Classification of fish species from different ecosystems using the near infrared diffuse reflectance spectra of otoliths

Irina M. Benson[1], Beverly K. Barnett[2], and Thomas E. Helser[1]

[1]National Marine Fisheries Service, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115

[2]National Marine Fisheries Service, Southeast Fisheries Science Center, 3500 Delwood Beach Rd., Panama City Laboratory, Panama City, FL 32408

Abstract

Applications of Fourier transform near infrared (FT-NIR) spectroscopy in fisheries science are currently limited. Our analyses of otolith spectral data demonstrate the potential applicability of FT-NIR spectroscopy to otolith chemistry and spatial variability in fisheries science. The objective of this study was to examine the use of FT-NIR spectroscopy as a tool to differentiate among marine fishes in four large marine ecosystems. We examined otoliths from 13 different species, with 3 of these species coming from different regions. Principal component analysis (PCA) described the main directions along which the specimens were separated. The separation of species and their ecosystems may suggest interactions between fish phylogeny, ontogeny, and environmental conditions that can be evaluated using FT-NIR spectroscopy. In order to discriminate spectra across ecosystems and species, four supervised classification model techniques were utilized: soft independent modelling of class analogies (SIMCA), support vector machine discriminant analysis (SVMDA), partial least squares discriminant analysis (PLSDA), and *k*-nearest neighbor analysis (KNN). This study showed that the best performing model to classify combined ecosystems, all four ecosystems, and species was the KNN model, which had

an overall accuracy rate of 99.9%, 97.6%, and 91.5%, respectively. Results from this study suggest that further investigations are needed to determine applications of FT-NIR spectroscopy to otolith chemistry and spatial variability.

Keywords

Fish, otolith, near infrared spectroscopy, principal component analysis, soft independent modelling of class analogies, support vector machine discriminant analysis, partial least squares discriminant analysis, *k*-nearest neighbor analysis

Introduction

Fishery resources in many parts of the world are managed based on population models for which fish age data are an essential source of information. In particular, age data provide information on recruitment, growth, maturity, and resource productivity[1,2] and form the basis upon which the overfished status of a stock is assessed.[1,2] Fish age has historically been determined by microscopically counting pairs of annual opaque and translucent growth zones in a number of different hard structures, including scales, vertebrae, opercula, spines, and, most commonly, otoliths.[3,4]

As part of the inner ear of teleost fishes,[5] otoliths are involved in such functions as balance and hearing.[6] Some studies link otolith morphology to swimming, feeding, spatial distribution, and acoustic communication.[7] Although fish have three pairs of otoliths, the largest otoliths, the sagittae, are usually used for age determination and research. Fish otoliths have stimulated scientific interest because these biological structures can be utilized as biochronometers.[8] Otoliths, which are acellular and metabolically inert, begin forming prior to hatching and continue to grow in three dimensions throughout the life of the fish, even when somatic growth is non-existent.[9,10] Therefore, these biological structures contain a micro-chemical record of temporally resolved environmental histories of the water properties within which the animal lived throughout its life.[10] They reveal information on stock structure,[11,12] ontogenetic movement and migration patterns,[13-15] thermal histories,[16,17] metabolic activity,[18] and paleoclimate.[19,20]

55 Otoliths are composed of alternating mineral-rich and protein-rich bands, which are
56 deposited daily.[21,22] A mineral fraction consists mainly of calcium carbonate and a variety of
57 minor and trace elements.[13,23] The organic fraction, which ranges from 0.2 to 10% by otolith
58 weight, includes over 380 proteins, glycoproteins, lipoproteins, glycosaminoglycans, and
59 polysaccharides. [8, 23-26] In otoliths an organic fraction produces absorption bands in the near
60 infrared region. Organic compounds in this region are represented by overlapping overtone and
61 combination bands of a few functional groups, such as C-H (aliphatic), C-H (aromatic), C-O
62 (carboxyl), O-H (hydroxyl), and N-H (amine and amide).[27] Although the molecular overtone and
63 combination bands seen in the near infrared region are broad, multivariate calibration and
64 classification techniques are employed to extract the desired chemical information.[28] The
65 combination of second derivative followed by standard normal variate (SNV) was shown to be
66 successful in resolving overlapping bands in addition to scattering correction.[29,30]

67 Fourier transform near infrared (FT-NIR) spectroscopy is widely utilized by agricultural,
68 pharmaceutical, chemical, and other industries due to its rapid and non-destructive testing
69 capabilities. Recently, FT-NIR spectroscopy has been applied in a few studies of biodiversity
70 and ecological physiology studies.[31,32] Applications of FT-NIR spectroscopy in fisheries science
71 are currently limited, with five published studies having focused on biological structures such as
72 otoliths of marine fishes[33-35] or shark vertebrae.[36,37] Traditional approaches to determine fish age
73 from otoliths are often time-consuming and expensive, and can include observational
74 subjectivity. In contrast, FT-NIR spectroscopy methods have shown the potential to increase the
75 efficiency and improve the repeatability of ageing studies.[35]

FT-NIR spectroscopy has offered new possibility for classification of fish otoliths based on spectral properties. The ability of near infrared radiation to penetrate samples to a great depth[38] provides unique opportunity for capturing otoliths' capacity as biochronometers. We hypothesized that this technology could provide essential information for ecological studies and may be useful in determining diet composition based on spectral data from otoliths found in stomachs of marine predators. The objective of this study was to examine the use of FT-NIR spectroscopy as a tool to differentiate between 16 marine fishes representing 13 different species from four large marine ecosystems (Table 1, Figure 1). The specific aims were to (1) acquire spectral scans of otoliths, (2) select effective spectral regions, (3) calibrate and validate classification models that discriminate between ecosystems and fish species, and (4) compare the models to select the optimal one by comparing the overall accuracy estimates.

Materials and methods

Spectral data acquisition and preprocessing

FT-NIR spectra were acquired from the sagittal otoliths (n = 3,703) of 16 marine fishes representing 13 different species sampled from four large U.S. marine ecosystems (Table 1, Figure 1). Otoliths were blotted and air dried prior to scanning. Diffuse reflectance measurements from all otoliths were collected on either Bruker TANGO R or MPA II FTNIR spectrometers with integrating spheres. Each otolith was covered with a gold-coated reflector

stamp and scanned at a 90° compass orientation with a concave-up position (Figure 2). Spectra were collected between 11,500 and 4,000 cm^{-1} at a resolution of 16 cm^{-1}. For each sample, 64 scans were co-added and converted to a final absorbance spectrum (Figure 3A). OPUS software v. 7.5 (Bruker Optics, Ettlingen, Germany) was used for spectral acquisition.

Multivariate Data Analysis

Spectral data were preprocessed with second derivative (Savitzky-Golay method, 2nd order polynomial, 21 points), SNV, and mean centering (Figure 3B-C). Chemometric software Solo v8.7 (Eigenvector Research, Inc., Manson, WA, USA) was used for data preprocessing, exploratory analysis using principal component analysis (PCA), data split, and multivariate classification including soft independent modelling of class analogies (SIMCA), support vector machine discriminant analysis (SVMDA), partial least squares discriminant analysis (PLSDA), and *k*-nearest neighbor analysis (KNN).

Exploratory Data Analysis

In order to learn about the data distribution and grouping, PCA was used to evaluate the extent of spectral variability across species and ecosystems. PCA considers all variables and linearly transforms the original data into new orthogonal latent variables (principal components, PC). Each PC is defined by a loading vector and has maximum variance of the scores.[39] For this study, PCA score plots were used to explore the sample distribution and analyze grouping patterns in the data by plotting the first two or three PCs, which often represent most of the variability in the data. PC loadings were used to select optimal wavenumbers for discriminating species and ecosystems.

Data Split

In order to validate classification models, each species data set was split into a calibration and a prediction set using the Kennard-Stone method, which selects a subset of samples that provide uniform coverage over the predictor space and includes exterior samples in the calibration set (Table 1).[40] The calibration set was used to generate classification models and the prediction set was used to assess the predictive performance of models.

Multivariate Classification

Four supervised classification model techniques, SIMCA, SVMDA, PLSDA, and KNN, were used for evaluating spectral variability across the data from different species and ecosystems. The main task of the spectra classification analysis is to summarize the multivariate data structure of the groups in order to establish rules for correctly assigning samples with unknown group membership.[39] The correct classification rate for each group was reported as the ratio of correctly classified samples with the total number of samples in the prediction set (Ratio) and rate (%) of correctly classified samples (Tables 2 - 4). The overall accuracy estimate was also reported for each model.

SIMCA is a classification technique that applies PCA decomposition to each group separately and identifies variables that are important for group assignment.[39,41] PLSDA is a linear classification method that calculates separate PLS regression vectors and corresponding predicted values to determine group membership.[28, 42] SVMDA is a non-linear method that computes an optimal direction to discriminate between groups. It maps the original data into a transformed space where linear boundaries can be constructed in order to maximize the margin

between classes, which is the distance between boundary and the nearest data point of each class.[39,43] KNN is a nonparametric method based on calculating the distance between each sample and its *k*-nearest neighbors and uses these closest *k* objects to estimate group membership of a new object.[28,39,44] For our study $k = 3$ was selected.

Results and discussion

Exploratory Data Analysis

All combinations of species were analyzed with four PCs from the PCA. PC loading indicated that the wavenumbers from 7,486.34 cm^{-1} to 6,645.36 cm^{-1} and from 6,035.24 to 4,015.25 cm^{-1} had the largest contributions to PCs. Therefore, these wavenumbers were selected for final data analysis (Figure 4). The selected wavenumbers cover first overtones of N-H, O-H, and C-H bonds, stretching vibrations, and combinations.[37]

Most variation in the spectral data was described by PC1, PC2, and PC3. PC1 (81.48% of captured variance) is the main direction along which the specimens separated (Figure 5A). Two data clusters were clearly separated along the PC1 axis. Eastern Bering Sea and North Pacific Ocean specimens represented most of the samples in two elongated clusters, with similar within-class variance, on the positive side of the PC1 axis (Figure 5A). Gulf of Mexico and North Atlantic Ocean specimens were represented in both an elongated cluster, with similar within-class variance, in the middle of the plot and a small tight cluster on the negative side of

the PC1 axis (Figure 5A). The zoomed-in view of the negative-side cluster showed Gulf of Mexico and North Atlantic Ocean clusters overlapping, with Gulf of Mexico species located to the left of North Atlantic Ocean species (Figure 5A-B).

PC2 (7.85% of captured variance) and PC3 (5.17% of captured variance) are the main directions along which the specimens separated according to the latitudinal variation of the ecosystems from which species were collected (Figure 5A). All data clusters overlapped along the PC2 and PC3 axes (Figure 5A-B). Eastern Bering Sea and North Pacific Ocean specimens were displaced from each other with the higher scores along the PC2 axis corresponding to the North Pacific Ocean species and the higher scores along the PC3 axis corresponding to the eastern Bering Sea (Figure 5A). Gulf of Mexico and North Atlantic Ocean specimens were also displaced vertically from each other with the lower scores along the PC3 axis corresponding to the Gulf of Mexico (Figure 5B).

Clustering of species was observed for all ecosystems. The difference between eastern Bering Sea and North Pacific Ocean species was visible when PCA scores were plotted against the first three PCs (Figure 6A-B). Red snapper from all regions grouped with the eastern Bering Sea and North Pacific Ocean species, while gag from all regions grouped with vermilion snapper and the other North Atlantic Ocean species (Figure 6A-B). Acadian redfish and haddock from the northern part of the North Atlantic Ocean were represented by separate clusters (Figure 6B). The difference between the eastern and western Gulf of Mexico species and North Atlantic Ocean species from the southern U.S. waters of the Atlantic Ocean was small with faint clustering along the PC3 axis (Figure 6B).

Multivariate Classification

Classification models were applied to discriminate between species grouped by the ecosystems. Classification by combined ecosystems was the most successful. Discrimination between the group that included eastern Bering Sea and North Pacific Ocean specimens and the group that included Gulf of Mexico and North Atlantic Ocean specimens showed high overall accuracy for all models (Table 2). After examining the four classification models, KNN and SVMDA were the best performing models with the KNN model (99.9% accuracy) slightly outperforming the SVMDA model (99.8% accuracy). Classification success by combined ecosystems can be explained by species genetic divergence. The transarctic interchange of marine organisms between the northern Pacific and Atlantic oceans happened about 3.5 million years ago during the middle and late Pliocene.[45] Significant genetic distances between modern populations of marine fish in the two oceans are thought to take millions of years to develop.[45]

Classification by four ecosystems had various degrees of success for different models and predicted group membership (Table 3). The accuracy of predictions for eastern Bering Sea and North Pacific Ocean specimens for all four models ranged from 98.6% to 100%. The accuracy of predictions for Gulf of Mexico specimens for all four models ranged from 64.7% to 98.2%. North Atlantic Ocean specimens were the hardest to predict, with accuracy of prediction being 26.1% for SIMCA, 60.5% for PLSDA and SVMDA, and 96.6% for KNN. After examining the four classification models, KNN was the best performing model with the overall accuracy of 97.6%.

These classification results are consistent with environmental variations in all four ecosystems that may be due to the general direction of ocean currents. The southward-flowing

California Current and the northward-flowing Alaska Current receive different volumes and flows of warm water from the same source, eastward-flowing North Pacific Current (Figure 1).[46,47] The Alaska Current continues into the Alaskan Stream, which transports water into the Bering Sea.[46,47] Difference of the eastern Bering Sea group from North Pacific Ocean specimens can be explained by annual sea ice formation and melt events, which have strong impacts on marine biogeochemical cycles in the Bering Sea.[46,48] In the Gulf of Mexico, the Loop Current, which brings warm water from the Caribbean, dominates oceanographic features (Figure 1).[49] The Loop Current leaves the Gulf and continues as the Gulf Stream parallel to the Atlantic coast (Figure 1).[49,50] The Gulf Stream separates from the shelf edge near Cape Hatteras and creates a boundary between two oceanographic regimes where some members of a southern warm-temperate fauna and a northern cold-temperate fauna are known to move between two regimes.[50]

Classification models were also applied to discriminate between the species. Classification by species had various degrees of success for different models and predicted group membership (Table 4). Three models (SIMCA, KNN, and SVMDA) obtained good discrimination rates for species from the eastern Bering Sea, North Pacific Ocean, and the northern part of the North Atlantic Ocean. PLSDA was the worst performing model for these species with an accuracy of prediction of 82% for walleye pollock from the eastern Bering Sea and 55.6% for haddock from the northern part of the North Atlantic Ocean. All four models could not correctly classify all species from the Gulf of Mexico and the southern part of the North Atlantic Ocean. Vermilion snapper was correctly classified by both KNN and SVMDA but not by PLSDA (50% accuracy) and SIMCA (75% accuracy). The accuracy of predictions for gag grouper from the Gulf of Mexico for all four models ranged from 48% by PLSDA to 99.3% by SIMCA. The accuracy of predictions for gag grouper from the southern part of the North

225 Atlantic Ocean ranged from 6.3% by PLSDA to 100% by SVMDA. While red snapper from the
226 eastern part of Gulf of Mexico was predicted with a higher degree of accuracy (74.7% to 97.3%),
227 the prediction of red snapper from the western part of the Gulf of Mexico had worse accuracy
228 rates for all models (3.3% to 56.7%). The accuracy of predictions for North Atlantic red snapper
229 ranged from 12.8% by SIMCA to 95.7% by KNN. After examining the four classification
230 models, KNN was determined to be the best performing model to discriminate between species,
231 with an overall accuracy estimate of 91.5%. Most of the misclassifications for KNN happened
232 between the same species from different geographic areas (e.g. red snapper and gag grouper) or
233 different species from the same geographic area (e.g. walleye pollock, Pacific cod, and yellowfin
234 sole) (Figure 7).

235 Although precise environmental and biological factors contributing to spectral variability
236 of otoliths are unknown at this time, our study suggests that for the most part species that are
237 taxonomically and regionally closer share similar molecular constituents activated by near
238 infrared light. Taxonomic differences may reflect phylogenetic relationship, functional role, and
239 structure of otoliths.[51] Thomas et al.[8] revealed that most otolith proteins are highly conserved
240 across taxa. The authors also suggested that the diversity of otolith proteins reflect development
241 and physiological change over an individual's lifetime.[8] Regional differences are related to
242 environmental differences between different geographic locations.[51] The structure of some otolith
243 proteins is influenced by such external physical factors as temperature,[22,52] which affects protein
244 synthesis.[53] Essential amino acids in otoliths are related to fish diet and trophic food web
245 structure.[54] Therefore, it is possible that habitat, environment, and diet composition in different
246 ecosystems may interact with fish phylogenetic and ontogenetic development to influence
247 spectral differences in otoliths. FT-NIR spectroscopy of otoliths may prove to be a useful tool

not only for distinguishing the residence of fishes among habitats but also for investigating the effects of warming ocean water on the food web.

The results of this study show potential for providing a fast and reliable method of identifying fish species and populations. It is less time-consuming than otolith shape analysis, otolith microstructure analysis, or genetic research. In addition to saving time, FT-NIR spectroscopy does not require otolith destruction. It leaves otoliths intact and usable for other analyses. To strengthen the case for using FT-NIR spectroscopy for taxonomic classification of otoliths, further research is needed to determine an optimal otolith set with all possible spectral variability among specimens and within each age class. Future research may determine if scanning otoliths with FT-NIR spectroscopy can assist with differentiating fish species that are of similar appearance and difficult to taxonomically identify. Perhaps additional studies can also make inquiries into other fish structures like skin or muscle[55] to validate our findings of differentiation between fish species and populations.

Conclusion

Our analyses of otolith spectral data collected from 13 different marine fish species from four marine ecosystems demonstrate the potential applicability of FT-NIR spectroscopy to fisheries science. The separation of ecosystems and species by longitude and latitude may suggest interactions between phylogenetics, ontogeny, and environmental conditions that can be evaluated using FT-NIR spectroscopy. This study showed that the best performing model to

classify combined ecosystems, all four ecosystems, and species was a KNN model. Results from this study clearly suggest that further investigations are needed to determine applications of FT-NIR spectroscopy to otolith chemistry and spatial variability. FT-NIR spectroscopy of otoliths may prove to be a useful tool not only for distinguishing residence among ecosystems but for investigating the effects of warming ocean waters and on the food web.

Acknowledgments

We express sincere appreciation to everyone whose contributions made this research successful. The Age and Growth Program staff (AFSC) who scanned Pacific cod, walleye pollock, and yellowfin sole otoliths. Andrew Claiborne (WDFW) contributed and scanned Chinook salmon otoliths. Patrick McDonald (NOAA-PSMFC) contributed and scanned North Pacific hake otoliths. Mellissa Monk (SWFSC) contributed and scanned gopher rockfish otoliths. Emmanis Dorval (SWFSC), Julianne Taylor (IMS-UC-SWFSC), Dianna Porzio (CDFW), and Leeanne Laughlin (CDFW) contributed and scanned Pacific sardine and Pacific mackerel otoliths. Eric Robillard (NEFSC) contributed and scanned Acadian redfish and haddock otoliths. Jennifer Potts (SEFSC) and Andy Ostrowski (SEFSC) contributed and scanned vermilion snapper and gag grouper otoliths. Michelle Passerotti (USC) scanned red snapper otoliths. Jason Erickson (Bruker Optics) shared his scientific insight and expertise on near infrared spectroscopy. Charles Hutchinson (AFSC), Delsa Anderl (AFSC), and Jordan Healy (UW) shared their advice and expertise on otolith imaging. Jay Orr (AFSC) and Duane Stevenson (AFSC) offered their suggestions for the manuscript improvement. We thank the National Marine Fisheries Service Science Board for providing support for the FT-NIR spectroscopy scientific endeavor.

Declaration of conflicting interests

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This project was funded by the National Marine Fisheries Service FT-NIR Spectroscopy Age Estimation Strategic Initiative. The findings and conclusions in the paper are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service, NOAA. The use of trade, firm, or corporation names in this publication is for the convenience of the reader and does not imply endorsement by the National Marine Fisheries Service, NOAA.

References

1. Maunder M and Punt A. A review of integrated analysis in fisheries stock assessment. *Fish Res* 2013; 142: 61-74. DOI: 10.1016/j.fishres.2012.07.025.

2. Ono K, Licandeo R, Muradian M, et al. The importance of length and age composition data in statistical age-structured models for marine species. *ICES J Mar Sci* 2014; 72(1): 31-43. DOI: 10.1093/icesjms/fsu007.

3. Bagenal T and Tesch F. Age and Growth. In: Bagenal, T (ed) *Methods for Assessment of Fish Production in Fresh Waters.* 3rd ed. Oxford: IBP Handbook No. 3, Blackwell Science Publications, 1978, pp.101-136.

4. Chilton D and Beamish R. Age determination methods for fishes studied by the Groundfish Program at the Pacific Biological Station. Ottawa: Dept. of Fisheries and Oceans, 1982, p.102.

5. Campana S. Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Mar Ecol Prog Ser* 1999; 188: 263-297. DOI: 10.3354/meps188263.

6. Popper A and Lu Z. Structure-function relationships in fish otolith organs. *Fish Res* 2000; 46(1-3): 15-25. DOI: 10.1016/S0165-7836(00)00129-6.

7. Tuset V, Farré M, Otero-Ferrer J, et al. Testing otolith morphology for measuring marine fish biodiversity. *Mar Freshw Res* 2016; 67: 1037–1048. DOI: 10.1071/MF15052.

8. Thomas O, Swearer S, Kapp E, et al. The inner ear proteome of fish. *FEBS J* 2019; 286(1): 66-81. DOI: 10.1111/febs.14715.

9. Morales-Nin B, Swan S, Gordon J, et al. Age-related trends in otolith chemistry of *Merluccius merluccius* from the north-eastern Atlantic Ocean and the western Mediterranean Sea. *Mar Freshw Res* 2005; 56(5): 599-607. DOI: 10.1071/mf04151.

10. Sturrock A, Trueman C, Darnaude A, et al. Can otolith elemental chemistry retrospectively track migrations in fully marine fishes? *J Fish Biol* 2012; 81(2): 766-795. DOI: 10.1111/j.1095-8649.2012.03372.x.

11. Campana S, Fowler A, Jones C. Otolith elemental fingerprinting for stock identification of Atlantic cod (*Gadus morhua*) using laser ablation ICPMS. *Can J Fish Aquat Sci* 1994; 51(9): 1942-1950. DOI: 10.1139/f94-196.

12. Sluis M, Barnett B, Patterson W, et al. Application of otolith chemical signatures to estimate population connectivity of red snapper in the western Gulf of Mexico. *Mar Coastal Fish* 2015; 7(1): 483-496. DOI: 10.1080/19425120.2015.1088492.

13. Elsdon T and Gillanders B. Reconstructing migratory patterns of fish based on environmental influences on otolith chemistry. *Rev Fish Biol Fish* 2003; 13(3): 217-235. DOI: 10.1023/B:RFBF.0000033071.73952.40.

14. Siskey M, Wilberg M, Allman R, et al. Forty years of fishing: changes in age structure and stock mixing in northwestern Atlantic bluefin tuna (*Thunnus thynnus*) associated with size-selective and long-term exploitation. *ICES J Mar Sci: Journal du Conseil.* 2016; 73(10): 2518-2528. DOI: 10.1093/icesjms/fsw115.

15. Wright P, Régnier T, Gibb F, et al. Assessing the role of ontogenetic movement in maintaining population structure in fish using otolith microchemistry. *Ecol Evol* 2018; 8(16): 7907-7920. DOI: 10.1002/ece3.4186.

16. Darnaude A, Sturrock A, Trueman C, et al. Listening In on the Past: What can otolith δ18O values really tell us about the environmental history of fishes? *PLoS ONE* 2014; 9(10): e108539. DOI: 10.1371/journal.pone.0108539.

17. Sakamoto T, Komatsu K, Yoneda M, et al. Temperature dependence of δ 18 O in otolith of juvenile Japanese sardine: Laboratory rearing experiment with micro-scale analysis. *Fish Res* 2017; 194: 55-59. DOI: 10.1016/j.fishres.2017.05.004.

18. Chung M, Trueman C, Godiksen J, et al. Field metabolic rates of teleost fishes are recorded in otolith carbonate. *Commun Biol* 2019; 2(1). DOI: 10.1038/s42003-018-0266-5.

19. Patterson W. Oldest isotopically characterized fish otoliths provide insight to Jurassic continental climate of Europe. *Geology* 1999; 27(3): 199. DOI: 10.1130/0091-7613(1999)027<0199:OICFOP>2.3.CO;2.

20. Helser T, Kastelle C, Crowell A, et al. A 200-year archaeozoological record of Pacific cod (*Gadus macrocephalus*) life history as revealed through ion microprobe oxygen isotope ratios in otoliths. *J Archaeol Sci Rep* 2018; 21: 1236-1246. DOI: 10.1016/j.jasrep.2017.06.037.

21. Jolivet A, Fablet R, Bardeau J, et al. Preparation techniques alter the mineral and organic fractions of fish otoliths: Insights using Raman micro-spectrometry. *Anal Bioanal Chem* 2013; 405(14): 4787-4798. DOI: 10.1007/s00216-013-6893-2.

22. Thomas O and Swearer S. Otolith Biochemistry—A Review. *Rev Fish Sci Aquac* 2019; 27(4): 458-489. DOI: 10.1080/23308249.2019.1627285.

23. Thomas O, Ganio K, Roberts B, et al. Trace element–protein interactions in endolymph from the inner ear of fish: Implications for environmental reconstructions using fish otolith chemistry. *Metallomics* 2017; 9(3): 239-249. DOI: 10.1039/c6mt00189k.

24. Dauphin Y and Dufour E. Composition and properties of the soluble organic matrix of the otolith of a marine fish: *Gadus morhua* Linne, 1758 (Teleostei, Gadidae). *Comp Biochem Physiol A Mol Integr Physiol* 2003; 134(3): 551-561. DOI: 10.1016/S1095-6433(02)00358-6.

25. Avallone B, Fascio U, Balsamo G, et al. Morphogenesis of otoliths during larval development in brook lamprey, *Lampetra planeri*. *Ital J Zool* 2007; 74(3): 247-258. DOI: 10.1080/11250000701459301.

26. Degens E, Deuser W, Haedrich R. Molecular structure and composition of fish otoliths. *Mar Biol* 1969; 2: 105—113. DOI: 10.1007/BF00347005.

27. Xiaoboa Z, Jiewen Z, Povey M, et al. Variables selection methods in near-infrared spectroscopy. *Anal Chim Acta* 2010; 667(1-2): 14-32. DOI: 10.1016/j.aca.2010.03.048.

28. Næs T, Isaksson T, Fearn T, et al. *A user-friendly guide to multivariate calibration and classification*. Chichester: NIR Publications, 2002. p. 344.

29. Huang J, Romero-Torres S, Moshgbar M. Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy. *Am Pharm Rev* 2010. https://www.americanpharmaceuticalreview.com/Feat ured-Articles/116330-Practical-Considerations-inData-Pre-treatment-for-NIR-and-RamanSpectroscopy

30. Conrad A, Rodriguez-Saona L, McPherson B, et al. Identification of *Quercus agrifolia* (coast live oak) resistant to the invasive pathogen *Phytophthora ramorum* in native stands using Fourier-transform infrared (FT-IR) spectroscopy. *Front Plant Sci* 2014; 5(521): 1-9. DOI: 10.3389/fpls.2014.00521.

31. Vance C, Tolleson D, Kinoshita K, et al. Near infrared spectroscopy in wildlife and biodiversity. *J Near Infrared Spectrosc* 2016; 24(1): 1-25. DOI: 10.1255/jnirs.1199.

32. Jintao X, Liming Y, Yufei L, et al. Noninvasive and fast measurement of blood glucose in vivo by near infrared (NIR) spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc* 2017; 179: 250-254. DOI: 10.1016/j.saa.2017.02.032.

33. Wedding B, Forrest A, Wright C, et al. A novel method for the age estimation of saddletail snapper (*Lutjanus malabaricus*) using Fourier Transform-near infrared (FT-NIR) spectroscopy. *Mar Freshw Res* 2014; 65(10): 894-900. DOI: 10.1071/MF13244.

34. Robins J, Wedding B, Wright C, et al. *Revolutionising fish ageing: using near infrared spectroscopy to age fish.* Report for the Department of Agriculture, Fisheries and Forestry, Brisbane, April 2015. CC BY 3.0. http://www.frdc.com.au/Archived-Reports/FRDC%20Projects/2012-011-DLD.pdf

35. Helser T, Benson I, Erickson J, et al. A transformative approach to ageing fish otoliths using Fourier transform near infrared spectroscopy: a Case study of eastern Bering Sea walleye pollock (*Gadus chalcogrammus*). *Can J Fish Aquat Sci* 2019; 76(5): 780-789. DOI: 10.1139/cjfas-2018-0112.

36. Rigby C, Wedding B, Grauf S, et al. The utility of near infrared spectroscopy for age estimation of deepwater sharks. *Deep-sea Res Part 1 Oceanogr Res Pap* 2014; 94: 184-194. DOI: 10.1016/j.dsr.2014.09.004.

405 37. Rigby C, Wedding B, Grauf S, et al. Novel method for shark age estimation using near
406 infrared spectroscopy. *Mar Freshw Res* 2016; 67(5): 537-545. DOI: 10.1071/MF15104.

407 38. ~~49.~~ Padalkar M and Pleshko N. Wavelength-dependent penetration depth of near infrared
408 radiation into cartilage. *Analyst* 2015; 140 (7): 2093–2100. DOI: 10.1039/C4AN01987C.

409 39. Varmuza K and Filzmoser P. *Introduction to multivariate statistical analysis in*
410 *chemometrics*. Boca Raton, FL: Francis & Taylor, CRC Press, 2009. p.336.

411 40. Kennard R and Stone L. Computer aided design of experiments. *Technometrics* 1969; 11(1):
412 137-148. DOI: 10.1080/00401706.1969.10490666.

413 41. Wold S. Pattern recognition by means of disjoint principal components models. *Pattern*
414 *Recogn* 1976; 8(3): 127-139. DOI: 10.1016/0031-3203(76)90014-5.

415 42. Barker M and Rayens W. Partial least squares for discrimination. *J Chemom* 2003; 17(3):
416 166-173. DOI: 10.1002/cem.785.

417 43. Cristianini N and Shawe-Taylor J. *An Introduction to Support Vector Machines and Other*
418 *Kernel-based Learning Methods*. Cambridge: Cambridge University Press, 2000. p. 189.
419 DOI: 10.1017/CBO9780511801389.

420 44. Altman N. An Introduction to kernel and nearest-neighbor nonparametric regression. *The Am*
421 *Stat* 1992; 46(3): 175-185. DOI: 10.1080/00031305.1992.10475879.

422 45. Palumbi S and Kessing B. Population biology of the trans-arctic exchange: mtDNA sequence
423 similarity between Pacific and Atlantic sea urchins. *Evolution* 1991; 45: 1790-1805. DOI:
424 10.1111/j.1558-5646.1991.tb02688.x.

425 46. Benson I, Kastelle C., Helser T, et al. Age interpretation in eulachon (*Thaleichthys pacificus*)
426 as suggested by otolith microchemical signatures. *Environ Biol Fishes* 2019; 102: 629–643.
427 DOI: 10.1007/s10641-019-00858-7.

428 47. Stabeno P, Farley E, Kachel N, et al. A comparison of the physics of the northern and
429 southern shelves of the eastern Bering Sea and some implications for the ecosystem. *Deep-*
430 *Sea Res II* 2012; 65-70: 14–30. DOI: 10.1016/j.dsr2.2012.02.019.

431 48. Vancoppenolle M, Meiners K, Michel C, et al. Role of sea ice in global biogeochemical
432 cycles: emerging views and challenges. *Quat Sci Rev* 2013; 79: 207–230. DOI:
433 10.1016/j.quascirev.2013.04.011.

434 49. Chen Y. Fish resources of the Gulf of Mexico. In: Ward C (ed) *Habitats and Biota of the*
435 *Gulf of Mexico: Before the Deepwater Horizon Oil Spill*. New York: Springer, 2017, pp.
436 869-1038.

437 50. Hare J, Churchill J, Cowen R, et al. Routes and rates of larval fish transport from the
438 southeast to the northeast United States continental shelf. *Limnol Oceanogr* 2002; 47(6):
439 1774–1789. DOI: 10.4319/lo.2002.47.6.1774.

440 51. Chang M-Y and Geffen A. Taxonomic and geographic influences on fish otolith
441 microchemistry. *Fish Fish* 2012; 14: 458–492. DOI: 10.1111/j.1467-2979.2012.00482.x.

442 52. Różycka M, Wojtas M, Jakób M, et al. Intrinsically disordered and pliable starmaker-like
443 protein from medaka (*Oryzias latipes*) controls the formation of calcium carbonate
444 crystals. *PLoS One* 2014; 9(12): e114308. DOI: 10.1371/journal.pone.0114308.

445 53. Hussy K, Mosegaard H, Jessen F. Effect of age and temperature on amino acid composition
446 and the content of different protein types of juvenile Atlantic cod (*Gadus morhua*) otoliths.
447 *Can J Fish Aquat Sci* 2004; 61: 1012-1020. DOI: 10.1139/f04-037.

448 54. McMahon K, Fogel M, Johnson B, et al. A new method to reconstruct fish diet and
449 movement patterns from δ13C values in otolith amino acids. *Can J Fish Aquat Sci* 2011;
450 68(8): 1330-1340. DOI: 10.1139/F2011-070.

451 55. O'Brien N, Hulse C, Pfeifer F, et al. Near infrared spectroscopic authentication of seafood. *J*
452 *Near Infrared Spectrosc* 2013; 21: 299-305. DOI: 10.1255/jnirs.1063.

Table 1. Ecosystems, fish species, and sample size.

Ecosystems	Map Key	Species	Cal. SN	Pred. SN
Eastern Bering Sea (EBS)	EBS.1	walleye pollock (*Gadus chalcogrammus*)	150	150
	EBS.2	Pacific cod (*Gadus macrocephalus*)	150	150
	EBS.3	yellowfin sole (*Limanda aspera*)	150	150
North Pacific Ocean (NPO)	NPO.1	Chinook salmon (*Oncorhynchus tshawytscha*)	150	150
	NPO.2	North Pacific hake (*Merluccius productus*)	150	150
	NPO.3	gopher rockfish (*Sebastes carnatus*)	100	13
	NPO.4	Pacific sardine (*Sardinops sagax*)	100	12
	NPO.5	Pacific mackerel (*Scomber japonicus*)	150	109
U.S. Gulf of Mexico (GOM)	GOM.1	red snapper (*Lutjanus campechanus*) – W GOM	150	150
	GOM.2	gag grouper (*Mycteroperca microlepis*)	150	150
	GOM.3	red snapper (*Lutjanus campechanus*) – E GOM	150	150
North Atlantic Ocean (NAO)	NAO.1	vermilion snapper (*Rhomboplites aurorubens*)	150	4
	NAO.2	red snapper (*Lutjanus campechanus*)	150	47
	NAO.3	gag grouper (*Mycteroperca microlepis*)	100	16
	NAO.4	haddock (*Melanogrammus aeglefinus*)	150	9
	NAO.5	Acadian redfish (*Sebastes fasciatus*)	150	43

Cal. SN – calibration set sample number, Pred. SN – prediction set sample number.
E GOM – Eastern U.S. Gulf of Mexico, W GOM – Western U.S. Gulf of Mexico.

Table 2. Classification by combined ecosystems comparing performance of four models: soft independent modelling of class analogies (SIMCA), partial least squares discriminant analysis (PLSDA), support vector machine discriminant analysis (SVMDA), and *k*-nearest neighbor analysis (KNN).

Classification by combined ecosystems	Model											
	SIMCA			PLSDA			SVMDA			KNN		
	Ratio	%	Overall accuracy estimate	Ratio	%	Overall accuracy estimate	Ratio	%	Overall accuracy estimate	Ratio	%	Overall accuracy estimate
EBS and NPO	884/884	100	94.6%	871/884	99.7	98.5%	884/884	100	99.8%	884/884	100	99.9%
GOM and NAO	490/569	86.1		560/569	98.4		566/569	99.5		568/569	99.8	

EBS – Eastern Bering Sea, NPO – North Pacific Ocean, GOM – U.S. Gulf of Mexico, NAO – North Atlantic Ocean.

Table 3. Classification by ecosystems comparing performance of four models: support vector machine discriminant analysis (SVMDA), partial least squares discriminant analysis (PLSDA), soft independent modelling of class analogies (SIMCA), and *k*-nearest neighbor analysis (KNN).

Classification by ecosystems	Model											
	SVMDA			PLSDA			SIMCA			KNN		
	Ratio	%	Overall accuracy estimate	Ratio	%	Overall accuracy estimate	Ratio	%	Overall accuracy estimate	Ratio	%	Overall accuracy estimate
EBS	450/450	100	85.8%	450/450	100	85.9%	448/450	99.6	93%	450/450	100	97.6%
NPO	433/434	99.8		428/434	98.6		430/434	99.1		434/434	100	
NAO	72/119	60.5		72/119	60.5		31/119	26.1		115/119	96.6	
GOM	291/450	64.7		298/450	66.2		442/450	98.2		419/450	93.1	

EBS – Eastern Bering Sea, NPO – North Pacific Ocean, NAO – North Atlantic Ocean, GOM – U.S. Gulf of Mexico.

Table 4. Classification by fish species results comparing performance of four models: partial least squares discriminant analysis (PLSDA), soft independent modelling of class analogies (SIMCA), support vector machine discriminant analysis (SVMDA), and *k*-nearest neighbor analysis (KNN).

Classification by species	Map Key	Model											
		PLSDA			SIMCA			SVMDA			KNN		
		Ratio	%	Overall accuracy estimate	Ratio	%	Overall accuracy estimate	Ratio	%	Overall accuracy estimate	Ratio	%	Overall accuracy estimate
walleye pollock	EBS.1	123/150	82	66.6%	143/150	95.3	84.9%	145/150	96.7	88.2%	142/150	94.7	91.5%
Pacific cod	EBS.2	145/150	96.7		146/150	97.3		145/150	96.7		148/150	98.7	
yellowfin sole	EBS.3	135/150	90		146/150	97.3		148/150	98.7		148/150	98.7	
Chinook salmon	NPO.1	150/150	100		150/150	100		150/150	100		150/150	100	
North Pacific hake	NPO.2	150/150	100		148/150	98.7		150/150	100		150/150	100	
gopher rockfish	NPO.3	13/13	100		13/13	100		13/13	100		13/13	100	
Pacific sardine	NPO.4	12/12	100		12/12	100		12/12	100		12/12	100	
Pacific mackerel	NPO.5	109/109	100		107/109	98.2		109/109	100		109/109	100	
red snapper - W GOM	GOM.1	13/150	8.7		5/150	3.3		22/150	14.7		85/150	56.7	
gag grouper	GOM.2	72/150	48		144/145	99.3		136/150	90.7		147/150	98	
red snapper - E GOM	GOM.3	143/150	95.3		146/150	97.3		139/150	92.7		112/150	74.7	
vermilion snapper	NAO.1	2/4	50		3/4	75		4/4	100		4/4	100	
red snapper	NAO.2	11/47	23.4		6/47	12.8		40/47	85.1		45/47	95.7	
gag grouper	NAO.3	1/16	6.3		12/16	75		16/16	100		12/16	75	
haddock	NAO.4	5/9	55.6		9/9	100		9/9	100		9/9	100	
Acadian redfish	NAO.5	39/43	90.7		43/43	100		43/43	100		43/43	100	

E GOM – East U.S. Gulf of Mexico, W GOM – West U.S. Gulf of Mexico.

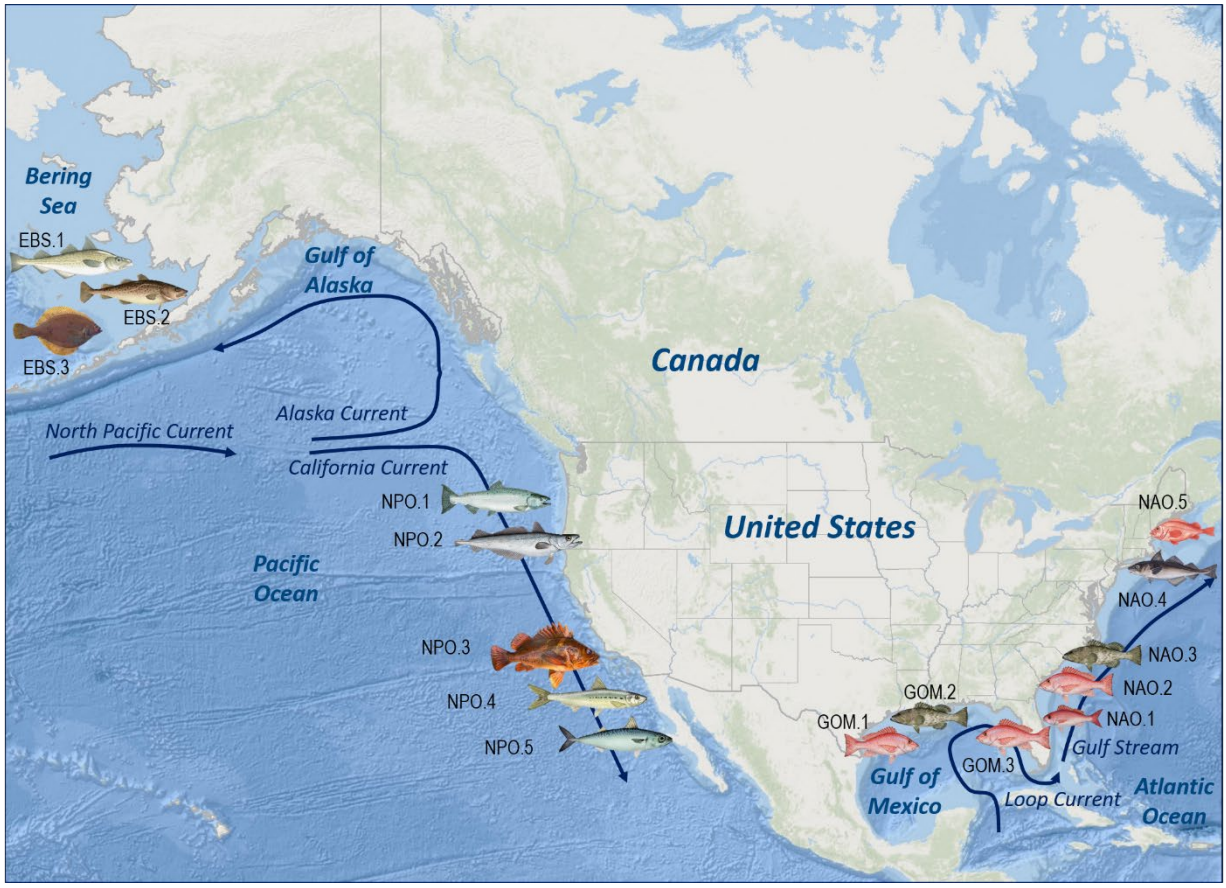

Figure 1. Location of fish collection sites. Complete list of species is shown in Table 1.

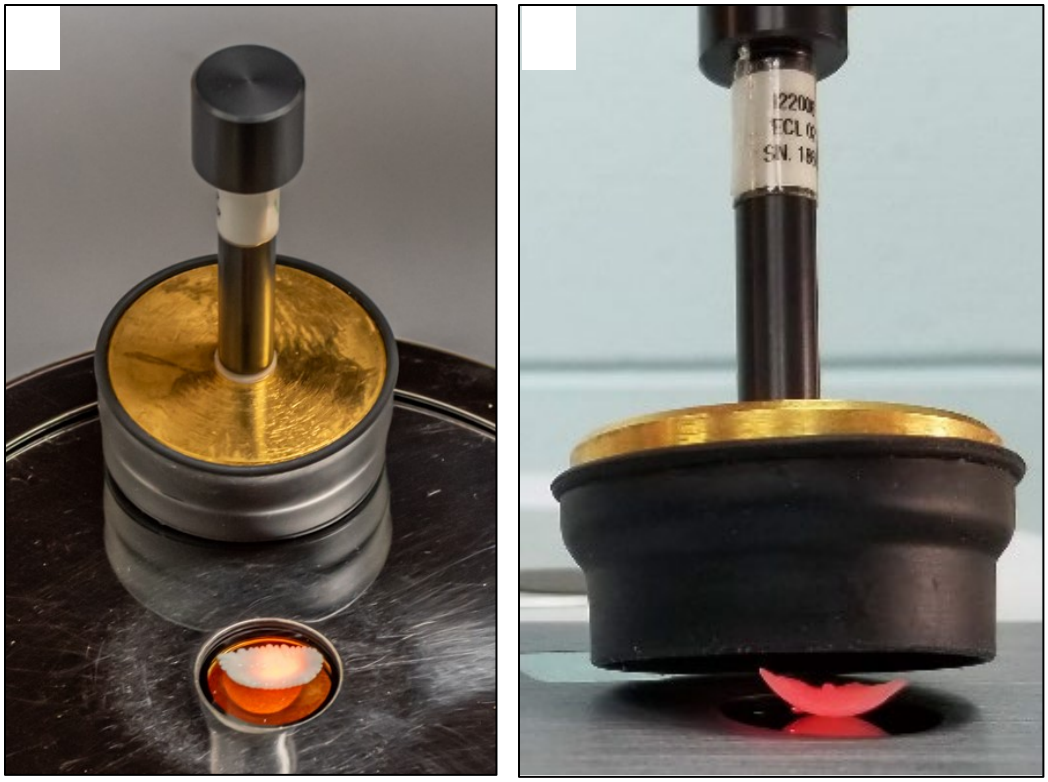

Figure 2. (A) Top and (B) side view of an otolith placed on the integrating sphere at a 90° compass orientation with a concave-up position. The gold-coated reflector stamp covers the sample during scanning to reduce stray light infiltration.

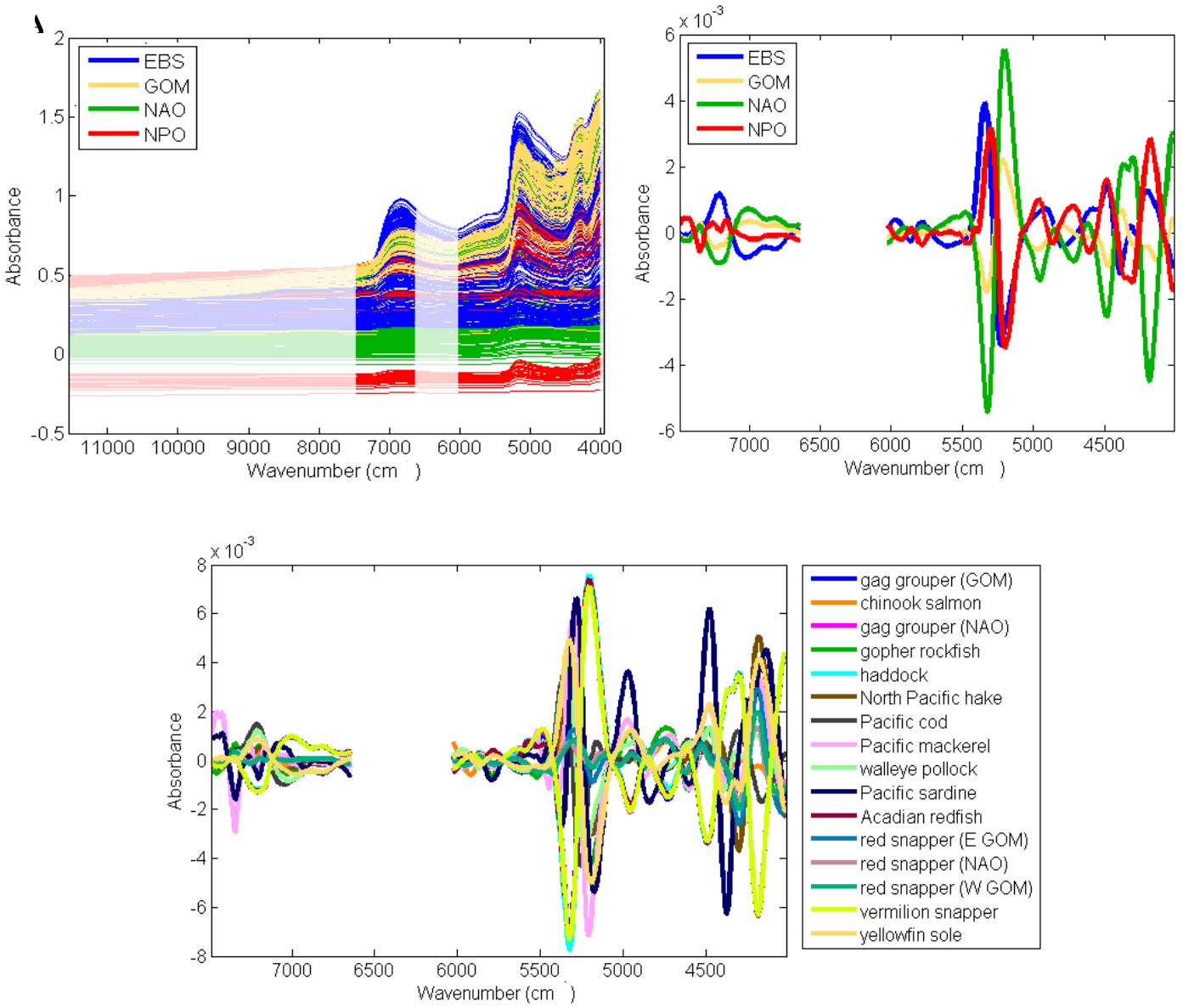

Figure 3. (A) Raw otolith FT-NIR spectra colored by regions. Whited-out areas indicate spectral regions that were excluded from consideration in final data analysis. (B) Second derivative and SNV-transformed mean spectra colored by regions. (C) Second derivative and SNV-transformed mean spectra colored by species. GOM – U.S. Gulf of Mexico, NPO – North Pacific Ocean, NAO – North Atlantic Ocean, EBS – Eastern Bering Sea, E GOM – East U.S. Gulf of Mexico, W GOM – West U.S. Gulf of Mexico.

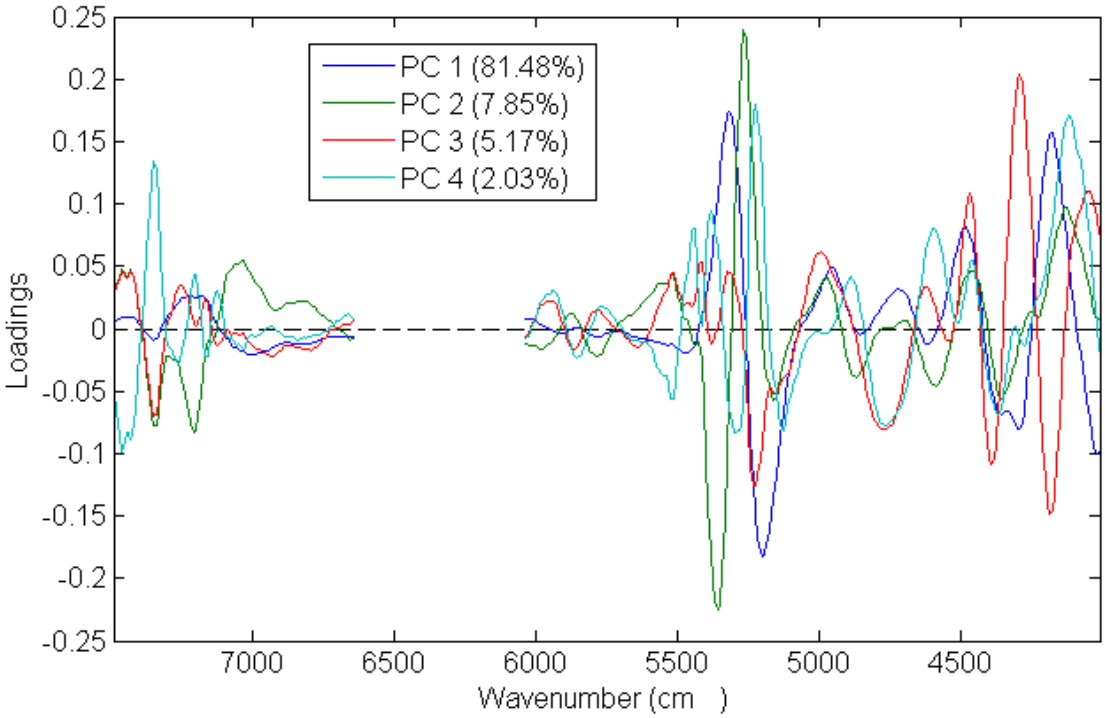

Figure 4. Principal component (PC) loadings.

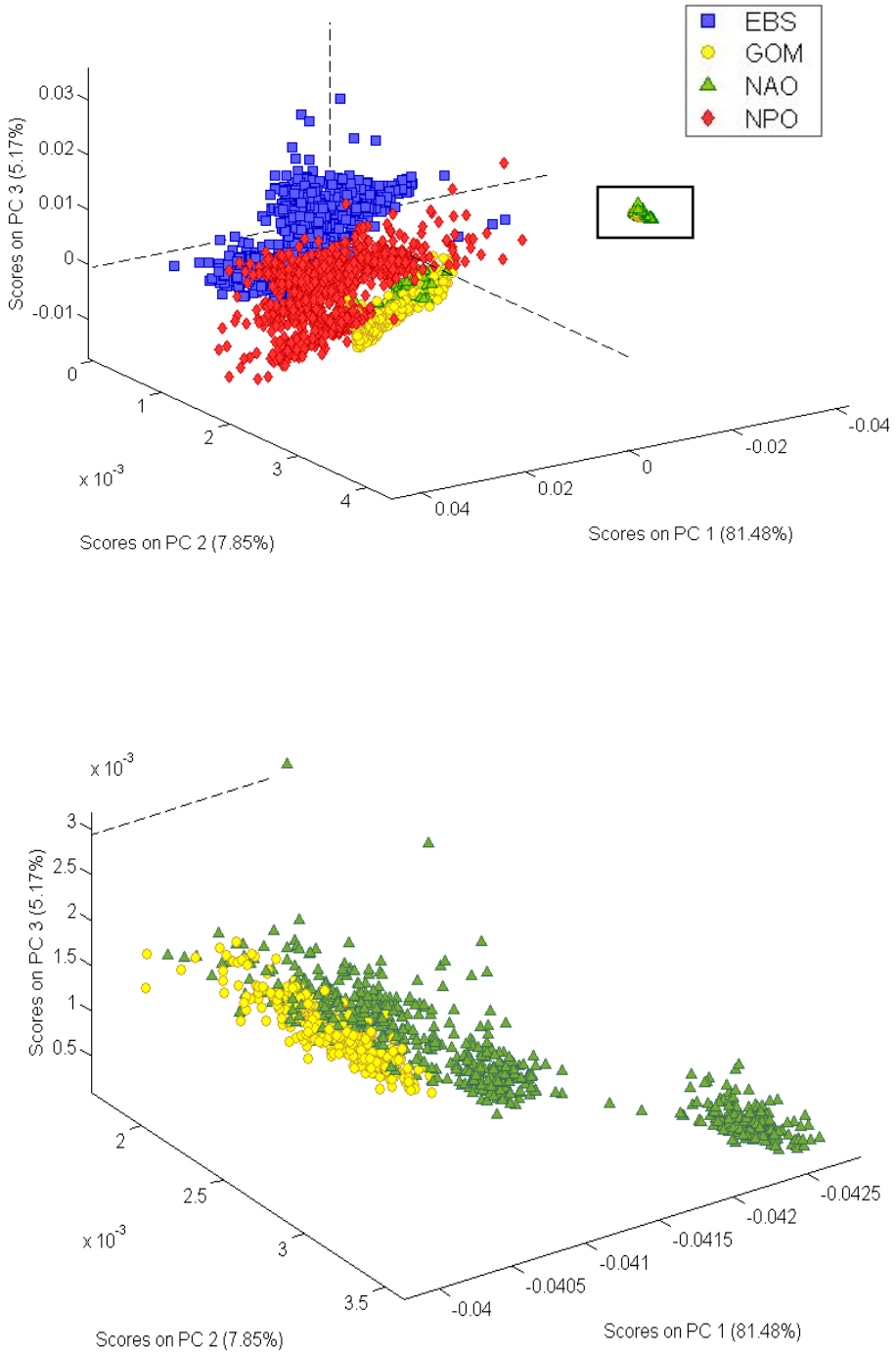

Figure 5. (A) 3D view of the PCA scores of otolith FT-NIR spectra colored by regions and (B) zoomed-in view of the black rectangular area of the plot. EBS – Eastern Bering Sea, GOM – U.S. Gulf of Mexico, NAO – North Atlantic Ocean, NPO – North Pacific Ocean.

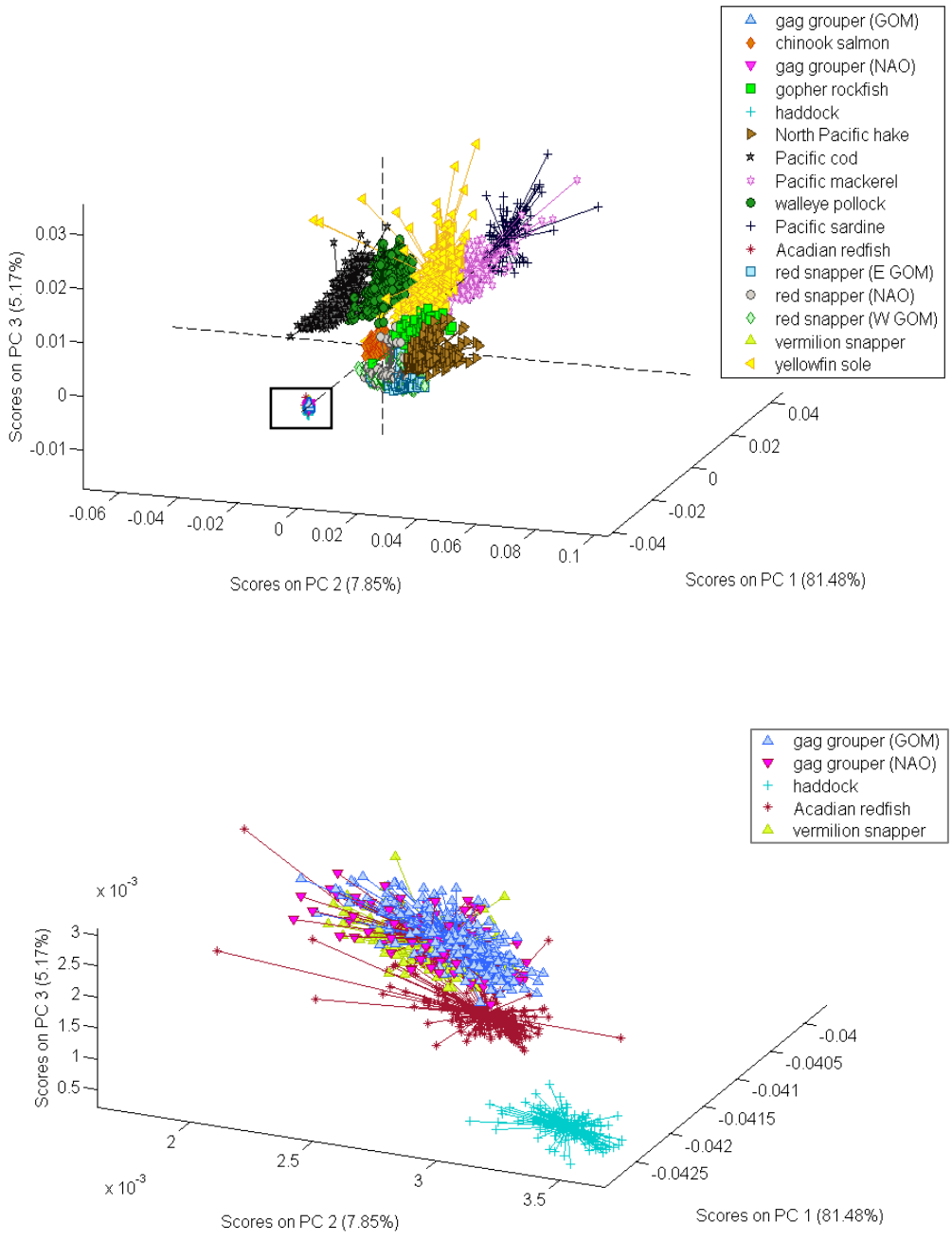

Figure 6. (A) 3D view of the PCA scores of otolith FT-NIR spectra colored by species and (B) zoomed-in view of the black rectangular area of the plot. GOM – U.S. Gulf of Mexico, NAO – North Atlantic Ocean, E GOM – East U.S. Gulf of Mexico, W GOM – West U.S. Gulf of Mexico.

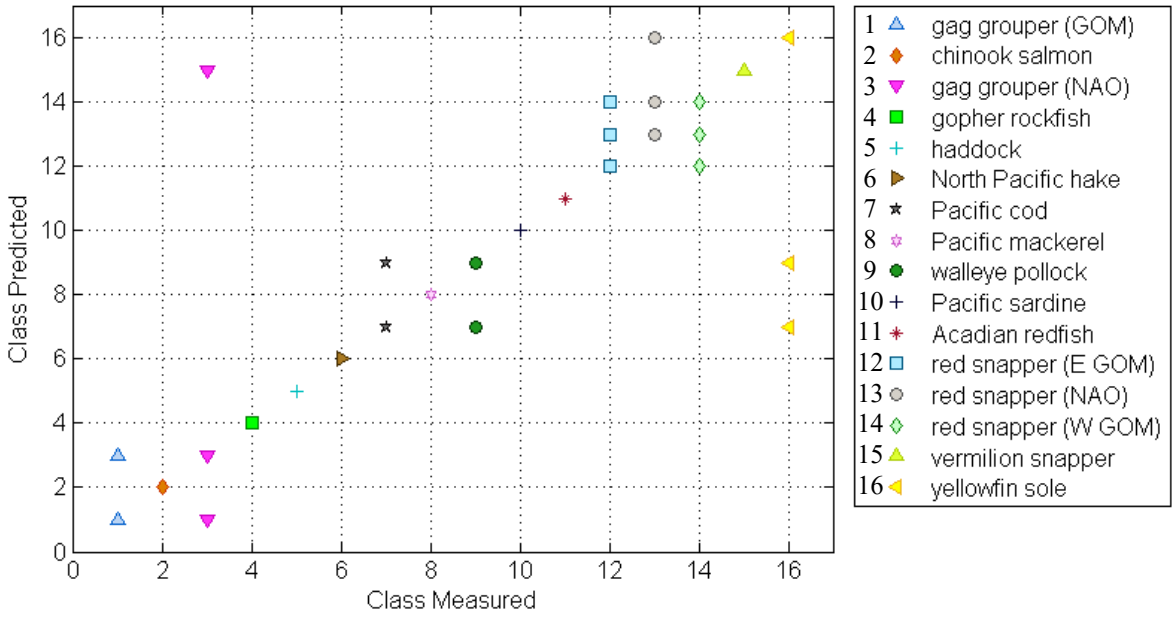

Figure 7. Species classification by *k*-nearest neighbor (KNN) modeling results.