracy

The classification accuracy of cod otoliths increased with increasing number of harmonics but stayed relatively constant for six and more harmonics (Fig. 4a). One exception is QDA, where the accuracy slightly decreased with a higher number of harmonics. In comparison, the accuracy continued to increase for herring otoliths with increasing number of harmonics (Fig. 4b).

The accuracy differed significantly between classifiers, except for QDA and KNN for cod otoliths as well as LDA and KNN for herring (Table 2). For both species, SVM resulted in the highest classification accuracy (Fig. 4), even when herring data were sequentially split into two-class subsets (Fig. S4). LDA resulted in slightly but significantly lower accuracy for cod (Fig. 4a, Table 2).

The 4-fold cross-validation using SVM (best classifier) and 13 harmonics (accounting for 99% variance of the otolith shape) resulted in an accuracy of 79.54% for cod and 74.13% for herring (Table 3). For cod, the misclassification rate was equal in both stocks (~10%). For herring, the highest misclassification occurred between GB and CSS herring (~7%). Misclassification among the other herring components was low (<1%).

The relative importance of individual Fourier descriptors was consistent among statistical classifiers for both species (Fig. 5), except for CART. CART and RF both rely on the importance of only a few descriptors (~8 or less), while the other classifiers rely on the importance of a higher number of Fourier descriptors.

4. Discussion

Presented review of the literature showed that the application and comparison of alternative classifiers to discriminate fish groups based on their otolith shape is limited. In this study, stock-specific differences in otolith shapes for cod and herring could be detected, which enables the assignment of individual fish to its respective stock of origin. Moreover, a comparison of different statistical classifiers suggested that ML algorithms, in particular

SVM, can improve the accuracy in stock discrimination approaches using the shape of otoliths.

4.1 Literature review of the use of statistical classifiers

The literature review emphasized that traditional DA was used in most of the studies for the classification of fish groups based on the elliptical Fourier descriptors of otolith shape, while application of alternative classifiers was less common. For example, Zhang et al. (2016) used random forest to discriminate stocks of the Japanese Spanish mackerel (*Scomberomorus niphonius*) based on Fourier descriptors of otolith shapes, but no comparison with other classifiers was reported. Mapp et al. (2017) used nine ML algorithms for fish stock separation of two clupeid species using otolith shapes. However, the study of Mapp et al. (2017) was not focused on the absolute classification accuracy, but on the applicability of morphometric approaches that incorporate size information. No comparison with traditional classifiers, like linear discriminant analysis, was made in Mapp et al. (2017), while Jones and Checkley (2017) showed that RF algorithms were superior to DA during classification of fish individuals into different taxonomic groups based on the morphological descriptors and elemental compositions of otoliths.

Studies comparing more than one statistical classification algorithm indicated that the success of fish classification can be significantly improved by alternative classifiers (Torres et al. 2000). These findings stress the need for the comparison of different classifiers, i.e., different approaches should be explored so that the best method is used in order to achieve the best possible assignment. More accurate assignment of individual fish allows for more robust estimation of the contribution of different fish stocks within the mixing areas (i.e., a mixed stock scenario, Hüssy et al. 2016). Accurate estimates of mixing levels can help to understand how movement and mixing affect stock dynamics and provide the quantitative basis for annual stock assessments and scientific advice (Horbowy 2005; Taylor et al. 2011).

4.2 Otolith shape variability

Our results support the previous studies showing that Baltic cod stocks can be successfully discriminated based on the elliptical Fourier analysis of otolith outlines (Paul et al. 2013; Hüssy et al. 2016). Significant differences in otolith shape were also reported for other stocks and spawning populations of cod, e.g., the northeast Arctic and Norwegian coastal cod (Stransky et al. 2008a), Faroe Plateau cod (Cardinale et al. 2004) or Icelandic cod (Petursdottir et al. 2006). Mean shapes reconstructed on the calculated Fourier descriptors indicated that the otolith outline of WBC and EBC differ in the large-scale shape characteristics (mainly length–width relationship), where otoliths from the western stock are wider and rounder than those from the eastern stock, which is in line with previous observations (Paul et al. 2013; Hüssy et al. 2016). Differences in circularity and rectangularity of otoliths were also reported in other cod stocks (Campana and Casselman 1993; Cardinale et al. 2004).

Similarly, discrimination methods based on the analysis of otolith outlines were applied to separate populations of herring in the Northern Atlantic (e.g., Burke et al. 2008; Libungan et al. 2015). Our study revealed differences in otolith shape between herring components. Most of the differences were based on the relationships between the length and width of the whole otolith. NSS and CBNC have wider otoliths, but the rostrum of NSS herring otoliths is clearly longer. Confusion matrices of the cross-validated models (Table 3) indicated that a relatively large number of individuals from the CSS and GB were mis-assigned, suggesting similarity in otolith shape. This result supports the current assessment approach, where both spawning components are considered as one stock (WBSS) because of the high level of overlap (ICES

2018b). Although selected herring spawning components were discriminated with a high level of accuracy, further studies need to include other stock components in this region, such as the autumn spawners and the southern component of CBH (ICES 2018a).

The differences in the shape of fish otoliths, for both fish species, may be associated with a combination of environmental and genetic drivers (Cardinale et al. 2004; Vignon and Morat 2010). To explore how these factors influence otolith shape, further analyses are needed, including experimental and laboratory studies with appropriate control of the potentially confounding variables (e.g., Berg et al. 2018). However, even without the mechanistic understanding of the sources of shape variability, these results support the applicability of Fourier analysis of otolith shape in stock discrimination routines and assessment of fish stocks (Cadrin et al. 2014). The use of otoliths as indicator of stock identity has been previously advocated because otoliths are routinely collected for aging in traditional fish monitoring, providing a robust and cost-effective method for stock discrimination (Campana and Casselman 1993; Cardinale et al. 2004).

4.3 Assessment of statistical classifiers

There were significant differences in accuracy between the six statistical classifiers tested. The highest accuracy of fish classification was achieved by SVM, one of the rapidly developing ML classifiers. Accuracy of the SVM model trained on cod data was only 0.9% higher than of the second best performing classifier (LDA), but differences were significant. However, the accuracy of the SVM trained on herring data was 7% to 20% higher than the other classifiers. Good performance of the SVM algorithm, as well as other ML algorithms, has been previously shown in discrimination studies of stocks, species or higher taxonomic levels of fishes based on their otolith shapes (Reig-Bolaño et al. 2010a; Benzinou et al. 2013; Zhang et al. 2016; Mapp et al. 2017).

These findings suggest that ML algorithms are a good alternative to traditional classifiers and can help to improve the accuracy of routine fish stock discrimination using the shape of the otolith. Although SVM achieved the highest accuracy in this study, we strongly advise to test a range of statistical classifiers in discrimination studies, because the selection of the best performing algorithm can be case-specific, and depends e.g., on the number of classes, similarity between groups, or type and number of variables in the dataset (Fernández-Delgado et al. 2014).

Caution is however warranted. The proposed benchmark of different statistical classifiers should be conducted only in systems with well-defined units. The ability of ML classifiers to find structures and clusters in the data needs to be considered with caution. Application of the ML algorithms for the discrimination of fish groups, where training baselines are not validated (e.g., by genetics or by sampling spawning individuals in their respective spawning area), may potentially lead to confusing results and recognition of subgroups, which may not represent the real biological or management units. The practical problems of managing natural resources with poorly defined units continue to be an important issue (Geffen 2009). For these reasons, the definition of robust baselines for the training of classification algorithms is a crucial point in the development of operational discrimination systems (Cadrin et al. 2014; Hüssy et al. 2016; Schade et al. 2019).

4.4 Study limitations and future implications

In this study, a simple approach was applied, using only Fourier descriptors of otolith shapes as predictors of fish stock affiliation. The focus was exclusively on the differences of statistical classifier accuracies on the length-normalized descriptors of otolith shape (Hüssy et al. 2016). However, incorporation of other potentially informative variables, such as shape indices or routinely collected information on length-at-age, and sex of individual fish can further improve the predictive abilities of classification algorithms (Burke et al. 2008; Mapp et al. 2017). Further, alternatives to reconstruct the otolith shape like wavelet transformation or curvature scale space representation should be reconsidered. Fourier descriptors focusing on periodic phenomena (Harbitz and Albert 2015) might be more suited for cod otoliths that are almost elliptical. For more complex otolith shapes with very localized landmarks, like herring otoliths, wavelet transformation could be better-suited (Sadighzadeh et al. 2014). Besides otolith shapes, ML algorithms were already used successfully in other stock discrimination fields, e.g., population genetics (Guinand et al. 2002), otolith microchemistry (Mercier et al. 2011), hydroacoustics (Robotham et al. 2010) or parasitology (Perdiguero-Alonso et al. 2008), even though the application is still rare.

In our study, the analysis of Fourier power spectrum indicated that 13 harmonics were needed to explain 99% of the variance in the otolith shape both for cod and herring. Interestingly, high accuracy for the cod assignment was already obtained with only 5 to 6 harmonics, suggesting that additional higher-frequency harmonics do not incorporate much information for the discrimination of these stocks. These results are in line with the analysis of variable importance which showed that lower-rank descriptors (D5, D1 - describing a global form of otoliths) were the most powerful predictors in all models. The broadly applied practice to include only a certain subset of harmonics (e.g., first *N* harmonics needed to describe 99% of shape variance) may not be optimal in the context of classification model performance. For fish species with simple otolith shapes, a reduced number of Fourier harmonics may be advantageous. Conversely, the inclusion of a larger number of harmonics in classification systems developed for species with more complex otolith structures, like herring, can help to achieve a better quality of classification models. In our study, a steady improvement of model accuracy with increasing number of harmonics was observed for SVM and RF, trained on the herring dataset. In the case of increasing dimensionality, the ML algorithms clearly outperform traditional classifiers due to their ability to integrate information from many variables without the high risk of overfitting (Breiman 2001; Ben-Hur et al. 2008). Improvement of the ML models accuracy can also be obtained by the elimination of non-informative variables during the model building (e.g., Smoliński 2019). Furthermore, heterogeneous ensemble techniques combining predictions of different model types could also be applied to improve the classification of fish stocks. Such an approach could help to minimize model-specific errors in class predictions and to obtain a more robust assignment of the fish origin.

The ability of SVM and other ML algorithms to model complex and non-linear patterns without any assumptions is of great importance in many biological applications (Noble 2006). Therefore, the variable transformations are not needed for the application of these algorithms, which make the pre-processing more straightforward and faster. Moreover, variables with non-normal distribution (typically required for the traditional parametric models) do not need to be excluded after an unsuccessful transformation, preventing from the loss of information potentially valuable for the discrimination of fish groups (Mercier et al. 2011).

Future operationalization of developing stock discrimination methods needs profound analyses of the level of temporal variability of within- and between-group differences, particularly in otolith shapes. The presented results are based on the samples collected within a short period of time, limiting the influence of the year-classes and long-term environmental effects on otolith shape. However, if the temporally stable character of fish otolith shapes can be confirmed for particular stocks, it may enable continuous enlargement of databases. In consequence, better performance of ML algorithms can be achieved, because their classification accuracy typically boosts with increasing size of training datasets.

4.5 Conclusions

Our study emphasized the potential for applying novel ML algorithms to improve the accuracy of classification systems based on the otolith shape of fish. We recommend conducting comparisons of different statistical classifiers in systems of well-identified stock structures using validated baselines. When temporal mixing of different fish stocks or stock components occurs, as with Baltic cod and herring in the Northeast Atlantic, possible improvements of stock discrimination processes by modern classifiers may be of great importance. More accurate assignment of fish individuals may help to more precisely estimate the contribution of different fish stocks within the mixing areas and in consequence, provide a more reliable quantitative basis for annual stock assessments and scientific advice.

Acknowledgments

We thank Thomas Naatz (MS “JULE”) for providing cod samples from SD 23 and all staff members involved in sampling during research and monitoring cruises. We are grateful also to the technical staff at the Thuenen Institute of Baltic Sea Fisheries for photographing cod otoliths. We thank Tomas Gröhsler for providing herring otoliths from Greifswald Bay. Institute of Marine Research technicians are thanked for their contribution in collecting and photographing otoliths of the NSS and CSS herring components. We also acknowledge Audrey J. Geffen, Uwe Krumme, Richard D. M. Nash and two anonymous reviewers for the input and comments on this manuscript. FMS was partly funded by the European Maritime and Fisheries Fund (EMFF) of the European Union (EU) under the Data Collection Framework (DCF, Regulation 2017/1004 of the European Parliament and of the Council). FB was funded by the Research Council Norway project 254774 (GENSINC).

References

Begg, G.A., Friedland, K.D., and Pearce, J.B. 1999. Stock identification and its role in stock assessment and fisheries management: An overview. Fish. Res. **43**: 1–8. doi:10.1016/S0165-7836(99)00062-4.

Ben-Hur, A., Ong, C.S., Sonnenburg, S., Schölkopf, B., and Rätsch, G. 2008. Support vector machines and kernels for computational biology. PLoS Comput. Biol. **4**: e1000173. doi:10.1371/journal.pcbi.1000173.

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B **57**(1): 289–300. doi:10.2307/2346101.

Benzinou, A., Carbini, S., Nasreddine, K., Elleboode, R., and Mahé, K. 2013. Discriminating stocks of striped red mullet (*Mullus surmuletus*) in the Northwest European seas using three automatic shape classification methods. Fish. Res. **143**: 153–160. doi:10.1016/j.fishres.2013.01.015.

Berg, F., Almeland, O.W., Skadal, J., Slotte, A., Andersson, L., and Folkvord, A. 2018. Genetic factors have a major effect on growth, number of vertebrae and otolith shape in Atlantic herring (*Clupea harengus*). PLoS One **13**(1): 1–16. doi:10.1371/journal.pone.0190995.

Bonhomme, V., Picq, S., Gaucherel, C., and Claude, J. 2014. Momocs: Outline analysis using R. JSS J. Stat. Softw. **56**(13).

Breiman, L. 2001. Random forests. Mach. Learn. **45**: 5–32. doi:10.1023/A:1010933404324.

Brian, R., Venables, B., Bates, D.M., Firth, D., and Ripley, M.B. 2015. Package MASS.

Burke, N., Brophy, D., and King, P.A. 2008. Shape analysis of otolith annuli in Atlantic herring (*Clupea harengus*); a new method for tracking fish populations. Fish. Res. **91**: 133–143. doi:10.1016/j.fishres.2007.11.013.

Cadrin, S.X., Kerr, L.A., and Mariani, S. 2014. Stock identification methods: Applications in fishery science. *In* 2nd edition. Elsevier. doi:10.1016/B978-0-12-397003-9.01001-8.

Campana, S.E., and Casselman, J.M. 1993. Stock discrimination using otolith shape analysis. Can. J. Fish. Aquat. Sci. **50**(5): 1062–1083. doi:10.1139/f93-123.

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. Can. J. Fish. Aquat. Sci. **61**(2): 158–167. doi:10.1139/F03-151.

Cortes, C., and Vapnik, V. 1995. Support-vector networks. Mach. Learn. **20**(3): 273–297. doi:10.1007/BF00994018.

Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. 2014. Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. **15**: 3133–3181.

Finn, J.E., Burger, C. V., and Holland-Bartels, L. 1997. Discrimination among populations of sockeye salmon fry with Fourier analysis of otolith banding patterns formed during incubation. Trans. Am. Fish. Soc. **126**: 559–578.

Fisher, R. 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. **7**(2): 179–188.

Geffen, A. 2009. Advances in herring biology: from simple to complex, coping with plasticity and adaptability. ICES J. Mar. Sci. **66**(8): 1688–1695.

Guinand, B., Topchy, A., Page, K., Burnham-Curtis, M., Punch, W., and Scribner, K. 2002. Comparisons of likelihood and machine learning methods of individual classification. J. Hered. **93**(4): 260–269. doi:10.1093/jhered/93.4.260.

Guisan, A., and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Modell. **135**: 147–186. doi:10.1016/S0304-3800(00)00354-9.

Haines, J., and Crampton, J.S. 2000. Improvements to the method of Fourier shape analysis as applied in morphometric studies. Paleontology **43**(4): 765–783.

Hall, P., Park, B.U., and Samworth, R.J. 2008. Choice of neighbor order in nearest-neighbor classification. Ann. Stat. **36**(5): 2135–2152. doi:10.1214/07-AOS537.

Harbitz, A., and Albert, O.T. 2015. Pitfalls in stock discrimination by shape analysis of otolith contours. ICES J. Mar. Sci. **72**(7): 2090–2097. doi:10.1093/icesjms/fsv048.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The elements of statistical learning - Data mining, inference. Springer Verlag. doi:10.1007/b94608.

Hemmer-Hansen, J., Hüssy, K., Baktoft, H., Huwer, B., Bekkevold, D., Haslob, H., Herrmann, J., Hinrichsen, H., Krumme, U., Mosegaard, H., Nielsen, E.E., Reusch, T.B.H., Storr-Paulsen, M., Velasco, A., von Dewitz, B., Dierking, J., and Eero, M. 2018. Genetic analyses reveal complex dynamics within a marine fish management area. Evol. Appl. **12**: 830–844. doi:10.1111/eva.12760.

Horbowy, J. 2005. Assessing Baltic herring stocks with a model that incorporates migration. Fish. Res. **76**: 266–276. doi:10.1016/j.fishres.2005.06.013.

Hüssy, K., Mosegaard, H., Albertsen, C.M., Nielsen, E.E., Hemmer-Hansen, J., and Eero, M. 2016. Evaluation of otolith shape as a tool for stock discrimination in marine fishes using Baltic Sea cod as a case study. Fish. Res. **174**: 210–218. doi:10.1016/j.fishres.2015.10.010.

ICES. 2015. Report of the benchmark workshop on Baltic cod stocks (WKBALTCOD). 2-6 March 2015, Rostock, Germany.

ICES. 2018a. Report of the workshop on mixing of western and central Baltic herring stocks (WKMixHER). 11-13 September, Gdynia, Poland.

ICES. 2018b. Report of the herring assessment working group for the area south of 62^{o}N (HAWG), 29-31 January 2018 and 12-20 March 2018, ICES HQ, Copenhagen, Denmark.

ICES. 2019a. Baltic fisheries assessment working group (WGBFAS). ICES Scientific Reports 1:20. 651 pp.

ICES. 2019b. Benchmark workshop on Baltic cod stocks (WKBALTCOD2). ICES Scientific Reports. 1:9. 310 pp. doi:http://doi.org/10.17895/ices.pub.4984.

ICES. 2019c. ICES Spatial Facility. Available from http:/gis.ices.dk [accessed 29 May 2019].

Jones, W.A., and Checkley, D.M. 2017. Classification of otoliths of fishes common in the Santa Barbara Basin based on morphology and chemical composition. Can. J. Fish. Aquat. Sci. **74**(8): 1195–1207. doi:10.1139/cjfas-2015-0566.

Karatzoglou, A., Smola, A., and Hornik, K. 2015. Package ‘ kernlab .’

Kell, L.T., Crozier, W.W., and Legault, C.M. 2004. Mixed and multi-stock fisheries: Introduction. ICES J. Mar. Sci. **61**(8): 1330. doi:10.1016/j.icesjms.2004.09.002.

Kerr, L.A., Hintzen, N.T., Cadrin, S.X., Clausen, L.W., Dickey-Collas, M., Goethel, D.R., Hatfield, E.M.C., Kritzer, J.P., and Nash, R.D.M. 2017. Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. ICES J. Mar. Sci. **74**(6): 1708–1722. doi:10.1093/icesjms/fsw188.

Kuhl, F.P., and Giardina, C.R. 1982. Elliptic Fourier features of closed contour. Comput. Graph. Image Process. **18**: 236–258.

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. **28**(5): 1–26. doi:10.1053/j.sodo.2009.03.002.

Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. R news **2**: 18–22. doi:10.1177/154405910408300516.

Libungan, L.A., Óskarsson, G.J., Slotte, A., Jacobsen, J.A., and Pálsson, S. 2015. Otolith shape: a population marker for Atlantic herring *Clupea harengus*. J. Fish Biol. **86**: 1377–1395. doi:10.1111/jfb.12647.

Loh, W.Y. 2011. Classification and regression trees. Wiley Iterdisciplinary Rev. Data Min. Knowl. Discov. **1**: 14–23. doi:10.1201/9781315139470.

Lord, C., Morat, F., Lecomte-Finiger, R., and Keith, P. 2012. Otolith shape analysis for three

Sicyopterus (Teleostei: Gobioidei: Sicydiinae) species from New Caledonia and Vanuatu. Environ. Biol. Fishes **93**: 209–222. doi:10.1007/s10641-011-9907-y.

Mapp, J., Hunter, E., Van Der Kooij, J., Songer, S., and Fisher, M. 2017. Otolith shape and size: The importance of age when determining indices for fish-stock separation. Fish. Res. **190**: 43–52. doi:10.1016/j.fishres.2017.01.017.

Mercier, L., Darnaude, A.M., Bruguier, O., Vasconcelos, R.P., Cabral, H.N., Costa, M.J., Lara, M., Jones, D.L., and Mouillot, D. 2011. Selecting statistical models and variable combinations for optimal classification using otolith microchemistry. Ecol. Appl. **21**(4): 1352–1364. doi:10.1890/09-1887.1.

Nasreddine, K., Benzinou, A., and Fablet, R. 2009. Shape geodesics for the classification of calcified structures: Beyond Fourier shape descriptors. Fish. Res. **98**: 8–15. doi:10.1016/j.fishres.2009.03.008.

Natterer, M., and Neumann, S. 2008. GNU image manipulation program. The GIMP Team.

Noble, W.S. 2006. What is a support vector machine? Nat. Biotechnol. **24**(12): 1565–1567. doi:10.1038/nbt1206-1565.

Olden, J.D., Lawler, J.J., and Poff, N.L. 2008. Machine learning methods without tears: a primer for ecologists. Q. Rev. Biol. **83**(2): 171–193. doi:10.1086/587826.

Parisi-Baradad, V., Lombarte, A., Garcia-Ladona, E., Cabestany, J., Piera, J., and Chic, O. 2005. Otolith shape contour analysis using affine transformation invariant wavelet transforms and curvature scale space representation. Mar. Freshw. Res. **56**: 795–804. doi:10.1071/MF04162.

Paul, K., Oeberst, R., and Hammer, C. 2013. Evaluation of otolith shape analysis as a tool for discriminating adults of Baltic cod stocks. J. Appl. Ichthyol. **29**: 743–750. doi:10.1111/jai.12145.

Perdiguero-Alonso, D., Montero, F.E., Kostadinova, A., Raga, J.A., and Barrett, J. 2008.

Random forests, a novel approach for discrimination of fish populations using parasites as biological tags. Int. J. Parasitol. **38**(12): 1425–1434. doi:10.1016/j.ijpara.2008.04.007.

Petursdottir, G., Begg, G.A., and Marteinsdottir, G. 2006. Discrimination between Icelandic cod (*Gadus morhua* L.) populations from adjacent spawning areas based on otolith growth and shape. Fish. Res. **80**: 182–189. doi:10.1016/j.fishres.2006.05.002.

Ponton, D. 2006. Is geometric morphometrics efficient for comparing otolith shape of different fish species? J. Morphol. **267**: 750–757. doi:10.1002/jmor.10439.

R Core Team. 2018. R: A language and environment for statistical computing. Vienna, Austria.

Reig-Bolaño, R., Marti-Puig, P., Lombarte, A., Soria, J.A., and Parisi-Baradad, V. 2010a. A new otolith image contour descriptor based on partial reflection. Environ. Biol. Fishes **89**: 579–590. doi:10.1007/s10641-010-9700-3.

Reig-Bolaño, R., Marti-Puig, P., Rodriguez, S., Bajo, J., Parisi-Baradad, V., and Lombarte, A. 2010b. Otoliths identifiers using image contours EFD. Adv. Intell. Soft Comput. doi:10.1007/978-3-642-14883-5_2.

Robotham, H., Bosch, P., Gutiérrez-Estrada, J.C., Castillo, J., and Pulido-Calvo, I. 2010. Acoustic identification of small pelagic fish species in Chile using support vector machines and neural networks. Fish. Res. **102**: 115–122. doi:10.1016/j.fishres.2009.10.015.

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. Fish. Res. **155**: 59–63. doi:10.1016/j.fishres.2014.02.024.

Schade, F., Weist, P., and Krumme, U. 2019. Evaluation of four stock discrimination methods to assign individuals from mixed-stock fisheries using genetically validated baseline samples. Mar. Ecol. Prog. Ser.: Accepted.

Smoliński, S. 2019. Incorporation of optimal environmental signals in the prediction of fish recruitment using random forest algorithms. Can. J. Fish. Aquat. Sci. **76**: 15–27. doi:https://doi.org/10.1139/cjfas-2017-0554.

Stransky, C. 2014. Morphometric outlines. *In* Stock identification methods: Applications in fishery science, 2nd edition. Elsevier Academic Press. pp. 129–140.

Stransky, C., Baumann, H., Fevolden, S.E., Harbitz, A., Høie, H., Nedreaas, K.H., Salberg, A.B., and Skarstein, T.H. 2008a. Separation of Norwegian coastal cod and Northeast Arctic cod by outer otolith shape analysis. Fish. Res. **90**: 26–35. doi:10.1016/j.fishres.2007.09.009.

Stransky, C., Murta, A.G., Schlickeisen, J., and Zimmermann, C. 2008b. Otolith shape analysis as a tool for stock separation of horse mackerel (*Trachurus trachurus*) in the Northeast Atlantic and Mediterranean. Fish. Res. **89**(2): 159–166. doi:10.1016/j.fishres.2007.09.017.

Taylor, N.G., McAllister, M.K., Lawson, G.L., Carruthers, T., and Block, B.A. 2011. Atlantic bluefin tuna: A novel multistock spatial model for assessing population biomass. PLoS One **6**(12): e27693. doi:10.1371/journal.pone.0027693.

Torres, G.J., Lombarte, A., and Morales-Nin, B. 2000. Sagittal otolith size and shape variability to identify geographical intraspecific differences in three species of the genus *Merluccius*. J. Mar. Biol. Assoc. UK **80**(02): 333–342. doi:10.1017/S0025315499001915.

Tracey, S.R., Lyle, J.M., and Duhamel, G. 2006. Application of elliptical Fourier analysis of otolith form as a tool for stock identification. Fish. Res. **77**: 138–147. doi:10.1016/j.fishres.2005.10.013.

Vieira, A.R., Neves, A., Sequeira, V., Paiva, R.B., and Gordo, L.S. 2014. Otolith shape analysis as a tool for stock discrimination of forkbeard (*Phycis phycis*) in the Northeast Atlantic. Hydrobiologia **728**: 103–110. doi:10.1007/s10750-014-1809-5.

Vignon, M., and Morat, F. 2010. Environmental and genetic determinant of otolith shape revealed by a non-indigenous tropical fish. Mar. Ecol. Prog. Ser. **411**: 231–241. doi:10.3354/meps08651.

Weist, P., Schade, F.M., Damerau, M., Barth, J.M.I., Dierking, J., André, C., Petereit, C., Reusch, T., Jentoft, S., Hanel, R., and Krumme, U. 2019. Assessing SNP-markers to study population mixing and ecological adaptation in Baltic cod. PLoS One **14**(6): e0218127. doi:10.1371/journal.pone.0218127.

Zhang, C., Ye, Z., Li, Z., Wan, R., Ren, Y., and Dou, S. 2016. Population structure of Japanese Spanish mackerel *Scomberomorus niphonius* in the Bohai Sea, the Yellow Sea and the East China Sea: evidence from random forests based on otolith features. Fish. Sci. **82**(2): 251–256. doi:10.1007/s12562-016-0968-x.

Zhang, C., Ye, Z., Panhwar, S.K., and Shen, W. 2013. Stock discrimination of the Japanese Spanish mackerel (*Scomberomorus niphonius*) based on the otolith shape analysis in the Yellow Sea and Bohai Sea. J. Appl. Ichthyol. **29**(2): 368–373. doi:10.1111/jai.12084.

Zhuang, L., Ye, Z., and Zhang, C. 2014. Application of otolith shape analysis to species separation in Sebastes spp. from the Bohai Sea and the Yellow Sea, northwest Pacific. Environ. Biol. Fishes **98**(2): 547–558. doi:10.1007/s10641-014-0286-z.

Zuur, A.F., Ieno, E.N., and Elphick, C.S. 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol. Evol. **1**: 3–14. doi:10.1111/j.2041-210X.2009.00001.x.

Figures

Fig. 1. Distribution of sampling locations of cod and herring. The shape and color of the points indicate the fish species and stock component, respectively. Size of the point shows the number of fish analyzed from the given location. The map was created based on the layer of ICES statistical areas (ICES 2019c).

Fig. 2. Cumulative Fourier power (PF_c) calculated for cod and herring showing examples of reconstructions of otolith outline with different numbers of harmonics. The box represents the interquartile range (IQR) with the median (midline) and the first and third quantiles at the bottom and top of the box, respectively. Lower and upper whiskers are restricted to 1.5 x IQR and black dots represent outliers.

Fig. 3. Principal component analysis (PCA) conducted on the Fourier coefficients of otolith shape for cod (a) and herring (b). The levels of variance explained by the first PCA axes are shown on the axes. The morphospace plotted over the observations represents theoretical shapes reconstructed based on the PCA scores.

Fig. 4. Classification accuracy of different statistical models based on different numbers of Fourier harmonics of otolith shapes. Lines represent median accuracy, shades 10^{th} and 90^{th} percentile. Models in the legend were arranged according to the median accuracy of classification on the dataset with highest number of harmonics.

Fig. 5. Variable (Fourier descriptors) relative importance obtained for cod (a) and herring (b) from otolith shape classification models.

Tables

Table 1. Summary of analyzed samples including fish species, stocks, components, capture years, sample size (N), mean total fish length (TL) ± standard deviation (SD), mean and range of age. *Due to age reading difficulties of eastern Baltic cod (EBC), age was only determined for western Baltic cod (WBC) captured in SD22 and SD23. NA= not available.

Species	Stock	Component	Years	N	Mean TL ± SD [cm]	Mean age	Age range
Cod	EBC		2015, 2016	243	43.11±5.24	NA	NA
Cod	WBC		2015, 2016	264	47.71±9.86	2.89*	1-6*
Herring	WBSS	CSS	2006, 2012, 2017	157	29.25±1.40	5.40	5-6
Herring	NSS	NSS	2018	207	31.08±1.63	5.20	5-6
Herring	CBH	CBNC	2017	170	19.39±1.81	5.51	5-6
Herring	WBSS	GB	2018	238	27.63±1.32	5.28	5-6

Table 2. Comparison of algorithm accuracies for cod (a) and herring (b). Dataset for the comparison was built on 13 harmonics (accounting for 99% variance of the otolith shape). Estimates of the difference (% accuracy) are reported in the upper diagonals, while p-values (with Bonferroni adjustment) for the hypothesis of no difference are reported in the lower diagonals.

a) cod	LDA	QDA	KNN	CART	RF	SVM
LDA		6.63	7.29	12.18	3.50	-0.90
QDA	< 0.001		0.66	5.55	-3.13	-7.53
KNN	< 0.001	0.0689		4.88	-3.79	-8.20
CART	< 0.001	< 0.001	< 0.001		-8.67	-13.08
RF	< 0.001	< 0.001	< 0.001	< 0.001		-4.41
SVM	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	

b) herring	LDA	QDA	KNN	CART	RF	SVM
LDA		-4.69	-0.48	8.80	-2.96	-11.20
QDA	< 0.001		4.21	13.49	1.73	-6.50
KNN	0.204	< 0.001		9.28	-2.48	-10.71
CART	< 0.001	< 0.001	< 0.001		-11.76	-19.99
RF	< 0.001	< 0.001	< 0.001	< 0.001		-8.24
SVM	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	

Table 3. Cross-validated (4-fold, repeated 100 times) confusion matrix obtained for cod and herring stocks with the best classifier (support vector machines) on Fourier descriptors of otolith shape. Entries are percentual average cell counts across resamples. Average accuracy (sum of diagonal cells): cod = 79.54; herring = 74.13.

	Reference					
Prediction	EBC	WBC	CSS	NSS	CBNC	GB
EBC	**38.0**	10.5				
WBC	9.9	**41.6**				
CSS			**10.7**	0.9	0.7	6.3
NSS			0.9	**23.1**	2.2	0.8
CBNC			0.4	2.3	**17.7**	1.1
GB			8.4	0.4	1.4	**22.6**

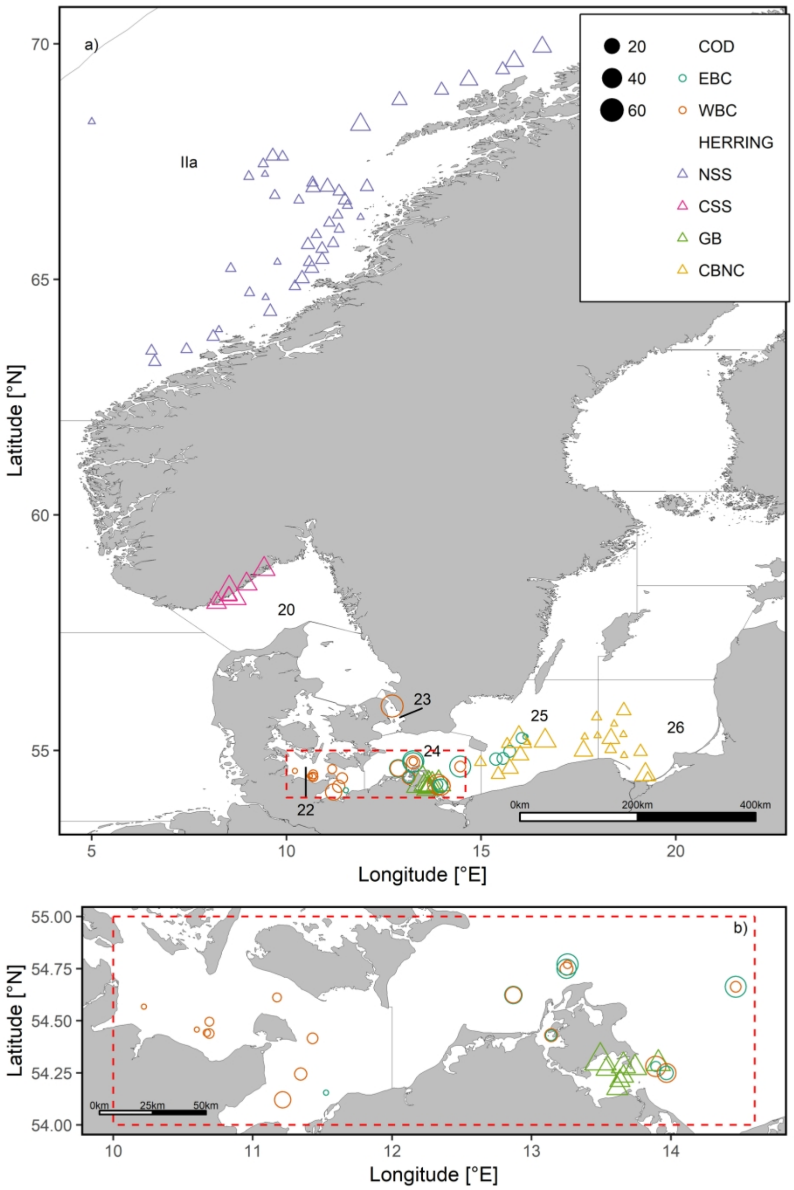

Fig. 1. Distribution of sampling locations of cod and herring. The shape and color of the points indicate the fish species and stock component, respectively. Size of the point shows the number of fish analyzed from the given location. The map was created based on the layer of ICES statistical areas (ICES 2019c).

152x228mm (300 x 300 DPI)

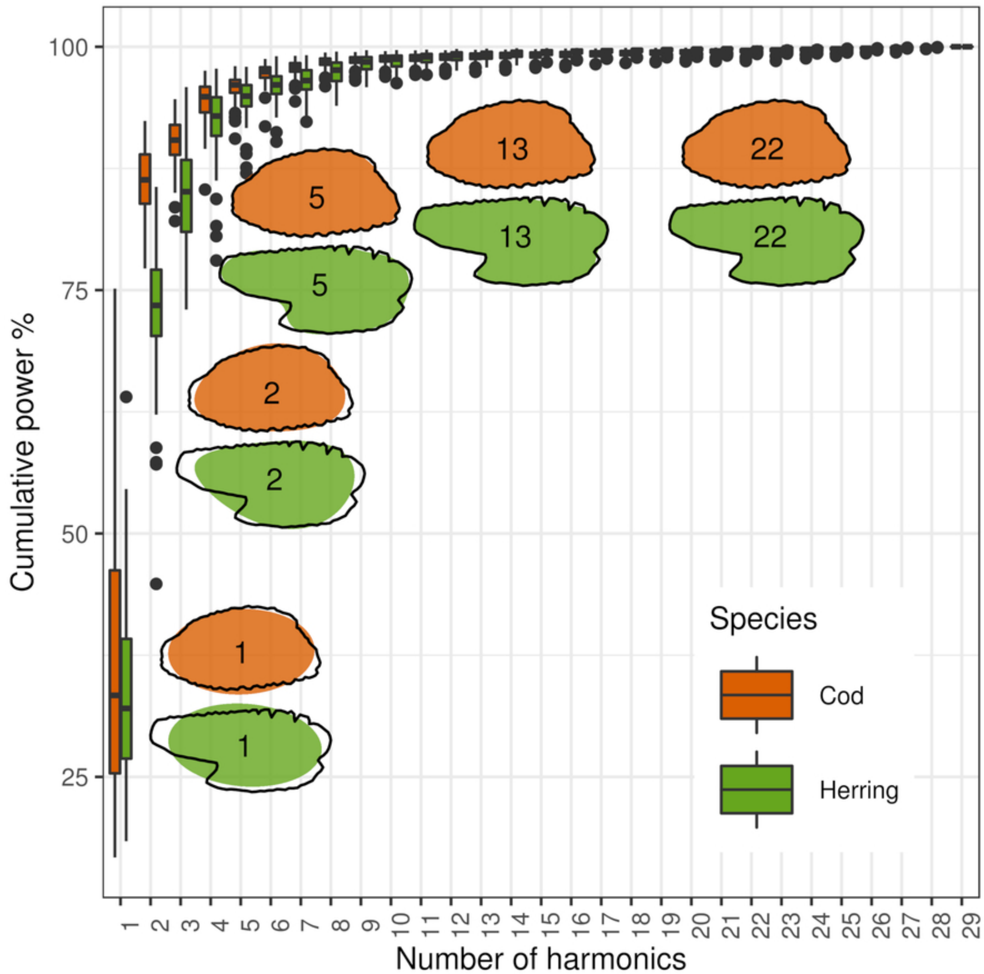

Fig. 2. Cumulative Fourier power (PF_c) calculated for cod and herring showing examples of reconstructions of otolith outline with different numbers of harmonics. The box represents the interquartile range (IQR) with the median (midline) and the first and third quantiles at the bottom and top of the box, respectively. Lower and upper whiskers are restricted to 1.5 x IQR and black dots represent outliers.

76x76mm (300 x 300 DPI)

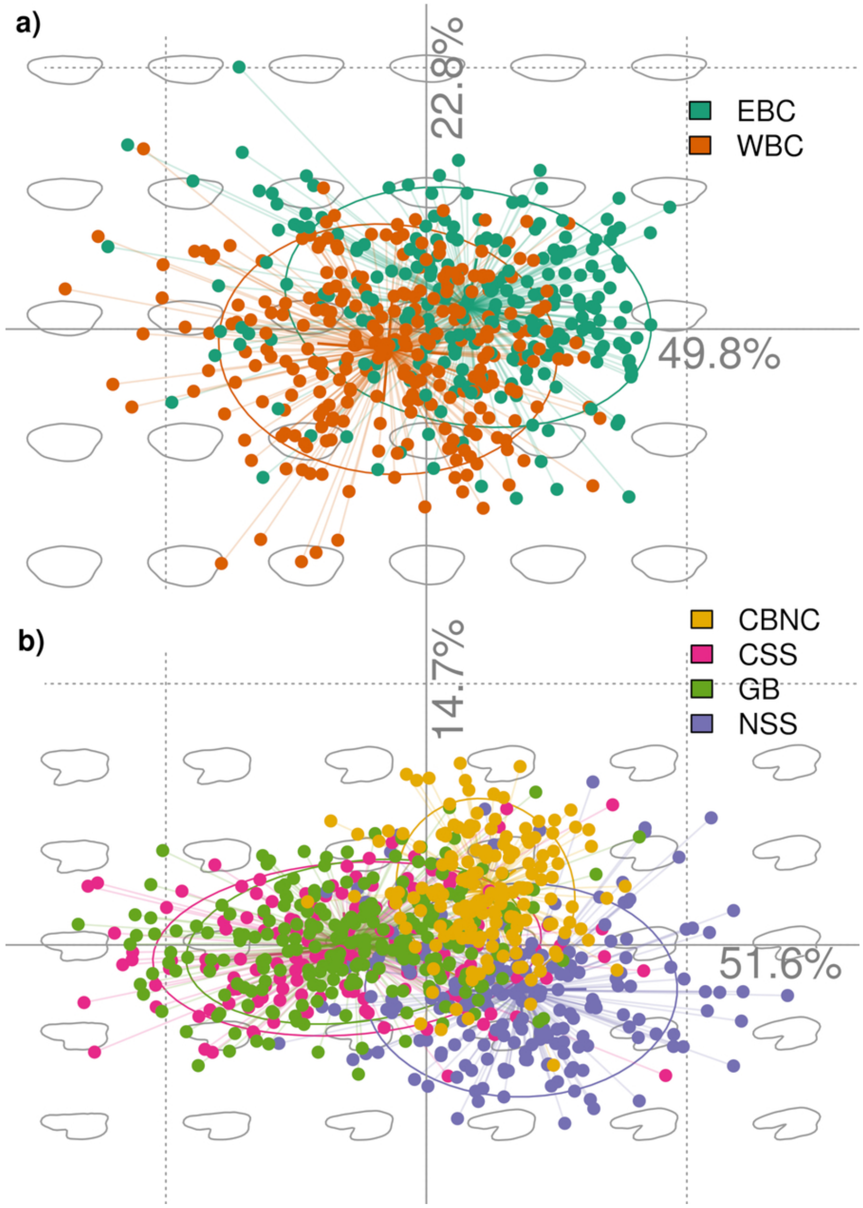

Fig. 3. Principal component analysis (PCA) conducted on the Fourier coefficients of otolith shape for cod (a) and herring (b). The levels of variance explained by the first PCA axes are shown on the axes. The morphospace plotted over the observations represents theoretical shapes reconstructed based on the PCA scores.

76x106mm (300 x 300 DPI)

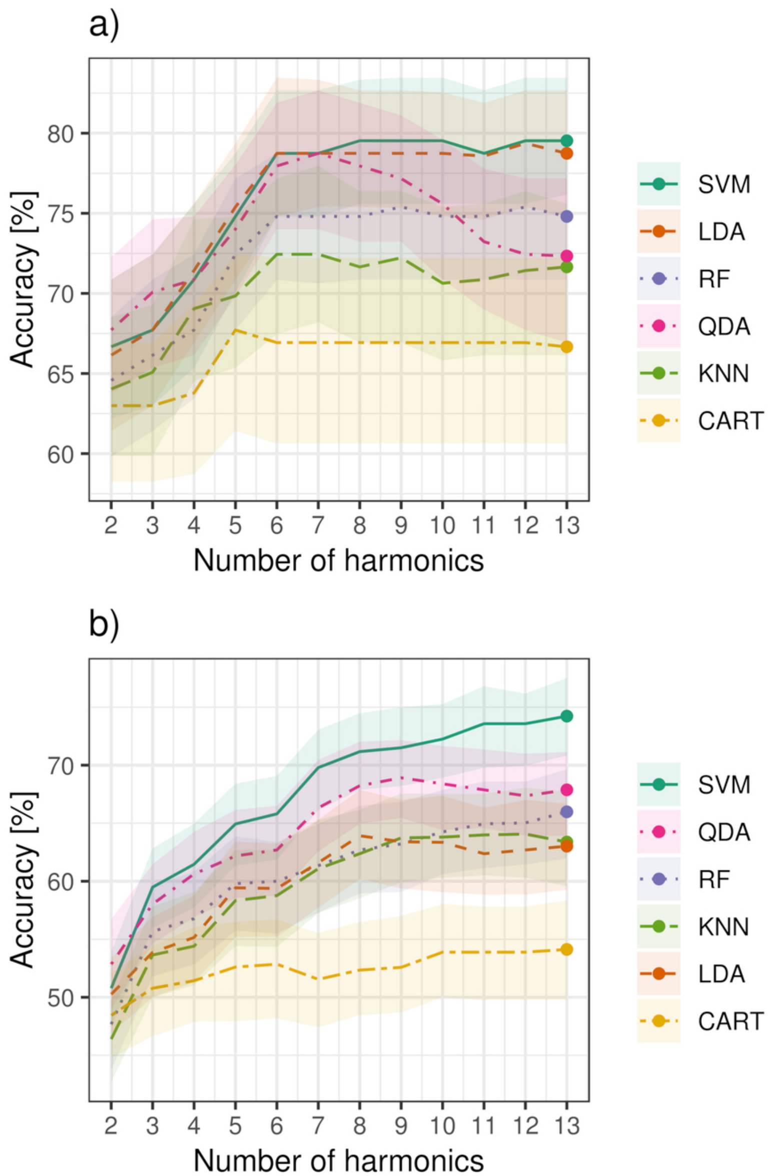

Fig. 4. Classification accuracy of different statistical models based on different numbers of Fourier harmonics of otolith shapes. Lines represent median accuracy, shades 10th and 90th percentile. Models in the legend were arranged according to the median accuracy of classification on the dataset with highest number of harmonics.

76x114mm (300 x 300 DPI)

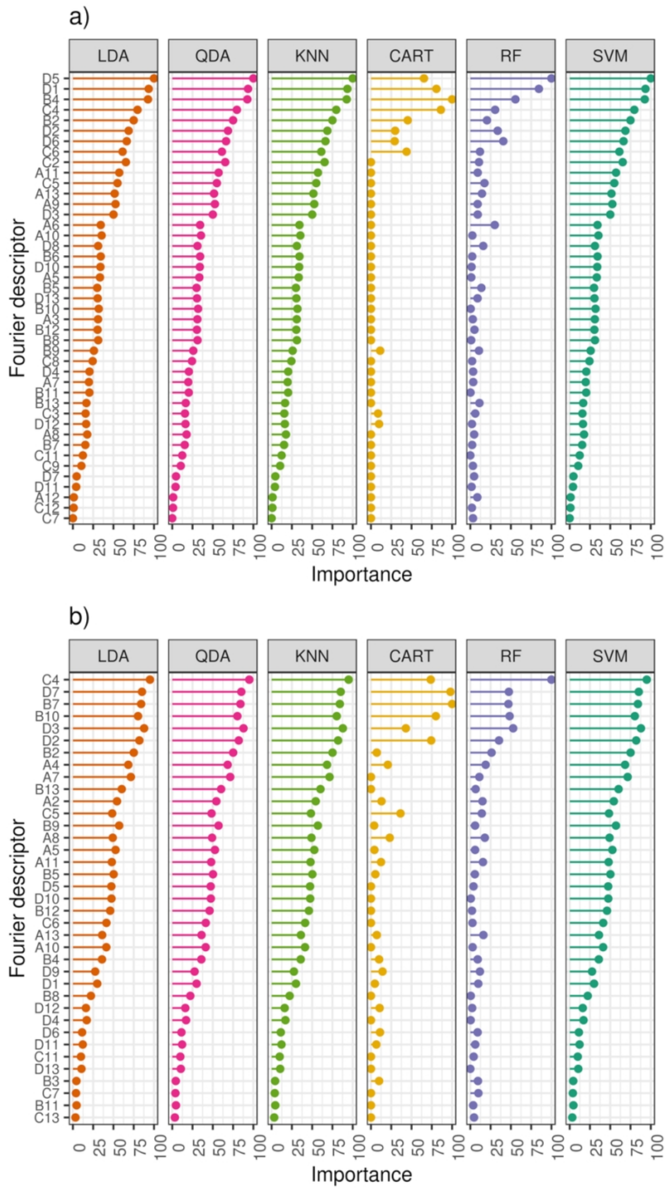

Fig. 5. Variable (Fourier descriptors) relative importance obtained for cod (a) and herring (b) from otolith shape classification models.

76x137mm (300 x 300 DPI)