AFSC PROCESSED REPORT 2024-01

NOAA FISHERIES

Alaska Fisheries Science Center
Resource Ecology and Fisheries Management Division
Age and Growth Program

Proceedings of the Fourth Research Workshop on the Rapid Estimation of Fish Age Using Fourier Transform Near Infrared Spectroscopy

JANUARY 2024

U.S. Department of Commerce | National Oceanic and Atmospheric Administration | National Marine Fisheries Service

AFSC Processed Report

This document should be cited as follows:
Matta, M. E., and Helser, T. E. (editors). 2024. Proceedings of the Fourth Research Workshop on the Rapid Estimation of Fish Age Using Fourier Transform Near Infrared Spectroscopy. AFSC Processed Rep. 2024-01, 524 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

This document is available online at: https://repository.library.noaa.gov/

Proceedings of the Fourth Research Workshop on the Rapid Estimation of Fish Age Using Fourier Transform Near Infrared Spectroscopy

Mary Elizabeth Matta and Thomas E. Helser (editors)

Resource Ecology and Fisheries Management Division
Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA 98115

January 2024

PREFACE

Increasing demand for fish age data to support stock assessments of the Nation's fisheries has necessitated investigation into alternative technologies to produce these data more efficiently. Among the advanced technologies the National Marine Fisheries Service (NOAA Fisheries) has explored in recent years, Fourier transform near infrared (FT-NIR) spectroscopy of otoliths represents a transformative, machine-based approach for rapidly estimating fish ages. This method is significantly faster and more reproducible than the traditional method of age estimation, which relies on a human reader's visual interpretation of growth patterns in hard structures. In contrast, FT-NIR spectroscopy functions at the molecular level, measuring absorbance associated with vibrations of covalent bonds in the NIR wavenumber range. Predictive analytics or deep learning tools are then used to establish a predictive relationship between otolith NIR absorbance and fish age.

In 2019, NOAA Fisheries funded a 5-year Strategic Initiative (SI) in support of research and development toward system deployment of FT-NIR spectroscopy for fish age determination. To date, the Alaska Fisheries Science Center (AFSC) has hosted four FT-NIR spectroscopy workshops, inviting staff from the regional NOAA age determination laboratories and experts in the field from around the world to share their research. These proceedings summarize presentations made at the fourth SI research workshop entitled, "*Rapid Estimation of Fish Age Using Fourier Transform Near Infrared Spectroscopy*", held at the AFSC in Seattle, WA, April 3-7, 2023. Presentations given at the workshop are included in this document as extended abstracts, and in many cases, full research papers. The studies presented herein reflect the tremendous progress made since the start of the SI towards application development and

implementation of FT-NIR spectroscopy for fish age determination and integration into the stock assessment process. Case studies on the application of FT-NIR spectroscopy for predicting fish age and other biological characteristics such as reproductive status and body condition are presented for a number of fish taxa and regions of the United States. In several studies, deep machine learning was used to predict age and advance the ability to quantify uncertainty not available to classical chemometric methods. Several extended abstracts highlight research towards developing a flexible simulation framework that can be applied to explore uncertainty in model-based age predictions for a range of scenarios and species. In another, the authors describe the envisioned full integration of FT-NIR spectroscopy into the production ageing process, providing a framework for system deployment in the laboratory. Finally, three case studies are presented that evaluate the sensitivity of stock assessment model outputs to the substitution of age data products generated from the FT-NIR spectroscopy approach. These are just a few of the works in this document, which represents the most comprehensive collection yet of the application of FT-NIR spectroscopy in fisheries science.

CONTENTS

1. NOAA Fisheries Fourier Transform Near Infrared Spectroscopy Fish Ageing Strategic Initiative: Past, Present, and Future

Thomas E. Helser

Chair, NOAA Fisheries FT-NIR Spectroscopy SIDT
Program Manager
AFSC Age & Growth Program

Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

INTRODUCTION

The National Oceanic and Atmospheric Administration (NOAA) National Marine Fisheries Service (Fisheries) has undertaken a nationally coordinated research-and-development (R&D) effort across seven biological laboratories to investigate the use of Fourier transform near infrared (FT-NIR) spectroscopy of otoliths for fish age estimation (Helser et al. 2019a). NOAA Fisheries funded a 5-year strategic initiative (SI) entitled "A revolutionary approach for improving age determination efficiency in fish using FT-NIR spectroscopy" that began with funding in Fiscal Year (FY) 2020 with the overarching goal of improving precision and efficiency within the NOAA Fisheries fish ageing enterprise. To achieve this objective, the FT-NIR Strategic Initiative Development Team (SIDT), comprised of NOAA scientists from all seven biological laboratories, convened a planning workshop entitled, "*Rapid Estimation of Fish Age Using Fourier Transform Near-Infrared Spectroscopy*" at the Alaska Fisheries Science Center in Seattle, WA, April 11-12, 2019 (Helser et al. 2019a). This collaborative effort involved the participation of national and international scientists and industry experts in FT-NIR spectroscopy and resulted in a framework and "roadmap" toward operationalizing the FT-NIR spectroscopy ageing technology across NOAA Fisheries science centers.

Among the advanced technologies NOAA Fisheries has been exploring in recent years (https://www.fisheries.noaa.gov/insight/advanced-technologies), FT-NIR spectroscopy represents a transformative, machine-based approach for rapidly estimating fish ages for over 200,000 hard structures, such as otoliths, that annually support fish stock assessments and fishery management plans. Multiples of that number of ageing requests are made, and in many cases, otoliths are collected. Nevertheless, the capacity to fulfill this demand remains unmet. The

central principle governing fish age determination based on the growth patterns visible in otoliths has changed little over the last century. While ageing fish using scales can be documented as far back as the early 1700s, Johannes Reibisch was the first to describe a protocol for ageing fish using otoliths in 1899 (Jackson 2007). Microscopic counting of the annual growth zones in otoliths was the foundation of the ageing method then and is still the basis of age estimation today. With the exception of advancements in microscopy and otolith preprocessing techniques, the standard practice today remains essentially unchanged. The entire process of generating ages can be expensive, time consuming, and labor intensive, especially for species with long lifespans or complex otolith processing methods. Age data have been referred to as one of the most expensive sources of data collected for stock assessments.

FT-NIR spectroscopy ageing of fish is fundamentally different from traditional microscopic ageing based on visual ring counts. It functions at the molecular level, measuring absorbance associated with vibrations of covalent bonds (O-H, C=O, C-H, C-N, and N-H) in the NIR wavenumber range (4,000 to 12,500 cm^{-1}). Predictive analytics or deep learning is employed, enabling the establishment of a mathematical relationship between otolith NIR absorbance and fish age. The overarching challenge of the SI was to transfer a mature, yet continuously evolving technology from other fields like medicine, agriculture, and pharmaceutical industries to unique conditions and problems presented in fishery science. It is my belief that the SIDT has risen to those challenges, both conceptually and technologically, although some challenges were completely unexpected.

Over the past several years, our R&D indicates that machine-based ageing using FT-NIR spectroscopy of otoliths holds substantial promise of a transformative impact on the NOAA Fisheries fish ageing enterprise. The irrefutable proof lays in the science underpinning the use of

FT-NIR spectroscopy for age prediction as demonstrated by a number of published studies across fish taxa and regions in the United States. (Helser et al. 2019a, Helser et al. 2019b, Healy et al. 2021, Arrington et al. 2022, Passerotti et al. 2020a, Passerotti et al. 2022b, Benson et al. 2023), with others in Australia (Wedding et al. 2014, Rigby et al. 2016, Wright et al. 2021), and more are yet to be published (including a number of studies described in these proceedings). Taken together, we are between the "Proof of Concept Development" and "Application Development" stages, a pathway describing the nine technical readiness levels (TRL) for technologies in machine learning systems described by Lavin et al. (2022).

Time horizons of attaining operational readiness vary depending on domain, but the general thinking is that five years or more are required to advance innovative technologies to deployment with a TRL-9 (Lavin et al. 2022). An article in Forbes Magazine under Leadership Strategy (https://www.forbes.com/sites/chuckswoboda/2020/06/15/leading-innovation-is-messy-so-get-over-it/?sh=2b9f47ae7434) says that "innovation is messy and success requires embracing the unexpected." We certainly did not expect the global SARS-CoV-2 pandemic that began on March 11, 2020 (https://www.who.int/europe/emergencies/situations/covid-19); the very first year of SI Funding. Needless to say, the pandemic had a profound impact on our SI R&D. A "significant delay" was the consensus of the SIDT, which was communicated to the NOAA Fisheries Science Board and resulted in extending SI funding for an additional year (FY2025).

Despite the 2+ year setback from the pandemic, substantial progress of the SI has been made. However, that progress has not been shared equally among the science centers. The reasons for varying progress among the ageing laboratories, in my estimation, stem from several reasons, although others may be in play. Ageing laboratories have different levels of staff and resources available to dedicate towards the FT-NIR spectroscopy SI, while at the same time

continue to provide production ages for annual stock assessments. The AFSC and Panama City ageing laboratories invested in FT-NIR spectroscopy instruments prior to the SI funding and were able to spearhead proof-of-concept development and initiate otolith scanning sooner than other laboratories. Moreover, during the SARS-CoV-2 pandemic, Bruker FT-NIR spectroscopy instruments could not be fielded and set up at other laboratories because infection mitigation procedures differed in various regions of the country, resulting in varying access to laboratory facilities. If the impression that one gets, from reading these proceedings, is that AFSC and Panama City have made greater advancements in SI R&D, it is not for lack of trying from SIDT members at other ageing laboratories. The central message to the reader from my perspective is that good progress toward operationalization has been made, although much work is yet to be done across all science centers. Within the pages of this research proceedings document, the reader will hopefully comprehend the depth and breadth of our R&D of FT-NIR spectroscopy and its application to fish ageing and beyond.

Our April 2023 FT-NIR spectroscopy fish ageing workshop featured three main themes that formed the backbone of the "roadmap" towards operational readiness: (1) application development, (2) application implementation, and (3) stock assessment integration (Helser et al. 2019a). Application development focused on the broad-scale applicability of the technology to achieve stated goals for age prediction from otolith spectra in real-world scenarios with robust, verifiable data. In application implementation, the goals were to investigate the delivery of new data generated by predictive models, define and evaluate model performance, establish protocols for data processing and quality control, and standardize operating procedures and best practices to guide decisions related to sample size, precision, and performance measures. The stock assessment integration theme sought to evaluate the sensitivity of population model output of

critical management outcomes and benchmarks from incorporating FT-NIR spectroscopy data products. Taken together, each element of the hierarchy builds upon the previous tier with "discovery switch-backs" helping to refine and improve the technical knowledge of the system.

Within the pages of this document, "Proceedings of the Fourth Research Workshop on the Rapid Estimation of Fish Age Using Fourier Transform Near Infrared Spectroscopy", we present the accumulated knowledge of the FT-NIR spectroscopy SIDT R&D efforts across the three themes. Presentations given at the workshop are presented in this document as extended abstracts, and in many cases, full research papers likely to appear in the scientific literature. Case studies within the application development theme are presented describing the successful application of FT-NIR spectroscopy for predicting fish age across a number of fish taxa and regions of the U.S. In several of those applications, deep machine learning, using convolutional neural networks (CNN), was employed to improve age prediction and advance the ability to quantify uncertainty not available to classical chemometric methods. Within the application implementation theme, several extended abstracts highlight research toward developing a flexible simulation framework that can be applied to explore uncertainty on model-based age predictions for a range of scenarios and species. In another abstract, the authors describe the end-to-end integration of the FT-NIR spectroscopy production ageing process as it would be envisioned for system deployment in the laboratory. Finally, within the stock assessment integration theme, three case studies are presented which evaluate the sensitivity of stock assessment model (e.g., Stock Synthesis) outputs and critical management quantities and benchmarks to the substitution of age data products generated from the FT-NIR spectroscopy approach.

On behalf of the SIDT and other scientific contributors, I am proud to present the collected works and accomplishments made on the strategic initiative, particularly under challenging circumstances. The abstracts mentioned represent just a few of the works of our efforts to explore the application of FT-NIR spectroscopy in fisheries and ecological science, several of which I consider cutting edge. I believe the goals of the SI are achievable with good probability of success, but more work is needed to get us over the finish line.

2. Preliminary Analysis of Fourier Transform Near Infrared (FT-NIR) Spectroscopy and Convolutional Neural Network (CNN) Models to Predict Ages of Yellowfin Sole

Todd T. TenBrink, Irina M. Benson, Brenna C. Hsieh, Mary E. Matta,

and Thomas E. Helser

Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

ABSTRACT

We present a preliminary analysis of two alternative methods to the traditional microscopic approach to predict ages for yellowfin sole (*Limanda aspera*). Yellowfin sole is the most abundant flatfish (Family Pleuronectidae) along the eastern Bering Sea continental shelf and is the target of the largest flatfish fishery in the world. For this study, we used microscope-based reference ages from fish collected from 2012 to 2017 (age range: 1 to 43 years; n = 4,808) to test the utility of alternative age determination methods, Fourier transform near infrared (FT-NIR) spectroscopy using partial least squares regression (PLS) models and convolutional neural networks (CNN). The performance of the PLS models for predicting yellowfin sole ages resulted in r^2 values of 0.90, with root mean square errors (RSME) of approximately 2.2 years. The CNN handled the non-linearity of the data slightly better than PLS models, with r^2 values of 0.92 (RMSE = 1.97 years) and 0.91 (RMSE = 2.03 years) for the training and test datasets, respectively. The CNN's age predictions were comparable with the traditional microscope-based age estimates up to approximately 25 years and were most impacted by absorbance at specific wavelength ranges and by otolith weight.

INTRODUCTION

Effective fisheries management is the key to sustainability and long-term success of exploited fish populations (Melnychuk et al. 2021). Age composition data used in age-structured stock assessment models is a critical component of fish stock assessments. These ages are incorporated into assessments to inform key model parameters such as natural mortality, recruitment, and growth, and support the calculation of biological reference points, which are benchmarks used to evaluate the status of a stock. With an effort of broadening assessments of fish stocks to a more holistic approach that includes ecosystem management, climate effects are also being recognized for single-species assessments (e.g., Townsend et al. 2019). These effects may have negative consequences on the age and size structure of populations (Ohlberger et al. 2022). Impacts to stocks from both fisheries and climate change have resulted in an increased demand in both the quality and quantity of age composition data. This demand has led to an effort to pursue complementary or alternative methods to the traditional microscope approach of age estimation. Traditional ageing involves the use of collecting otoliths (ear stones), calcium carbonate structures located behind the brain in bony fishes, and enumerating annual bands or growth zones to estimate an age, often after some process of otolith preparation. In some species, it can be difficult to interpret these growth zones, and preparation methods do not always provide clarity of the age reading surface. The traditional approach is time consuming, potentially expensive, and requires experienced age readers.

Methods for automating the ageing process from otolith samples have been investigated. The use of machine-based technologies to streamline, increase the efficiency, and improve repeatability of age determination in fisheries science has recently been attempted. The approach

of using Fourier transform near infrared (FT-NIR) spectroscopy has been employed as a method to generate ages from otoliths of marine fish (e.g., Robins et al. 2015, Helser et al. 2019b, Passerotti et al. 2020b). FT-NIR spectroscopy is used in a wide variety of industries including pharmaceutical, chemical, petrochemical, and agricultural manufacturing. It has applications ranging from raw material identification to quality control monitoring and determination of product composition. FT-NIR spectroscopy itself is a vibrational spectroscopy technique based on the interaction of electromagnetic energy of a specific frequency range with the covalent bonds in the near infrared region. The bonds associated with different chemical functional groups (C–H, N–H, and O–H) absorb energy at unique and characteristic frequencies, and the relative amount of each functional group present is proportional to the amount of energy absorbed. Quantitative analysis of FT-NIR spectra is often performed using partial least squares regression (PLS) models, a classic linear method applied to spectral data and that one that works well on this type of data as it reduces the high dimensionality of an x-matrix of spectra down to its principal components (Norgaard et al. 2000). In addition to FT-NIR spectroscopy, the use of machine learning algorithms in the form of artificial neural networks has also garnered attention in fisheries science (e.g., Chen and Wang, 1999; Dempsey et al. 2020). Artificial neural networks are computational structures mimicking cerebral neural networks consisting of layers or computational units (Cook 2020). Algorithms of convolutional neural networks (CNN) were found to be useful for FT-NIR spectral data (Cui and Fearn 2018). Furthermore, recent studies have shown that CNN have been successful in the age prediction of fish using otolith images (Ordoñez et al. 2020). Benson et al. (2023) developed a multimodal convolutional neural network (MMCNN) capable of predicting fish ages using FT-NIR spectra and corresponding biological and geospatial data. Here, we present a study on the utility of FT-NIR spectroscopy

coupled with CNN models to predict ages from otoliths of yellowfin sole (*Limanda aspera*) with a comparison to reference ages from traditional methods.

METHODS

Species Description

Yellowfin sole is a medium-sized and moderately long-lived flatfish (Family Pleuronectidae) inhabiting eastern North Pacific waters from British Columbia, Canada, to the Chukchi Sea. In Alaska, it is the most abundant flatfish along the eastern Bering Sea continental shelf and is the target of the largest flatfish fishery in the world with annual catches exceeding 125,000 metric tons (Spies et al. 2022). Yellowfin sole is managed as a single stock in the Bering Sea-Aleutian Islands (BSAI) management area (Spies et al. 2022). Yellowfin sole is assessed annually through age-structured modeling. Each year, approximately 1,300 otoliths from yellowfin sole are aged by the Age and Growth Program of the Alaska Fisheries Science Center. This species has an age range from 1 to 43 years from a historic collection of approximately 54,000 ages (https://apps-afsc.fisheries.noaa.gov/refm/age/stats/max_age.htm). Ages are incorporated into the assessment of the yellowfin stock to support calculations of management reference points, such as acceptable biological catch (ABC) and overfishing level (OFL).

Reference Data and Spectral Acquisition

Otoliths from yellowfin sole specimens were collected by the Alaska Fisheries Science Center during the summer on bottom trawl research surveys of the eastern Bering Sea from 2012 to 2017 (Lauth et al. 2019). Reference ages from these specimens (n = 4,808) were determined by traditional methods following protocols in Matta and Kimura (2012). Otoliths were scanned and spectra acquired using a Bruker Tango-R spectrometer. Scanning of each otolith occurred after blotting dry glycerol-thymol preservative and recording otolith weights (nearest 0.001 g). Orientation and placement of each otolith was consistent for every sample following initial testing of different presentation methods. The effects of otolith side (blind and eyed) and accessories (apertures and stamps) applied to the Tango-R spectrometer scan (aperture) window were evaluated. It was determined that either otolith side placed within a chrome ring and covered with a flat bottom chrome stamp provided the best prediction results (Fig. 2.1). Spectra were collected on the Tango-R spectrometer primarily between 11,500 and 4,000 cm^{-1}. Samples were analyzed at a resolution of 16 cm^{-1} with 64 replicate scans.

Data Splitting and Preprocessing

Preprocessing of raw spectral data was a fundamental process subsequent to classification model building and involved several pretreatment methods to remove significant noise and other factors. Preprocessing was conducted using the chemometric software OPUS™ (version 7.8, Bruker Optics). Spectral data were preprocessed with a first derivative Savitsky-Golay transform (17 point smooth). To evaluate the extent of spectral variability across collection years and detect

outliers, we used principal component analysis (PCA) and Hotelling's T^2 (Rodionova and Pomerantsev 2020). Data splitting was conducted using the Kennard-Stone (KS) algorithm (Kennard and Stone 1969). This procedure maximized the Euclidean distance based on the importance of the principal components, measuring the most representative samples. The training and test datasets were split 80% to 20%, respectively. The training set was used to train the model, and included the full range of reference ages and spectral variation. The test set consisted of unobserved samples not included in the training set.

Modeling Approaches

For this study, we used two modeling approaches to predict ages of yellowfin sole. Spectral data and reference ages were fit using PLS models. Additionally, CNN was considered, as it has the potential to handle nonlinearity in data better than the traditional chemometrics approach (i.e., PLS; Benson et al. 2023). For the PLS models, Solo 8.7 software (Eigenvector Research, Inc., Manson, WA, USA) was used to select wavenumbers and generate the regression models. Wavenumber selection was completed using the interval PLS regression algorithm (Norgaard et al. 2000) with interval size 10 and automatic selection of the interval number. To implement the CNN algorithms, we employed Python 3.7.0 using TensorFlow v. 2.5.0 with implementation of the Keras API (i.e., application programming interface; Chollet 2015). To build the CNN model, input features of biological data (fish length) and geospatial data (gear depth, gear bottom temperature, latitude) during the study period (2012-2017) were introduced along with the spectral data. To interpret the predictions, the Python SHAP (SHapley Additive exPlanations) library was used to understand the importance of the variables in the trained model

for predictions (Lundberg and Lee 2017). Model predictive performances were compared using the coefficient of determination (r^2) and root mean square error (RMSE; Helser et al. 2019b).

To compare the accuracy between traditional microscope-based ages and those ages predicted by our modeling approaches, we reported the frequency of relative bias over all ages and mean bias by age class. Traditional bias was calculated as a difference between two trained age readers' microscope-based age estimates for the same specimen. Relative bias (B) was calculated as a difference between age predicted by the models (rounded to the nearest integer; $B^{MODEL} = Age^{MODEL} - Age^{Traditional}$) and traditional microscope-based age estimation ($B^{Traditional} = Age^{Reader1} - Age^{Reader\,2}$; Helser et al. 2019b).

RESULTS AND DISCUSSION

For the PLS models, r^2 values for both the training and test datasets were 0.90 (Table 2.1). The RMSE for the training and test datasets suggest the PLS models predicted ages within 2.2 years, based on an age range up to 43 years (Table 2.1). The non-linearity of the data, which might be the result of the measured spectral data associated with the otoliths, indicated that CNN might be a more robust method to deal with this relationship (Fig. 2.2). The CNN performed slightly better than PLS (Table 2.1; Fig. 2.3). CNN appears to be the better method for predicting yellowfin sole ages than PLS. The optimal CNN model resulted in a coefficient of variation (r^2) of 0.92 and 0.91 for the training and test datasets, respectively. The RMSEs associated with these datasets were approximately 2.0 years. Variable importance from the CNN model suggested that absorbance associated with 6,000 to 4,000 cm^{-1} wavenumbers had the highest influence on model predictions followed by otolith weight (Fig. 2.4). Otolith weight was highly correlated

with fish age, resulting in greater influence on the model than fish length. Given the results of CNN, the comparison between the traditional age estimates suggested that CNN had similar precision up to ages of 25 years (Fig. 2.5).

FT-NIR spectroscopy of otoliths coupled with machine learning techniques such as CNN can predict fish age with comparable precision to traditional age-based methods. Precision in this case was comparable ±1 years up to approximately 25 years, which represented >90% of the ages associated with the reference age dataset. Further use of CNN should be explored. Benson et al. (2023) demonstrated that integrating the entire range of wavenumbers of FT-NIR spectra and corresponding biological and geospatial data for use with CNN can predict ages of walleye pollock (*Gadus chalcogrammus*) comparable to classical PLS and yield good or slightly better precision than traditional microscope-based ages.

Yellowfin sole is currently considered a healthy stock and not in danger of being overfished; however, current environmental conditions in the eastern Bering Sea (e.g., elevated bottom temperatures observed cyclically) might impact fish growth and condition potentially altering the size and age structure of the population (Spies et al. 2022). In the stock assessment of yellowfin sole, the growth of males and females are modeled separately and inform assessment model parameters, such as natural mortality and recruitment. Furthermore, the use of annual observed population mean weight-at-age from the trawl survey allows for time-varying growth in the age-structured assessment model, given the observed correlation between bottom temperature and growth. Future analysis will need to account for these variables observed for yellowfin sole. If CNN are to be explored further, models separated by sex would be recommended.

Table 2.1. -- Modeling prediction results for yellowfin sole (*Limanda aspera*) age using the coefficient of determination (r^2) and root mean square error (RMSE) from partial least squares regression (PLS) and convolutional neural networks (CNN). Training and test numbers are based on a data-split of 80-20%.

Model	Dataset	Age range	*N*	r^2	RMSE
PLS	Training	1 - 43	3,851	0.901	2.251
	Test	2 - 34	957	0.903	2.156
CNN	Training	1 - 43	3,851	0.924	1.971
	Test	2 - 34	957	0.914	2.027

Figure 2.1. -- Yellowfin sole (*Limanda aspera*) otolith sample presentation on the scanning window of the Tango-R (Bruker) spectrometer. A chrome ring is inserted inside the cover mount with the otolith placed inside the chrome ring directly on the window. A chrome stamp is then placed over the sample prior to instrument scanning and spectral acquisition.

Figure 2.2. -- Model results from partial least squares regression (PLS) of the training and test datasets for predicting ages of yellowfin sole (*Limanda aspera*).

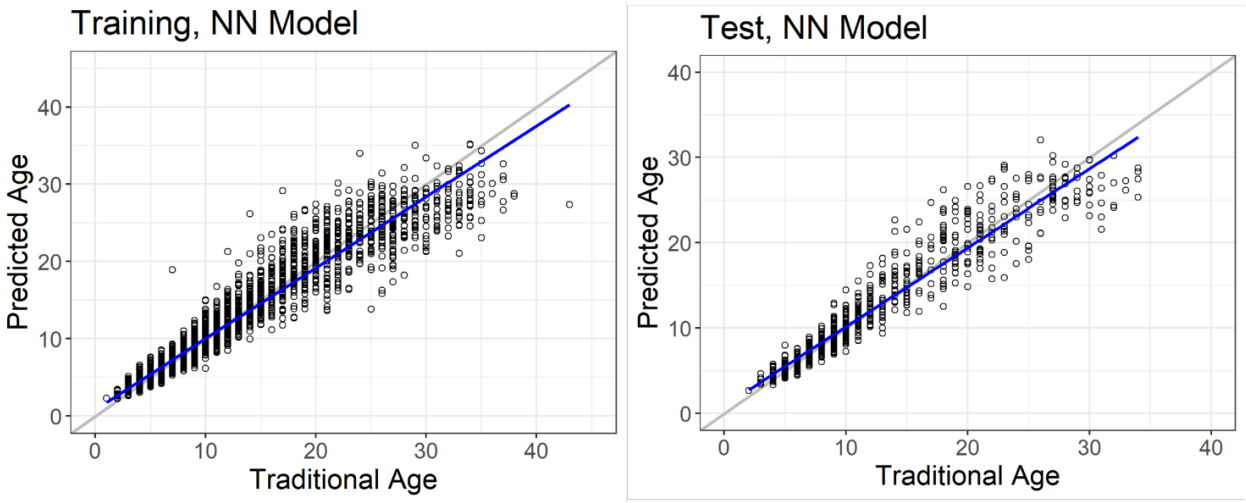


Figure 2.3. -- Model results from convolutional neural networks (CNN) of the training and test datasets for predicting ages of yellowfin sole (*Limanda aspera*).

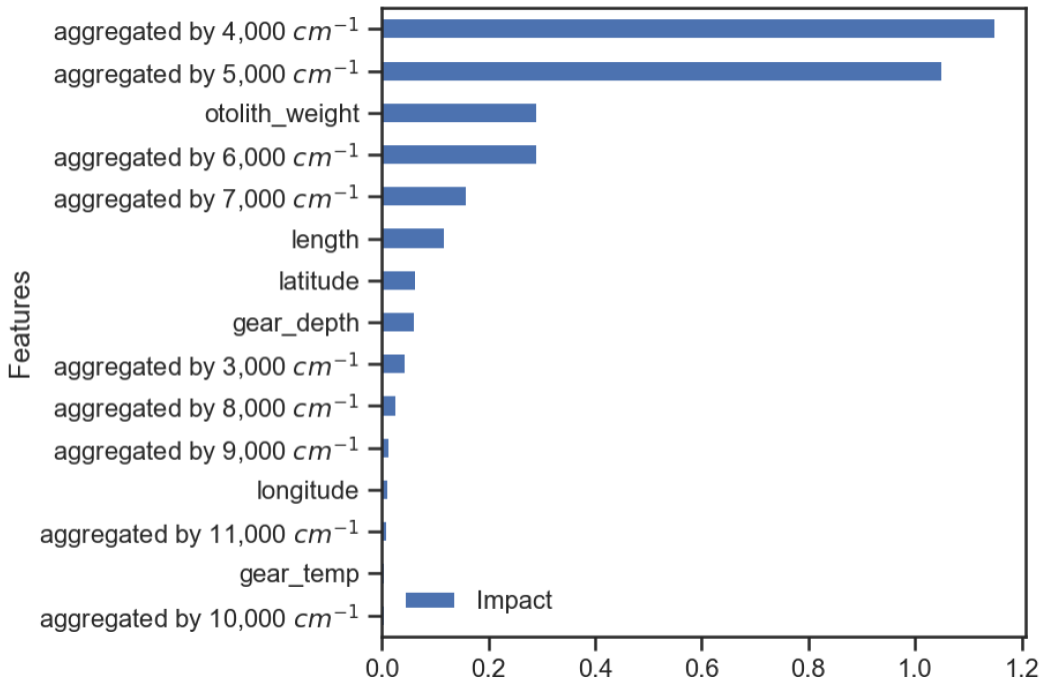


Figure 2.4. -- The relative importance of convolutional neural network (CNN) input features for predicting ages of yellowfin sole (*Limanda aspera*). Wavenumbers are aggregated by 1,000 cm^{-1}. The y-axis indicates the feature name in order of importance from top to bottom. The x-axis indicates the average of the absolute Shapley value of each feature.

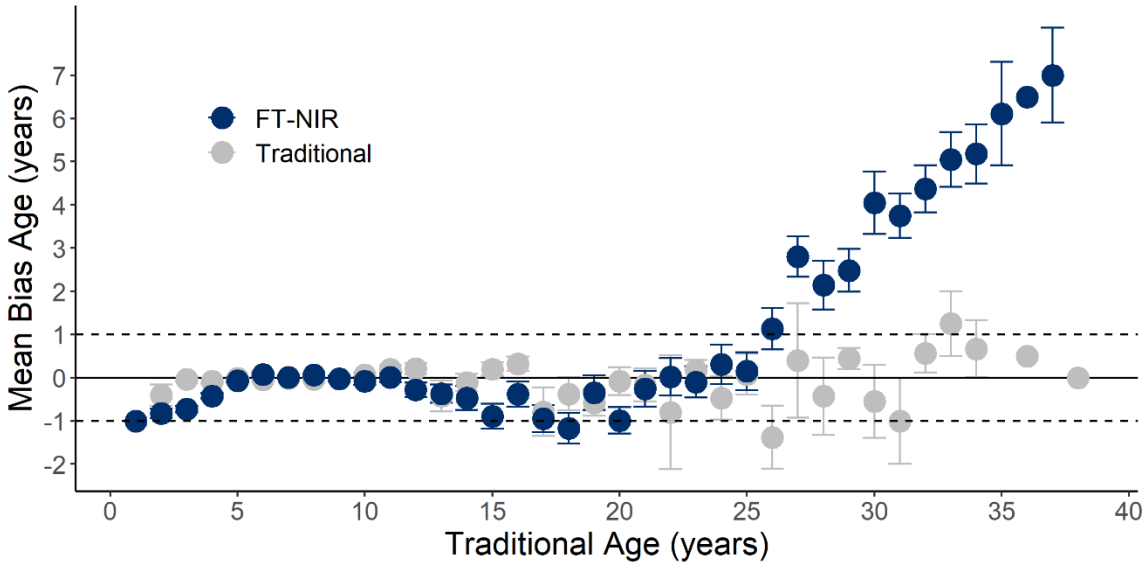


Figure 2.5. -- Frequency of mean bias by age class between the convolutional neural network (CNN) based on FT-NIR spectral data and the traditional microscope-based ages for yellowfin sole (*Limanda aspera*). The error bars show one standard error interval around the mean. The solid line represents a 100% agreement line, while the dashed line represents ±1 year.

3. A Preliminary Analysis and Review of FT-NIR Spectroscopy and DNA CpG Site Methylation for Fish Age Prediction

Laurel S. Lam[1,2,] Thomas E. Helser[3], Krista M. Nichols[2], Benjamin T. Mayne[4],

Irina M. Benson[3], Brenna C. Hsieh[3], Oliver F. Berry[4], Simon N. Jarman[5],

and Kevin W. McNeel[6]

[1] Pacific States Marine Fisheries Commission
205 SE Spokane Street, Suite 100
Portland, OR USA

[2] Northwest Fisheries Science Center
NOAA, National Marine Fisheries Service
2725 Montlake Boulevard E
Seattle, WA USA

[3] Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

[4] The Commonwealth Scientific and Industrial Research Organisation (CSIRO)
54 Fairway
Crawley, Western Australia AUSTRALIA

[5] School of Biological Sciences
The University of Western Australia
54 Fairway
Crawley, Western Australia, AUSTRALIA

[6] Age Determination Unit
Alaska Department of Fish and Game
1255 W 8th Street
Juneau, AK USA

ABSTRACT

Production ageing of groundfishes for stock assessments relies on validated methods that have high accuracy and precision, and are scalable to large sample sizes. Traditional methods are the most reliable but increasingly inefficient in the face of growing stock assessment demands. Lingcod (*Ophiodon elongatus*) have traditionally been aged using fin rays, a process that is both time intensive and costly in terms of materials, space, and personnel, yet has the highest accuracy, readability, and minimal between-reader bias compared to other age structures. However, emerging technologies that are cheaper, faster, and comparable in accuracy and precision are showing promise. Here, we assessed the performance of two alternative, model-based methods of predicting lingcod age, DNA methylation and Fourier transform near infrared (FT-NIR) spectroscopy. Average coefficient of variation (ACV), percent agreement, and bias from the two alternative methods were compared to traditional methods of counting annuli using fin rays and otoliths (break and burn). Results demonstrated that ages estimated by DNA methylation and FT-NIR spectroscopy have slightly higher ACV but are in close agreement (greater than 95% agreement within 2 years) with fin ray and otolith-determined ages. Between the two alternative methods, the FT-NIR spectroscopy model performed slightly better than the DNA methylation model. Considering the potential for scalability, increased efficiency, and reduced cost, these methods have significant potential to be used in production ageing for lingcod stock assessments.

INTRODUCTION

Lingcod (*Ophiodon elongatus*) are a demersal top predator found along the U.S. West Coast, and play an important role ecologically in nearshore and shelf-slope habitats, and economically in commercial and recreational fisheries. The coastwide lingcod stock is federally managed as part of the Pacific Coast Groundfish Fishery Management Plan (FMP), where stock assessors use age-structured models to estimate population abundance, stock productivity, and future harvest goals. Fishery-independent and -dependent lingcod age data throughout their range are vital in age-structured stock assessments models and are provided by federal and state agencies dedicated to the production ageing of groundfish for management purposes.

Lingcod are traditionally aged using hardened and cross-sectioned fin rays from the second dorsal fin (Beamish and Chilton 1977), a method that has been validated using known-age fish and consistently has the highest accuracy, readability, and minimal between-reader bias compared to the use of other hard structures (e.g. otoliths, scales, vertebrae) (McFarlane and King 2001, Claiborne et al. 2014). In spite of this, the fin ray method is also the most time intensive and costly in terms of materials, space, and personnel due to the extensive preparation process needed before the fin ray can be aged under a microscope. This preparation process includes drying the fin, hardening the fin using cyanoacrylate glue, sectioning the fin into equal lengths, and mounting the sections onto slides (Beamish and Chilton 1977). During production ageing for an assessment year, upwards of 1,000 lingcod fin rays are required and can take a dedicated age team between 3-6 months, depending on the number of personnel and lab space available (N. Atkins, PSMFC, pers. comm.). In comparison, ageing the same number of otoliths (break and burn) takes 1-1.5 months, though with varying accuracy and precision depending on

geographic location (K. McNeel, ADFG, unpublished report). In recent years, alternative, model-based methods of ageing fish using DNA methylation and Fourier transform near infrared spectroscopy (FT-NIR) spectroscopy have been explored across a variety of fish taxa with promising results. DNA methylation and FT-NIR spectroscopy have broad applications in medicine, pharmaceuticals, agriculture, and food production; however, their applications to the field of fisheries and resource management is relatively new with respect to production ageing for stock assessment purposes. Here, we examined the performance of these two model-based ageing methods on a commercially harvested, temperate marine fish from the North Pacific Ocean.

DNA methylation is a key epigenetic mechanism associated with aging and refers to the addition of a methyl group to cytosine residues within cytosine-phosphate-guanine (CpG) sites. Generally, there is an age-dependent hypermethylation of specific CpG sites, also known as "clock sites". Because DNA methylation changes are clock type (i.e., they follow a certain rate and affect certain loci), methylation of specific loci can be isolated to create a model, or genetic clock, that can be used to age certain species (Bird 1993, Jung and Pfeifer 2015). DNA methylation can be readily measured using DNA sequencing, making it an attractive option to be used for wild populations where large sample sizes are often needed (Heather and Chain 2016). Methods in obtaining DNA samples (via skin, tissue, or blood) can also be non-lethal, aiding in conservation efforts for protected species (Mayne et al. 2021). Moreover, biomarkers for age have been shown to be transferable across distantly related taxa (e.g., mammals, reptiles, birds, and fish), thereby increasing efficiency for identifying, isolating, and sequencing optimal CpG sites and eliminating the need to sequence entire genomes (De Paoli-Iseppi et al. 2017, Mayne et al. 2020, Lu et al. 2023).

Near infrared spectroscopy, widely used in agriculture, pharmaceutical and petrochemical industries, and most recently in terrestrial ecology (Vance et al. 2014) and fishery science (Helser et al. 2019b), functions by exciting covalent bonds (O-H, C=O, C-H, C-N, and N-H) at the molecular level using NIR light and measuring absorbance in the wavenumber range 4,000 to 12,500 cm^{-1}. The light interaction with the sample results in measurable vibrational frequencies represented by spectral signatures associated with molecular combinations and overtones that make up compounds in the sample (Conzen 2014, Siesler et al. 2002). Otoliths are composed of alternating layers of calcium carbonate and otolin, an organic protein structure that comprises 4-5% of an otolith and accumulates as fish age (Campana 1999). In otoliths, these spectral signatures derived from their molecular composition are a proxy for fish age. The use of FT-NIR spectroscopy of otoliths to estimate fish age has been reported for a number of species (Wedding et al. 2014, Helser et al. 2019b, Passerotti et al. 2020b). This approach is based on generating a predictive model between the spectra of the ageing structure and fish age, which, once calibrated and validated, is able to produce age estimates at many times the rate compared to traditional methods (Helser et al. 2019b).

The objective of this study was to 1) develop an epigenetic clock for lingcod using known age-associated CpG sites in zebrafish (*Danio rerio*), 2) develop a predictive model between otolith spectra from FT-NIR spectroscopy and reference ages to age lingcod, and 3) compare the performance of these two alternative methods of predicting lingcod age (DNA methylation and FT-NIR spectroscopy) with two traditional methods (fin rays and otoliths) using metrics commonly used in age structure comparisons studies: percent agreement, average percent error, and average coefficient of variation. Our goal was to investigate the application of FT-NIR

spectroscopy and DNA methylation as alternative methods of production ageing lingcod for stock assessment purposes.

METHODS

Sampling Methods

Lingcod fin rays and otoliths from California were collected between 2015 and 2017 for a total of 735 paired samples. Otoliths were initially scanned by the Alaska Fisheries Science Center (AFSC) using FT-NIR spectroscopy, then sent to be aged microscopically by the Alaska Department of Fish and Game (ADFG) Age Determination Unit (ADU) in Juneau, Alaska. Lingcod gill tissue was collected for DNA extraction (n = 206) from fish caught off southeast Alaska to the U.S.-Mexico border (excluding Canadian waters). Of the 206 individuals with genetic material, 70 also had otoliths, collected making for a maximum of 70 paired samples. Two hundred and two individuals out of the 206 fish in the genetic subset had paired fin rays collected.

Traditional Method: Fin Ray

Prior to age estimation, fin rays were dried, hardened using cyanoacrylate glue, cut into 2 mm sections using a high-speed saw, and mounted onto microscope slides (Beamish and Chilton 1977), a process that can take up to seven days for a batch of 100 fin rays. Ages were determined by counting the number of annuli under a compound microscope (40× magnification) using

transmitted light. Early annuli can be difficult to discern or become resorbed, but can be verified using known mean radii measurements for the first, second, and third annuli (Cass and Beamish 1983, McFarlane and King 2001, Laidig et al. 2001). These methods of fin ray preparation and age estimation follow Beamish and Chilton (1977) and the Committee of Age Reading Experts manual (CARE 2006), and are currently used by the Washington Department of Fish and Wildlife, Oregon Department of Fish and Wildlife, and Northwest Fisheries Science Center.

Traditional Method: Otolith (Break and Burn)

The ADFG ADU has used sagittal otoliths to estimate lingcod age since 1995. Using the standard break and burn protocol (Williams and Bedford 1974), otoliths were broken through the core to produce two halves. The anterior half of the left otolith is preferred, and the final age estimate is determined from the ventral side of the cross-section. Annual growth bands are counted using reflected light on a stereoscopic microscope with standard 0.63-6.0X magnification with a lower powered eyepiece (16X). Location of the first annulus is verified using a size criterion derived from measurements of otoliths from young-of-the-year and one-year-old lingcod (unpublished ADFG manuscript obtained from K. McNeel, Age Determination Unit, Juneau, Alaska).

Alternative Method: DNA Methylation

Using lingcod DNA sequenced from gill tissue and known age-associated CpG sites from zebrafish (*Danio rerio*), we used a genome pairwise alignment method to identify conserved age-

associated CpG sites in the lingcod genome. These sites were targeted for primer design in the multiplex polymerase chain reaction (PCR) assay (Mayne et al. 2020). Multiplex PCR was designed for primers to target and isolate conserved age-associated CpG sites using PrimerSuite (Lu et al. 2017) and to measure DNA methylation at these sites. CpG sites were sequenced using designed primers and a model was fitted to predict age from DNA methylation. To generate the model to predict age from DNA methylation, samples were randomly assigned to either a training or testing data set with a 70/30 split, respectively. Age in years was log transformed prior to fitting the data with an elastic net linear regression model (Friedman et al. 2010). The glmnet function in the glmnet R package was used to apply the elastic net regression model and was set to a 10-fold cross validation and the α-parameter was set to 0.5, similar to other studies (Horvath 2013, Mayne et al. 2020). The testing data set was used as a validation of the model and to test for accuracy. The performance of the model was measured using Pearson correlations and mean absolute errors (MAE). A similar Pearson correlation and MAE between training and testing datasets indicates similar performance, a lack of overfitting, and high reproducibility (Mayne et al. 2020). We also determined the performance of the model within age ranges to determine if the model was biased to specific age classes.

Alternative Method: FT-NIR Spectroscopy

For spectroscopic analysis, unaltered whole otoliths were removed from their vials and gently cleaned with Kimwipes to remove contaminants before placement on the 22 mm sample window of a Bruker MPA II FT-NIR spectrometer. Otoliths were covered with a gold stamp and analyzed using diffuse reflectance on an integrated sphere and spectra collected at 16 cm^{-1}

resolution at 8 wavenumber intervals with 64 co-scans, scanner velocity of 7.5 kHz and measuring absorbance values from 12,500 to 4,000 cm^{-1}. Spectral data were pre-processed using 1st derivative and a Savitsky-Golay (2nd order polynomial, 17 points) smoothing filter. Spectral variability across collection years and detection of outliers was evaluated using principal component analysis (PCA) and measures such as Hotelling's T^2 and Q-statistic. Helser et al. (2019b) reported that scan time to acquire otolith spectra ranges between 40-60 seconds. Spectral regions that exhibited the highest correlation with changes in the reference values (age) were isolated using PCA. We used the chemometric software Solo 8.7 (Eigenvector Research, Inc., Manson, WA, USA) for exploratory analysis using PCA, the subsequent data split into training and test data sets for data processing and model generation, and to select the optimal model and used partial least squares (PLS) regression (Chen and Wang 2001) to develop a quantitative NIR model from the reference values (fin ray ages and otolith ages, respectively) to predict age. To cross validate the model, each calibration sample is temporarily removed from the dataset, then a PLS model is created from the remaining samples and the sample that was temporarily removed is predicted as an unknown. The difference between the reference value and the predicted value is determined, then the sample is returned to the dataset. This is repeated for each sample. Model performance and robustness was determined using the coefficient of determination (r^2), which assesses predictive accuracy of the model, and the root mean square error of cross validation (RMSECV), an indicator of model accuracy (Helser et al. 2019b).

Ageing Error

Ageing error was quantified and compared between traditional and alternative methods (fin rays vs. DNA methylation, fin rays vs. FT-NIR spectroscopy, otoliths vs. DNA methylation, otoliths vs. FT-NIR spectroscopy). Ageing bias was examined graphically using age bias plots and statistically using symmetry tests to identify systematic biases, such as over- or under-ageing, and where these biases may be occurring (e.g., in older ages or younger ages) (Campana et al. 1995). Precision, or repeatability, was quantified using percent agreement, average percent error (APE), and average coefficient of variation (ACV) (Campana 2001). APE and ACV are functionally interchangeable when there are only two age readers, but ACV is preferred when there are more than two readers as ACV is based on the standard deviation divided by the number of times the fish was aged. It is considered more accurate than APE as variance is a better estimator of precision than the absolute difference between reads and tends to be more unbiased and consistent (Chang 1982). While the median ageing error across studies is an ACV of 7.6 (or a 5.5 APE), there is no acceptable standard value of precision as readability is influenced by the species and the structure being aged (Campana 2001).

RESULTS

Epigenetic Clock Model Performance

We found 11 CpG sites out of the initial 29 zebrafish CpG sites were needed as the minimum number of predictors with the highest performance for use in the lingcod epigenetic model. Two separate models were created based on the reference dataset used to calibrate the

model, one using fin ray ages as reference (n = 202) and one using otolith ages as the reference (n = 68). A high correlation was found in the training data set between the fin ray reference ages and predicted ages (Fig. 3.1A, Pearson correlation = 0.93, $p < 2.20 \times 10^{-16}$). The high correlation was also maintained in the testing data set (Fig. 3.1B, Pearson correlation = 0.92, $p < 2.20 \times 10^{-16}$). This suggests a lack of overfitting as the training and testing data set had similar performance. This was further demonstrated with no statistical difference found between the absolute error rates between the training (median = 1.06 years) and the testing (median = 1.14 years) data sets (Fig. 3.1, $p = 0.868$, *t*-test, two-tailed). No difference in absolute error rates was found between sexes in both the training ($p = 0.1806$, *t*-test, two-tailed) and testing data sets ($p = 0.544$, *t*-test, two-tailed). The otolith-referenced epigenetic clock similarly had high correlation in the training data (Fig. 3.2A, Pearson correlation = 0.79, $p < 2.10 \times 10^{-11}$) and the testing data set (Fig. 3.2B, Pearson correlation = 0.83, $p < 1.30 \times 10^{-5}$). There was no statistical difference found between the absolute error rates between the training (median = 1.34 years) and the testing (median = 1.14 years) data sets (Fig. 3.2, $p = 0.455$, *t*-test, two-tailed).

FT-NIR Spectroscopy Model Performance

Optimal models from PLS regression ranged from 7 to 9 latent variables with the most informative region for age prediction being within the 7,000 to 4,000 cm^{-1} wavenumber region. Similar to the epigenetic models above, two models were generated based on the reference data set used. The FT-NIR spectroscopy age prediction model using the fin ray ages as reference (n = 735) yielded high performance, with 93% of the variability in estimated ages explained by the model where 68% of estimated ages should fall within 0.82 years of expected age ($r^2 = 0.93$,

RMSECV = 0.82 years). The model using otolith ages as reference (n = 70) was comparable in predictive accuracy, with 90% of the variability in estimated ages explained by the model and 68% of estimated ages falling within 1.2 years of expected age (r^2 = 0.90, RMSECV = 1.2 years).

Comparison of Relative Accuracy/Precision Between Traditional and Alternative Methods

Bias plots, percent agreement, and ACV were determined for each ageing method pair: fin rays and epigenetics (fin ray reference); fin rays and FT-NIR spectroscopy (fin ray reference); otoliths and epigenetics (otolith reference); and otoliths and FT-NIR spectroscopy (otolith reference). Between the epigenetic method and FT-NIR spectroscopy (using fin ray reference), the FT-NIR spectroscopy predictive model had higher percent agreement (50% agreement, 92.4% agreement within one year, 95% agreement within 2 years) and slightly lower error compared to the epigenetic predictive model (37.1% agreement, 85.6% agreement within one year, 98.8% agreement within 2 years) (Table 3.1). The fin ray referenced FT-NIR spectroscopy model showed less bias and no evidence of systematic under- or over-ageing (Fig. 3.3) compared to the epigenetic model.

Between the epigenetic method and FT-NIR spectroscopy using the otolith reference, the FT-NIR spectroscopy predictive model similarly had higher percent agreement (35.7% agreement, 80.0% agreement within one year, 95.7% agreement within two years) and lower error compared to the epigenetic predictive model (20.6% agreement, 58.8% agreement within one year, 86.85 agreement within two years) (Table 3.1). Both the epigenetic and FT-NIR spectroscopy models appeared to be slightly biased towards over-ageing rather than under-

ageing; however the FT-NIR spectroscopy model outperformed the epigenetics model overall (Fig. 3.3).

DISCUSSION AND CONCLUSIONS

Understanding the applications, benefits, and limitations of alternative, model-based methods in the production ageing of groundfish is important for the future of stock assessments and fisheries management agencies. While thousands of ageing structures across multiple groundfish species are collected in fisheries-independent and -dependent surveys, only a fraction are regularly aged due to resource constraints, such as time and personnel. In this study, we demonstrated that DNA methylation and FT-NIR spectroscopy appear to be promising alternative methods of ageing lingcod with accuracy and precision comparable to traditionally used methods and thus have the potential to greatly increase ageing throughput and efficiency during time sensitive assessment years.

There are many benefits to using DNA methylation for ageing, namely that epigenetic clocks are not influenced by temperature as somatic growth tends to be (Anastasiadi and Piferrer 2019), and methods are continually refined, speeding up the initial model-building process across diverse taxa (Mayne et al. 2020). Here, an epigenetic clock was developed for lingcod, further illustrating that conserved CpG sites between evolutionary divergent species (e.g., zebrafish) still maintain age-association. Both fin ray and otolith reference datasets were able to generate high-performing predictive models given the available reference data. The model using otoliths had higher absolute error, possibly due to smaller sample size compared to the fin ray referenced model. While both models could have benefitted from more thorough sampling of older age

classes, as model performance was still quite high, increasing sample sizes might not result in a dramatic increase in r^2. This highlights the limitations of utilizing machine-learning based models. While epigenetic ageing has the advantages of being non-lethal and high throughput, the model cannot be extrapolated to ages outside of its calibrated range (Mayne et al. 2021). This is the case with other machine learning based models where performance is often reduced outside of calibration data. Unlike otoliths where annual growth rings may be consistent throughout the life of the fish, epigenetic ageing is limited to the original model's training data. Additionally, the model's performance decreased with increasing age as younger ages are typically more available than older ones, especially in fished populations (Beamish et al. 2006).

The use of FT-NIR spectroscopy in ageing fish for stock assessments has been investigated in temperate and tropical species (e.g., walleye pollock, red snapper, barramundi) for purposes of production ageing and has yielded promising results (Helser et al. 2019b, Passerotti et al. 2020b, Wright et al. 2021). FT-NIR spectroscopy is advantageous for being rapid and non-destructive to the ageing structure and is highly repeatable with as good or better precision to that of traditional methods. Moreover, recent developments illustrate FT-NIR spectroscopy coupled with machine learning can improve model prediction (Benson et al. 2023). Both FT-NIR spectroscopy models created here for lingcod were comparable in performance, though the fin ray calibrated model had higher accuracy and lower bias. Overall, both FT-NIR spectroscopy models were higher in precision than the epigenetics models in terms of ACV and percent agreement. However, as FT-NIR spectroscopy is also a secondary method based on predictive analytic models, it is similarly limited due to the availability and accuracy of the reference data used to calibrate the model. Also, both alternative methods may require predictive model recalibration to account for unseen variability or when factors affecting biological

processes or the underlying biochemical mechanisms change. In such cases, reference data using traditional microscopic methods will continually need to be collected (though at a much smaller percentage) to monitor changes in precision or drift from non-stationary processes, such as fishing pressure, changing ocean conditions, and spatio-temporal variability (Heino et al. 2015, Lam et al. 2021, Fennie et al. 2023).

Between the two alternative methods investigated here, DNA methylation and FT-NIR spectroscopy, FT-NIR spectroscopy may at present be a more efficient approach for high throughput, production ageing requirements for groundfish stocks that are non-threatened and where otoliths are regularly collected from survey and fishery sources, though this can change in the future as epigenetic ageing methods improve in efficiency or as stock statuses change. The FT-NIR spectroscopy model using fin ray reference data had the highest percent agreement and lowest error among all comparisons and would be the most promising alternative method for use along the U.S. West Coast; however, traditional, microscopic ageing will still serve as an indispensable primary method to monitor consistency and precision in the reference data should predictive model recalibration be necessary.

Table 3.1. -- Precision between traditional age determination methods (fin ray and otolith examination) and alternative age determination methods (epigenetics and Fourier transform near infrared spectroscopy, FT-NIRS) for lingcod (*Ophiodon elongatus*) was calculated using average coefficient of variation (ACV), average percent error (APE) and percent agreement with exact agreement (% Agreement). While the median ageing error across studies is an ACV of 7.6 (or a 5.5 APE), there is no acceptable standard value of precision as readability is influenced by the species and aged structure (Campana 2001). Percentage of age estimates with agreement within one year (± 1 yr) and within two years (± 2 yr) are also shown.

	n	ACV	APE	% Agreement	± 1 yr	± 2 yr
Fin ray vs. Epigenetics	202	9.0	6.4	37.1	85.6	95.0
Fin ray vs. FT-NIRS	735	8.8	6.2	50.0	92.4	98.8
Otolith vs. Epigenetics	68	12.8	9.0	20.6	58.8	86.8
Otolith vs. FT-NIRS	70	8.3	5.9	35.7	80.0	95.7

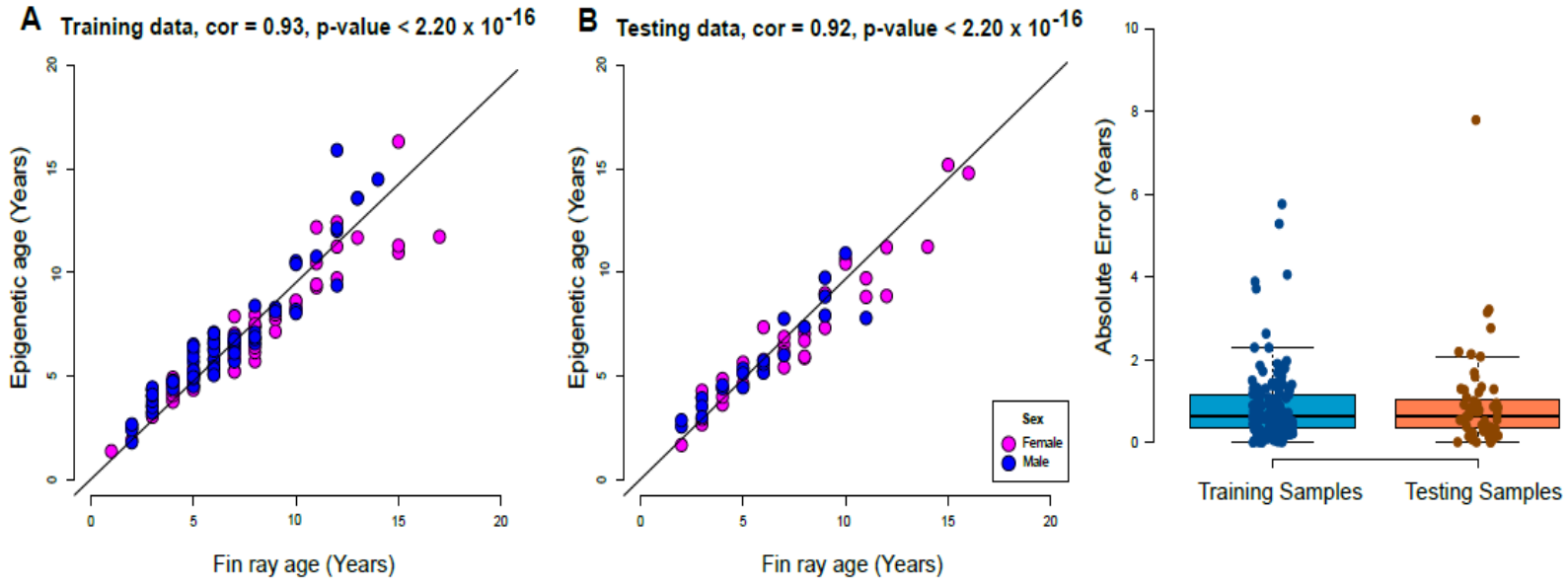


Figure 3.1. -- Performance of the epigenetic model using lingcod (*Ophiodon elongates*) fin ray reference ages was measured by comparing Pearson correlations and mean absolute errors between the A) training data set and B) testing data set. The 1:1 age agreement line is shown as a black diagonal line on Figure 3.1A and 3.1B. Significant *p*-value ($p < 0.05$) indicates that fin ray ages and predicted ages from the epigenetic model are correlated. Similar Pearson correlation and mean absolute error between training and testing datasets indicates similar performance between training and testing data sets.

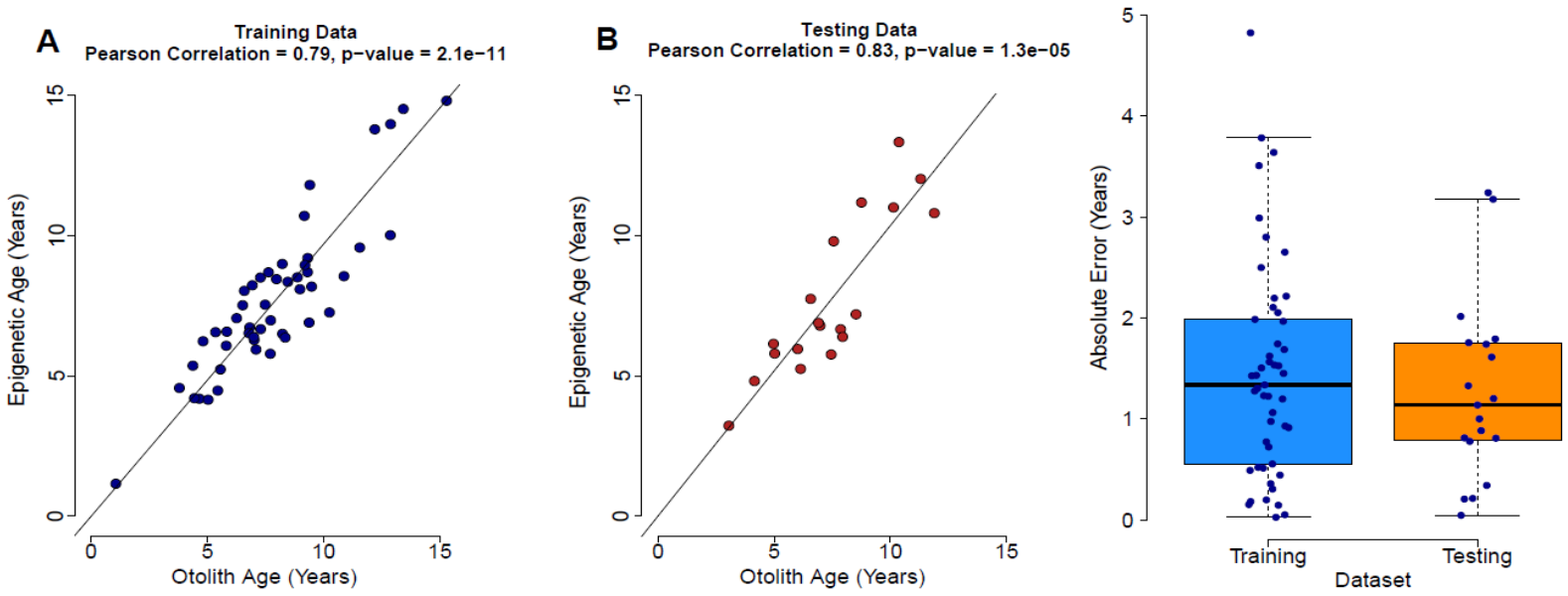


Figure 3.2. -- Performance of the epigenetic model using lingcod (*Ophiodon elongates*) otolith reference ages was measured by comparing Pearson correlations and mean absolute errors between the A) training data set and B) testing data set. The 1:1 age agreement line is shown as a black diagonal line on Figure 3.1A and 3.1B. Significant *p*-value ($p < 0.05$) indicates that fin ray ages and predicted ages from the epigenetic model are correlated. Similar Pearson correlation and mean absolute error between training and testing datasets indicates similar performance between training and testing data sets.

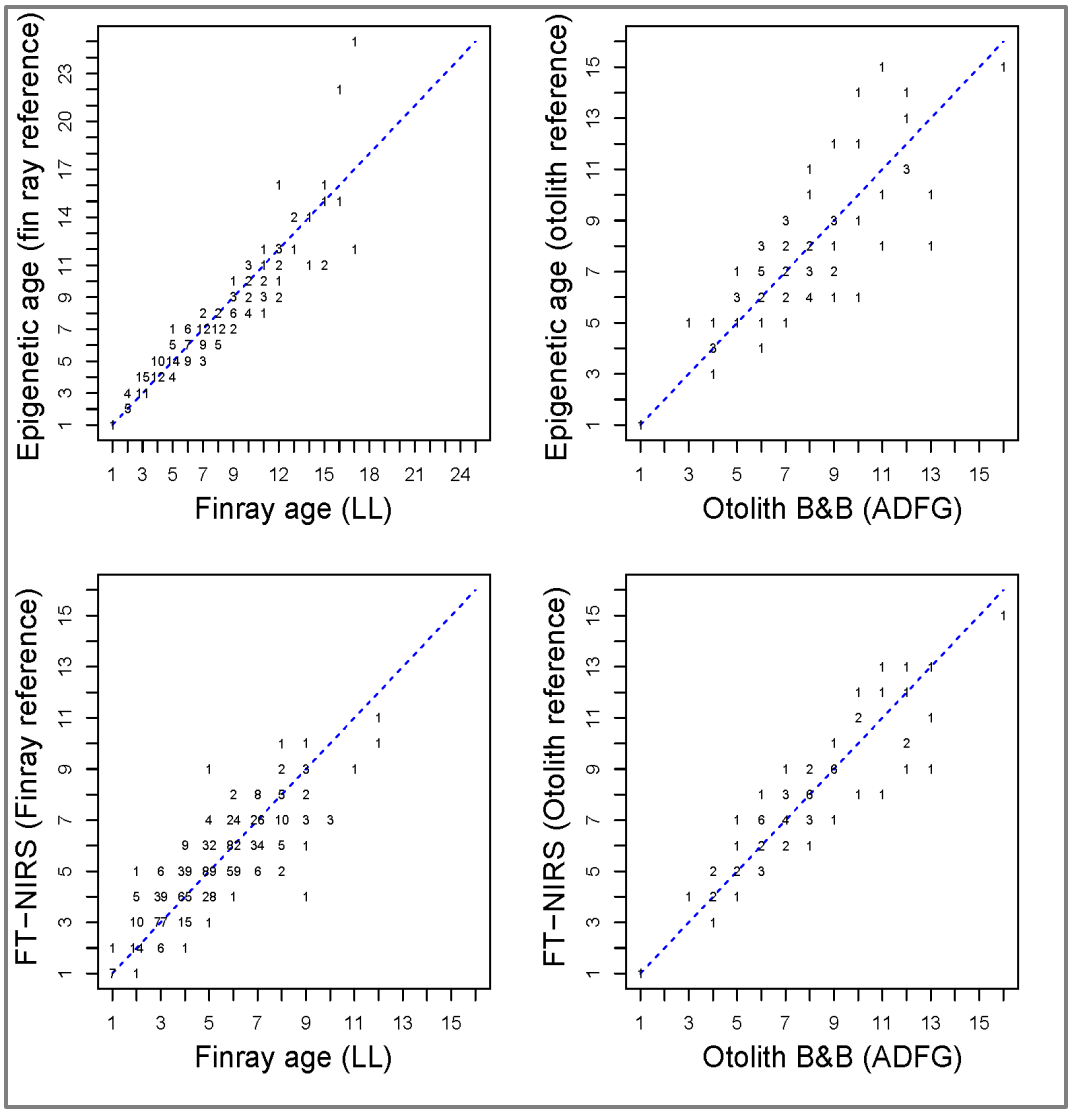


Figure 3.3. -- Age agreement plots between traditional lingcod (*Ophiodon elongates*) age determination methods (fin ray and otolith examination) and alternative age determination methods (epigenetics and FT-NIRS) to identify systematic under- or over-ageing and symmetry. The 1:1 age agreement line is shown as a blue dashed diagonal line. Plotted numbers along the line represent the frequency of paired samples that had 100% age agreement. Numbers above and below the 1:1 agreement line indicate the frequency of samples where age estimates differed. In a symmetric age-agreement table, disagreements in age would be expected to fall randomly on either side of the 1:1 line.

4. Applying FT-NIRS Predictive Ageing to Different Genetic Stocks of White Grunt (*Haemulon plumierii)* Within the U.S. South Atlantic

Jamie M. Clark [1, 2], Andy D. Ostrowski [1], Jennifer C. Potts [1], Beverly K. Barnett [3], Steven B. Garner [2, 3], and Walter D. Rogers [1, 2]

[1] Southeast Fisheries Science Center
NOAA, National Marine Fisheries Service
101 Pivers Island Road
Beaufort, NC USA

[2] Cooperative Institute for Marine and Atmospheric Studies
University of Miami
4600 Rickenbacker Causeway
Miami, FL USA

[3] Southeast Fisheries Science Center
NOAA, National Marine Fisheries Service
3500 Delwood Beach Road
Panama City, FL USA

INTRODUCTION

Age data are an essential component of modern fisheries stock assessments because they are utilized to estimate growth, longevity, maturation, mortality, track cohorts, and characterize landings compositions to understand population dynamics of state and federally-managed fisheries. The Southeast Fisheries Science Center collectively receives up to 85,000 otolith samples annually and is responsible for providing age data to three different management councils in multiple regions. Otoliths (inner ear stones) are the primary fish hard part collected to estimate fish age, but traditional ageing methods are time-consuming because samples require processing, sectioning, and visually counting the incremental annuli using microscopes. Thus, the process of production ageing requires a large investment of time and staff resources to provide data for stock assessment. Fourier transform-near infrared (FT-NIR) spectroscopy has recently been applied to fish ageing with the potential to dramatically increase ageing efficiency without any loss in productivity. NOAA Fisheries staff are continuing to evaluate the feasibility of FT-NIR spectroscopy for accurately predicting fish ages compared to traditional ageing methods.

FT-NIR spectroscopy has been utilized in the pharmaceutical, chemical, and agricultural industries for decades, but has only recently been applied to fish ageing (Robins et al. 2015; Helser et al. 2019b). FT-NIR spectroscopy is a non-destructive method that transmits near infrared light through a sample, thereby exciting its chemical bonds. Absorbance measurements are collected from the scanned object to form a spectral signature, which quantifies the presence and concentration of organic bonds and is visualized as a continuous spectrum across the relevant wavelength (cm^{-1}) range (i.e., the spectral signature). Peaks in spectral absorbance relate

to the chemical composition of the sample in a consistent way. Thus, when spectral data are paired with a representative sample of observed ages, a predictive model can be built to predict age from spectral data alone without the need for sectioning and ageing thousands of otoliths. However, this technique has only been applied to a handful of species and requires further investigation. Accuracy, potential bias, and the frequency of model updating must be evaluated before age estimates predicted with FT-NIR spectroscopy can be used at a production scale for stock assessment.

We examined white grunt (*Haemulon plumierii*) as a candidate species for FT-NIR spectroscopy because it has relatively high precision among traditional age estimates, adequate sample availability, and an upcoming stock assessment scheduled in 2027. White grunt is a reef-associated coastal fish found in the western Atlantic from the Chesapeake Bay south through the Gulf of Mexico and the Caribbean to the coast of Brazil. Previous studies have identified four genetically distinct stocks of white grunt: (1) the Carolinas, (2) Southeast Florida and the Florida Keys, (3) the eastern Gulf of Mexico, and (4) the Caribbean Sea (O'Donnell et al. 2019). Age data indicate that fish from Southeast Florida grow much slower and attain smaller sizes than fish collected from North Carolina and South Carolina (Potts & Manooch 2001). Fish from the Gulf coast of Florida reach comparable size-at-age to fish from the Atlantic coast of Florida, yet both are smaller at age than fish from the Carolinas (Murie and Parkyn 2005). White grunt supports important U.S. commercial and recreational fisheries (Potts 2000), but there has never been a stock assessment using age-based methods. Additionally, previously planned assessments have been delayed for various reasons, one being uncertainty in stock definition (O'Donnell et al. 2019). The application of FT-NIR spectroscopy to ageing white grunts could provide critical age

data to support the upcoming assessment and efficiently provide age data in the future, given the 26,033 otolith samples currently awaiting processing at the SEFSC Beaufort laboratory.

The objectives of this study were to examine the efficacy of FT-NIR spectroscopy for 1) predicting age of white grunt and 2) predicting age for genetically distinct stocks of white grunt in the U.S. South Atlantic (i.e., the Carolinas and South Florida). Our specific goals were to (1) develop standardized practices for FT-NIR spectroscopy across various NOAA labs; (2) assess the overall efficacy of FT-NIR spectroscopy (i.e., accuracy and precision) by comparing ages predicted with FT-NIR spectroscopy models to observed ages for white grunt; and 3) determine if single or separate age prediction models are required for estimating ages of white grunt stocks in the Carolinas and South Florida.

METHODS

In total, 1,930 Atlantic white grunt otoliths (left otoliths were mostly used; right otolith was used if left was missing or damaged) were weighed, scanned with a TANGO-R spectrometer (Bruker), sectioned on a Hillquist high-speed saw, and aged under a Nikon stereo-microscope with transmitted light. Fish were assigned to each genetic stock based on landing location because the majority of samples were fishery-dependent and fishing location was not known. Assigning fish to a genetic stock by landing location is the most accurate metric without a known fishing location. However, misassignment of genetic stock is possible since fish from the opposite stock could be landed in the opposite region. Fish landed in NC, SC, and GA were assigned to the Carolinas stock (n = 863). Fish landed between Cape Canaveral and Key West, FL, were assigned to the South Florida stock (n = 1,065). A subset of otoliths from each stock

was randomly selected (~10 samples per age class; roughly 10% of the entire sample size) and comprised the calibration set used to build the FT-NIR spectroscopy predictive model. The remaining set of otoliths comprised the validation set and was used to compare accuracy and precision between "true ages" and ages estimated with FT-NIR spectroscopy. Age estimates provided by the primary reader were compared with age estimates from an expert reader to compare accuracy and precision of the observed ages, but consensus ages were not estimated if ages differed between readers. The primary reader's age estimates were considered the "true ages" when compared to ages predicted with FT-NIR spectroscopy. Two partial least squares (PLS) regression models, one for each genetic stock, were fitted to paired observed versus predicted ages in R (R Core Team 2021). The best calibration model was then applied to each stock's validation set with unknown ages and compared to the true ages to assess model performance. Comparison metrics included r^2, root mean square error of cross-validation (RMSECV), root mean square error of prediction (RMSEP), residual percentage difference (RPD), and average percent error (APE).

RESULTS AND DISCUSSION

Morphometric data in this study agreed with previous work indicating that white grunt from the Carolinas grew to a larger size and older age than South Florida white grunt (Fig. 4.1). Our data also indicated that otolith weight was greater for white grunts from the Carolinas than those from South Florida, which was not previously known (Fig. 4.2). In addition, there were differences in the quality of age estimates between the two stocks. Reader agreement among observed ages for white grunt from the Carolinas was higher and error was much lower (APE =

2.8%, Fig. 4.3A) than fish from South Florida (APE = 5.8%, Fig. 4.3C). When compared with traditional age estimates, ages predicted with FT-NIR spectroscopy models were less accurate and were positively biased at the youngest age class and negatively biased at older ages for both stocks. Results from comparisons of calibration versus validation age data sets indicated a better correlation between spectral signatures and predicted age for the Carolinas stock (r^2 = 0.91, Fig. 4.3B) than the South Florida stock (r^2 = 0.85; Fig. 4.3D). For the Carolinas model, RMSEP was 1.5, which indicates the model-predicted age had a mean difference of ±1.5 years from the observed age; the RPD value of 3.3 indicates a good model fit (RPD > 3). For the South Florida model, RMSEP was 1.3, but the RPD was below the accepted performance threshold (RPD = 2.6). Predicted ages for fish from the South Florida stock had higher APE when compared to the true ages (Fig. 4.3C-D), which was likely due to higher ageing error in the traditional ages used to build the predictive models, greater variation in size-at-age, and fewer age classes among South Florida white grunt. Fish from the South Florida stock experience less seasonality in temperature regimes than fish from the Carolinas stock, which likely affects otolith banding clarity causing decreased precision in traditional age estimates. We recommend including more young (age-0 and 1) and old fish (>age-12), if available, into each calibration model to reduce ageing bias. We also recommend using a second age reader to maximize ageing precision through consensus ages for the true ages used to build the predictive models. Overall, we observed a moderately good correlation between (1) true ages and spectral signatures and (2) true ages and model-predicted age, but additional age data and reducing imprecision in input data will greatly increase our ability to more accurately predict fish age with FT-NIR spectroscopy.

We recommend that significant improvement in the precision between predicted and observed age estimates is needed before ages predicted with FT-NIR spectroscopy are of

sufficient quality for use in production ageing for stock assessment. Reduced precision in age estimates will increase uncertainty around stock assessment model parameter estimates and affect assessment outputs. Ageing bias at the youngest and older ages will also negatively impact assessment model outputs due to cohort smoothing and imprecision, respectively. However, ageing bias at older age classes may have little effect on assessment model outputs if bias occurs at ages comprising the plus group of the assessment model. Ageing bias can be somewhat corrected in stock assessment models by specifying expected ages and age-specific error in the ageing error matrix, but increased imprecision around age estimates can erroneously alter parameter estimates affected by size-at-age when fitting to multiple data sources in complex assessment models. Additional research is needed to improve the precision of predicted ages for fish from the South Florida stock. We also recommend further analyses be done to understand the influence of sexually dimorphic growth on age estimates, which was not explored in this study but is known to affect size-at-age estimates for this species (Potts 2000). Overall, genetic differences between stocks that affect size-at-age, growth, or maximum age are important to consider when building age prediction models. FT-NIR spectroscopy shows great promise as a method to produce indirect age estimates at high volume with great efficiency and acceptable precision but requires additional research before applying it to the production ageing of white grunt.

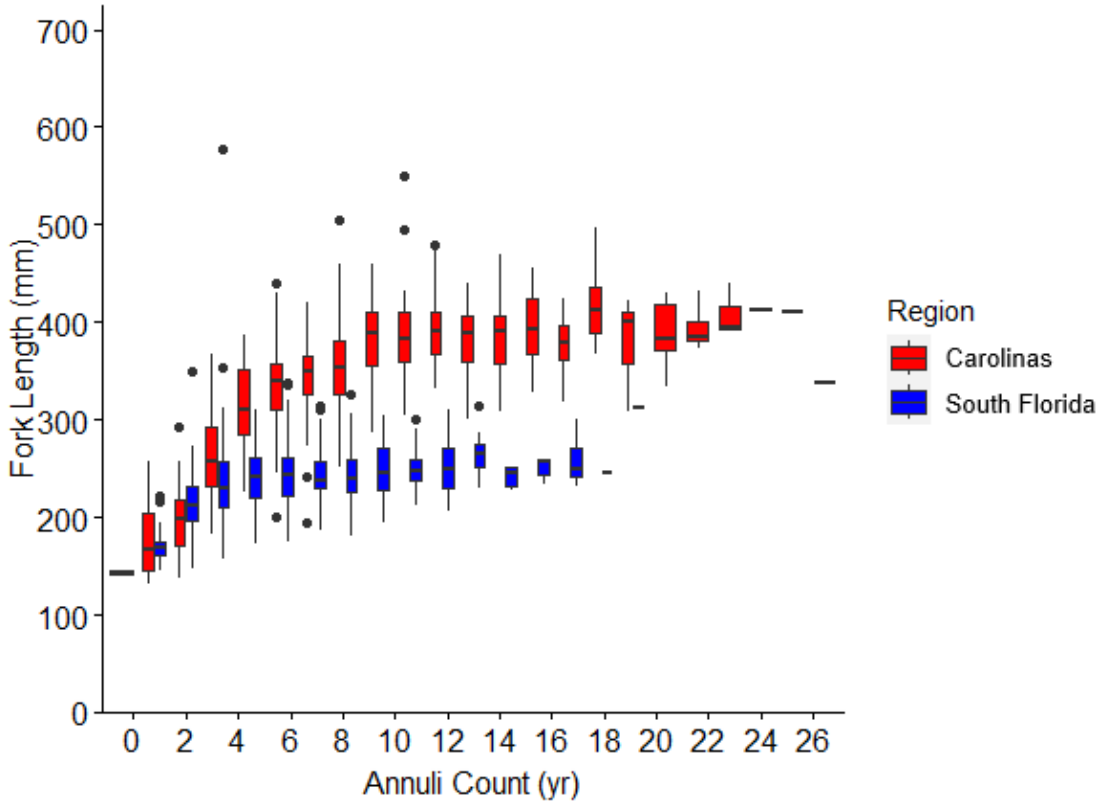


Figure 4.1. -- Boxplots of length (fork length mm)-at-age (yr) for the two different genetic stocks (Carolinas, red and South Florida, blue) of white grunt (*Haemulon plumierii*) in the U.S. South Atlantic. Horizontal lines within the boxes indicate median values, box ends indicate the 25th and 75th quartiles, whisker ends indicate 1.5*interquartile range, and points indicate outlying values.

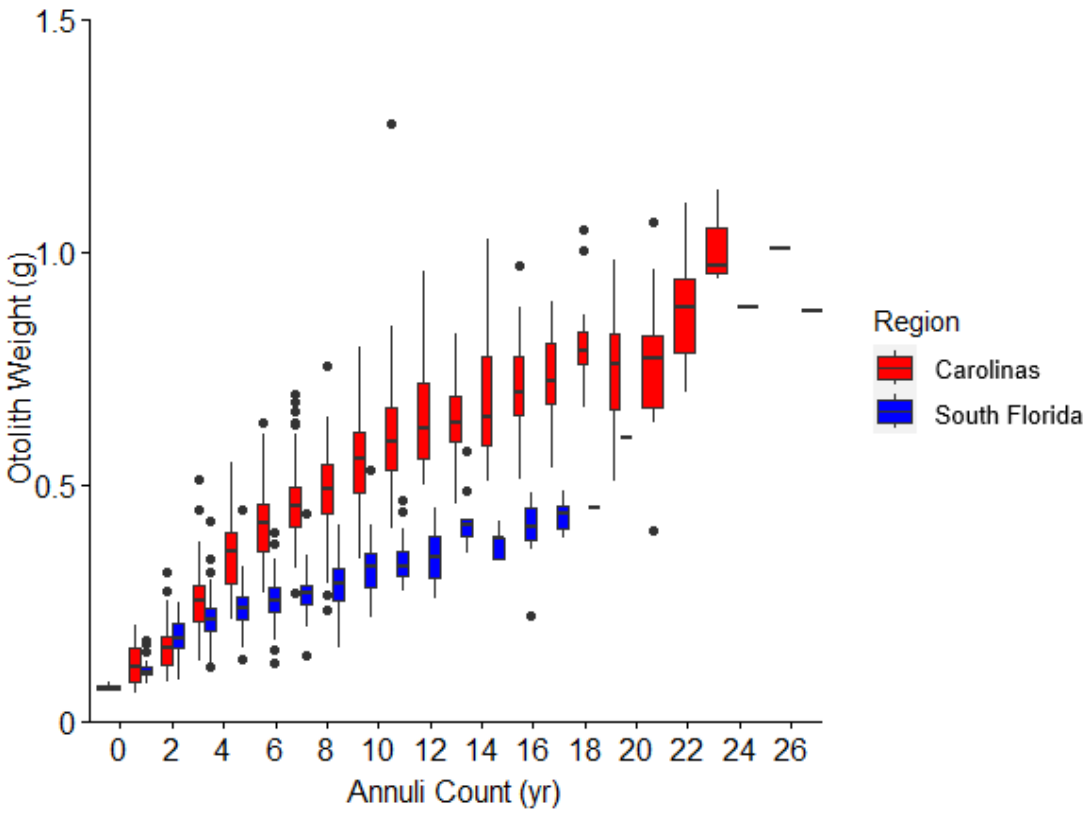


Figure 4.2. -- Boxplots of otolith weight (g)-at-age (yr) for the two different genetic stocks (Carolinas, red and South Florida, blue) of white grunt (*Haemulon plumierii*) in the U.S. South Atlantic. Horizontal lines indicate median values, box ends indicate the 25th and 75th quartiles, whisker ends indicate 1.5*interquartile range, and points indicate outlying values.

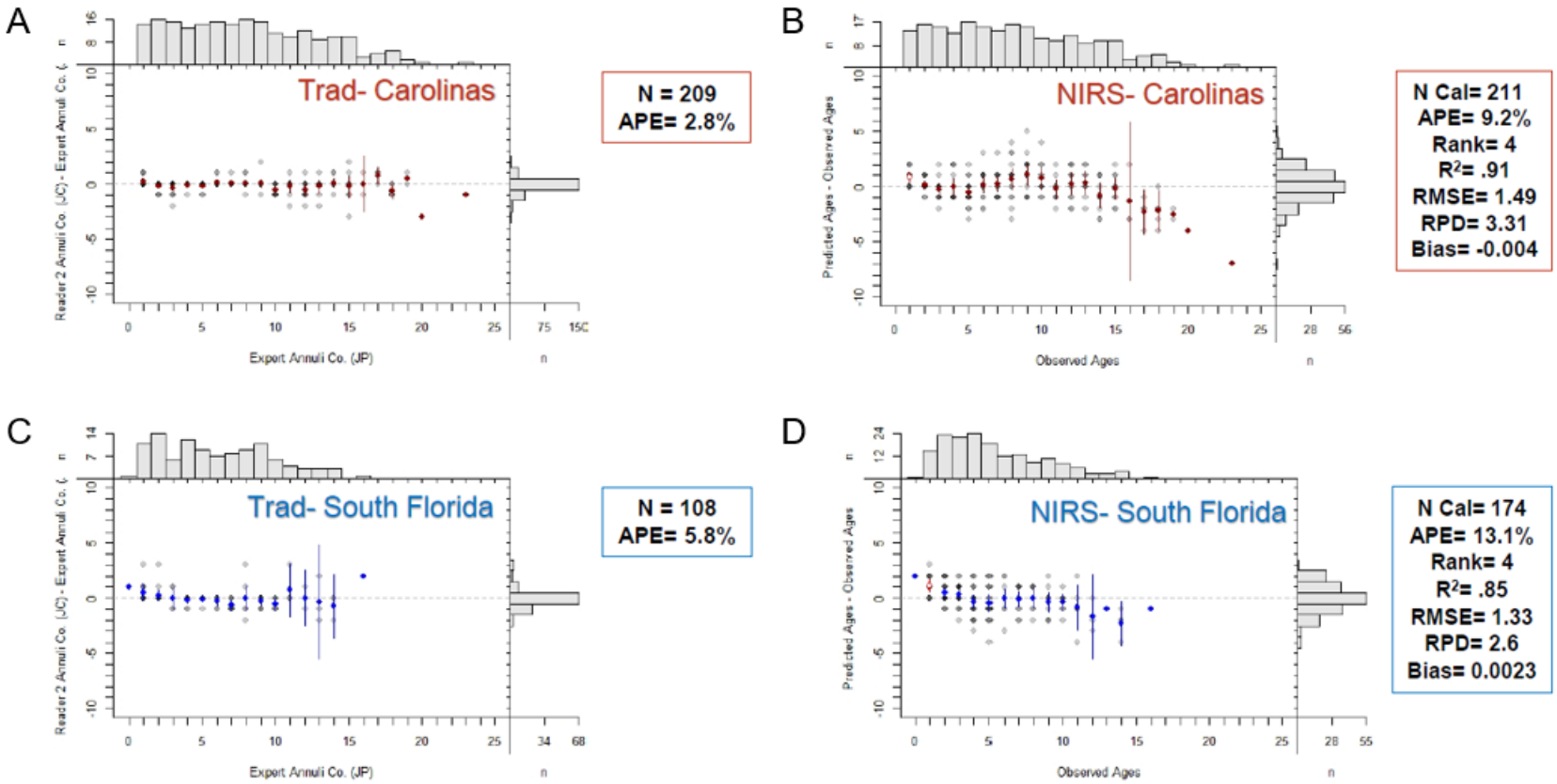


Figure 4.3. -- Comparison plots of traditional age estimates (primary vs. expert reader age (Panels A and C) and age estimates predicted with FT-NIR spectroscopy (Panels B and D) for each white grunt (*Haemulon plumierii*) stock. Sample sizes and average percent error (APE) are shown for each panel. Red and blue points indicate mean values for the Carolinas and South Florida stocks, respectively, while gray-scale points indicate individual age estimates with a transparency of 1/10 (e.g., 10 observations of the same age difference would produce a fully darkened point). Vertical lines indicate the 95% confidence interval of the mean. The histogram above each panel indicates the sample count for the expert or observed ages by age class; the histogram to the right of each panel indicates the sample count for the difference in age estimates between the primary and expert reader or observed and predicted age.

5. A Novel Approach for Determining the Reproductive Status of Walleye Pollock (*Gadus chalcogrammus*) Using Raman Spectroscopy

Sandi K. Neidetcher[1], Morgan B. Arrington[2], Thomas E. Helser[1], Esther D. Goldstein[1], and Irina M. Benson[1]

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

[2] Cooperative Institute for Climate, Ocean, and Ecosystem Studies (CICOES)
University of Washington
3737 Brooklyn Ave NE
Seattle, WA USA

INTRODUCTION

Knowledge of the reproductive status of fishes is vital to the sound management of fish stocks. Spawning stock biomass (SSB) provides an estimate of the proportion of the stock contributing to reproduction and is used to determine stock status and abundance to set catch quotas. These estimates are obtained through the construction of maturation schedules (spawning ogives) based on field-based observations of ovarian development and provide estimates of the proportion of reproductively mature females in a stock or population. In addition to providing estimates of spawning potential, knowledge of reproductive strategy and the geography and phenology of spawning can greatly benefit our understanding of interactions between a species and its environment (e.g., identification of early life stage distribution and dispersal to nurseries where young fish develop). In recent decades, climate-driven warming trends have been shown to impact fish distribution, phenology, and growth rates (Hollowed et al. 2013). In addition, reproductive patterns in fish, such as gonad maturation and ovulation, are often tied to temperature thresholds (Alix et al. 2020). Understanding the impacts of climate change on the reproductive biology of fish will increase the need for data and for tools with accuracy and efficiency in these collections.

Histological analysis, considered the most accurate method of determining oocyte staging (West 1990), involves the microscopic examination of ovary tissues that have been fixed, stained, and mounted on a slide. This process allows for the identification of microscopic structures and conditions that occur as an ovary develops. However, histological processing is effort intensive and incurs costly processing fees from specialized laboratories. To conduct histological analysis, ovaries are often collected in the field and preserved in formalin, a

hazardous chemical. The cost associated with histological processing, the challenges of working with formalin in the field, and the time required for processing and slide analysis are often limiting factors to this approach.

Vibrational spectroscopy is an emerging technology in ecological research that may provide an alternative approach for determining maturity and reproductive status of fishes. Several spectroscopy techniques have been applied to ovarian tissue. Examples include the use of Fourier transform near infrared (FT-NIR) spectroscopy (10,000 to 4,000 cm^{-1}) to differentiate non-spawning from spawning-capable fish (TenBrink et al. 2022) and Raman spectroscopy to discriminate egg quality and viability by tracking macromolecule fluctuations during specific development stages post-fertilization (Ishigaki et al. 2016).

Ovary tissue, like most biological tissues, is composed primarily of various proteins, lipids, carbohydrates, and phosphates with levels that vary throughout development. As the ovary matures, changes observed through histological analysis include structural development and material accumulations, which reflect changes in the molecular compositions. Raman spectroscopy has unique characteristics that may make it an optimal spectroscopy method for detecting changes in the levels and configuration of the molecular components through reproductive maturation. In Raman spectroscopy, a laser is used to irradiate a material across a range of wavenumbers, and highly sensitive lenses are used to collect the light that is scattered back. Scattered light is analyzed and provides a spectrum "fingerprint" with patterns or bands unique to the molecular composition and configuration of the sample under interrogation. This can be useful when unknown molecular constituents in biological tissues need to be identified or quantified. Raman spectroscopy is also less sensitive to water than other spectroscopy techniques and therefore may be more useful for biological tissue samples, as they contain large amounts of

water (Butler et al. 2016). The use of Raman spectroscopy as a tool to access ovary maturation would allow for more rapid, and in-situ, evaluations while reducing the use of hazardous formalin and expensive histological processing.

In this study, we evaluate the potential for Raman spectroscopy to provide information on the maturity and reproductive status of walleye pollock (*Gadus chalcogrammus*), referred to as "pollock" hereafter. Pollock is a semi-pelagic, schooling fish widely distributed in the North Pacific Ocean and supports highly lucrative commercial fisheries in the Bering Sea and the Gulf of Alaska. Reproductive information used in pollock stock assessments is currently based on macroscopic evaluation of their gonads (Williams 2007). Visual maturity estimates are collected at the point of capture and used to estimate maturation status and spawning phenology. Because pollock are a multi-batch spawning species, ovaries may contain oocytes at varying levels of maturation and are often difficult to stage macroscopically. Differentiating between immature, developing, and spent fish is particularly difficult for this species. In addition, recent research has identified shifts in the phenology of spawning (Rogers and Dougherty 2019) and in distribution on the spawning grounds for pollock (Eisner et al. 2020). Improved monitoring of this species would be beneficial to evaluate changes in reproduction that may impact stock assessment estimates and management.

We hypothesize that 1) mature pollock ovaries with histologically visible yolk development (hereinafter referred to as "yolked" ovaries) can be differentiated from pollock ovaries with little to no visible yolk development (hereinafter referred to as "non-yolked" ovaries) based on the Raman spectra of their oocytes; 2) within non-yolked ovaries, mature but not developing and mature but recently spent ovaries can be differentiated from immature or early developing ovaries based on the Raman spectra of their oocytes. Further, we explore the

ability for this tool to differentiate among progressive reproductive stages in yolked ovaries for potential application in spawning phenology studies.

METHODS

Ovary specimens were collected aboard the NOAA ship *Oscar Dyson* during the Midwater Assessment and Conservation Engineering (MACE) Program of the Alaska Fisheries Science Center's acoustic-trawl pollock stock assessment surveys. Samples were collected in the Gulf of Alaska (GOA) during February and March 2017 (Fig. 5.1; McCarthy et al. 2018). Ovaries were removed from the fish, weighed, and placed in a mesh bag and preserved in 10% formalin. Additional collections of 1 and 2 year old pollock were included in this study due to the rare capture of immature pollock during the prespawning cruises. For these collections, young-of-the-year pollock were beach seined and reared through the first two years of life at an aquaculture facility at Little Port Walter, NOAA. In preparation for histological analysis and Raman spectra collection, a cross-section of tissue from each ovary was processed and stained by histology laboratories. We collected one subsample of tissue (<1 g) from a location adjacent to the histology tissue sample for spectral data acquisition.

Immature (IMM) fish are those which have not aged to first maturation. Developing stage (DEV) fish show increased oocyte sizes and the development of cortical alveoli and may include very early signs of yolk development. Vitellogenic stages (VIT) include increasing yolk concentrations as well as the migration of the nucleus. The prespawning stage (PSWN) includes a brief period in the annual cycle when yolk droplets, along with the nucleus break down, or coalesce. Hydration, or the rapid increase in size due the uptake of water, marks the early

spawning stage (SWN), which further progresses with ovulation as oocytes are released from the follicles and an increase of empty or post-ovulatory follicles (POFs) is observed. The Partial Spent stage (PSNT) accounts for the batch spawning strategy of pollock and where vitellogenic stages occur with POFs. Spent stage (SNT) ovaries are composed mostly of POFs and possibly a few atretic yolk stages that failed to ovulate. Mature but not developing (MND) ovaries were identified as having spawned previously, but which show no signs of development in the capture season. Ovary stage assignments were made using the most predominant oocyte structures present or by the presence of combinations described above. In addition, each specimen was identified with either the presence or absence of visible yolk development regardless of other staging identification.

We acquired Raman spectral data from ovary tissue using a Raman Process BallProbe with a 780 nm laser and a wavenumber range of 50-3300 cm^{-1} (MarqMetrix). The spherical sapphire lens on the fiber optic probe was positioned to touch the sample during each measurement. The laser power on the samples was 400 mW with an exposure time of 1000 ms. Ten repeat measurements were collected and averaged to create one representative spectrum per sample. All fluorescent lights were turned off during Raman spectra collection.

Spectra Collection and Processing

Prior to data analysis, we subtracted a blank spectrum (measured with >2 ft between the probe and closest object) from each sample spectrum to eliminate peaks due to the sapphire in the fiber optic probe (Raml et al. 2011). Spectra were truncated into the fingerprint region (200-1800 cm^{-1}) and high frequency region (2600-3250 cm^{-1}). We then preprocessed raw spectral data

using extended multiplicative signal correction (EMSC) to normalize spectra (Martens and Stark 1991) and performed polynomial subtraction using the modified polyfit method to remove fluorescent background (Lieber and Mahadevan-Jansen 2003). For the fingerprint region, we parameterized EMSC and modified polyfit with seventh-order polynomials. For the high frequency region, we parameterized EMSC with a fifth-order polynomial and modified polyfit with a third-order polynomial. Savitzky-Golay smoothing was applied to both regions to reduce unwanted noise in spectra (Delwiche and Reeves 2010). We parameterized Savitzky-Golay smoothing with a second-order polynomial and 25 points.

Statistical Analyses

We used principal component analysis (PCA) as an exploratory analysis for data visualization and to identify spectral outliers. Detection and elimination of outliers was performed based on robust calculation of outlier limits (Pomerantsev and Rodionova 2014). Values that fell outside a significance level of 0.01 were considered outliers and removed from the data set ($n = 2$).

We used a two-step analysis to classify samples as biologically mature or immature from the Raman spectra of their ovaries. First, we used a partial least-square (PLS) regression analysis followed by a linear discriminant analysis (LDA) to classify yolked and non-yolked samples (Boulesteix 2004). A PLS regression is a multivariate method commonly used in chemometric analysis for dimensionality reduction and to linearize spectral data (Wold et al. 1884, Boulesteix 2004). In PLS, the independent (here, spectral data) and dependent (here, yolk presence as determined by histological analysis) data matrices are decomposed into a set of scores and

loadings (hereinafter referred to as latent variables). The latent variables are then used as predictors in LDA. Linear discriminant analysis performs well on a smaller number of uncorrelated variables (Dudoit et al. 2002). In LDA, the independent variables (here, latent variables from PLS) are assumed to have normally distributed probability density functions with respect to the dependent variable. Class membership is determined from the class with maximum posterior probability. All samples classified as yolked in the PLS-LDA were considered biologically mature. A second PLS-LDA was applied to classify the non-yolked samples as either biologically mature (if histology indicated it was MND or SNT) or immature (if histology indicated it was IMM or DEV).

To evaluate the ability to further predict the spawning progression of walleye pollock ovaries with visual yolk development from their Raman spectra, we used a PLS-LDA to classify yolked specimens to their histologically determined reproductive stage (VIT, PSWN, SWN, PSNT). Visible yolk at barely discernible levels can sometimes be seen in DEV ovaries and in small amounts of residual, yolked oocytes in SNT ovaries. For these stages, yolk occurs at minute levels compared to the yolk contained in VIT, PSWN, SWN, or PSNT ovaries, and were predicted as non-yolked. Here, yolked samples with histologically determined stages of DEV or SNT were removed from this analysis due to insufficient sample sizes ($n = 1$).

Variable importance in projection (VIP) scores were calculated for each wavenumber to identify the spectral regions most important for classification in each model (Goldstein et al. 2021). The calculation of VIP includes the covariance between independent and dependent variables and reflects how much a variable contributes to describing both independent and dependent data (Anderson and Bro 2010). A VIP score of less than one indicates a less important variable.

Cross-validation was used to determine the optimal number of latent variables (Boulesteix 2004) and the predictive ability of the three models on new data. For all categories, we report specificities, or the accuracy of rejecting a sample from an incorrect category, sensitivities, or the accuracy of assigning a sample to the correct category, and overall model accuracy and area under the curve (AUC). Data analysis was conducted in R statistical computing software version 4.1.0 (R Core Team 2021).

RESULTS AND DISCUSSION

We summarize PLS-LDA classification results and whether they support the ability to classify fish as biologically mature or immature from the Raman spectra of their ovaries, and further, to classify yolked specimens to their histologically determined reproductive stage (VIT, PSWN, SWN, PSNT). In total, 226 specimens were included in this analysis (Table 5.1).

Maturity Status

The Raman spectra of biologically mature ovaries with yolk development were differentiable from the spectra of ovaries with little to no yolk development (Fig. 5.3, Table 5.2). Spectral data showed clear separation between yolked and non-yolked samples along latent variable (LV) 1 and LV2 from PLS used for dimensionality reduction (Fig. 5.4). The optimal number of latent variables from PLS used for LDA was 8, which explained 96.34% of variance in the spectral data matrix and 88.72% of variance in the response data (here, yolked vs. non-yolked). All but two samples were correctly classified in the PLS-LDA (99.0% accuracy; AUC

of 0.99) (Table 5.3). Wavenumber regions 450-558, 776-818, 838-851, 1192-1214, 1296-1426, 1474-1510, 1582-1628, 1751-1778, and 2794-2831 cm^{-1} contributed substantially to variation in spectra related to the response variable (VIP > 1) (Fig. 5.3, Table 5.2). All samples classified as yolked were considered mature. Ovaries with no yolk development can signify a biologically immature fish, but may also be a mature fish under certain circumstances. The differentiation between maturity statuses of non-yolked ovaries is historically difficult to determine using existing rapid methods (visual maturity).

The Raman spectra of mature but non-yolked ovaries (MND or SNT) were differentiable from the spectra of non-yolked immature ovaries (IMM or DEV (Fig. 5.3). There was separation between non-yolked immature and mature ovaries along LV1 and LV2 (Fig. 5.4). The optimal number of LVs from PLS used for LDA was 5, which explained 95.03% of variance in the spectral data matrix and 71.09% of variance in the response data (here, mature vs. immature). Of samples classified as non-yolked, 98.0% of mature samples were correctly classified by PLS-LDA and 88% of immature samples (AUC = 0.98) (Table 5.3). Wavenumber regions 437-553, 666-755, 770-784, 873-924, 1062-1130, 1242-1299, and 1455-1465 cm^{-1} contributed most substantially to variation in spectra related to the response variable (VIP > 1) (Fig. 5.3, Table 5.2). Overall from the two-step analysis, 43 out of 50 (86.0 %) total biologically immature samples were corrected classified and 175 out of 176 (99.4 %) total biologically mature samples.

Yolked Stage Progression

There was less differentiation in Raman spectra among yolked samples by histologically determined reproductive stage. Progression in stages with yolk present was visible along LV1

and LV5, especially for SWN and PSNT samples, but with overlap among stages (Fig. 5.4). The optimal number of LVs from PLS was 9, which explained 94.90% of variance in the spectral data matrix but only 42.63% of variance in the response data. The highest classification accuracy was for PSNT samples (86.0%; AUC = 0.92) followed by SWN samples (85.0% accuracy; AUC = 0.93) and VIT samples (85.1% accuracy; AUC = 0.95) (Table 5.3). The PSWN samples had the lowest classification accuracy (78.9% accuracy; AUC = 0.90). Wavenumber regions 324-339, 422-528, 1573-1587, 2873-2877 cm^{-1} contributed most substantially to variation in spectra related to the response variable (VIP > 1) (Fig. 5.3, Table 5.2).

The use of Raman spectroscopy to identify molecular changes in biological tissue is complex, as each molecular component contributes to information captured in the spectra, and the specific functional groups present in the biological tissue are not always known (Talari et al. 2015). The chemical composition of pollock ovaries has not yet been fully investigated. In addition, our application of this technology to identify shifts in molecular composition during ovary maturation is also complex, as the individual components of the ovary change over the course of maturation. However, patterns in the spectral data and associated molecular assignments in published literature provide evidence supporting our model results. Gorbatenko and Lazhentsev (2016) analyzed the biological composition of pollock ovaries and showed that lipids, proteins, and carbohydrates varied within a broad range of maturation. These authors observed the lowest values of these molecular components in post-spawning individuals, while the highest lipid and protein content of the gonads was recorded from females at the advanced yolk stages.

Our results suggest that spectroscopy can provide an effective tool in differentiating 1) yolked from non-yolked stages and 2) immature fish from mature fish, an important metric for

fisheries management. Differentiating yolked stages would also benefit researchers in tracking shifts in the capture location, timing, and periodicity of spawning marked by the presence of later stage yolked oocytes and partially spent ovaries.

Table 5.1. -- Length distribution (in cm) by ovary stage (IMM = immature, DEV = developing, VIT = vitellogenesis, PSWN = prespawning, SWN = spawning, PSNT = partially spent, SNT = spent, and MND = mature, no development) for walleye pollock (*Gadus chalcogrammus*) collected from the Gulf of Alaska.

	IMM	DEV	VIT	PSWN	SWN	PSNT	SNT	MND	Total
Length									
15	2								2
20	9								9
25	6								6
30	21								21
35	4								4
40		1	5	5	1	3	7	8	30
45		7	26	19	18	18	10	11	109
50			9	4	8	4	3	4	32
55				1	2	2	2		7
60			1		3		2		6
Total	42	8	41	29	32	27	24	23	226

Table 5.2. -- Variable importance in projection (VIP) wavenumber (cm^{-1}) ranges for each model identifying the spectral regions most important for classification (VIP > 1).

Yolked vs. non-yolked	Non-yolked: Immature vs. Mature	Yolked: Stage Progression
	227-229	
		280-300
330-409		324-339
	374-407	
		409-428
435-558	437-553	495-528
		613-623
	657-760	
		751-770
776-854	770-784; 787-824	795-863
	873-924	
		996-1016
1087-1113	1062-1130; 1144-1175	1026-1173
1192-1214	1183-1216	1189-1212
1296-1426	1242-1299	1397-1428
1474-1510	1426-1512	1476-1483
	1523-1539	
1555-1560; 1592-1628		1532-1642
1680-1728; 1749-1778	1606-1692	1677-1778
2616-2624		
2651-2654	2634-2659	2637-2655
	2674-2679	2665-2679
		2690-2695
2794-2831		2700-2703; 2715-2724; 2777-2781; 2800-2802
2869-2909	2902-2931	2861-2865; 2872-2913
2926-3004	2947-3000	2957-2964
	3013-3022	
3067-3132	3072-3081	3064-3092
3147-3190		3104-3109; 3128-3155
		3190-3200
	3214-3217	3211-3221
		3231-3232
		3236-3238

Table 5.3. -- Results of classification of walleye pollock (*Gadus chalcogrammus*) reproductive state based on Raman spectra of ovaries. The ratio of correctly predicted samples relative to total sample size, the sensitivity (true positive rate), the specificity (true negative rate), the balanced accuracy, and area under the curve (AUC) relative to the no information rate of the model predictions from cross validation. Results shown as proportions of total samples.

Analyses	Class	Ratio	Sensitivity	Specificity	Balanced Accuracy (%)	AUC	No information rate
1. Yolk presence	non-yolked yolk present	95:96 129:130	0.99	0.99	99.1	0.99	0.58
2. Non-yolked: Maturity status	immature mature	43:49 46:47	0.88	0.98	92.8	0.98	0.51
3. Yolked: Stage Progression	vitellogenic prespawning spawning partial spent	33:41 20:29 23:32 21:26	0.80 0.69 0.72 0.81	0.90 0.89 0.98 0.91	85.1 78.9 85.0 86.0	0.95 0.90 0.93 0.92	0.32 0.32 0.32 0.32

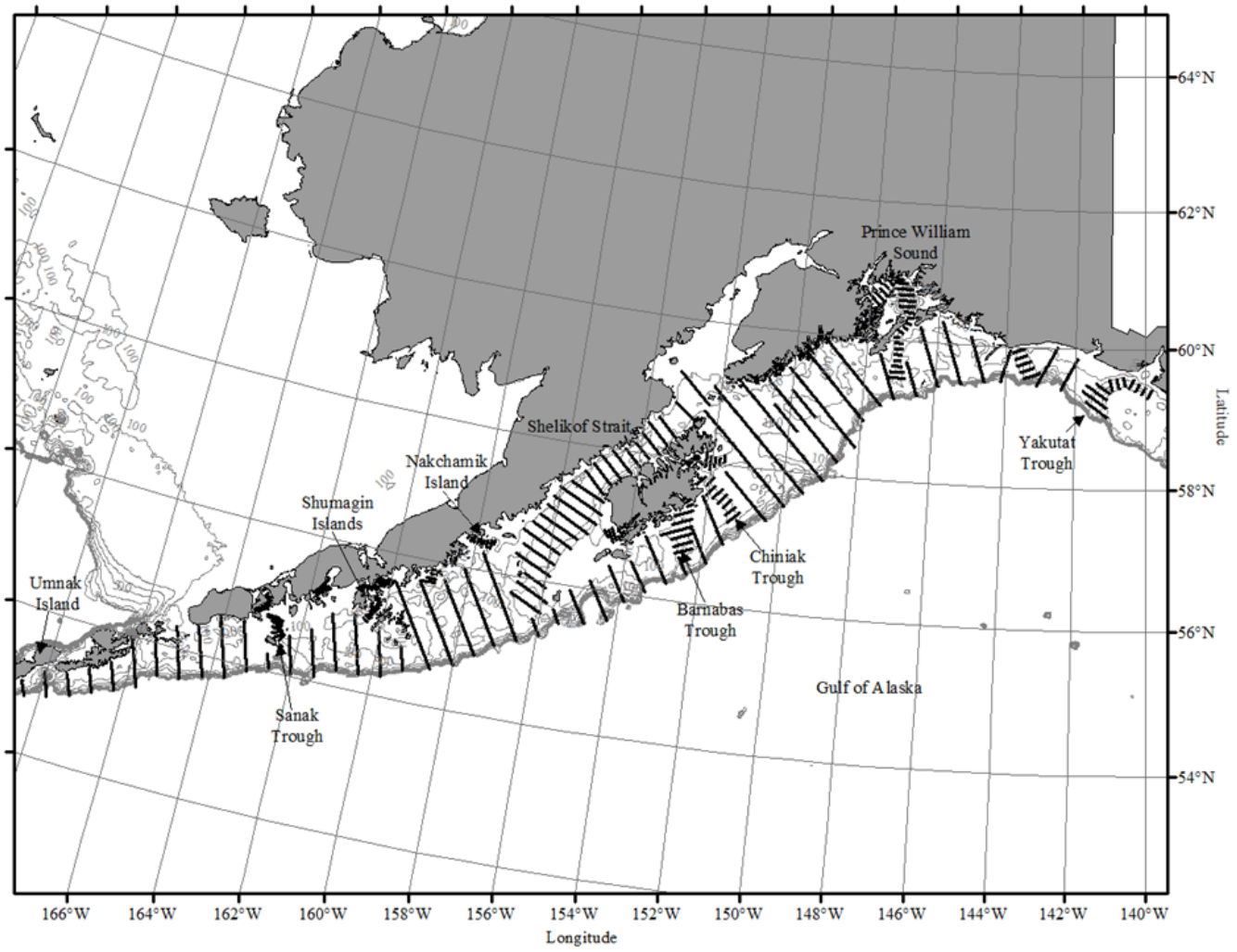
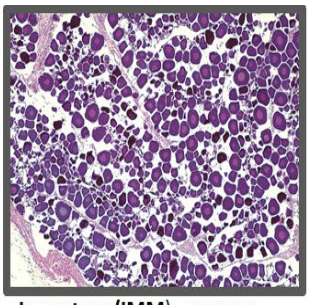
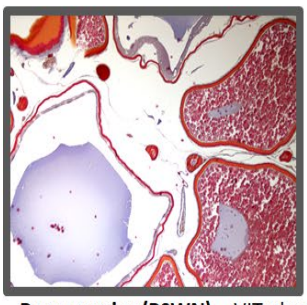


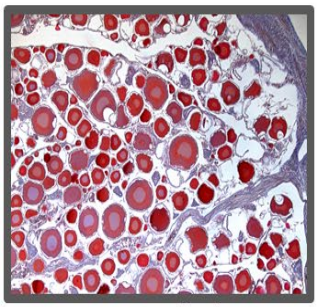
Figure 5.1. -- Transect line locations for the Alaska Fisheries Science Center's (AFSC) acoustic-trawl walleye pollock (*Gadus chalcogrammus*) stock assessment surveys.



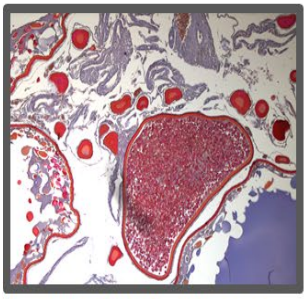
Immature (IMM)– reserve fund, tightly packed oocytes, little tunica, thin wall



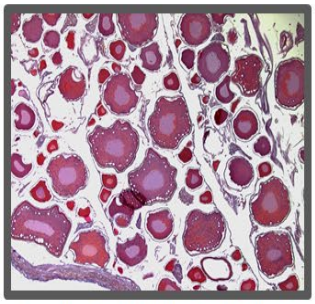
Prespawning (PSWN) – VIT plus hydration



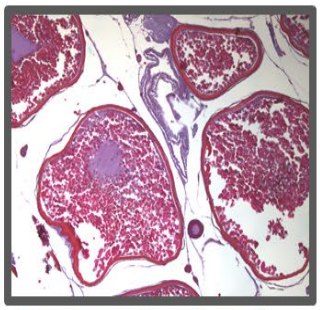
No development (ND) – reserve fund, more tunica, thick wall



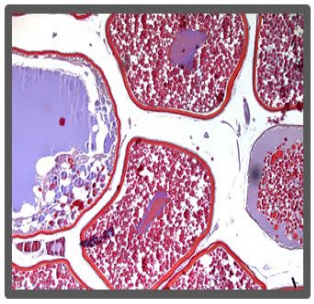
Spawning (SWN) – VIT, some hydration, plus post ovulatory follicles



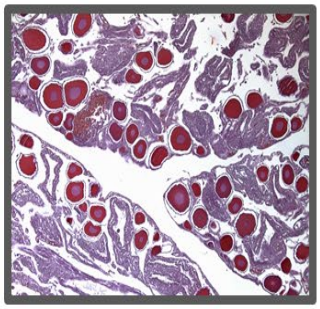
Developing (DEV) – Cortical Alveoli



Partial Spent (PSNT) – VIT (no coalescence or hydration) plus post ovulatory follicles



Vitellogenesis (VIT) – early to late vitellogenesis, nuclear migration, coalescence



Spent (SNT) – early post ovulatory follicles, residual VIT resorbing

Figure 5.2. -- Images reflecting oocyte structures and components used to assign ovary maturation stages in walleye pollock (*Gadus chalcogrammus*). Immature oocytes are characterized by small, tightly grouped previtellogenic oocytes; they enter the maturation cycle and begin to develop with cortical alveoli and early vitellogenesis (or yolk accumulation; as yolk continues to accumulate, the oocyte increases exponentially in size and is considered spawning-capable. Oocytes continue to increase in size during vitellogenesis, followed by coalescence and hydration during the pre-spawning stage. Evidence of spawning includes the presence of hydrated oocytes and post-ovulatory follicles (POFs; the structural support of oocytes during development). After spawning, the remaining follicle structures and residual oocytes that remain are resorbed during the spent stage. A resting or reproductively inactive stage (non-spawning; is often identified as the length of time between spent and developing where the ovary appears inactive or with signs of late resorption, and early perinuclear development.

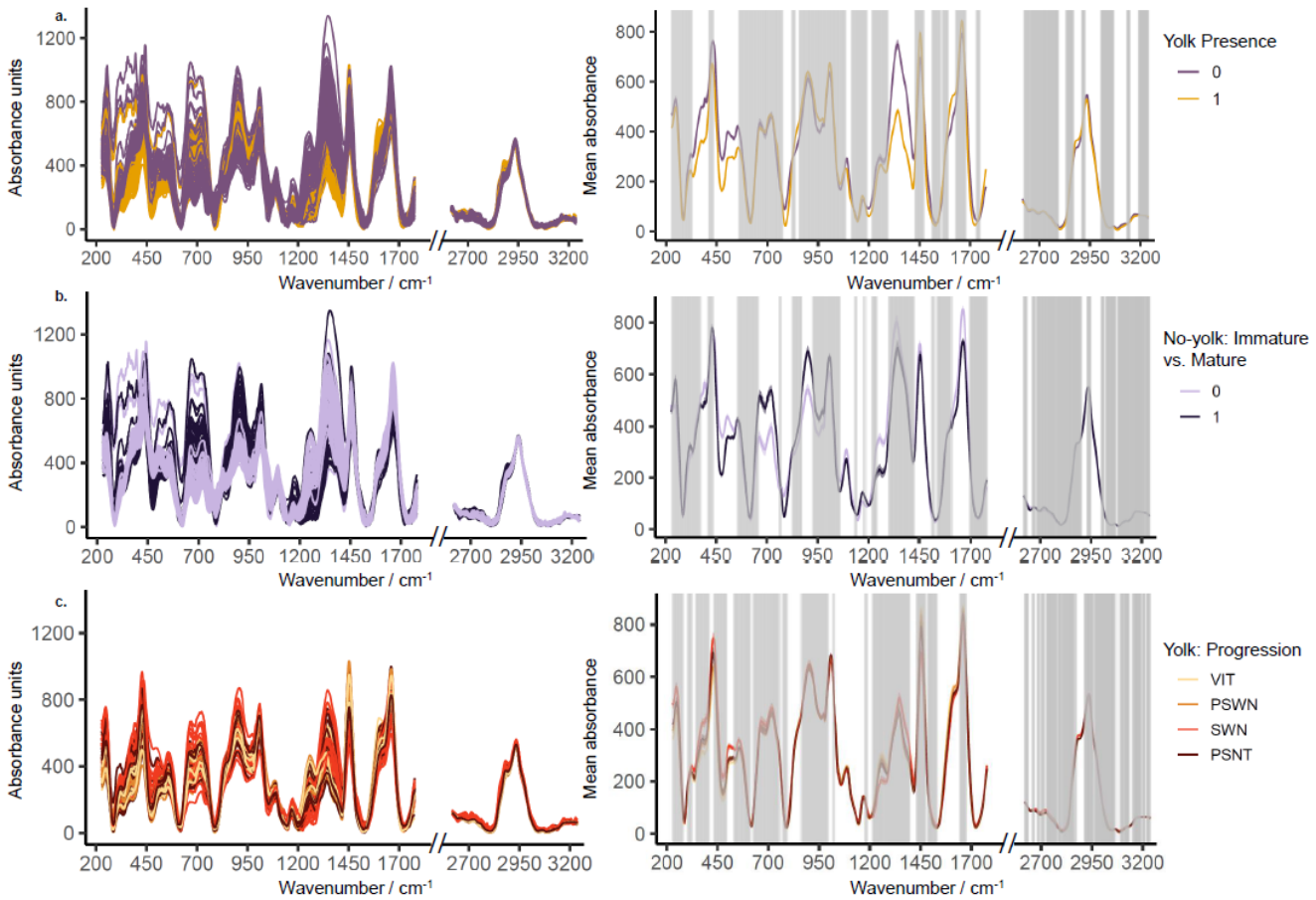


Figure 5.3. -- Raman spectra of walleye pollock (*Gadus chalcogrammus*) ovaries preprocessed with extended multiplicative scatter correction, polynomial subtraction, and Savitzky-Golay smoothing. The left column shows all spectra, and the right column shows average spectra and standard error for a) yolked vs. non-yolked samples, b) the maturity status of samples predicted as non-yolked, and c) the finer-scale stage progression for samples predicted as yolked (VIT = vitellogenesis, PSWN = prespawning, SWN = spawning, and PSNT = partial spent). Wavenumbers identified as important for projection (VIP score > 1) with respect to the response variables are shown highlighted in the white spaces. All other wavenumbers are grayed out (VIP scores < 1).

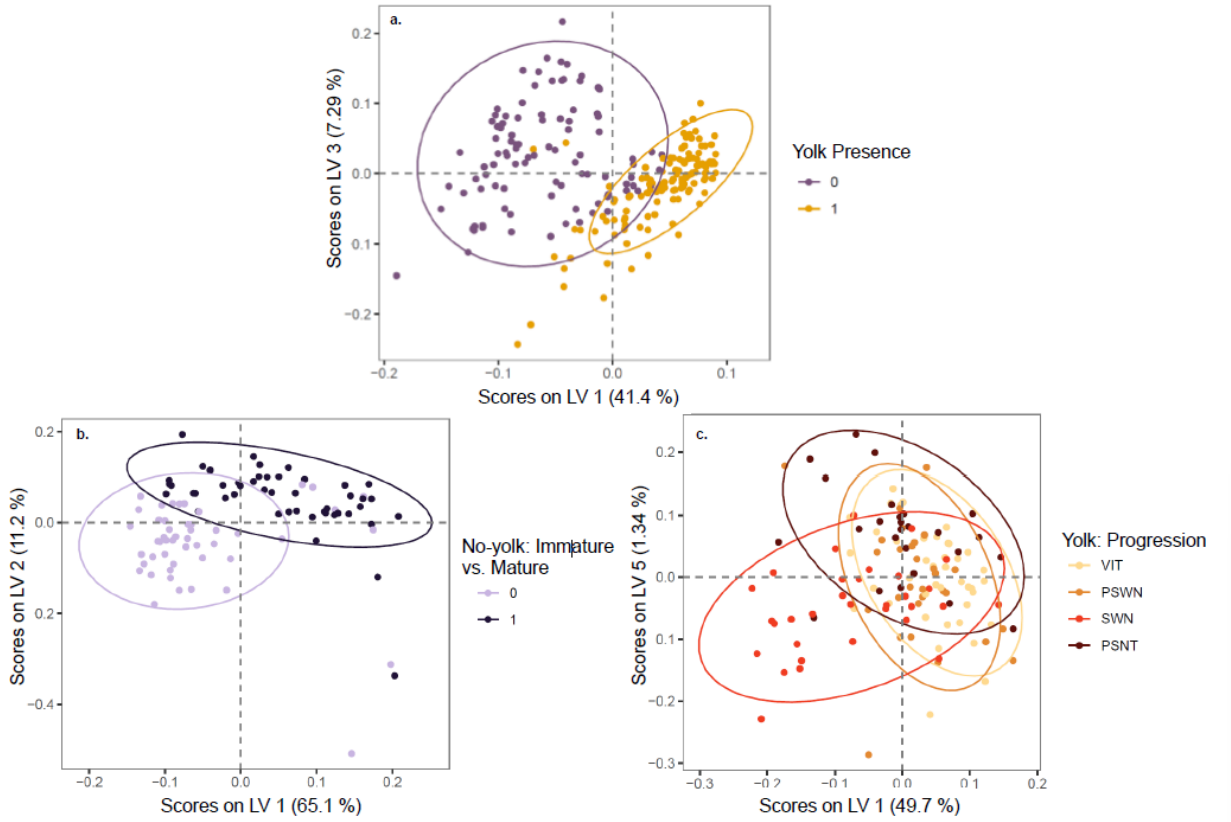


Figure 5.4. -- Scores on latent variables with 95% confidence ellipses shown from partial least squares dimensionality reduction for a) yolked vs. non-yolked samples b) the maturity status of samples predicted as non-yolked and c) the spawning phenology of samples predicted as yolked (VIT = vitellogenesis, PSWN = prespawning, SWN = spawning, and PSNT = partial spent) based on Raman spectra of walleye pollock (*Gadus chalcogrammus*) ovaries. The latent variables with the most differentiation by response variable shown for each analysis.

6. Fourier Transform Near Infrared Spectroscopy Ageing of Finfish and Shark Species in the Northwest Atlantic

Alexander M. Rubin[1, 2], Michelle S. Passerotti[2], and Eric M. Robillard[3]

[1] IBSS Corporation
1110 Bonifant Street
Silver Spring, MD USA

[2] Northeast Fisheries Science Center
NOAA, National Marine Fisheries Service
28 Tarzwell Drive
Narragansett, RI USA

[3] Northeast Fisheries Science Center
NOAA, National Marine Fisheries Service
166 Water Street
Woods Hole, MA USA

INTRODUCTION

Fourier transform near infrared (FT-NIR) spectroscopy is an emerging method to estimate age of marine fishes. In this method, hard structures traditionally used for ageing fish (e.g., otoliths and vertebrae) are irradiated with light from the near infrared (NIR) spectrum. Organic molecules comprising these structures absorb NIR light at different wavenumbers depending on their chemical, and to a lesser extent, physical, composition. Throughout the lifespan, the physical and chemical characteristics of these structures change, and these changes are reflected in NIR absorbance spectra. FT-NIR spectroscopy has been previously shown to be an effective method at predicting ages of individuals through the use of partial least squares (PLS) regression models relating NIR absorbance spectra to fish age, with models trained on known-age samples of otoliths or vertebrae (Arrington et al. 2022, Healy et al. 2021, Passerotti et al. 2020b, Rigby et al. 2016). This method has been applied to shorter-lived teleost species (Healy et al. 2021, Passerotti et al. 2020b) and smaller elasmobranch species (Arrington et al. 2022), but few studies have used it to age long-lived teleosts and no studies to date have used it to age large-bodied sharks.

We assessed the efficacy of FT-NIR spectroscopy to predict the age of two hard-to-age species in the northwest Atlantic: Acadian redfish (*Sebastes fasciatus*) and blue shark (*Prionace glauca*) using otoliths and vertebrae, respectively. We also investigated the optimal sample presentation for blue shark vertebrae given their large size. Finally, we evaluated the efficiency of FT-NIR spectroscopy scanning with respect to processing times associated with traditional ageing of both species.

METHODS

Samples

Acadian redfish sagittal otoliths were collected by the NOAA Fisheries Northeast Fisheries Science Center (NEFSC) bottom trawl survey from 2016-2019 and were stored dry after collection. A total of 1,891 otoliths were selected for FT-NIR spectroscopy analysis. All otoliths were cleaned by wiping with 70% ethanol using a lab tissue, and blotted dry before scanning.

Blue shark vertebrae were collected by the NEFSC Apex Predators Program from 1994-2018. A total of 184 whole vertebrae and 188 sectioned vertebrae were selected for FT-NIR spectroscopy analysis. All samples were cleaned of bulk tissue with a scalpel and cleaned of any residual tissues by immersion in a bleach solution and rinsing with water. Afterward, vertebrae were stored in 70% ethanol, and were blotted dry before scanning. Vertebrae were not allowed to fully dry, so as not to jeopardize their use in future traditional ageing.

Traditional ages were generated by expert age readers for otolith and vertebra samples using secondary structures from the same fish (for Acadian redfish, the sectioned paired otolith, and for blue shark, a second vertebra from the same shark).

Spectroscopy

All samples were scanned on a Bruker TANGO-R near infrared spectrometer (Bruker, Billerica, MA). Otoliths were covered with a rubber-rimmed gold reflective stamp while scanning to reduce stray light. Vertebrae were scanned without the stamp due to size limitations.

All absorbance spectra were collected at a resolution of 64 scans every 16 wavenumbers. Acadian redfish were scanned sulcus side down with rostrum oriented horizontally on the sample window. Blue shark vertebrae were scanned in two different sample presentations: whole centra and half centra. Whole vertebrae were scanned with the focus centered over the sample window and the neural arches facing forward, and half centra were sectioned frontally through the focus and scanned with the cut side facing down, with half of the resulting "bowtie" centered over the sample window (Fig. 6.1). The models from the two vertebra presentations were compared to determine the optimal presentation for age predictions.

Data Analysis

Traditional ages were paired with FT-NIR spectra, and PLS regression models were developed for training and test sets of samples using OPUS software (Bruker Scientific). Training sets comprised a subset of randomly selected samples from the original sample collections (n = 500 for Acadian redfish, n = 88 whole centra and 94 half centra for blue shark), with samples selected proportionally to the overall age distribution of each sample set. Sample sizes for Acadian redfish models were based on preliminary work that found age prediction did not meaningfully improve with calibration set sample sizes above 500 (Rubin, unpublished data). Test sets comprised the remaining samples for each species. Acadian redfish spectra were preprocessed using a Savitsky-Golay first derivative transform with 17-point smoothing, and blue shark spectra were preprocessed using a standard normal variate transformation. The models were assessed using the coefficient of determination (r^2), root mean square error of cross validation (RMSECV) and of prediction (RMSEP), residual prediction deviation (RPD), model bias, and slope.

RESULTS

FT-NIR spectroscopy models predicted age to within 1.8 years and 1.7 years for Acadian redfish training and test sets, respectively (Table 6.1). FT-NIR spectroscopy predicted ages showed strong correlation with traditional ages ($r^2 = 0.937$), although regression fit was worse at older ages (Fig. 6.2). Average scanning time for otoliths was 1.1 minutes per sample, with an additional 0.75 minutes required for cleaning otoliths prior to scanning.

Blue shark FT-NIR spectroscopy models predicted age to within a similar prediction range (1.5-1.8 years) as otoliths but with poorer correlation of predicted and traditional age ($r^2 =$ 0.790 for whole centra, 0.812 for half centra, Fig. 6.2). Models trained on absorbance spectra of half centra predicted ages better than whole centra models (RMSECV = 1.82 years for whole centra and 1.68 years for half centra; RMSEP = 1.56 years for whole centra and 1.5 years for half centra, Table 6.1). Average scanning time for whole centra was also 1.1 minutes, but half centra required sectioning, which took an average of 5-10 minutes per vertebra, making the use of half centra for FT-NIR spectroscopy substantially slower.

DISCUSSION

Preliminary findings indicate that FT-NIR spectroscopy may be useful for predicting age in Acadian redfish and blue shark using rapid scans of otoliths and vertebrae, respectively. However, while otoliths required little preparation for scanning, blue shark vertebrae often exceeded the physical size of the spectrometer scanning window and required sectioning to facilitate better prediction results. Additional work to find optimal sample presentation for large shark vertebrae is needed. Prediction results for sharks were not as promising as those for

Acadian redfish otoliths, which may also stem from high subjectivity in vertebral ages and the lack of age validation for blue shark.

The primary advantage of using FT-NIR spectroscopy is its processing speed. While scanning time was short across both types of ageing structures, otolith cleaning added an average of 0.75 minutes per sample and vertebra sectioning added 5-10 minutes per sample. These additional processing steps reduce the efficiency gains of FT-NIR spectroscopy for ageing, although overall times are still slightly better than traditional ageing processes. Further model development and improvement in sample presentation for shark vertebrae are needed to fully vet FT-NIR spectroscopy as a potential replacement for traditional ageing.

Table 6.1. -- Fourier transform near infrared spectroscopy calibration and validation model results for Acadian redfish (*Sebastes fasciatus*) and blue shark (*Prionace glauca*). RMSECV = root mean square error of the cross validation, RMSEP = root mean square error of the prediction, and RPD = residual prediction deviation.

	Calibration						Validation					
Data	*n*	r^2	RMSECV	RPD	Bias	Slope	*n*	r^2	RMSEP	RPD	Bias	Slope
Acadian redfish	500	93.86	1.81	4.04	-0.00775	0.942	1391	93.72	1.73	3.99	0.00384	0.966
Blue shark whole centra	88	74.6	1.82	1.98	-0.0318	0.768	96	79.03	1.56	2.18	-0.008	0.805
Blue shark half centra	94	78.21	1.68	2.14	0.01	0.797	94	81.15	1.5	2.32	-0.202	0.902

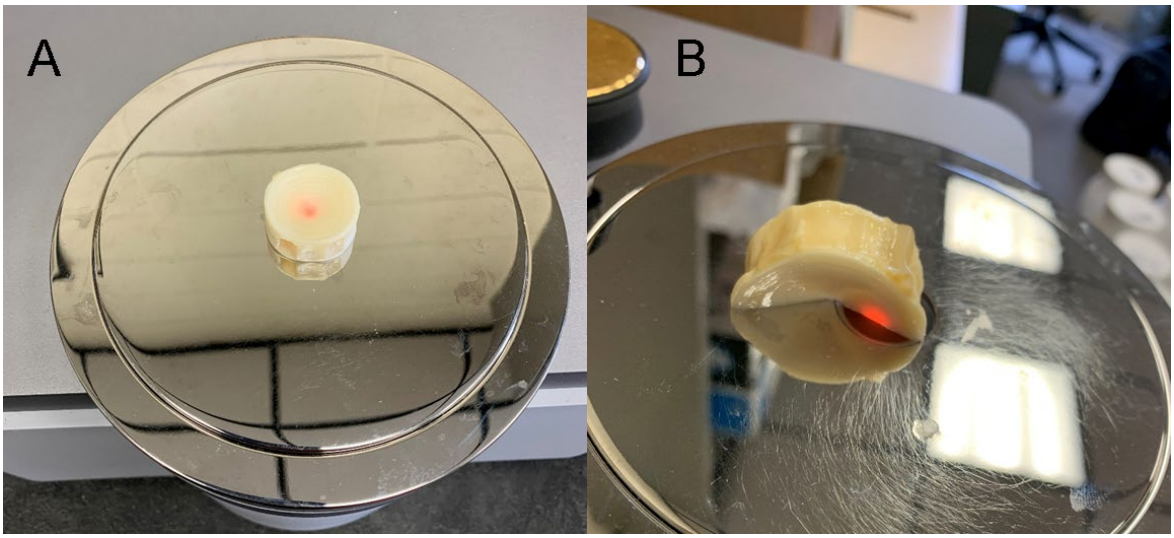


Figure 6.1. -- (A) Whole and (B) half blue shark (*Prionace glauca*) vertebral centra centered over the spectrometer scanning window.

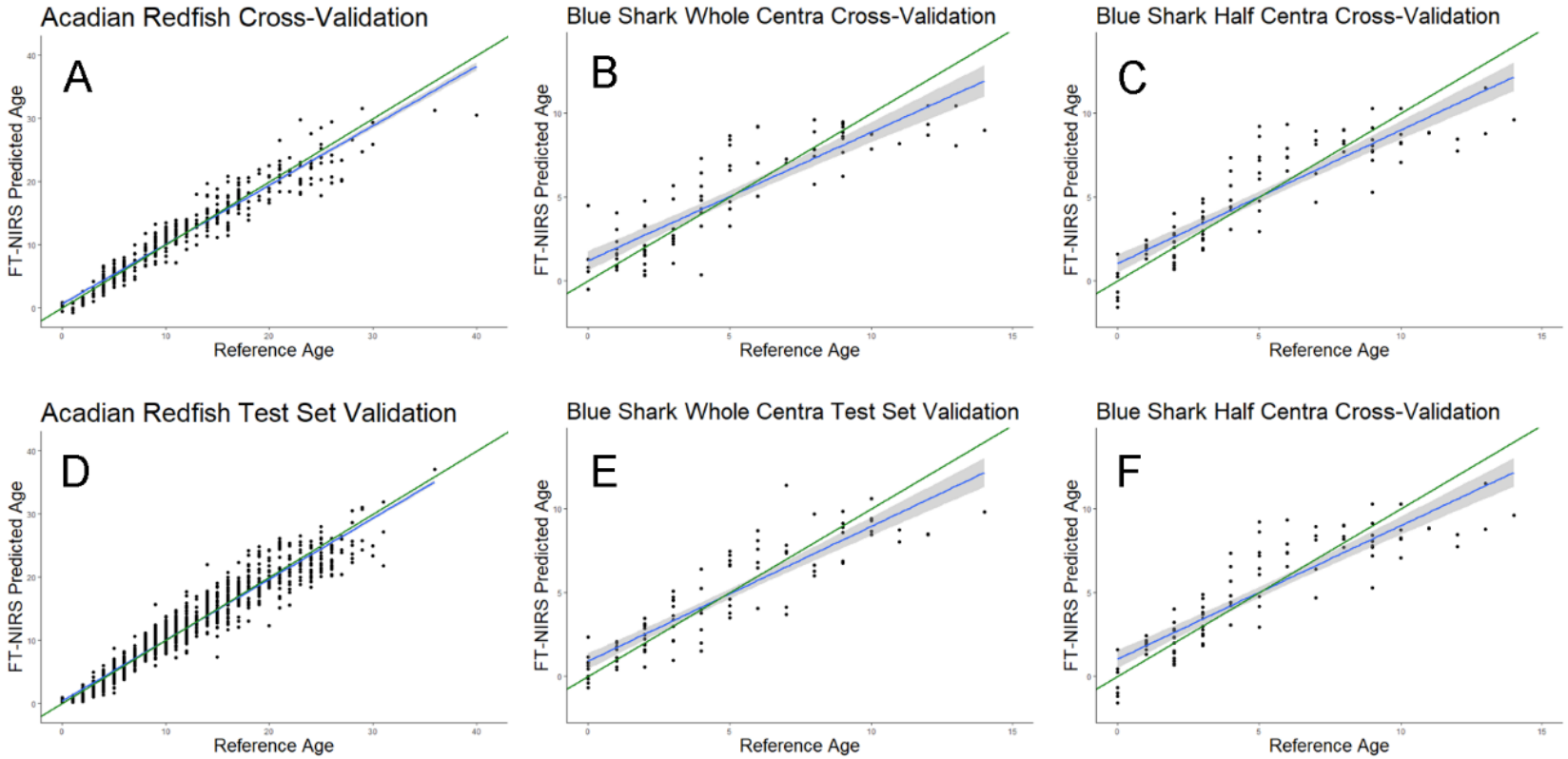


Figure 6.2. -- Cross validation models for (A) Acadian redfish (*Sebastes fasciatus*), (B) blue shark (*Prionace glauca*) whole centrum scans, and (C) blue shark half centrum scans, and test set validation models for (D) Acadian redfish, (E) blue shark whole centrum scans, and (F) blue shark half centrum scans. The green line represents a 1:1 relationship and the blue line is the regression line between traditional and Fourier transform near infrared spectroscopy (FT-NIRS) predicted age for each model. Gray shading represents the 95% confidence interval for the regression model.

7. Exploration of Fourier Transform Near Infrared Spectroscopy for the Shortbelly Rockfish (*Sebastes jordani*), an Ecologically Important Forage Fish, off the West Coast Region

Jessica J. M. Choi[1,2] and Melissa H. Monk[1]

[1] Southwest Fisheries Science Center
NOAA, National Marine Fisheries Service
110 McAllister Way
Santa Cruz, CA USA

[2] Fisheries Collaborative Program
Institute for Marine Ecosystems and Climate
University of California Santa Cruz
110 McAllister Way
Santa Cruz, CA USA

INTRODUCTION

Shortbelly rockfish are an ecologically important forage fish off the West Coast region of North America. The species' lifespan can be up to 22 years, and the maximum length is 35 cm (13.7 in) (Pearson et al. 1991). This species of rockfish can be found from La Perouse Bank, Vancouver Island, BC, to San Benito Island, Baja California. In the water column, adults can be found at depths from 91 to 491 m (300 to 1620 ft) but occur most commonly at 150 to 200 m (495 to 665 ft) (Love et al. 2002). They are generally a pelagic midwater schooling species. Shortbelly rockfish mature early; 50% of females are mature by age 3 and 100% by age 4 (Wyllie-Echeverria 1987). This species of rockfish is important to its ecosystems as it is a forage fish, often prey for salmon and other rockfishes.

Shortbelly rockfish are caught in significant numbers in regional surveys, including the West Coast Groundfish Bottom Trawl Survey (WCGBTS), the primary goal of which is to estimate abundance to support stock assessments of chilipepper and several other species of slope and shelf rockfish species along the coast (widow, chilipepper, bocaccio, and several others). The trawl survey is also used to estimate young-of-year ages (in days) and to evaluate regional and interannual variability in growth and recruitment.

The shortbelly rockfish stock was last assessed in 2007 and has further data to update models. The previous 2007 assessment was modeled using Stock Synthesis 2, which takes age and size into account. The model treats cohorts as a collection of fish that has a mean and variance of size-at-age (Field et al. 2007).

This study was initiated to explore the feasibility of using Fourier transform near infrared (FT-NIR) spectroscopy for age determination of shortbelly rockfish. Methods of using shortbelly otoliths for the FT-NIR spectroscopy Tango machine were decided upon due to having an

abundance of already-aged shortbelly otoliths. Shortbelly rockfish were chosen as the focal species of this study because they are fairly short-lived for a *Sebastes species*. Shorter-lived species have been shown to have better age prediction results from otolith spectra. Samples came from a region-wide distribution along the West Coast from a single survey source. This study will further analyze changes in growth by latitude, as well as changes in growth by sex. Samples came from the same source with the same age readers, resulting in less aging bias.

In addition, we explored the potential for relationships between growth and latitudinal indices. We also explored the suitability of how little an otolith could weigh to produce spectral scans of optimal and reproducible results.

METHODS AND RESULTS

The shortbelly rockfish otolith samples sourced from WCGBTS covered the entire West Coast region and were collected from 2011-2013 (n = 466). The available metadata included individual fish weight, length, sex, capture date, location, and depth. Shortbelly rockfish otoliths were aged at the Southwest Fisheries Science Center in Santa Cruz, CA, by Don Pearson.

This study's shortbelly rockfish samples were divided into North, Central, and South regions from Alaska to California in an effort to represent growth along the West Coast region. The study followed a standardized protocol that was specific to this species and included otolith preservation methods as well as species-specific accessory use when using the TANGO FT-NIR machine. We found that there is in fact a weight limitation of how 'small' an otolith could weigh for this study. A Teflon disc with a 4-mm aperture was used to scan shortbelly rockfish otoliths.

We also evaluated time efficiency of FT-NIR scanning as an aging method relative to traditional methods for shortbelly rockfish. For the traditional 'break and burn' method, one

otolith takes 5 to 10 minutes to prepare and age, resulting in 20 otoliths aged in ~1 hour. Using the Tango, one otolith scan takes ~45 seconds, resulting in 20 otoliths aged in ~15 minutes. If calibrated for the species, this method can effectively increase the number of otoliths aged.

Concluding the initial timeframe of the study, the stages of modeling are still exploratory. Goals are to see which models produce the best analysis results given the data we have.

FUTURE DIRECTIONS

Future research objectives include the following:

- Include sex as a predictor of age to determine if it is a significant factor.
- Many *Sebastes* species exhibit dimorphic growth. Evaluate whether separate PLS models are needed for each sex.
- Quantify ageing error between traditional ages and FT-NIR spectroscopy.
- Continue exploring model selection and covariates.
- Explore cross-validation methods.
- Explore patterns in growth by latitude, which may have implications for species management.
- Determine whether there is increased error in age estimates by latitude. Rockfish are becoming more difficult to age in the southern region of the California Current, where seasonality is weaker, and intraseasonal and interannual variability is stronger in comparison to latitudinal regions moving farther north.

8. Investigating the Use of FT-NIR Spectroscopy to Age Gag Grouper (*Mycteroperca microlepis*), a Protogynous Hermaphroditic Species

Beverly K. Barnett[1], Irina M. Benson[2], Thomas E. Helser[2], Susan K. Lowerre-Barbieri[3], and Hayden S. Menendez[4]

[1] Southeast Fisheries Science Center
NOAA, National Marine Fisheries Service
3500 Delwood Beach Road
Panama City, FL USA

[2] Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

[3] School of Forest, Fisheries, and Geomatics Sciences
University of Florida
100 8th Avenue SE
St. Petersburg, FL USA

[4] Fish and Wildlife Research Institute
Florida Fish and Wildlife Conservation Commission
100 8th Avenue SE
St. Petersburg, FL USA

ABSTRACT

Fish age is a critical data input for estimating longevity, growth, maturity functions, and mortality in age-structured stock assessment models. However, providing these age estimates is a labor-intensive process and can often lead to delays in providing age data for an assessment. Fourier transform near infrared (FT-NIR) spectroscopy may provide a more rapid and objective method for providing age estimates from fish otoliths and perhaps move fish ageing closer to providing real-time data for stock assessments. In this study, FT-NIR spectroscopy was applied to otoliths from gag, *Mycteroperca microlepis*, collected from the U.S. Gulf of Mexico (GOM). Gag is a protogynous hermaphroditic species that supports important commercial and recreational fisheries in the GOM. Objectives of this study were to: 1) investigate FT-NIR spectroscopy to estimate fish age from otoliths of gag, and 2) explore potential patterns in FT-NIR spectra from a small set of otoliths for which histological sex had been assigned to investigate if FT-NIR spectra would differ among males and females. Whole otoliths were scanned with an FT-NIR spectrometer and traditional age estimates assigned from whole or sectioned otoliths. The spectra and corresponding age data were subset into calibration and test sets, and partial least squares regressions (PLS) fit to each of the two sets. Three types of ages (annulus count, calendar age, and fractional age) were input into separate PLS predictive models to investigate which age type would provide the best performing model. PLS models that used fractional ages slightly outperformed the other two age types for both the calibration and test sets (calibration n = 1,230: r^2 = 0.89, RMSECV = 0.888, test n = 809: r^2 = 0.90, RMSEP = 0.947). Results from principal component analysis of FT-NIR spectra by sex (n = 30) showed some separation between females and males; however, this was only a small set of data, so increasing sample size will be the next step to see if this pattern continues to hold. Additional research and

development are required before this technology can be fully implemented in production ageing for fisheries management, but results from ongoing case studies show promise.

INTRODUCTION

Fish age is a critical data input for estimating longevity, growth, maturity functions, and mortality in age-structured stock assessment models. However, these ages are often subject to poor reader precision, which can lead to ageing error being propagated through stock assessment models (Beamish and McFarlane 1983; Campana 2001). Furthermore, conventional methods (e.g., microscopic interpretation) for providing these age estimates is a labor-intensive and time-consuming process that can often lead to delays in providing age data for an assessment. Oftentimes, due to the time-consuming process of preparing age structures and estimating age, subsampling strategies are needed to provide a subset of age estimates for stock assessment. To meet the need for more stock assessments or to provide more real-time age data to support the Fishery Management Councils, more efficient alternatives to the conventional methods for providing age estimates are needed.

In recent years, Fourier transform near infrared (FT-NIR) spectroscopy has been applied to predict fish age from otoliths (Wedding et al. 2014, Helser et al. 2019b, Passerotti et al. 2020b) and skate vertebrae (Arrington et al. 2022), to classify fish species from different ecosystems (Benson et al. 2020), to assess fish condition (Goldstein et al. 2021), and to predict spawning status (TenBrink et al. 2022). While relatively new to applications in fish ecology, FT-NIR spectroscopy is a well-established method that has been used across a wide range of industries (e.g., agricultural, pharmaceutical, dairy products, etc.) to ensure quality control/quality assurance. It is a rapid, non-destructive process that requires little to no sample preparation. Applications of FT-NIR spectroscopy, whether in industry or fisheries science, rely on developing a calibration model that relates a feature of interest (e.g., analyte concentration, fish age, etc.) to an NIR spectrum. The calibration model is then applied to another set of data

(i.e., test) to determine how well the calibration model will predict unknown data. For predicting fish age, predictive calibration models rely on age estimates assigned by either an individual reader or from consensus ages to provide the reference age. Thus far, FT-NIR spectroscopy has been applied to predict ages for red snapper, *Lutjanus campechanus* (Passerotti et al. 2020b) and gray snapper, *Lutjanus griseus* (Barnett, unpublished data) from the Gulf of Mexico (GOM), and NIR-predicted ages from each species has shown promising results. Therefore, to further investigate FT-NIR's applicability for predicting fish age, otoliths from other species need to be scanned and predictive models built to determine the feasibility of applying this technology to predict ages for other species.

Gag, *Mycteroperca microlepis*, are a moderately long-lived species (maximum observed age 33 years; SEDAR 2021) that support economically important commercial and recreational fisheries in the GOM. As juveniles, gag generally settle into seagrass beds (Switzer et al. 2012) and then make an ontogenetic shift offshore to deeper reefs (Bullock and Smith 1991, Carruthers et al. 2015). However, gag are predominantly spatially segregated during most of the year until adult females and males come together to spawn (Lowerre-Barbieri et al. 2020). Gag are a protogynous hermaphrodite (i.e., born as female but change to male), and most commercially-landed fish are not sexed during dockside sampling since the fish are usually gutted at sea. Gag are federally managed and highly regulated in the GOM (GMFMC 2013), and their stock status is currently listed as overfished and undergoing overfishing (SEDAR 2021). Most otoliths from gag are aged whole, given that their annuli are more easily identified from whole otoliths rather than sectioned otoliths. Only otoliths from older fish and those with annuli that are difficult to interpret are typically sectioned. Therefore, it was of interest to predict fish age using FT-NIR spectra measured from whole otoliths of gag.

The objectives of this study were to 1) investigate FT-NIR spectroscopy to estimate fish age from otoliths of gag, and 2) explore potential patterns in FT-NIR spectra from a small set of otoliths for which histological sex had been assigned to investigate if FT-NIR spectra would differ among males and females.

METHODS

Sample Collection

Sagittal otoliths from gag were collected from the GOM for years 2017-2019 by fishery-dependent and fishery-independent sources. Otoliths were stored dry in coin envelopes and archived at the National Marine Fisheries Service, Panama City Laboratory (PCL). An additional set of gag otoliths, for which sex was histologically assigned, collected from 2018-2019 by Florida Fish and Wildlife Research Institute (FWRI) were sent to PCL to investigate if sex could be identified based on FT-NIR spectra. Whole otoliths were weighed to the nearest 0.1 mg. Most otoliths were aged whole, and remaining otoliths from older fish and those with annuli that were difficult to interpret were sectioned using a Hillquist Thin Section Machine (Hillquist Inc.). Calendar ages were based on annulus counts and marginal edge code assignment. For all fish with a wide translucent zone (i.e., translucent zone > 2/3 of new growth after the last observed opaque zone) and landed between 1 January and 30 June, annulus count was advanced by one year. Calendar age for all other fish was set equal to the annulus count. Fractional (i.e., biological) age was calculated using calendar age, date of peak spawning (1 March), and date of capture using the following equation:

$A_F = A_C + ((D_C - D_S)/365)$, where

A_F = fractional age (years)

A_C = calendar age (years)

D_C = date of capture

D_S = date of peak spawning

FT-NIR Spectroscopic Measurements

Otoliths were scanned using a Bruker Multi Purpose Analyzer II (MPA II) FT-NIR spectrometer. OPUS software (version 8.5, Bruker Optics, Ettlingen, Germany) was used to acquire diffuse reflectance measurements of all otoliths. Otoliths were removed from coin envelopes, placed on the integrating sphere in a concave up position, and covered with a gold-coated reflector stamp (see Benson et al. (2020) for more details). Spectral absorbances for all otoliths were acquired between 11,500 and 4,000 cm^{-1}, averaged across 64 scans, and with a scan resolution of 16 cm^{-1}. Spectral data for the ageing component of this study were preprocessed with first derivative, second order polynomial, and a 17-point Savitzky-Golay smoothing function. Otoliths received from FWRI were preprocessed using first derivative plus mean centering.

Data Analysis

For the ageing part of this study, a total of 2,060 samples were selected; however, 21 samples had to be excluded from further analysis due to adhering tissue remaining on the otolith surface. A linear regression was fit to fractional age and left otolith weight. The spectra and

corresponding age data were divided into calibration and test sets using OPUS software. The calibration set consisted of 60% of the samples (n = 1,230), and the test set consisted of 40% (n = 809). Partial least squares regressions (PLS) were fit to each of the two sets. Calibration models, which used a cross validation method (i.e., "leave-one-out"), were evaluated using the root mean square error of cross validation (RMSECV) to determine the predictive power on new data, and the test models were evaluated using the root mean square error of prediction (RMSEP) as a measure of prediction error. Three types of ages (annulus count, calendar age, or fractional age) were input into separate PLS predictive models to investigate which age type would provide the best performing model. Percent agreement for the two methods, traditional observed ages and FT-NIR predicted ages, was investigated using calendar ages. However, to compare between these two methods, FT-NIR predicted ages, which are output as continuous numbers, were rounded to an integer age. FT-NIR predicted ages with a decimal of 0.5 or higher were rounded up to the next integer (e.g., 3.5 rounded to 4), and those with a decimal age <0.5 were rounded down (e.g., 3.4 rounded to 3). Principal component analysis (PCA) using Solo+MIA 9.1 (Eigenvector Research, Inc., Manson, WA USA) chemometrics software was used to investigate patterns in the spectra from otoliths (n = 30) sent by FWRI.

RESULTS AND DISCUSSION

Calendar ages for the PCL samples ranged from 2 to 26 years, and fork lengths ranged from 320 to 1,226 mm (Fig. 8.1A). The linear regression showed a high correlation (r^2 = 0.86, n = 1,333) between fractional age and left otolith weight (Fig. 8.1B). PLS models that used fractional ages slightly outperformed the other two age types (annulus count, Figs. 8.2A and 8.2B, and calendar ages, Figs. 8.2C and 8.2D) for both the calibration and test sets (calibration, n

= 1,230: r^2 = 0.89, RMSECV = 0.888, test n = 809: r^2 = 0.90, RMSEP = 0.947; Figs. 8.2E and 8.2F). Most otoliths from gag are aged whole given that their annuli are more easily identified compared to sectioned otoliths from gag. Percent agreement analysis showed that the traditional reader age had a higher (>70%) agreement for 0 difference in calendar age than the FT-NIR-predicted age (~50%; Fig. 8.3). However, it is possible that percent agreement of the FT-NIR-predicted age could be affected by the age rounding process, given that some ages are rounded either up or down. To explore this further, a set of traditional reader ages using 3-year-old fish as an example (n = 57) were plotted against capture date represented as day of year (Fig. 8.4). Results from this exploratory analysis showed that rounding to the nearest integer caused these 3-year-old fish to be placed in four age classes: age-2 (n = 1); age-3 (n = 35); age-4 (n = 17), and age-5 (n = 4) (Fig. 8.4A). Exploring this rounding scenario further, this same set of 3-year-old fish (n = 57) with decimals <0.99 were rounded down to the next lower integer (e.g., 3.99 rounded to 3), which placed slightly more (n = 38) of the 3-year-olds in their respective 3-year age class (Fig. 8.4B). Since most stock assessments rely on calendar ages to assess recruitment and year class strength, a better method and understanding is needed for converting the FT-NIR-predicted ages (i.e., continuous numbers) to an integer age.

There were 14 females, 14 males, and two transitional gag included in the FWRI data set. A plot of absorbance by wavenumbers showed otoliths from males having higher absorbance than females (Fig. 8.5A). Spectra from otoliths of the two transitional fish were intermixed with spectra from the otoliths of females (Fig. 8.5A). Results from the PCA of FT-NIR spectra by sex showed some separation between histologically-assigned males and females (Fig. 8.5B). However, this was only a small sample set of 30 samples, so increased sample size will be the next step to see if this pattern continues to hold. One potential factor that may be driving this separation is otolith weight, since otoliths from males tended to be heavier compared to females

and transitional gag. Passerotti et al. (2020b) found that for most otoliths in their study, FT-NIR spectra predicted otolith weight within a milligram or less, but they went on to conclude that other factors may also be influential in driving this relationship. It is possible that the proteins incorporated into an otolith could provide markers for development and physiological change during a fish's life (Thomas et al. 2019). Indeed, otolith proteomics data suggests otoliths may archive sex-specific proteins and hormones (Oliver Thomas, personal communication). Since FT-NIR spectroscopy is often used to identify protein content from a sample (Ingle et al. 2016), it may be useful in differentiating sexes from otoliths based on sex-specific proteins incorporated into the organic matrix of an otolith (Rideout et al. 2023). A recent proteomics study on otoliths from Atlantic cod, *Gadus morhua*, revealed at least two proteins (*SPATA6* protein in males and *Vitellogenin-2-like* in females) that are sex-related biomarkers (Rideout et al. 2023). If the spectral patterns observed for females and males in the current study should continue to hold as more otoliths are scanned from gag that have a histological sex assigned, it may be possible that sex could be assigned from otoliths, particularly for commercially-landed fish that are gutted at sea and no field sex can be assigned by the dockside samplers. Sex ratios for gag are difficult to assign given that females and males exhibit spatial segregation until coming together to spawn (Lowerre-Barbieri et al. 2020). For a hermaphroditic species such as gag, it could provide a method to better understand the percentages of males and females that are being caught in the fishery.

Overall, FT-NIR predicted ages for gag collected from the GOM tended to be in good agreement with the traditional reader ages, which provide promising results, but more otoliths need to be scanned, particularly for younger and older ages. Since stock assessments generally rely on calendar ages to assess recruitment and year class strength, additional research and perhaps stock assessment simulations are needed to find a best practice for converting the FT-

NIR-predicted ages, which are continuous numbers, to an integer age. Most otolith spectra showed an increase in absorbance in specific wavenumber regions as age increased. It is possible that this increased absorbance may be due to proteins, which are incorporated into the otolith matrix (Thomas et al. 2019). To investigate this further, research on otolith proteomics for red snapper, *Lutjanus campechanus*, and walleye pollock, *Gadus chalcogrammus*, is currently underway. These two species were chosen given that their genomes have been sequenced. The goal for the proteomics research is to try and learn more about what FT-NIR spectroscopy measures from an otolith that is so closely related to age. Having a better understanding of what FT-NIR spectroscopy is measuring from an otolith will perhaps provide guidance for building better predictive models and provide knowledge about the uncertainty or error around an FT-NIR predicted age. Additional research and development are required before this technology can be fully implemented in production ageing for fisheries management, but results from ongoing case studies, such as those presented here for gag, show promise.

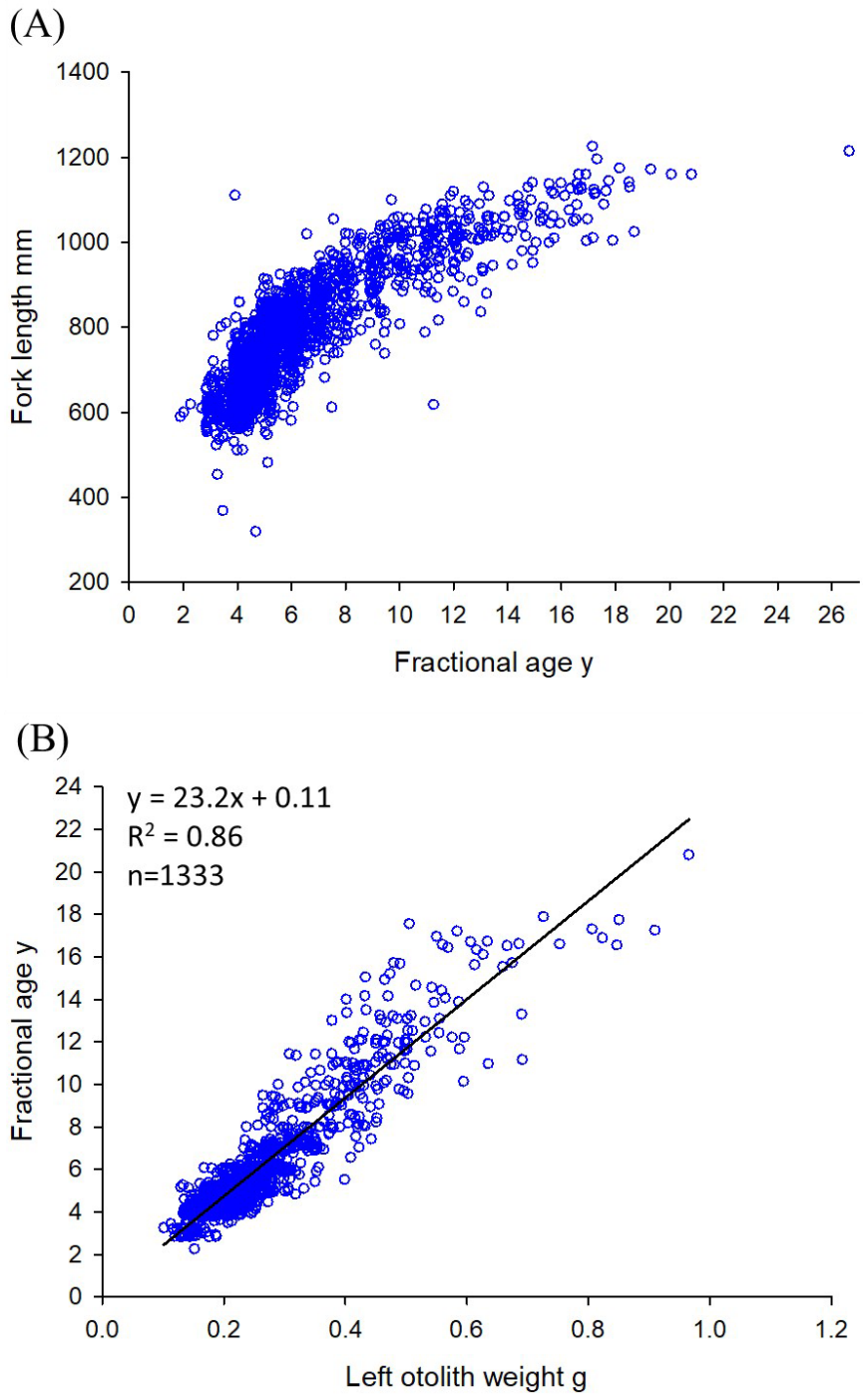


Figure 8.1. -- Scatterplots of A) fork length versus fractional age, and B) fractional age versus left otolith weight for Gulf of Mexico gag, *Mycteroperca microlepis*.

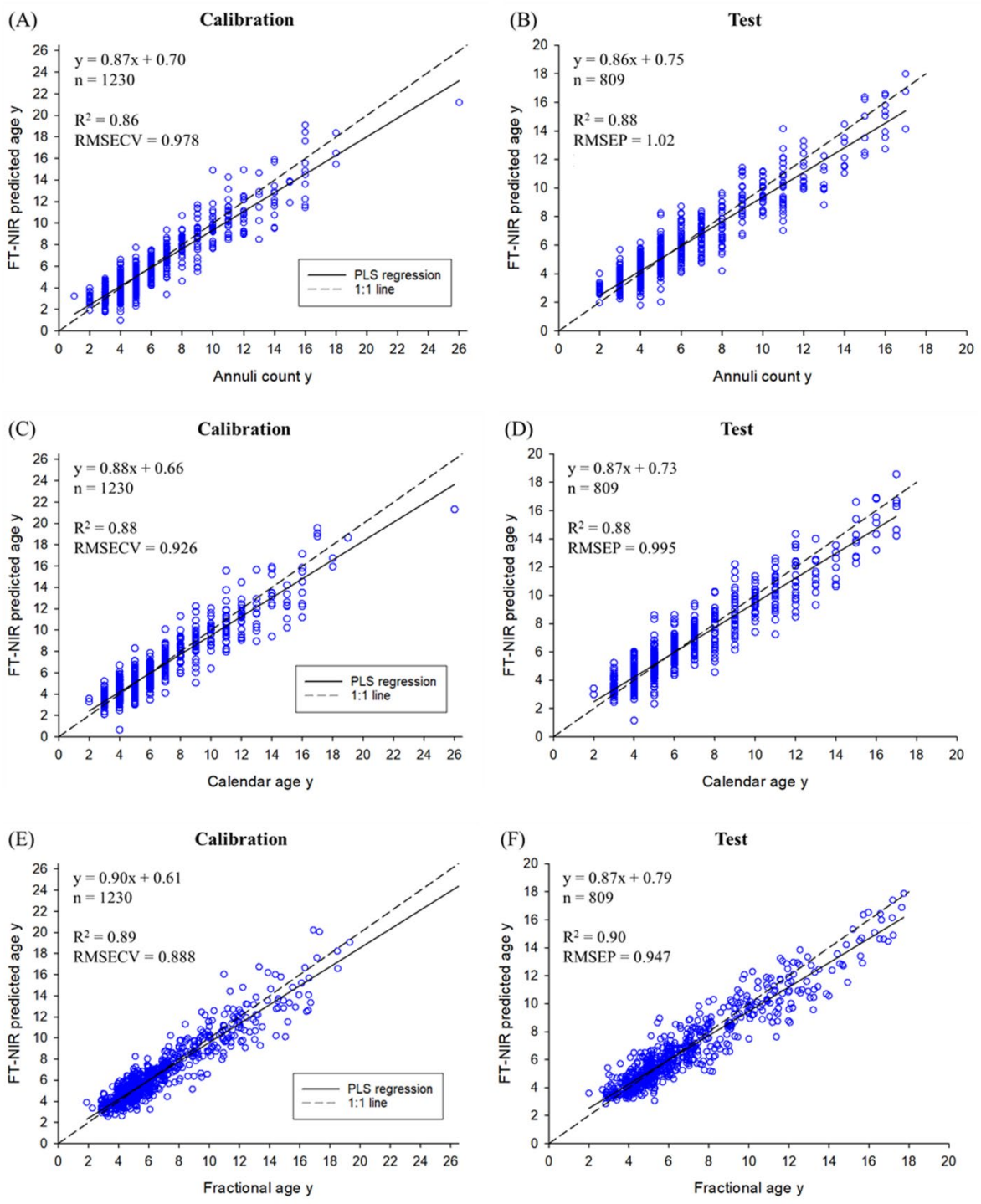


Figure 8.2. -- Results from partial least squares (PLS) regression for A) calibration and B) test sets using annuli count as reference age, C) calibration and D) test sets using calendar age as reference age, and E) calibration and F) test sets using fractional age as reference age for Gulf of Mexico gag, *Mycteroperca microlepis*. R^2 = coefficient of determination; RMSECV = root mean square error of cross validation; RMSEP = root mean square error of prediction.

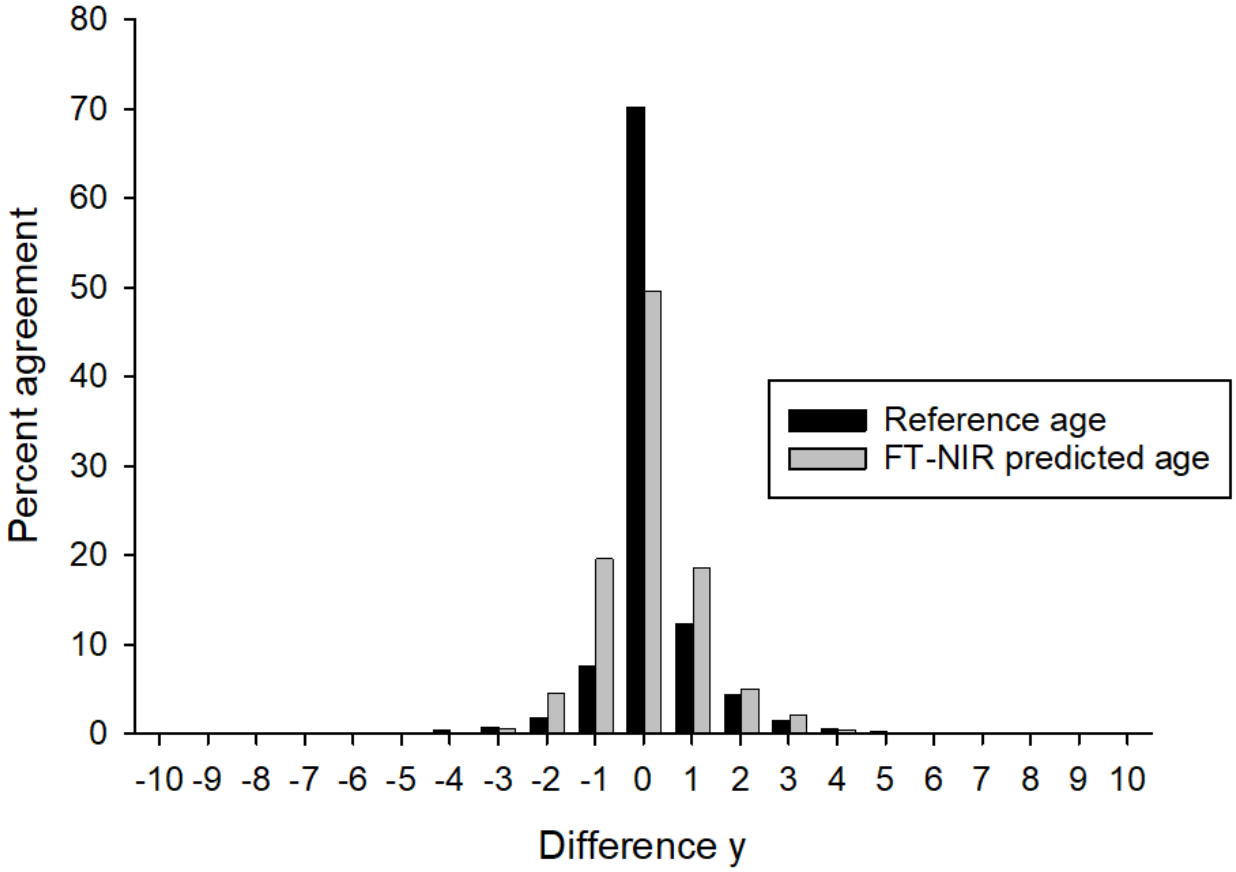


Figure 8.3. -- Percent agreement between two age readers and Fourier transform near infrared (FT-NIR) predicted age (gray bars) and reference age (black bars) for Gulf of Mexico gag, *Mycteroperca microlepis*.

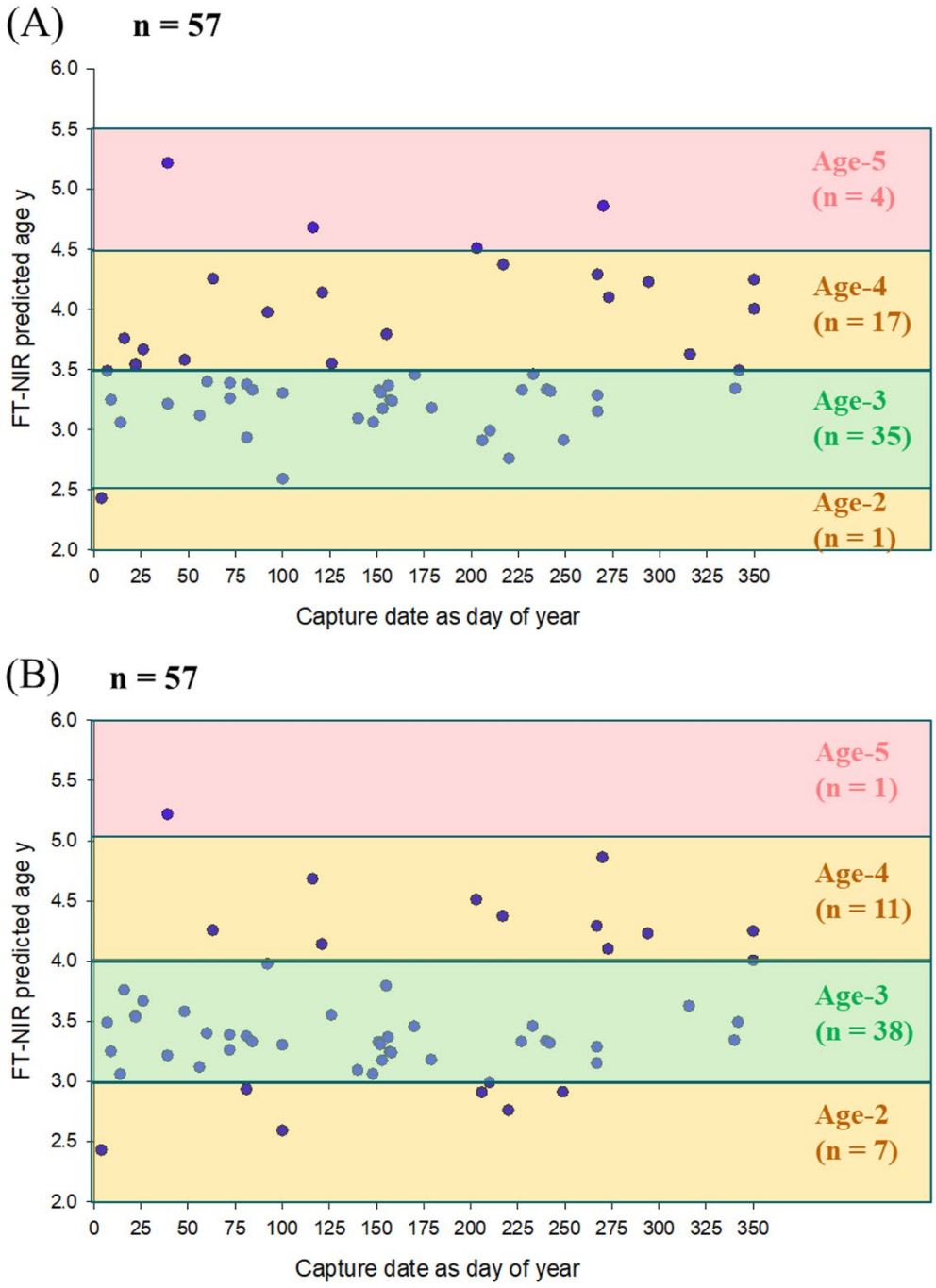


Figure 8.4. -- Exploratory analysis for a set of 3-year old gag (*Mycteroperca microlepis*) showing the potential impact of rounding Fourier transform near infrared (FT-NIR) predicted ages A) up (decimal age ≥ 0.5) or down (decimal age < 0.5), and B) rounding down (decimal age ≤ 0.99) to the next lower integer.

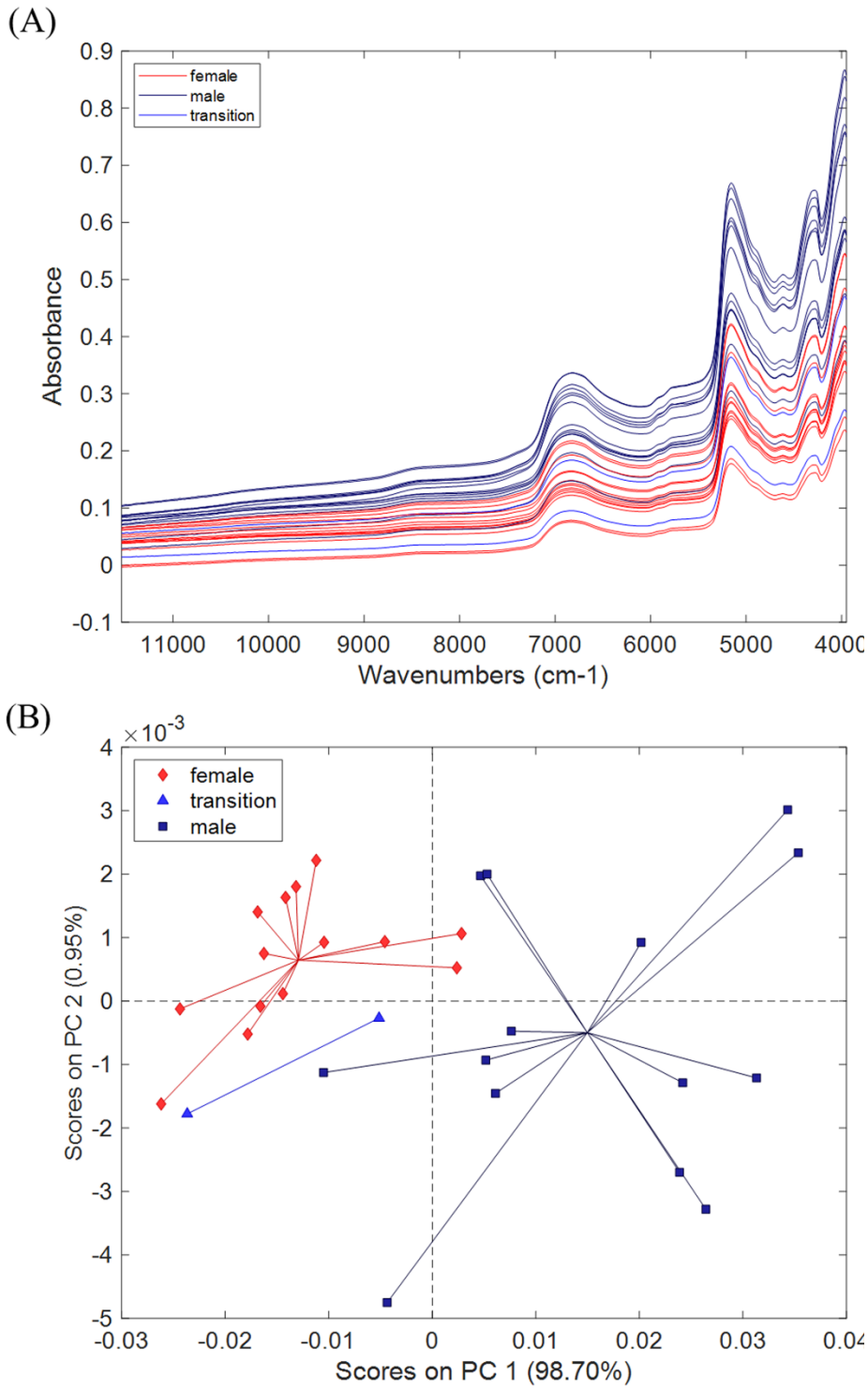


Figure 8.5. -- A) Fourier transform near infrared (FT-NIR) spectra, and B) principal component analysis of FT-NIR spectra from otoliths of gag (*Mycteroperca microlepis;* n = 30) colored by histologically-assigned sex.

9. Exploring the Use of FT-NIR Spectroscopy for Ageing Sablefish (*Anoplopoma fimbria*) and Pacific Hake (*Merluccius productus*) off the U.S. West Coast

John R. Wallace

Northwest Fisheries Science Center
NOAA, National Marine Fisheries Service
2725 Montlake Boulevard E
Seattle, WA USA

INTRODUCTION

Otolith scanning using Fourier transform near infrared (FT-NIR) spectroscopy was used to explore the chronological age of sablefish (*Anoplopoma fimbria*) (n = 1,357) and Pacific hake (*Merluccius productus*) (n = 2,337) caught off the West Coast of the contiguous U.S. in 2019. This non-destructive spectroscopic scanning procedure was directly compared to the slower traditional method of ageing (TMA). The TMA for both sablefish and Pacific hake is the standard 'break and burn' method for otoliths. Using FT-NIR spectroscopy for ageing otoliths was first espoused by Wedding et al. (2014) and further advocated by Helser et al. (2019b).

METHODS

Near infrared spectrometry produces spectra that show the intensity of the response of the molecular structure to different wavelengths on the infrared scale (Fig. 9.1). To start, an iterative partial least squares (iPLS) model was used to define ranges of spectra with the most information. Partial least squares (PLS) models applied to those informative iPLS ranges gave lack-luster results. Hence neural net models using the same data were researched, deployed in R, and directly compared to the PLS models.

Keras software (https://keras.io) that sits on top of Google's TensorFlow software (https://www.tensorflow.org) is an industry standard for neural net models and was selected for this task. A complete port of Keras is available as an R package (https://tensorflow.rstudio.com; RStudio is not needed). On the penultimate validation step, the main data was split into the training set and the test set, using a 2/3, 1/3 split. On the last validation step, a k-fold model with 10 levels was employed. For the k-folding, one tenth of the data was left untouched for testing

the 2/3, 1/3 model which was run on the remaining 90% of the data. This was repeated for each one tenth of data until each otolith had a predicted age using a model that did not contain the otoliths for which the age was being predicted. Starting with different pseudo random numbers, the entire k-folding process was repeated 20 times. Medians across predicted ages for each otolith were then taken as each complete k-fold model was added sequentially, for a total of 20 medians. Across all otoliths, various summary statistics that measure how close the median of the predicted ages were to the respective TMA were calculated. Along with the original order, five different randomized orders of the additions over the 20 complete k-folds were looked at to gauge the rate of improvement as the number of randomized k-folds added to the median increased.

RESULTS AND DISCUSSION

For sablefish, the original run order and five other randomized order additions of 20 complete k-fold models are shown in Figure 9.2. In the original run order, the 12th model addition had the best stats. The median over those 12 complete k-fold neural net models versus the TMA is shown in Figure 9.3. The root mean square error (RMSE) was 3.527 years and the sum of absolute differences (SAD) was 2,789.

For the relatively short-lived Pacific hake, the original run order and five other randomized order additions of 20 complete k-fold models are shown in Figure 9.4. In the original run order, the 20th model addition had the best stats. The median over those 20 complete k-fold neural net models versus the TMA is shown in Figure 9.5. The RMSE was 0.773 years and the SAD was 882.

In the Figures 9.3 and 9.5, note that the SAD equals the sum of the correctly matched otoliths (zeros, in red), plus 1 for each estimated age off by one year from the TMA (yellow), plus 2 for each estimated age off by two years (green) from the TMA, and so on. The SAD is well suited to rounded age data and works well to compare runs within a dataset. Note that standardized SAD and RMSE are closely correlated, as seen in Figures 9.2 and 9.4.

In the future not only will medians be calculated, but also corresponding lower (0.025) and upper (0.975) quantiles based on the 20 random replicates will be calculated, reported, and plotted as error bars in a credible interval type figure. The quantiles will reflect the NN models precision based on the 20 random replicates, not the accuracy to a TMA age.

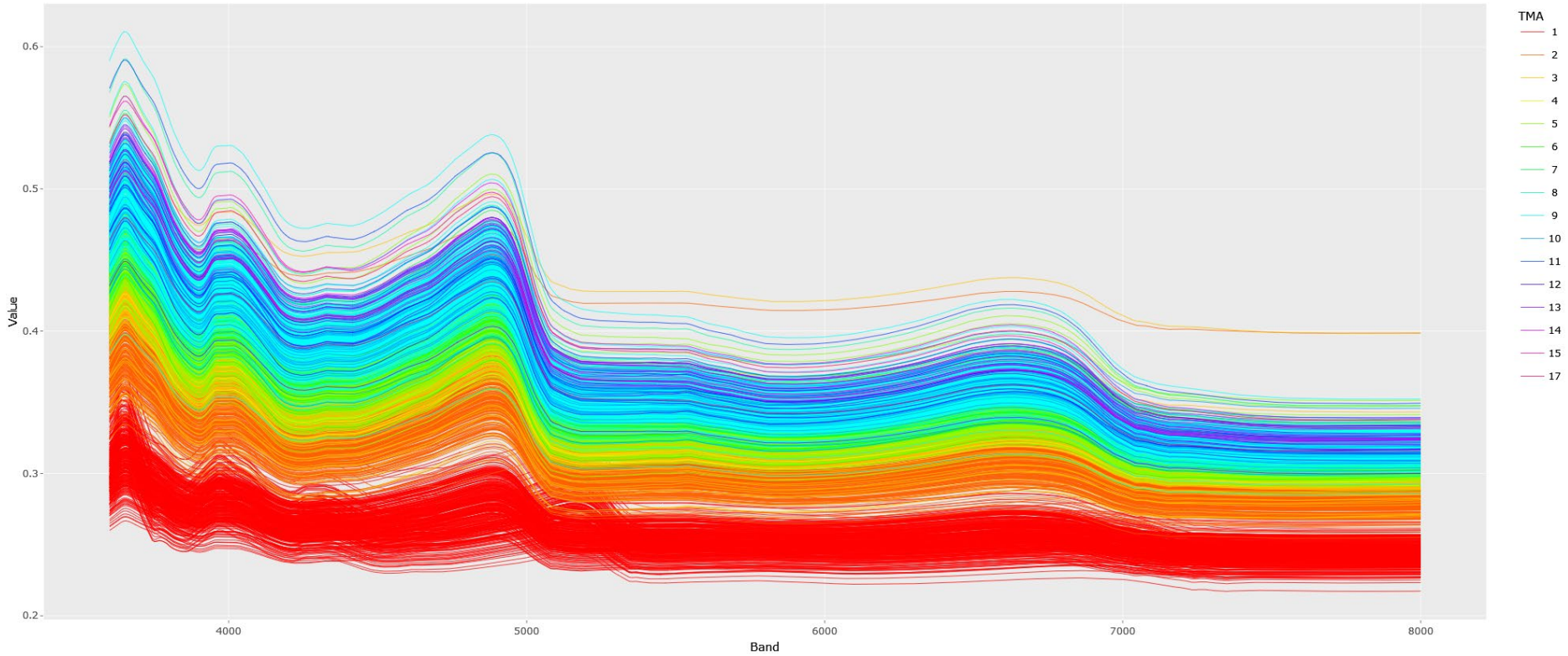


Figure 9.1. -- Raw spectra of 2019 Pacific hake (*Merluccius productus*) otoliths scanned on a Bruker MPA spectrometer. The x-axis represents wavenumber and the y-axis represents absorbance.

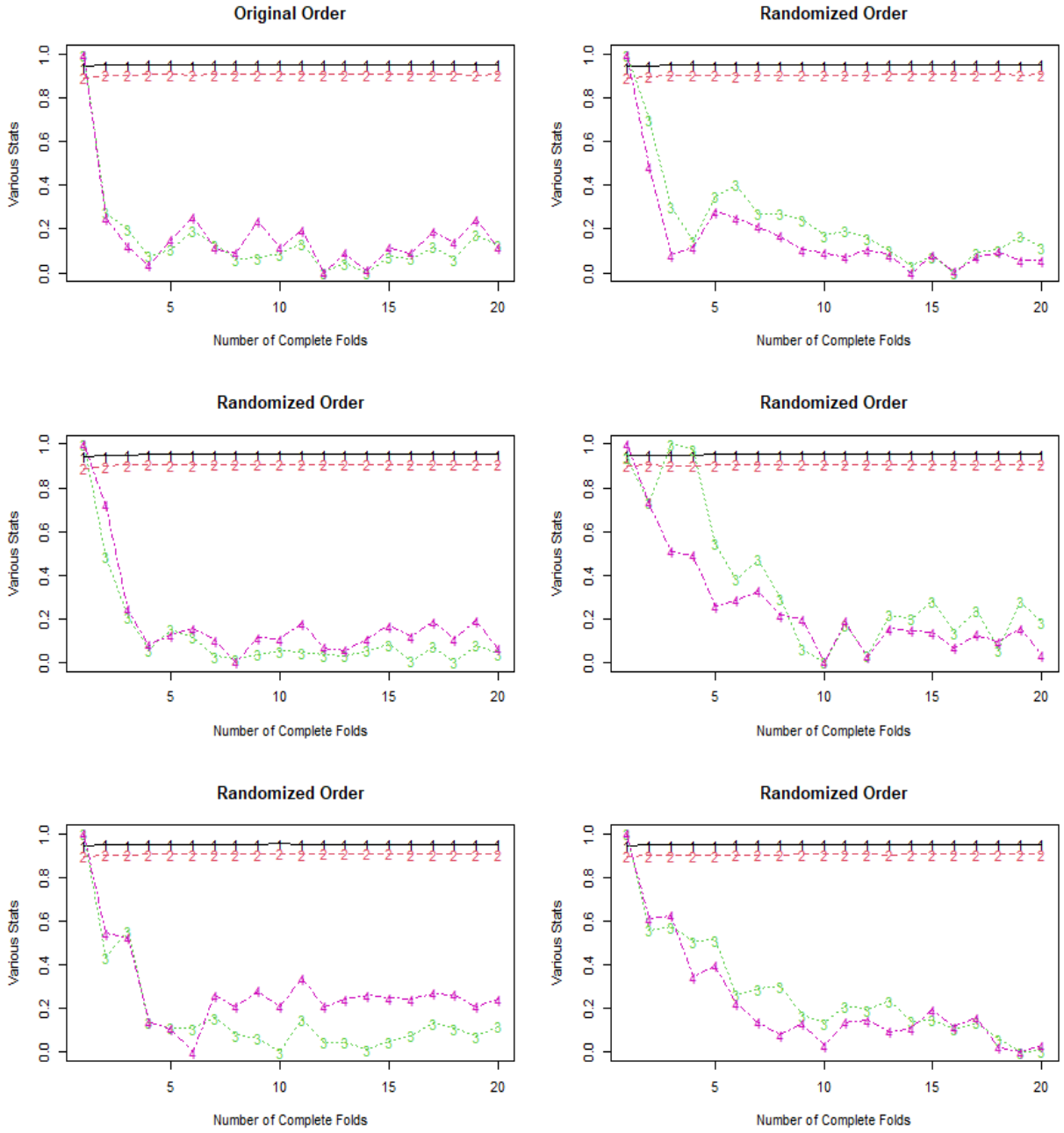


Figure 9.2. -- Randomized additions of 20 complete k-fold models for sablefish (*Anoplopoma fimbria*). Correlation values are shown in black, r^2 values are shown in red, the standardized root mean square error values are shown in green, and the standardized sum of absolute differences is shown in purple. In the original run order, the 12th model addition had the best statistics.

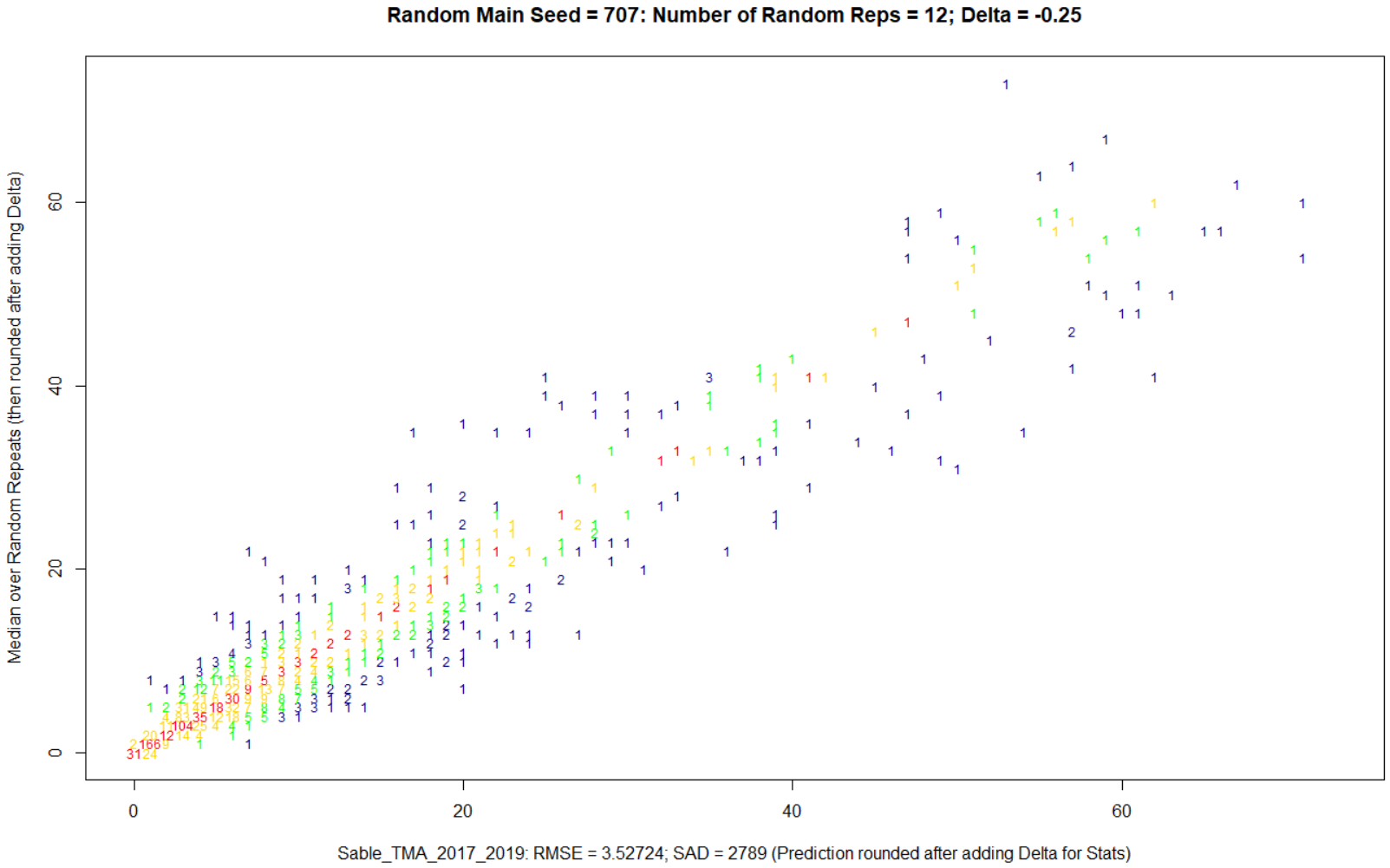


Figure 9.3. -- Median over the first 12 complete k-fold models for sablefish (*Anoplopoma fimbria*) in the original order. Using the first 12 models gave the best statistics (see Figure 9.2).

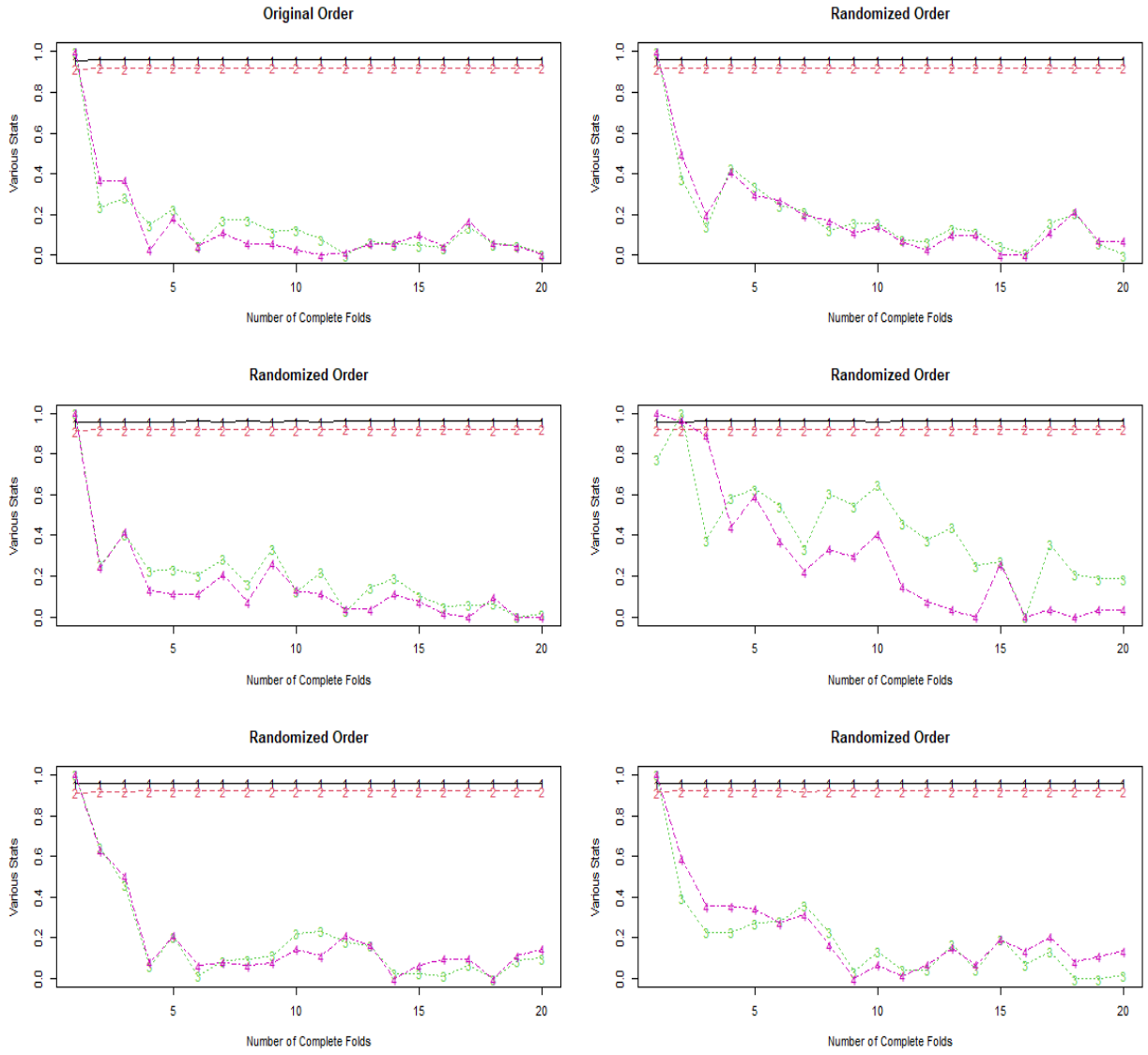


Figure 9.4. -- Randomized additions of 20 complete k-fold models for Pacific hake (*Mercluccius productus*). Correlation values are shown in black, r^2 values are shown in red, the standardized root mean square error values are shown in green, and the standardized sum of absolute differences is shown in purple. In the original run order, the 20th model addition had the best statistics.

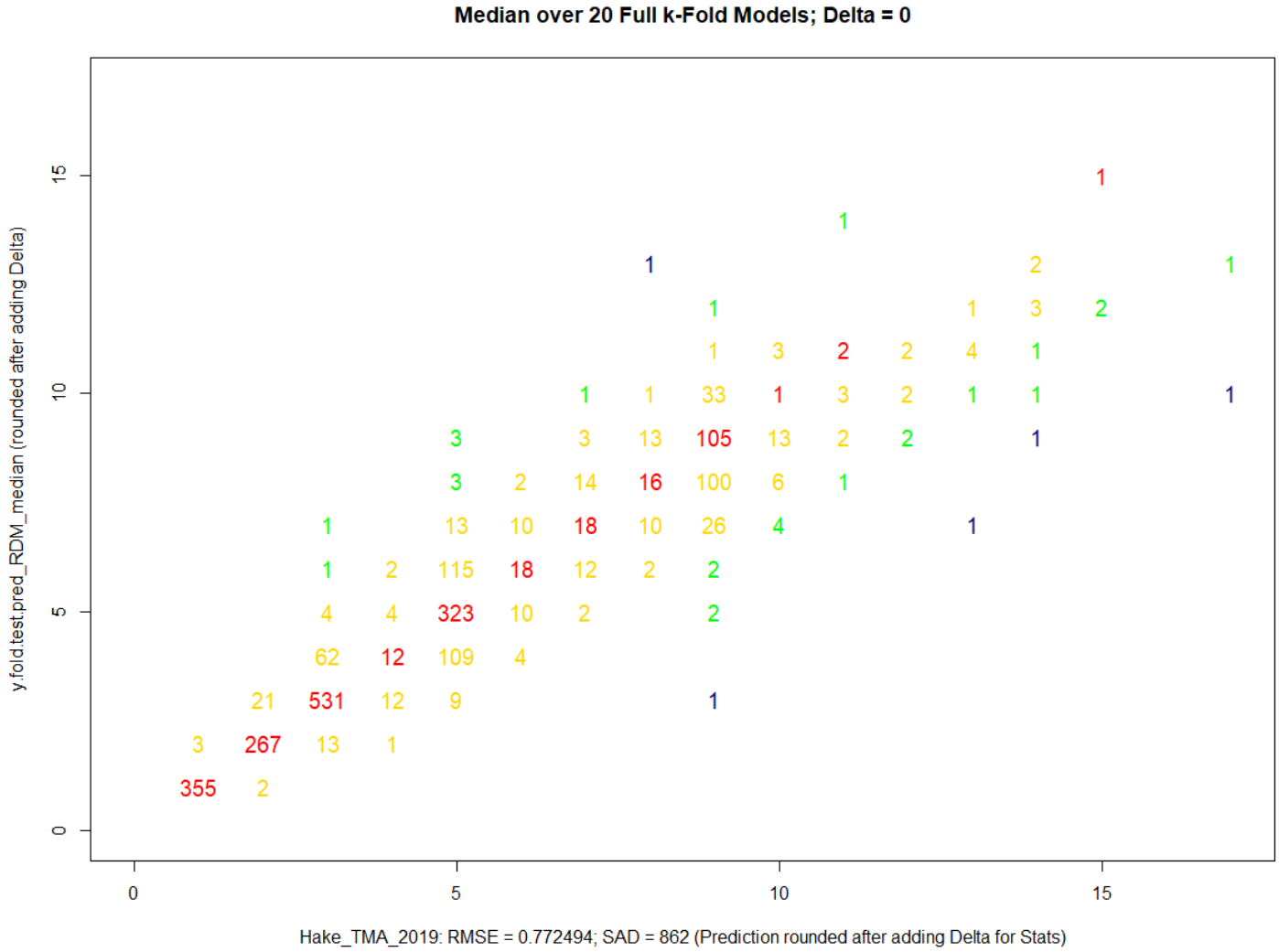


Figure 9.5. -- Median over all 20 complete k-fold models for Pacific hake (*Merccluccius productus*) in the original order. Using all 20 models gave the best statistics (see Figure 9.4).

10. Developing Spectroscopy Approaches to Measure Life History Characteristics of Fish Throughout Ontogeny

Esther D. Goldstein[1], Charles D. Waters[2], Heather K. Fulton-Bennett[2], Mary E. Matta[1], Thomas E. Helser[1], Johanna J. Vollenweider[2], Christine M. Hinds[3], Kevin W. McNeel[3], Craig R. Kastelle[1], Sandi K. Neidetcher[1], Dion S. Oxman[3], and Franz J. Mueter[4]

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

[2] Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
17109 Point Lena Loop Road
Juneau, AK USA

[3] Mark, Tag and Age Laboratory
Alaska Department of Fish & Game
10107 Bentwood Place
Juneau, AK USA

[4] College of Fisheries and Ocean Sciences
University of Alaska Fairbanks
17101 Point Lena Loop Road
Juneau, AK USA

INTRODUCTION

Changing environmental conditions necessitate adaptive management strategies that require timely and comprehensive collection of life history information to be effective (Holsman et al. 2019, Shotwell et al. 2022). Despite links with population productivity and recruitment, information on daily age, body condition, and reproductive biology is often limited in scope or inconsistently incorporated into management because data collection is laborious. The discrepancy between data requirements and collection limitations justify the exploration of new technologies to increase timely availability of data. Fourier transform near infrared (FT-NIR) spectroscopy and Raman spectroscopy are commonly used for food quality (Herrero 2008, Ferreira et al. 2013) and medical applications (Choo-Smith et al. 2002) to measure properties linked with molecular composition including lipid content and purity, but are rarely used for ecological or fisheries applications despite similarities in data needs.

Utilizing spectroscopy approaches to collect data for fisheries management applications requires the successful development of calibration models for key biological indices. Calibration models correlate spectra from targeted tissues or hard parts with known or well-quantified reference data such as age, body condition, or reproductive status. Prior to implementation, the predictive skill of calibration models must be tested using samples that were not included in the calibration. Furthermore, comprehensive calibration models for life history metrics require reference data that encompass ontogenetic shifts and energy allocation throughout ontogeny (Martin et al. 2017) to capture the entire data domain.

This work focused on two commercially and ecologically important fish species, walleye pollock (*Gadus chalcogrammus*) and Pacific cod (*Gadus macrocephalus*), to test applications of

FT-NIR and Raman spectroscopy approaches for biological data collection. Walleye pollock and Pacific cod both responded negatively to a recent marine heatwave in the Gulf of Alaska (GOA) due to increased energetic demands in warmer waters (Barbeaux et al. 2020, Rogers et al. 2021), highlighting the need for rapid and comprehensive data collection to monitor their populations in a changing environment. We selected key energetic, life history, and demographic characteristics that are time-consuming to collect using traditional approaches: daily and annual age, body condition, and reproductive maturity. We reared fish at the National Oceanic and Atmospheric Administration (NOAA), Alaska Fisheries Science Center's (AFSC's) Little Port Walter Research Station (LPW) starting from the first year of life over the course of several years to facilitate flexible and frequent sampling throughout ontogeny. We also obtained wild-caught specimens from the GOA to test and refine our calibration models. These comprehensive reference datasets will provide foundational information for the development of operational calibration models to assess life history parameters associated with two commercially important Alaska species and a framework for expanding spectroscopy applications to other taxa and ecosystems.

METHODS

Age-0 fish were collected in the summers of 2020-2022 in southeast Alaska and reared at LPW (Fig. 10.1) in indoor large round ponds and outdoor net pens under ambient conditions to emulate the natural environment. Data collection is ongoing, but sampling followed a longitudinal project design. Sampling frequency was weekly for young fish and reduced to biweekly or monthly as fish increased in age. Fish were sacrificed by submersion in tricaine

methanesulfonate (MS222), weighed (g), and measured (fork length, mm). Target biological tissues for reference data and spectroscopy scans were collected and preserved for later processing (Table 10.1). Additional wild-caught samples for model testing were obtained opportunistically from field surveys in the GOA conducted by NOAA AFSC's Resource Assessment and Conservation Engineering Division and Auke Bay Laboratory (Fig. 10.1).

Reference materials were scanned using a Bruker TANGO R or MPA II FT-NIR spectrometer (Bruker Optics, Ettlingen, Germany) with diffuse reflectance and an integrating sphere with a reciprocal wavelength range from 11,500 and 4,000 cm^{-1} to obtain FT-NIR spectral data. Raman data were collected with a MarqMetrix All-In-One high-performance Raman spectrometer with 785 nm wavelength excitation and a 12.7 mm diameter sapphire probe. Sample presentation differed among material types and included a modified aperture for FT-NIR scans of small otoliths (Matta et al. this volume-b) and a quartz vial configuration suitable for plasma and tissue samples (Fig. 10.2; Goldstein et al. 2021).

RESULTS AND DISCUSSION

Three cohorts of walleye pollock and two cohorts of Pacific cod were reared at LPW. Growth trajectories were similar among the three reared cohorts, but sizes of wild specimens from the same annual cohort differed regionally and differed from growth trajectories of reared fish (Fig. 10.3). Such differences may be attributed to hatch dates or environmental conditions but highlight the need to test calibration models built with reared fish against diverse specimen sources.

Spectra and reference data have been collected from a subset of samples, and preliminary analyses indicate notable differences in tissue spectral patterns and tissue-specific protocols for data acquisition. Spectroscopy data acquisition from blood plasma was limited by the volume available from an individual fish. This prohibited Raman data collection for some specimens by limiting submersion of the probe. Plasma volume was less of a constraint for the sample window configuration of the FT-NIR spectrometer (Fig. 10.2). Collection of reference data is still being conducted, but FT-NIR spectral data from female pollock plasma indicate spectral separation among fish of different body sizes (Fig. 10.4), suggesting value in pursuing hormone analyses, since body size of juvenile fish is inherently linked with growth and development.

Preliminary exploration of FT-NIR scans focusing on body condition from liver and muscle showed variation in spectral patterns among tissue types. Spectra were more complex (i.e., peaks and valleys were evident) for liver than for muscle, suggesting greater material complexity (Fig. 10.5). Variability in liver spectra also coincided with informative wavenumber regions for determining lipid content from homogenized Pacific cod tissues (Goldstein et al. 2021), suggesting that liver may be an appropriate tissue for building calibration models for body condition. FT-NIR absorbance spectra from liver and muscle showed only weak visual patterns related to Fulton's K, a coarse condition index based on length and weight relationships (Wuenschel et al. 2018; Fig.10.5). Correlations between Fulton's K and focal biochemical analyses are positive but weak for juvenile Pacific cod (Goldstein et al. 2021), suggesting that decoupling between Fulton's K and spectra does not necessarily negate potential correlations between spectra and energy density or lipid content. Additionally, comparisons of FT-NIR triplicate liver scans from different regions of the same liver suggested that for some specimens, there was greater variability among liver scan regions within individuals than among specimens.

This finding suggests that repeated scans or homogenization may be necessary to obtain representative liver scans (Fig. 10.6).

Most of the work to date has focused on reproduction and age, which have both yielded promising calibration model results. Daily age and otolith FT-NIR spectroscopy results are presented in Matta et al. (this volume-b). Raman results from ovary scans and histological reference data that used a subset of fish reared at LPW indicate Raman could be used successfully to differentiate among various ovary maturity stages (Neidetcher et al. this volume).

Preliminary analyses for the suite of metrics examined here reinforce important considerations for calibration model development of fish life history indices. Variation in scans and spectral data repeatability indicate the importance of standardizing and optimizing material preparation and presentation. Variation in body size among reared fish and their free-ranging counterparts emphasizes the importance of curating comprehensive calibration and test datasets prior to model implementation. Following these considerations, finalized and vetted calibration models will likely provide new routes to incorporate, develop, and update indices for ecosystem-based management and stock assessment.

Table 10.1. -- Description of data types and biological tissue collections from reared walleye pollock (*Gadus chalcogrammus*) and Pacific cod (*G. macrocephalus*). R denotes that Raman spectra were obtained for the scan material and F refers to Fourier transform near infrared spectral data collection.

Life history metric	Reference and spectroscopy scan material	Reference data collection	Preservation
Maturity: ovary	ovary (R, F)	Structures from histological slides	formalin
Maturity: hormones	blood plasma (R, F)	Enzyme-linked immunosorbent assay (ELISA) from plasma extracted from blood	frozen
Age: daily	otolith (F)	Otolith microstructure analysis	dry
Age: annual	otolith (F)	Otolith analysis	dry
Body condition: liver	liver (R, F)	Proximate composition analysis for percent composition of lipid, protein, and water	frozen
Body condition: muscle	muscle (R, F)	Bomb calorimetry for energy density	frozen

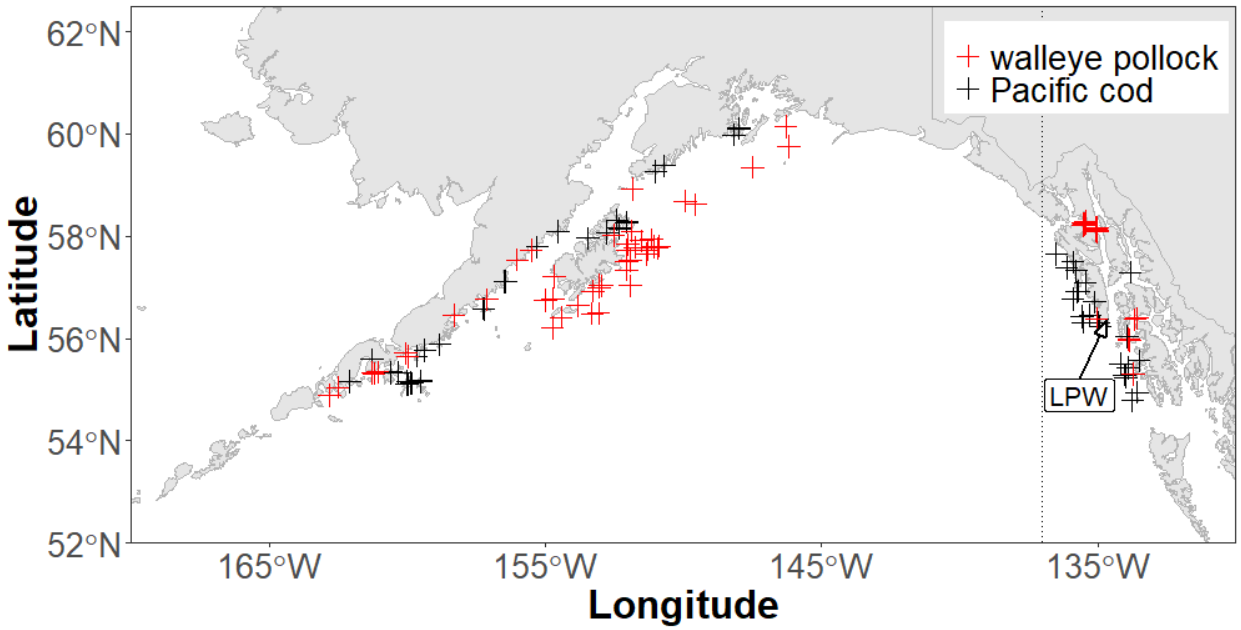


Figure 10.1. -- Location of Little Port Walter Research Station (LPW) and catch locations for a subset of wild-caught walleye pollock (*Gadus chalcogrammus*) and Pacific cod (*G. macrocephalus*) specimens from 2019-2022 for comparison with fish reared at LPW. The vertical dotted line represents the designation of eastern Gulf of Alaska (GOA) compared to central and western GOA for this study.

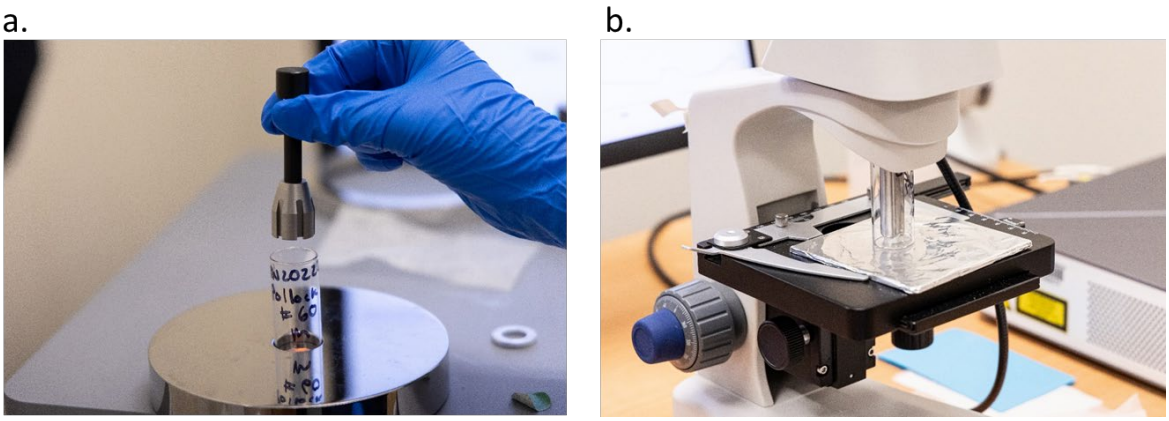


Figure 10.2. -- Soft tissue (ovary, muscle, liver) and plasma sample presentation for (a) FT-NIR spectroscopy and (b) Raman data acquisition. Photo credit: Delsa Anderl, AFSC (retired).

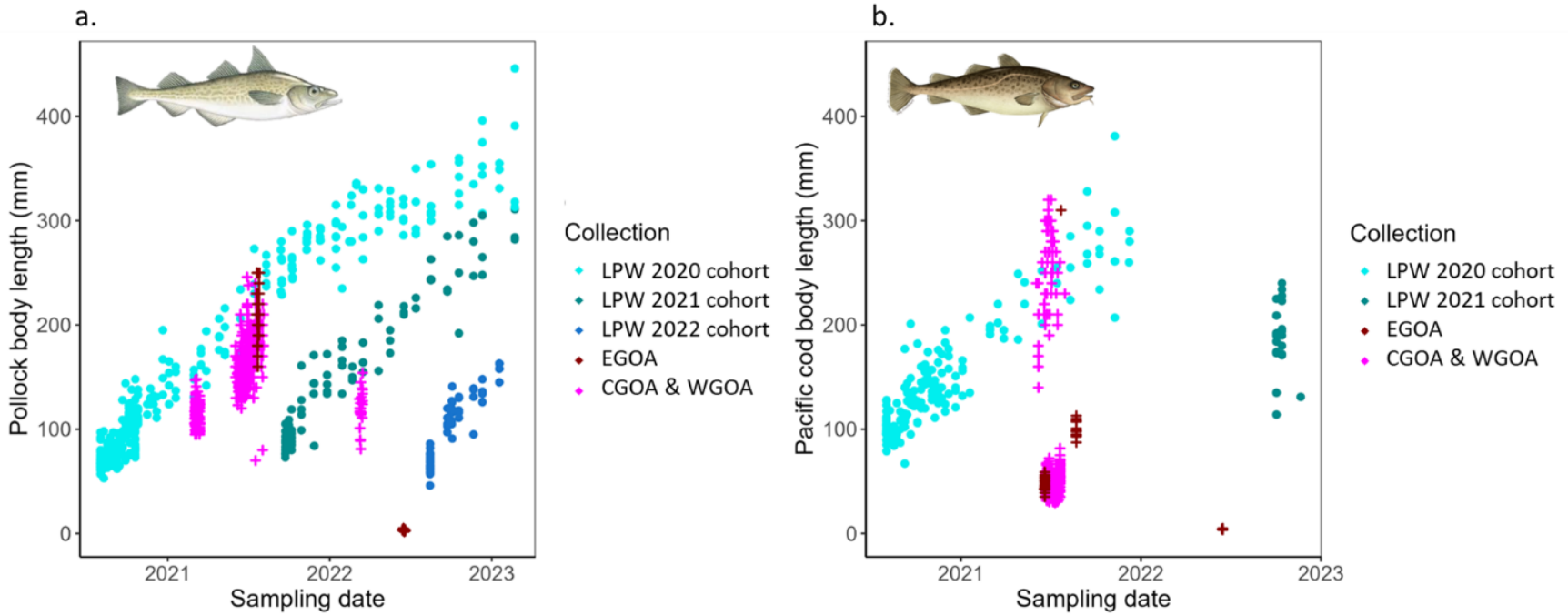


Figure 10.3. -- Growth trajectories and sampling dates of (a) walleye pollock *Gadus chalcogrammus* and (b) Pacific cod *G. macrocephalus* cohorts reared at Little Port Walter Research Station (LPW) and wild-caught fish that have been collected in the eastern Gulf of Alaska (EGOA) and the central (CGOA) and western GOA (WGOA; see Fig. 10.1). Wild-caught specimens were aged annually and correspond with annual cohorts that are being reared at LPW.

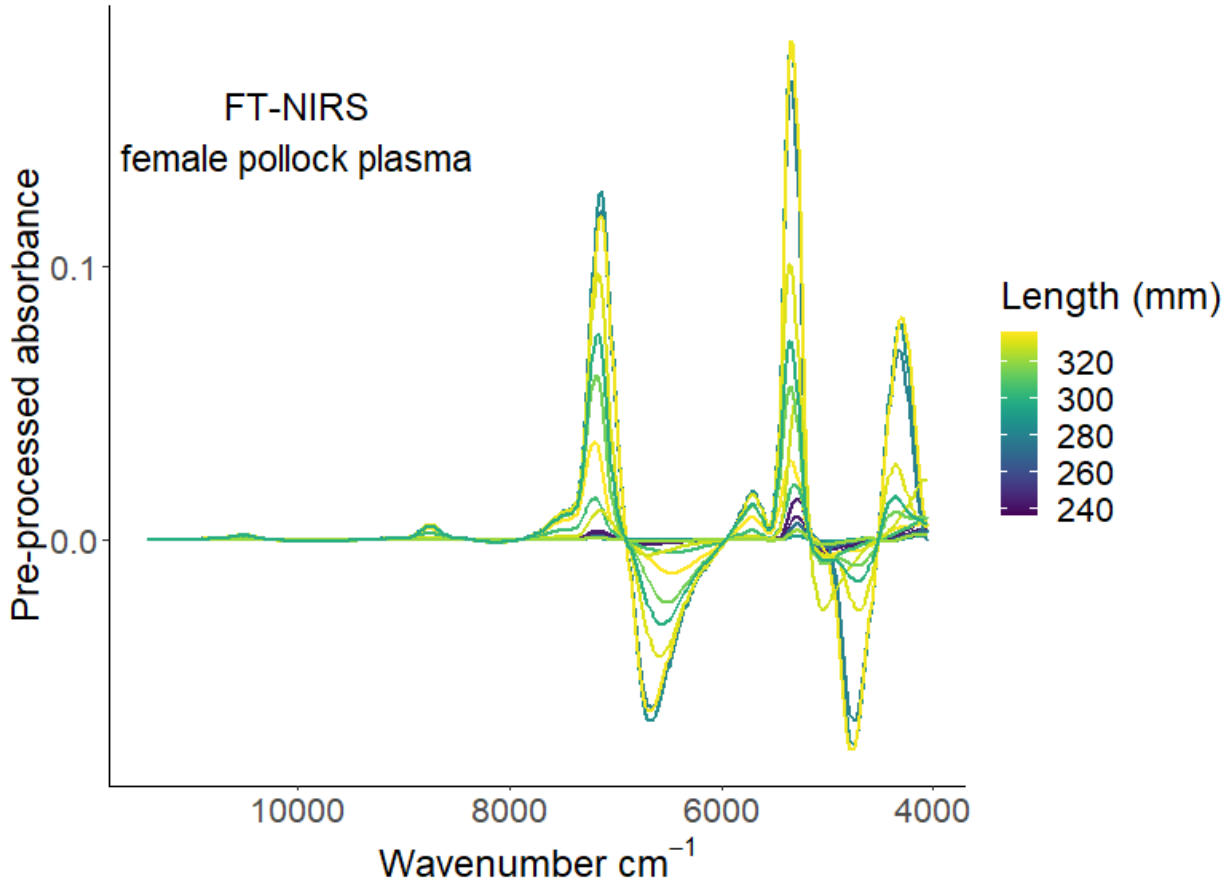


Figure 10.4. -- Spectra from Fourier transform near infrared spectrometer (FT-NIRS) scans of plasma from female walleye pollock (*Gadus chalcogrammus*) from a subset of fish. Color scale is fish fork length, since reference hormone data collection is in progress. Data were pre-processed with a first derivative Savitzky-Golay filter (polynomial order = 2, window size = 15).

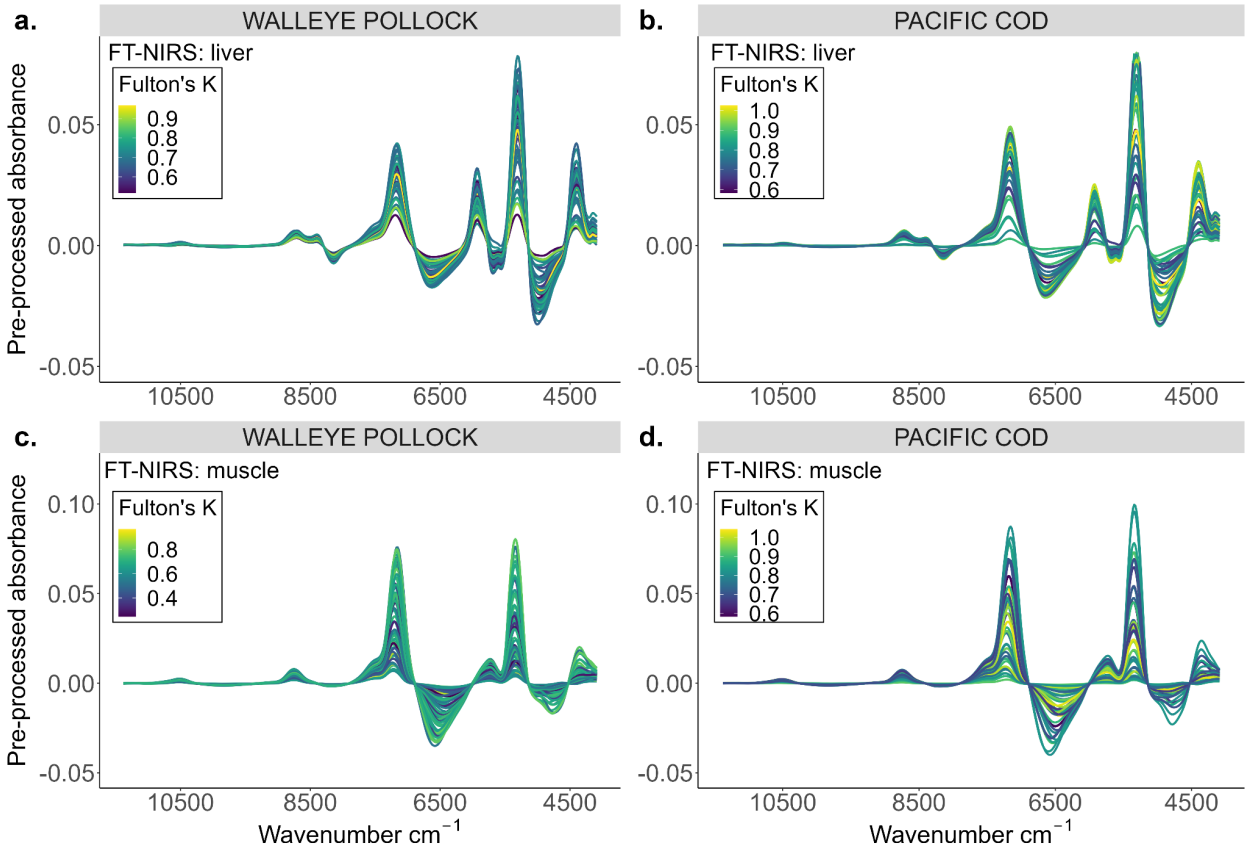


Figure 10.5. -- Fourier transform near infrared spectra (FT-NIRS) from (a, b) liver, and (c, d) muscle scans from reared walleye pollock (*Gadus chalcogrammus*) and Pacific cod (*G. macrocephalus*). Color scale is Fulton's condition index (Fulton's K; 100 * length/weight^3) as a coarse metric of body condition in the absence of reference biochemical data for the preliminary analyses. Note that the color scale changes for each panel, and some specimens differ between liver and muscle scans due to limitations in liver size that constrained FT-NIRS data collection. Data were pre-processed with a first derivative Savitzky-Golay filter (polynomial order = 2, window size = 19).

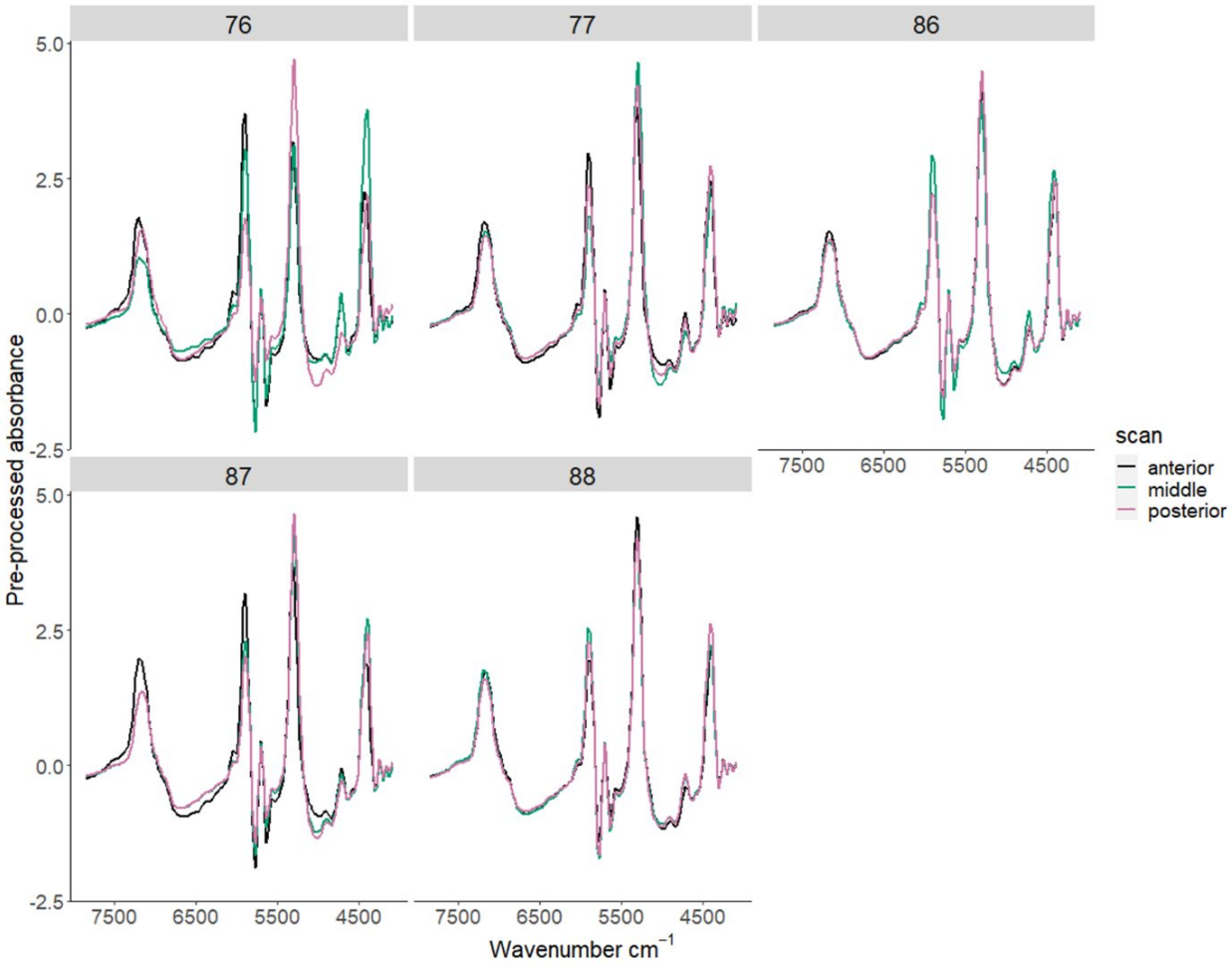


Figure 10.6. -- Triplicate Fourier transform near infrared spectral scans of liver tissue from five individual walleye pollock (*Gadus chalcogrammus*) specimens (panels) taken from the anterior, middle, and posterior regions of the liver. Data were pre-processed with a first derivative Savitzky-Golay filter (polynomial order = 7, window size = 19) and subsequently with Standard Normal Variate pre-processing in an effort to correct for any additional issues that may create differences in spectra among specimens, such as light scatter (Fearn 2008).

11. Rapid Daily Age Estimation of Juvenile Walleye Pollock in the Gulf of Alaska using FT-NIR Spectroscopy

Mary Elizabeth Matta[1], Esther D. Goldstein[1], Irina M. Benson[1], Heather K. Fulton-Bennett[2], Charles D. Waters[2], Brenna C. Hsieh[1], and Thomas E. Helser[1]

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

[2] Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
17109 Point Lena Loop Road
Juneau, AK USA

INTRODUCTION

Estimates of daily age can provide valuable insight and context for understanding early life history and recruitment dynamics of marine fishes. Specifically, daily ages may be used to determine the phenology of hatching and spawning and to estimate growth, which in turn can influence mortality and overwinter survival and ultimately, the recruitment success of a population (Rogers and Dougherty 2019, Yoklavich and Bailey 1990). However, obtaining daily age estimates is laborious, destructive, and requires specialized equipment and expertise in growth pattern interpretation, limiting its feasibility in terms of broader, production-scale applications. One innovative method that has shown promise to rapidly and non-destructively produce estimates of daily age and growth in juvenile fish is Fourier transform near infrared (FT-NIR) spectroscopy. A recent study on young-of-year (age-0) red snapper (*Lutjanus campechanus*) used FT-NIR spectroscopy to successfully predict fish age within 6 days of traditional, microscopic age estimates, with both methods resulting in comparable estimates of growth (Passerotti et al. 2020a). Here, we applied a similar approach to predict daily ages of juvenile walleye pollock (*Gadus chalcogrammus*), a commercially important species in Alaska waters.

METHODS

Young-of-year walleye pollock were caught in summer and fall 2020 and raised at NOAA's Little Port Walter Research Station in Southeast Alaska. Fish were sacrificed sequentially as a part of a larger longitudinal study to evaluate changes in growth and

development from the juvenile stage through maturation. Sampling was conducted at short intervals (weekly) at the beginning of the study and longer intervals (biweekly to monthly) thereafter.

A subset of young-of-year otoliths was aged using traditional, microscopic examination. Only fish sacrificed prior to October 1 were aged, due to compression in the daily rings that typically occurs near the otolith edge in fall as growth starts to slow, which can complicate pattern interpretation. Otoliths were thin-sectioned to a thickness of about 10 microns and read at least twice by two different analysts at 400-1000X magnification. Ages estimated microscopically were used to derive the median date of hatching for the 2020 cohort. This hatch date was applied to the remaining unaged fish to determine estimates of daily age. Only otoliths that had a Chang's coefficient of variation (CV; Chang 1982) less than 10% (Jones 2013, Passerotti et al. 2020a) were retained to predict the median hatch date for the rest of the sample.

All otoliths were scanned on a Bruker MPA-II FT-NIR spectrometer to produce a spectrogram for each otolith (Fig. 11.1). Due to the small size of the otoliths, we tested several different aperture and stamp combinations to optimize the signal-to-noise ratio (Passerotti et al. 2020a) of spectra from small otoliths in a related species, Pacific cod (*Gadus macrocephalus*). The best results in terms of predicting fish length from otolith spectra were obtained using Teflon discs with drilled 2-mm and 5-mm diameter apertures and a gold-coated reflector stamp covering the otolith. We opted for a 5-mm aperture diameter to accommodate the largest otoliths in our sample (Fig. 11.2).

Otolith spectral data were split randomly into 75% train and 25% test data sets and used to build partial least squares regression (PLS) predictive models between spectral data and fish daily age. Spectral regions dominated by noise were excluded (>7500 cm^{-1}) and data were

preprocessed using a first derivative Savitzky-Golay filter (17-point smooth, polynomial order = 3) (Fig. 11.1). An initial calibration model was constructed, including all the wavenumbers >7500 cm^{-1}. Wavenumber variables for inclusion in the final model were then selected using a combination of variable importance in projection (VIP) and jack-knifed coefficient *p*-values (α = 0.05) (Farrés et al. 2015).

RESULTS AND DISCUSSION

In total, we aged 58 otoliths using traditional methods. The overall precision (% CV) of age estimates was 7.92%. Acceptable levels of precision (<10% CV) were achieved for 36 of the otoliths; these otoliths were used to estimate the hatch date distribution of the sample. Estimated hatch dates ranged from April 16 to May 31, with a median date of May 2.

Preliminary results demonstrated satisfactory predictive capability of PLS models between otolith spectra and daily ages (Fig. 11.3). The root mean square error (RMSE) and r^2 of the cross validation (leave-one-out) data set were 9.34 days and 0.93, respectively. The RMSE and r^2 of the test data set were 12.09 days and 0.87, respectively. The percent RMSE, a measure that can be used to compare the error relative to the maximum observed age in a study, was very low at 4.02%, and comparable to results for juvenile red snapper and adult walleye pollock (Helser et al. 2019b; Passerotti et al. 2020a).

Ultimately, FT-NIR spectroscopy appears to be an effective method of estimating daily ages of juvenile walleye pollock. Future work will focus on increasing sample size, refining hatch date estimates of unaged fish, and incorporation of data from wild-caught fish and other

cohorts to test the model's ability to predict ages of fish from different areas, years, and over a wider range of ages.

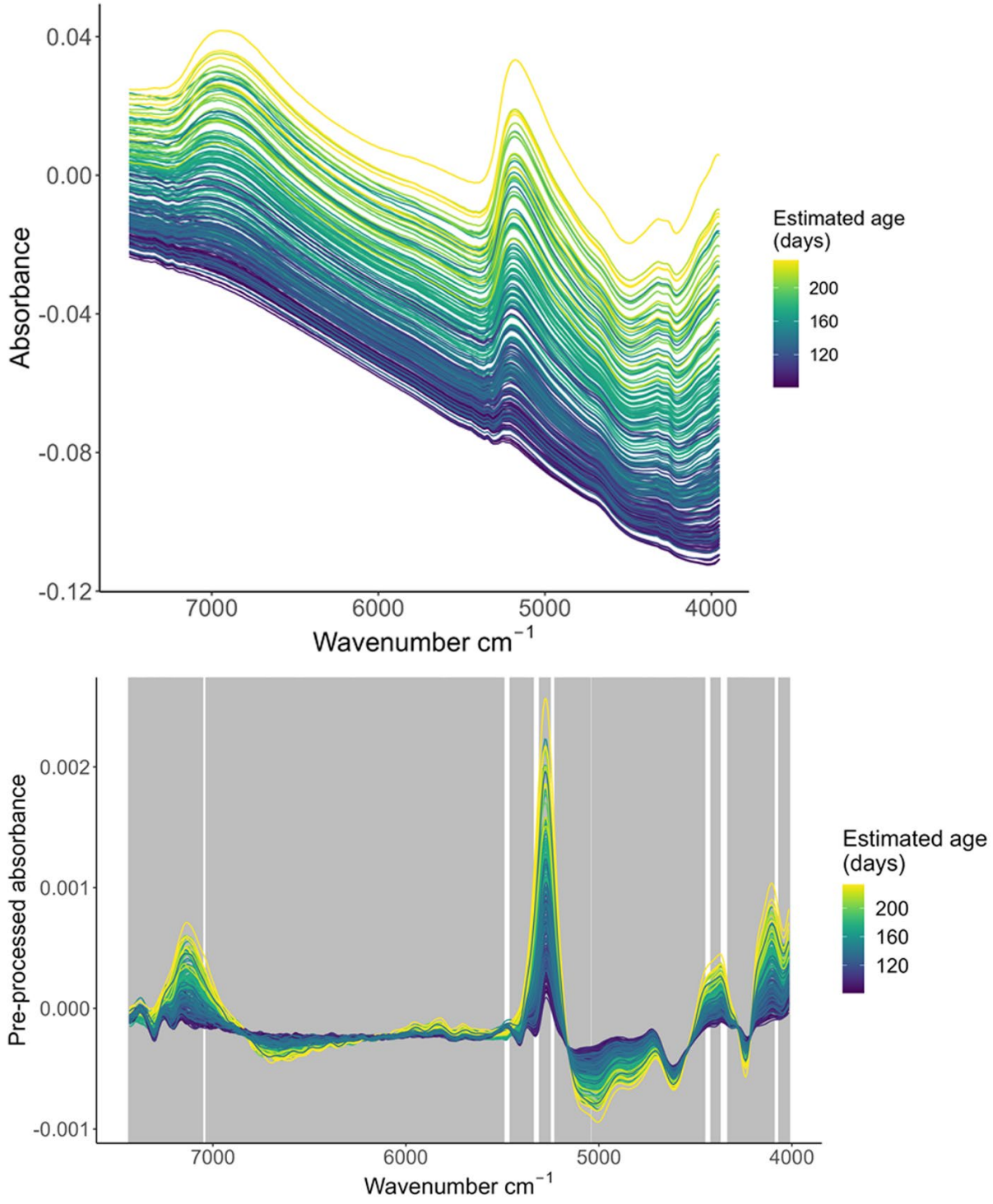


Figure 11.1. -- Raw (top) and preprocessed (bottom) spectrograms of age-0 walleye pollock (*Gadus chalcogrammus*) otoliths.

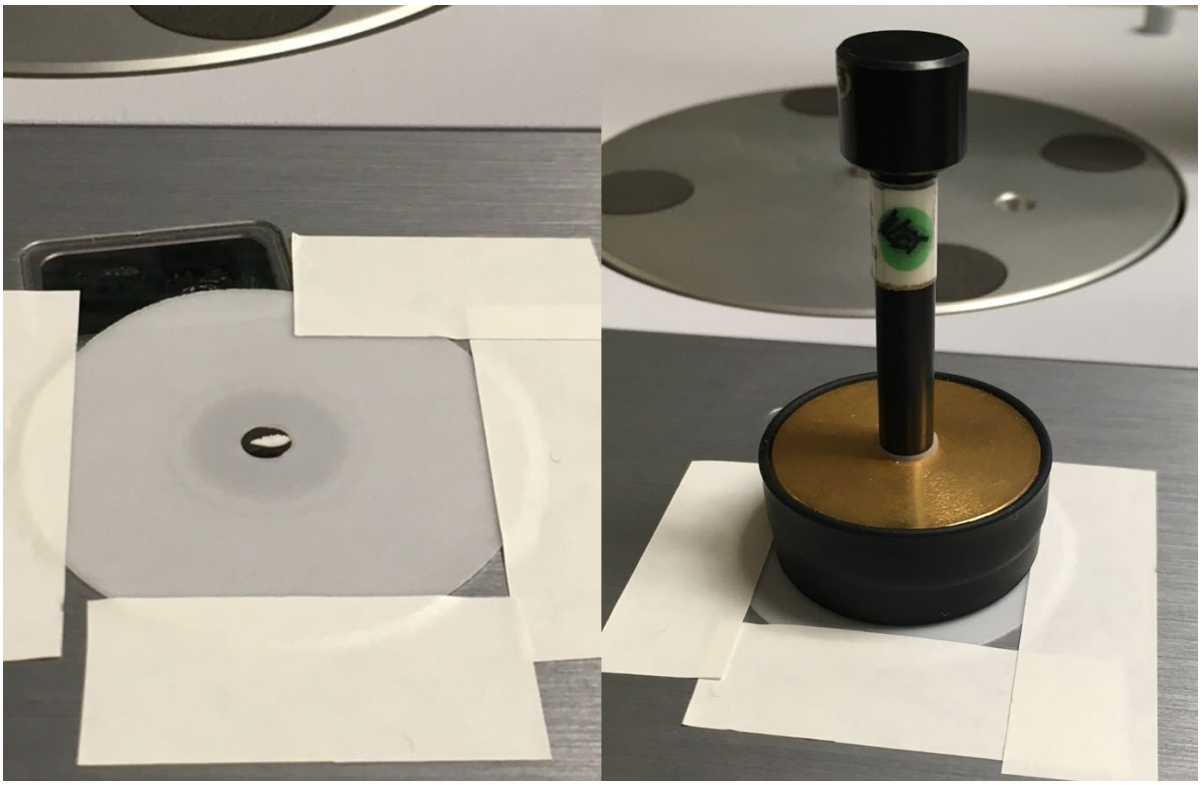


Figure 11.2. -- Accessories used to scan otoliths from age-0 walleye pollock (*Gadus chalcogrammus*). Left: Teflon disc with a drilled 5-mm diameter aperture to improve the signal-to-noise ratio. Right: Gold reflector stamp used to cover the otoliths.

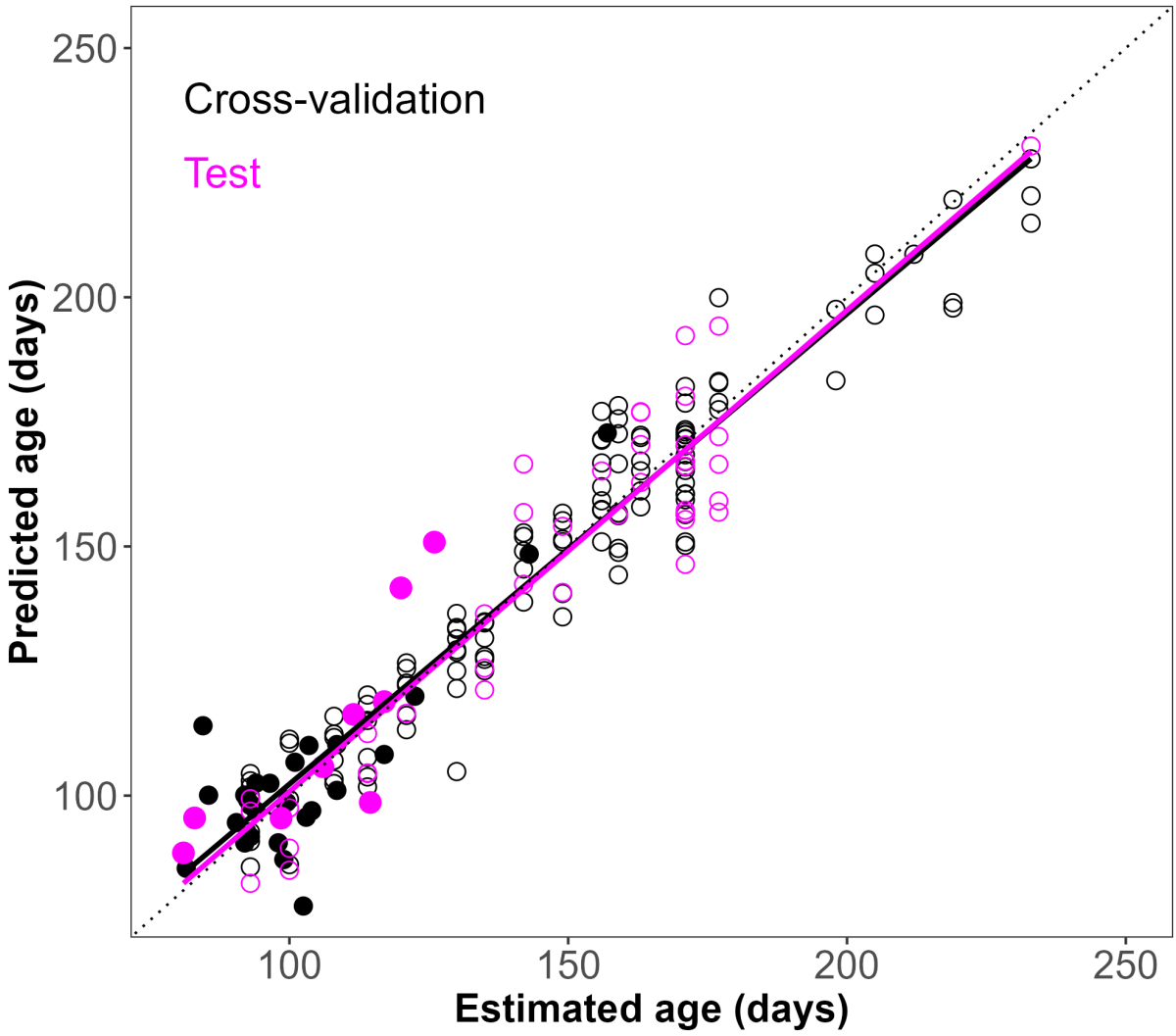


Figure 11.3. -- Partial least squares regression showing relationship between human-estimated age and model-predicted age of age-0 walleye pollock (*Gadus chalcogrammus*). Black symbols indicate cross-validation dataset and pink symbols indicate test dataset. Closed circles represent fish with ages that were estimated using traditional microscopic methods; open circles represent fish with ages based on the median hatch date of the sample and time of sacrifice.

12. Developing NIR Sampling Methodology for Modeling Species Discrimination in Live Catfish for Aquaculture

Carrie K. Vance, Ashmita Poudel, Li-Dunn Chen, Peter J. Allen, and Andrew J. Kouba

Mississippi State University
75 B.S. Hood Drive
Mississippi State, MS USA

INTRODUCTION

Near infrared (NIR) spectroscopy of otoliths for age determination and species classification in fish such as Pacific cod (*Gadus macrocephalus*), red snapper (*Lutjanus campechanus*), walleye pollock (*G. chalcogrammus*) and saddletail snapper (*L. malabaricus*) occurs post-mortem (Helser et al. 2019b, Benson et al. 2020). In other post-mortem studies, NIR spectroscopy has been used to assess the quality and nutritional composition of fish filets, classify fresh or frozen tissue, and predict microbial contamination or adulteration of market products such as fish meal (Liu et al. 2013). NIR spectroscopy has not been used to examine physiology and health of living fish for pre-harvest management decisions, as the prospect of using NIR spectroscopy for aquatic species *in vivo* is often dismissed due to fears of water signal interference. Yet, we have successfully used NIR spectroscopy to address physiology and health questions in live amphibians, terrestrial and aquatic, specifically for species discrimination, sex determination in monomorphic species or juvenile animals, reproductive status in female adults, and to detect subclinical levels of fungal disease (Vance et al. 2016a, Chen et al. 2022, Chen et al. 2023). Similar information may be obtainable for live fish in real time with NIR spectroscopy, particularly if coupled with machine learning algorithms. Here we develop methodology for NIR spectral collection of species-specific information from live catfish as a proof of principle in NIR spectroscopy for *in vivo* physiological assessment.

To examine if biochemical information for physiological traits of interest could be collected from live fish, we analyzed the discriminatory capacity of NIR spectroscopy for species classification of adult blue catfish (*Ictalurus furcatus*) and channel catfish (*Ictalurus punctatus*) catfish with intent to transfer the model predictions to discriminate fingerling fish for

aquaculture. We needed to develop a method for collecting NIR spectra from live adult catfish that could give us physiological information despite the presence of a thick mucus layer on the skin and handling in an aquatic environment. To this end, we first compared the NIR spectra prediction results for catfish that were or were not anesthetized with tricaine methanesulfonate (MS-222) to see if spectra from calm fish were more informative than those of stressed fish, and if there was a masking effect of the MS-222 anesthetic in the NIR spectrum. Second, we compared the spectral collection method for three important regions of the live fish. Spectra from the head could contribute to species differentiation, sexual dimorphism at breeding, and infectious disease such as *Edwardsiella ictaluri* bacterium, which causes enteric septicaemia. Spectra from the lateral line may also differentiate species, sexual maturity, and channel catfish virus disease, which is a disease with 100% mortality in fingerlings. Spectra from the vent may provide information on species, sexual maturity, and even translate to measuring gamete development. Finally, we tested six model algorithms (PCA-LDA, GLMnet, KMM, RF, XGBoost, and SVM) to see if predictive capacity for our trait of interest was dependent on the modeling approach.

METHODS

Adult blue and channel catfish (n = 20 per species) were grown in the aquaculture facility at Mississippi State University. Near infrared spectra were collected from fish that were not anesthetized and then again when they were anesthetized with MS-222. NIR spectra were collected from the head, lateral line, and vent (three spectra per site), using a portable ADS FieldSpec 3 + Indigo-Pro w/fiber optic source and a low-intensity contact probe with a diameter

quartz window (Fig. 12.1A). Each spectrum was collected over the range 350-2500 nm and was an average of 100 scans with a 34 μs integration time, totaling about 12 seconds of direct contact with the live fish. Spectra were then transformed with SNV + detrend and 1st derivative (GAP 25; segment 19), and SG smoothing (order 1, 12 symmetric kernel points) (Fig. 12.1B). Spectra from 32 fish (16/species) were used to generate the calibration and internal validation sets (5-fold cross-validation) and spectra from the remaining 8 fish (4/species) were the external validation set in the discriminative models for each collection site.

Six conventional machine learning modeling algorithms for discrimination of species were examined and compared using Tukey's honestly significant difference test ($p < 0.05$) and analysis of variance: 1) principal component analysis-linear discriminant analysis (PCA-LDA), which is a combined unsupervised-supervised dimensionality-reducing technique for discriminating observations amongst discrete classes, 2) generalized linear model with elastic net regression (GLMnet), which is a supervised linear regression method, 3) k-nearest neighbors (KNN), a supervised machine learning algorithm that predicts the class of an observation based on the majority-class of k-nearest neighbors, found best for fish species, 4) random forest (RF) search, a supervised algorithm that applies an ensemble approach when placing predictions based on repetitive splitting of features (branches) at random to split nodes into decision trees, 5) extreme gradient boosting (XGBoost), a supervised algorithm that employs an ensemble of trees for prediction, and 6) support vector machine (SVM), which is a linear supervised algorithm that plots all observations on a hyperplane comprised of all predictor features (Chen et al. 2023).

RESULTS

Classification of species using NIR spectroscopy across all six algorithms provided informative models with 77.3-97.5% accuracy (Table 12.1). Species discrimination was obtained with NIR spectra collected from the head, lateral line, and vent region of catfish with external validation average predictions for anesthetized fish with >90% accuracy, and for non-anaesthetized fish between 75% and 90% accuracy for all algorithms tested. Models built with spectra from anesthetized fish exhibited improved species classification in the head and vent regions, and MS-222 did not appear to interfere with species distinguishing biochemistry (Table 12.1). Species classification did not significantly differ by scanning region, although spectra from the lateral line were more informative than spectra from the head or vent regions in non-anesthetized fish, while head and vent spectra were most informative in anesthetized fish. The use of anesthesia may lead to decreased stress response and reduce noise, allowing for increased signal detection for predicting species designation, yet alternative means to collect spectra without stress may provide more rapid data collection and yield similar results.

CONCLUSIONS

Catfish species can be classified via NIR spectroscopy collected from the epidermis of living individuals with water and mucus layers. Spectral collection is improved when animals are calm or when collecting spectra from easily accessible regions (lateral line and head vs. vent), although developing alternatives to administering anesthesia would benefit field studies. Furthermore, spectral collection from different regions contributes information on different

physiologically based traits such as disease and reproduction, and all regions provide discriminatory information that can be tailored to the trait of interest. Finally, use of a computationally simple model (e.g., PCA-LDA) can provide sufficient information for basic physiological analysis in live fish, despite the presence of water.

Fingerlings from different catfish species have significant differences in not only the metabolic processes that affect their growth rates, nutrient composition, and marketability, but also in disease susceptibility and transmission. Currently, channel, hybrid, and blue fingerlings can only be discriminated through polymerase chain reaction amplification of genetic material as no distinguishing visual characteristics are evident in their physiology. Similar to our work in amphibians, NIR spectroscopy may also be able to detect pathogenic disease presence and load at early stages to enable isolation or treatment prior to spread or population outbreak, which can wipe out an entire crop. NIR spectroscopy could provide a spectral fingerprint based on the biochemistry of each species in real time, allowing investment decisions without termination.

Table 12.1. -- Discrimination of live channel and blue catfish by near infrared spectroscopy, with aggregated accuracies from five repeats of 5-fold cross validation from the external dataset of 8 fish left out of the prediction models.

	No Anesthesia			*Anesthesia*		
Model	**Head**	**Lateral Line**	**Vent**	**Head**	**Lateral Line**	**Vent**
Random Forest	80.3 ± 11.7%	**84.4 ± 13.5%**	81.0 ± 9.2%	**94.2 ± 7.7%**	90.0 ± 7.7%	**95.8 ± 2.6%**
XGBoost	78.9 ± 10.2%	**84.5 ± 13.7%**	83.6 ± 8.5%	**94.2 ± 5.7%**	93.9 ± 9.4%	**96.7 ± 3.1%**
GLMnet	78.0 ± 10.7%	**82.1 ± 11.7%**	81.5 ± 13.4%	**97.5 ± 5.0%**	95.8 ± 6.5%	**97.5 ± 2.0%**
SVM	81.7 ± 11.6%	**87.8 ± 15.5%**	79.3 ± 60.8%	**97.5 ± 5.0%**	95.0 ± 4.1%	**95.8 ± 4.6%**
LDA	**78.3 ± 11.6%**	77.2 ± 15.5%	78.1 ± 10.8%	**95.0 ± 4.9%**	93.3 ± 5.7%	**95.8 ± 4.6%**
KNN	77.3 ± 10.9%	**83.3 ± 8.6%**	73.7 ± 8.3%	**95.8 ± 6.5%**	92.5 ± 7.2%	**95.0 ± 4.1%**

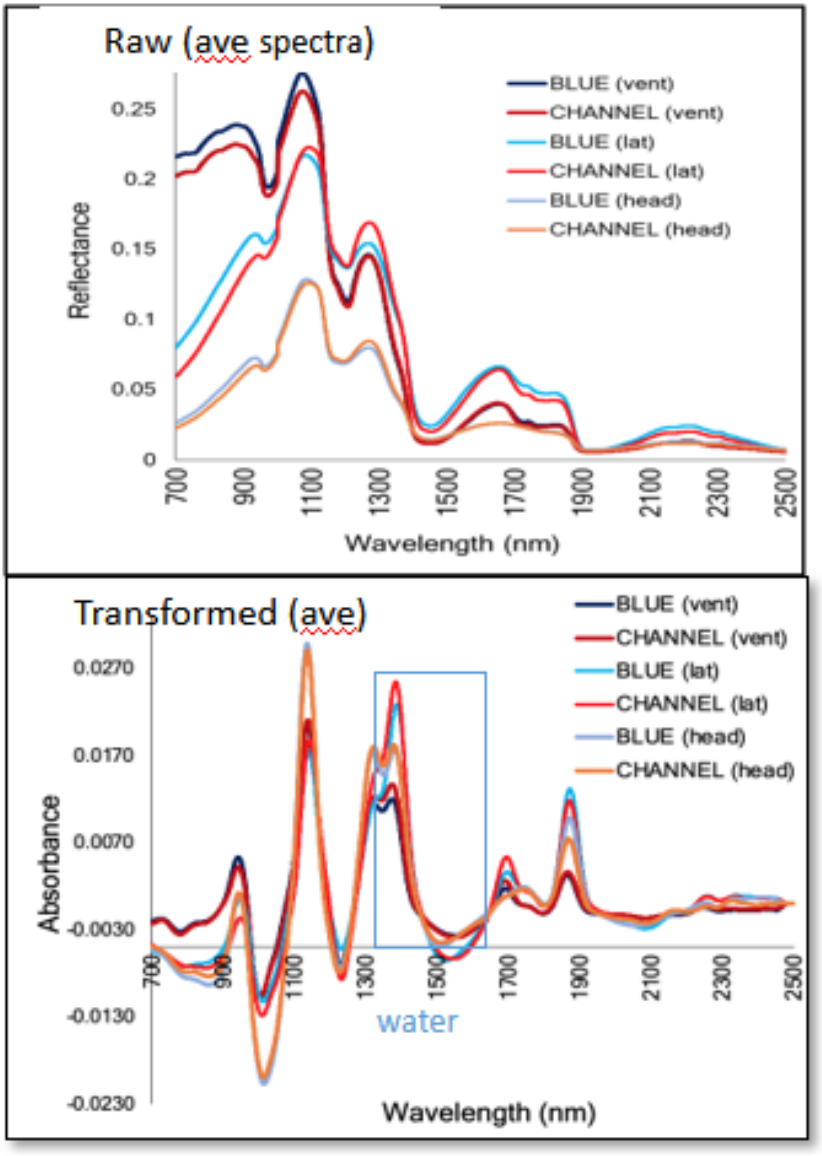


Figure 12.1. -- Average raw and processed spectra collected from two species of live catfish (*n* = 20 fish /species) at three positions on the body. Aquaphotomics performed at the water signal highlighted by the blue box in the transformed spectra has potential for additional classification power.

13. Fourier Transform Near Infrared (FT-NIR) Spectroscopy Discriminates Archived Otoliths of Newly Detected Cryptic Species, *Etelis carbunculus* and *Etelis boweni*

Kristen A. Dahl[1*], Joseph M. O'Malley[1], Beverly K. Barnett[2], William Kline[2], and Joseph B. Widdrington[3]

[1] Pacific Islands Fisheries Science Center
NOAA, National Marine Fisheries Service
1845 Wasp Boulevard, Building 176
Honolulu, HI USA

[2] Southeast Fisheries Science Center
NOAA, National Marine Fisheries Service
3500 Delwood Beach Road
Panama City, FL USA

[3] Southern Seas Ecology Laboratories
School of Biological Sciences
University of Adelaide
North Terrace, South Australia AUSTRALIA

INTRODUCTION

Etelis carbunculus Cuvier (Cuvier and Valenciennes 1828) is an important component of the bottomfish management unit species (BMUS) fisheries in the Pacific Territories of the United States (Langseth et al. 2019). There has long been speculation about whether *E. carbunculus* actually comprises two species due to its extensively wide range from Seychelles (western Indian Ocean) to Hawai'i, and geographic variation in certain morphological characters, such as maximum body size and otolith growth and morphology (Anderson 1981, Smith and Kostlan 1991, Wakefield et al. 2014, Williams et al. 2015). Historically, *E. carbunculus* was divided into *E. carbunculus* (Seychelles holotype 1828) and *E. marshi* (Hawaii holotype), but *E. marshi* was later designated as a junior synonym (i.e., a previously published name) of *E. carbunculus*, and not a unique species (Anderson 1981). Later, phenotypic differences in otolith morphology of *E. carbunculus* reported from different regions ignited a new debate around the possibility of a cryptic species pair (Smith and Kostlan 1991, Smith 1992). Recent genetic and morphological studies have provided conclusive evidence that *Etelis carbunculus* comprises two cryptic species, *E. carbunculus*, the pygmy ruby snapper, and *E. boweni*, the newly described giant ruby snapper (Loeun et al. 2014, Andrews et al. 2016, 2021). Although exceptionally similar in their morphology, the two species differ in coloration of the dorsal caudal fin tip and the shape of the opercular spine (Andrews et al. 2016).

Now that taxonomic revision of the species complex has been verified, life history research has revealed fundamental differences in the biology between *E. carbunculus* and *E. boweni*, including maximum body size, growth rate, and length at age (Williams et al. 2017, Wakefield et al. 2020). Due to their overlapping distributions across the Indo-Pacific, both

species have been previously misreported as a single species (*E. carbunculus*) in fisheries catch data and in biological studies (Williams et al. 2013, Wakefield et al. 2020). This complicates life history research, because the otolith collections that were thought to be solely *E. carbunculus* are now known to be contaminated with an unknown number of *E. boweni* otoliths. Collections of biological samples on these species for which misidentification is a potential issue go back decades, and fisheries age-structured assessments that incorporate life history data from the cryptic species alongside *E. carbunculus* may suffer from model misspecification. This is also true for data poor assessment approaches that rely heavily on otolith-derived growth estimates. Given the inherent vulnerability of Eteline snappers to overexploitation (Williams et al. 2013), it is imperative that methods be developed to recognize speciation within archived otolith collections to move forward with sustainable management.

Fourier transform near infrared (FT-NIR) spectroscopy is an emerging technology in fisheries and conservation biology (Wedding et al. 2014, Vance et al. 2016b, Helser et al. 2019b) which may offer an alternative to both morphometric and shape analyses. The technique relies on light from the near infrared region (12,800-4,000 cm^{-1} wavenumbers) of the electromagnetic spectrum and absorbance patterns that reflect a sample's organic chemical composition. These spectra can then be compared among samples using multivariate statistical analysis to discriminate known variables (e.g., age, stock, or species) of interest (Murray and Williams 1987). FT-NIR spectroscopy has recently gained attention with its ability to predict fish age from otoliths (Wedding et al. 2014, Helser et al. 2019a, 2019b, Passerotti et al. 2020b), as well as the potential to discriminate geographical differences expressed in otolith chemistry (Wedding et al. 2014, Benson et al. 2020). FT-NIR spectroscopy, if effective in distinguishing cryptic *Etelis* species, would provide a time efficient alternative to manual morphometrics using calipers or

image analysis, especially of species with fragile otoliths prone to breaking, and could be used to differentiate future cryptic species from archived collections. Furthermore, because FT-NIR spectroscopy scanning does not damage the sample (i.e., otoliths are scanned whole), otoliths may later be used for fish ageing following spectral scans and the determination of species.

Herein, we examine the utility of FT-NIR spectroscopy to distinguish between archived otoliths of cryptic species *E. carbunculus* and *E. boweni*. We assessed the efficacy of FT-NIR spectroscopy using voucher otoliths from the Southwest (SW) Pacific Ocean that were previously identified to species using phenotypic and genetic differences. Following calibration and validation of voucher (i.e., known) data, we applied optimal models to predict and classify archived *Etelis* sp. otoliths previously collected around Guam.

METHODS

Samples

Voucher otoliths from *E. carbunculus* and *E. boweni* were located at the Pacific Community Marine Specimen Bank archive based in New Caledonia (Smith et al. 2017) (Fig. 13.1). Fish were collected between July 2012 and April 2014 from research surveys and commercial fishers from seamounts and island reef slopes in the Exclusive Economic Zones (EEZs) of New Caledonia, Fiji, Tonga, and Vanuatu (range: -15.1°-24.4°S, 164.9°E-174.3°W) in the SW Pacific Ocean. Samples (n = 93) were chosen by fork length (FL) to capture the full size range of each species, as well as to maximize the overlap in their size distributions. Samples were selected to also minimize spatial variability, selecting samples as evenly from as few

regions as possible (i.e., similar numbers of each species for each EEZ). Voucher *Etelis* samples were distinguished phenotypically at the time of capture based on the presence (*E. boweni*) or absence (*E. carbunculus*) of a black margin on the upper lobe of the caudal fin and a rounded (*E. boweni*) or acutely sharp (*E. carbunculus*) opercular spine (Andrews et al. 2016). Fork length was measured to the nearest 1 mm, wet weight was recorded to the nearest 0.1 g, and the sagittal otoliths were removed, rinsed, and stored dry.

Presumed *E. carbunculus* otoliths were collected from the waters around Guam (13.4°N, 144.7°E) at the Guam Fishermen's Cooperative from 2009-2019 (n = 91) and were archived at the Pacific Islands Fisheries Science Center in Honolulu, Hawai'i. Currently, these samples comprise a potentially mixed collection of cryptic species, *E. carbunculus* and *E. boweni*. As above, fork length was measured to the nearest 1 mm, wet weight was recorded to the nearest 0.1 g, and the sagittal otoliths were removed, rinsed, and stored dry.

Spectral Measurements and Pre-processing

All otoliths were scanned using diffuse reflectance on a Bruker Multi Purpose Analyzer II Fourier transform near infrared spectrometer with a 22-mm diameter sample window and OPUS software (version 7.8; Bruker Scientific, Billerica, MA). The spectrum acquisition was performed from 10,000 cm^{-1} to 4,000 cm^{-1} (64 scans, resolution 16 cm^{-1}). Scans were averaged to produce a single representative spectrum for each sample.

Following acquisition, spectral data were uploaded into Solo 8.9.2 (2021), a chemometric software that uses a MatLab framework (Eigenvector Research Inc., Manson, WA) for data processing and model generation. Selected spectral regions were in the 7,500 to 4,000 cm^{-1}

range. Raw spectra were plotted to detect any spectral outliers in need of rescanning. Spectral data were pre-processed with 13-point Savitzky-Golay smoothing (2nd order polynomial), 13-point Savitzky-Golay first derivative transformation, and mean-centering, which produced the best separation among groups.

Principal Component Analysis

Exploratory data analysis based on principal component analysis (PCA) was performed to visualize trends in the voucher data and to evaluate the discriminatory possibility of the FT-NIR spectra between species. The Guam *Etelis* sp. samples were also explored using PCA, with samples ≥60 cm FL assigned to *E. boweni* based on known disparities in maximum size between species (Andrews et al. 2016). PCA on the pre-processed data matrix was obtained using two or three principal components, and cross-validation was performed using venetian blinds with ten splits and algorithm-SVD (singular value decomposition). Species groups were visualized using a scores plot. Hotelling's T^2 and Q residuals were inspected to detect any spectral outliers in need of rescanning.

Multivariate Classification

Partial least squares discriminant analysis (PLS-DA) was used to evaluate spectral variability from voucher *E. carbunculus* and *E. boweni* based on a known class membership (i.e., species) of the samples. To obtain the discrimination model, each *Etelis* species was defined as a separate class. PLS-DA models on voucher data were calibrated and validated using two approaches. First, the entire dataset was used to calibrate a discrimination model, and cross-

validation (CV) of the model was executed using 10 sub-validation sets of 8 samples each (venetian blinds cross-validation). For the second approach, the total sample set was partitioned into calibration, assembled from 70% of samples, and validation datasets, using the remaining 30% of samples. The split was done randomly using the Kennard-Stone method (Kennard and Stone 1969). The calibration dataset was used to generate the classification model (venetian blinds CV), and the validation dataset was used to evaluate predictive (i.e., classification) performance after assigning class membership to one of the two species.

In each case, automatic variable selection was used to find subsets of variables (i.e., wavenumber regions) that improve predictions (compared to all 949 variables). Goodness of fit for each model was based on the cross-validation (CV) error, r^2 (coefficient of determination), RMSECV (root mean square error of cross validation), and specificity (true negative rate) and sensitivity (true positive rate) of the model. Class error was calculated for each model as the average of the false positive and false negative rates and was viewed as a confusion matrix.

Lastly, PLS-DA models built on voucher samples were used to predict class (i.e., species) membership of the Guam *Etelis* sp. samples. The PLS-DA classification method builds a model on the calibration dataset (e.g., voucher) and, by default, assumes the two classes have equal prior probability, regardless of the relative number of samples of each class in the calibration dataset. However, if there is known information on the prior probability of each class in the population, it is possible to incorporate these priors into the model's classification so that the model would be more accurate for datasets randomly sampled from that population. The assumption that classes were evenly distributed in Guam samples was not met based on evidence of far more *E. carbunculus* (~80%) compared to *E. boweni* (~20%) in the dataset, thus prior probabilities of the PLS-DA model were modified using “priorprob”. Species predictions of

these samples were compared to quadratic discriminant analysis results from otolith morphometrics.

RESULTS

Despite outward similarities in phenotype, *E. carbunculus* reaches a much smaller maximum size compared to *E. boweni*. Fork length of voucher samples ranged from 25.0 to 58.0 cm for *E. carbunculus* (*n* = 38) and from 29.0 to 85.0 cm for *E. boweni* (*n* = 45) (Pacific Islands Fisheries Science Center 2023) (Fig. 13.2, Table 13.1). Mean (±SE) FL was 42.26 (±1.49) cm for *E. carbunculus* and 53.47 (±2.43) cm for *E. boweni* (Fig. 13.2, Table 13.1). Out of all 83 voucher samples, only 22% (*n* = 18) had whole, unchipped otoliths for morphometric measurements. The remaining otoliths (*n* = 65) had material missing from the right otoliths due to chipped or completely broken rostrums. Fork length of Guam samples ranged from 18.1 to 89.0 cm (Fig. 13.2, Table 13.1). Out of all 91 Guam samples, only 20% (*n* = 19) had otoliths pairs in whole condition for morphometric measurements (i.e., no rostrum breaks), whereas 24 samples had a whole right (*n* = 15) or left (*n* = 9) otolith. The remaining 48 samples (>50%) had material missing from both right and left otoliths from chips or breaks.

Spectral data showed distinct absorbance patterns (i.e., peaks and valleys at certain wavenumbers) based on the organic chemical composition of the otoliths (Fig. 13.3). Twelve samples in the voucher dataset and one sample in the Guam dataset were rescanned due to abnormally high peaks apparent in spectral absorbance, and the best spectra were selected for analysis. Distinct peaks among all spectra were observed at approximately 6,800 cm^{-1}, 5,160 cm^{-1}, and 4,300 cm^{-1} (Fig. 13.3). When voucher spectral data was pre-processed, species

separation became apparent in several wavenumber regions, with the largest differences observed around 7,100 cm^{-1}, 5,300 cm^{-1}, 4,400 cm^{-1}, and 4,000 cm^{-1} (Fig. 13.4).

The PCA on voucher data included all 949 wavenumber regions as variables, and showed significant clustering between species (Fig. 13.5), thus demonstrating the possibility of classifying *Etelis* sp. samples based solely on differences in their chemical compositions that are visible in the NIR region. Most of the variation in the multivariate morphometric data (99.4%) was described by PC1 (96.96%) and PC2 (2.48%) (Fig. 13.5). The two species clusters were clearly separated along PC1 and PC2; however there was some overlap present.

A total of 46 *E. boweni* and 39 *E. carbunculus* voucher sample spectra were used to calibrate PLS-DA models. Variable selection for each model selected 480 (fully cross-validated) and 623 (test split) wavenumber regions (out of 949) as the most important for discrimination between spectra. The fully cross-validated PLS-DA model was computed with four latent variables (LVs) and performed nearly perfectly for species classification, with a class error rate of 2.4%. The model correctly classified 83 voucher samples into *E. boweni* (n = 45) and *E. carbunculus* (n = 38) classes (Fig. 13.6A). Only one sample of each species was misclassified in cross-validation (Table 13.2). Both specificity (0.974) and sensitivity (0.978) were very high.

Comparable results were observed when voucher samples were split into calibration (70%, n = 60) and external validation (30%, n = 25) datasets (Fig. 13.6B). Class error of the calibration CV (venetian blinds) was 5.0%, and only one *E. boweni* and two *E. carbunculus* were misclassified (Table 13.2). The PLS-DA model also performed very well for true predictions on the external validation dataset and correctly predicted the species class of 24 voucher samples without information on the true species. Only one *E. carbunculus* was misclassified as *E. boweni*

(Table 13.2, Fig. 13.6B). The class error for the validation set was slightly lower than the calibration at 4.0%.

The fully cross-validated model (lowest class error rate) was used to test predictions on the archived Guam *Etelis* sp. samples. Prior probabilities were modified so that the model would not assume an equal probability of observing either species, as *E. boweni* appear to be rare in Guam (an assumption which was corroborated by morphometric results herein). The model predicted that 8% (7 of 91) of archived otoliths belonged to the newly identified *E. boweni*. Two Guam samples over 60 cm FL (GECC-1133, GECC-1317) were classified as *E. boweni* from predictions made on voucher-calibrated PLS-DA model, which we assume to be a correct species assignment (Fig. 13.7). Three Guam samples over 60 cm FL (GECC-0318, GVDP-0370, and GVDP-0503) were classified as *E. carbunculus* which we assume to be false, based on maximum reported FL of the species (Fig. 13.7). However, sample GVDP-0370 stood out as an outlier in the overall plot of Hotelling's T^2 and Q residuals, so prediction on this sample is unresolved. Five additional Guam samples were classified as *E. boweni*: GECC-0069 (50.4 cm), GECC-0591 (46.8 cm), GECC-0711 (39.9 cm), GECC-1494 (30.4 cm), and GVDP-0381 (35.7 cm) (Fig. 13.7).

DISCUSSION

This study explored the use of FT-NIR spectroscopy as a potential new method to distinguish the cryptic species pair. Differentiation of the cryptic species pair *Etelis carbunculus* and *Etelis boweni* was shown to be possible and highly reliable through examination of archived otoliths. FT-NIR spectral absorbance patterns of archived voucher otoliths from the SW Pacific

were distinct between species. Classification models applied using NIR spectral data were able to predict species with a high degree of accuracy, despite a relatively large spatial area of voucher specimen collection (±10° latitude and longitude) and regardless of whether otoliths were whole (i.e., unbroken). In addition, FT-NIR spectroscopy identified members of the newly described *E. boweni* species in the archived collection of *E. carbunculus* otoliths captured around Guam, which confirmed that the species' distributions overlap in this region. The identification of both *E. carbunculus and E. boweni* in the archived catch from Guam has important implications for fisheries management; therefore, it is imperative that the corresponding otolith collections are examined to ensure that the otoliths are assigned to the correct species.

FT-NIR spectroscopy was able to identify and separate the two cryptic *Etelis* species in this study with extremely high classification accuracy. Importantly, the model performed best when making predictions for voucher data, indicating spatial (or some other) variability existed within both *Etelis* species from these distant sampling locales. Further, while exploring the spectral data, we observed PCA clustering of fish based on region within voucher samples. Thus, FT-NIR spectroscopy may be a useful tool to also discriminate spatial differences (e.g., stock discrimination) in these species, and conceivably others. Robins et al. (2015) also reported evidence of spatial discrimination using this method. The best solution to avoid such geographic issues is to include samples from all locations in the calibration stage, in which case, the full scope of otolith microchemistry variation that may be encountered in the test dataset is captured by the model (Robins et al. 2015). Had our calibration models included some spectra from fish collected in Guam (for which species identification could be confirmed), it would have undoubtedly improved predictions.

Species predictions from best models suggested that archived otoliths from Guam belong primarily to *E. carbunculus* (92%: 84 of 91). Seven archived otoliths were predicted to belong to the newly identified *E. boweni*. There was a general lack of overlap in body sizes between the two *Etelis* species identified in Guam. Based on FT-NIR spectroscopy, samples classified as *E. boweni* included two of the largest individuals (>60 cm FL) and two smaller individuals (30.4 and 35.7 cm FL). Given what we know about maximum body sizes in *E. carbunculus* (<60 cm FL), it appears more spectra from these populations of fish are needed to provide reliable predictions of these cryptic species.

Predictive models for FT-NIR spectroscopy approaches are only as good as the calibration data provided to train models on. An important assumption of these analyses is that each voucher specimen was correctly identified to species. Voucher specimens were collected after recognition of the new cryptic species and the identification of the distinguishing external characteristics. However, the presence of the dark margin on the upper lobe of *E. boweni* caudal fin is not always easy to see and the 'sharpness' of the opercular spine is relative between the two species and not always obvious. While we are not calling into question the ability of the researchers to correctly identify the fish used in this study, the possibility of a misidentification cannot be discounted.

In summary, differentiation of voucher *Etelis carbunculus* and *Etelis boweni* specimens was shown to be possible and highly reliable through examination of FT-NIR spectroscopy. Beyond this cryptic species pair, FT-NIR spectroscopy is likely to provide the same level of accuracy regardless of species and be applicable to archived otoliths when other cryptic species are identified subsequent to collection. Case-specific, otolith morphometrics (i.e., shape analysis) may be the preferred method to determine species classification if otolith length is not a

distinguishing feature between the two species, otolith rostrums are not susceptible to breaking during extraction and transportation, and the archived sample size is not overwhelming; conversely, FT-NIR spectroscopy may provide greater accuracy and time efficiency.

Table 13.1. -- Descriptive statistics for fish fork length, fish weight, and seven otolith morphometric variables for voucher *Etelis carbunculus* and *Etelis boweni* from the SW Pacific and for *Etelis* sp. from Guam. Sample sizes (*n*) differ by variable depending on whether whole (i.e., unbroken) otoliths (*) were used, or whether field data existed for the sample. Morphometric abbreviations are noted in parentheses.

	SW Pacific voucher samples										Guam samples				
	Etelis carbunculus					*Etelis boweni*					*Etelis* sp.				
Metrics	Range	Mean	SE	*n*		Range	Mean	SE	*n*		Range	Mean	SE	*n*	
Fork length, cm	25.0-58.0	42.26	1.49	38		29.0-85.0	53.47	2.43	45		18.1-89.0	34.77	1.44	91	
Fish weight, kg	0.28-3.70	1.59	0.17	38		0.40-11.10	3.79	0.47	43		0.10-13.32	1.11	0.21	90	
Otolith length, mm (L)	10.6-14.4	11.80	0.48	8	*	11.6-17.4	15.35	0.67	10	*	7.7-20.6	10.89	0.38	43	*
Otolith width, mm (W)	4.69-8.85	6.25	0.16	38		3.88-7.61	5.84	0.14	45		3.50-8.20	5.13	0.11	91	
Otolith thickness, mm (T)	1.17-2.65	1.78	0.06	38		0.99-2.16	1.37	0.04	45		0.88-3.04	1.33	0.04	91	
Otolith area, mm^2 (A)	39.0-72.5	49.32	4.50	8	*	40.0-72.4	60.72	3.70	10	*	19.2-85.5	38.76	2.24	43	*
Otolith perimeter, mm (P)	30.2-48.7	36.35	2.40	8	*	33.6-72.4	43.74	2.10	10	*	20.7-60.2	31.09	1.23	43	*
Otolith sulcus groove width, mm (S)	0.78-1.28	1.04	0.02	38		0.59-1.17	0.85	0.02	45		0.58-1.39	0.90	0.02	91	
Otolith excisura major length, mm (E)	6.84-9.63	8.26	0.12	36		7.18-13.93	10.25	0.29	40		5.30-12.20	7.67	0.15	91	

Table 13.2. -- Confusion tables for cross-validation and external validation of partial least squares discriminant analysis models on spectral data obtained via Fourier transform near infrared spectroscopy fit using voucher *Etelis* species samples from the SW Pacific. Values in bold are the number of correct species assignments based on known species of the voucher dataset.

				Actual Class	
Data set	Model	*n*	Predicted as	*E. boweni*	*E. carbunculus*
Full cross-validation	Calibration	85	*E. boweni*	**45**	1
			E. carbunculus	1	**38**
Split-Test	Calibration	60	*E. boweni*	**32**	1
			E. carbunculus	2	**25**
	Validation	25	*E. boweni*	**12**	1
			E. carbunculus	0	**12**

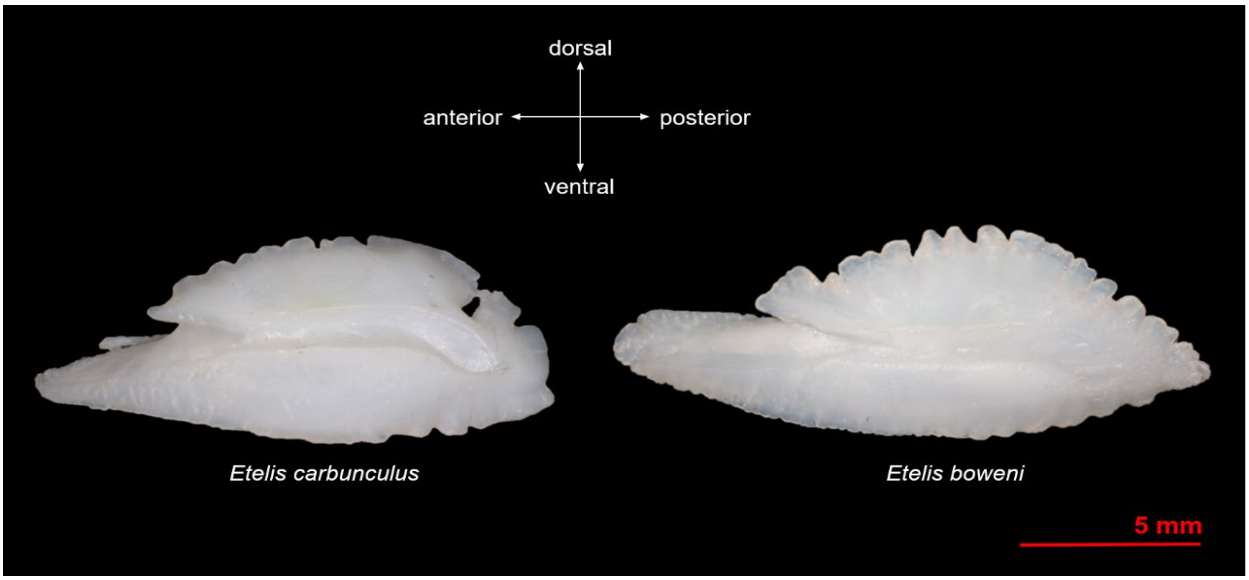


Figure 13.1. -- Proximal view of right sagittal otoliths from *Etelis carbunculus* (50 cm FL, left), and *Etelis boweni* (67 cm FL, right). Scale bar (in red) is 5 mm.

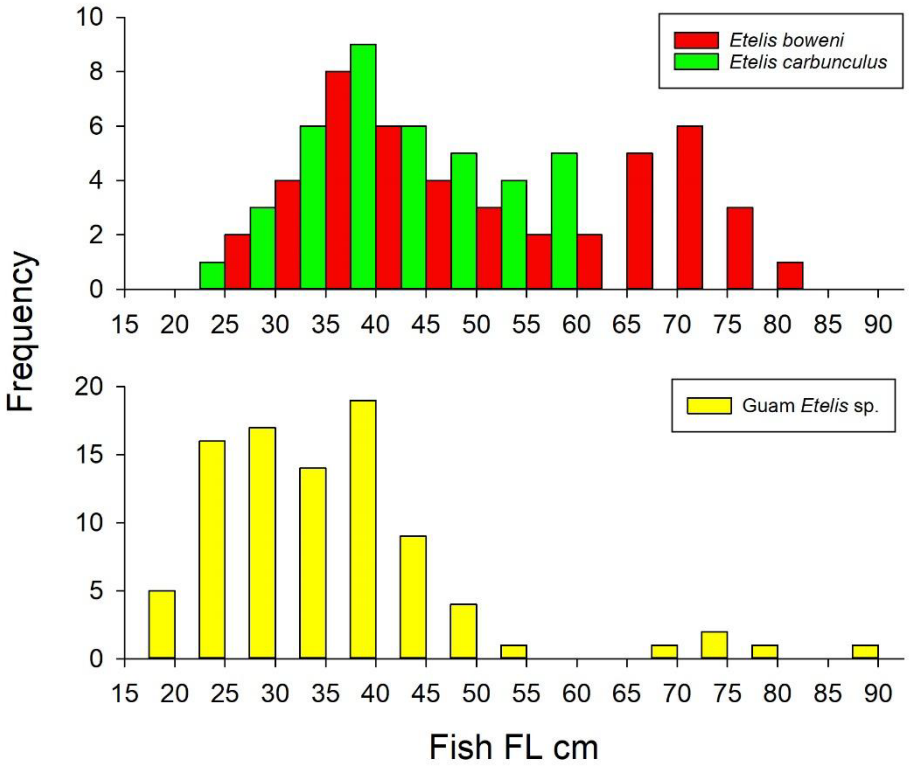


Figure 13.2. -- Length frequency distribution of voucher *Etelis carbunculus* (green, *n* = 39) and *Etelis boweni* (red, *n* = 46) specimens from the SW Pacific, and *Etelis* sp. (*n* = 91) previously collected from the Guam bottomfish fishery, which were assumed to be all *E. carbunculus* but are now known to be a mix of *E. carbunculus* and *E. boweni*.

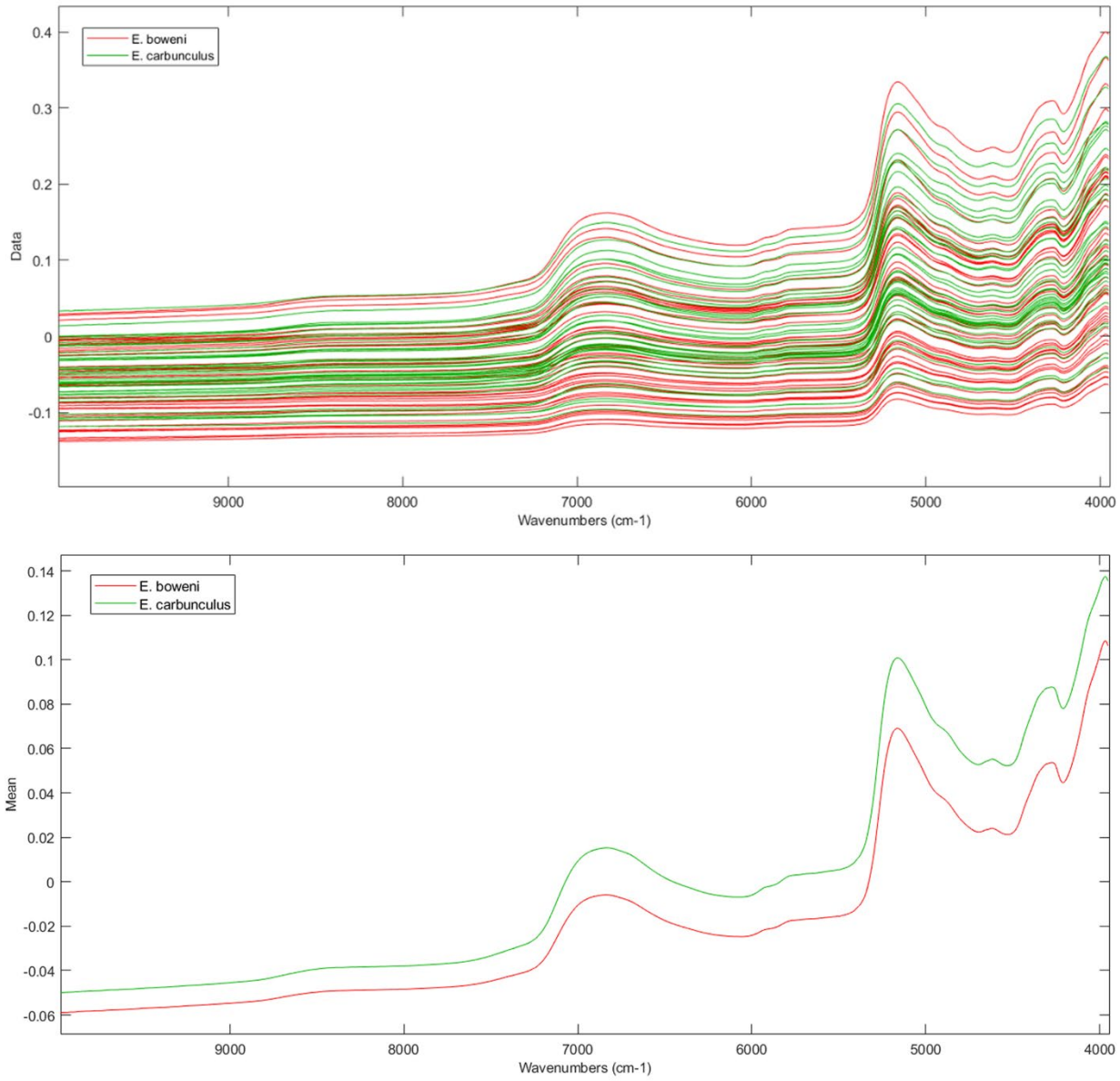


Figure 13.3. -- Raw (top) and average (bottom) Fourier transform near infrared spectra (10,000-4,000 cm^{-1}) of voucher otoliths by *Etelis* species (n = 85). Absorbance is on the y-axis for each plot.

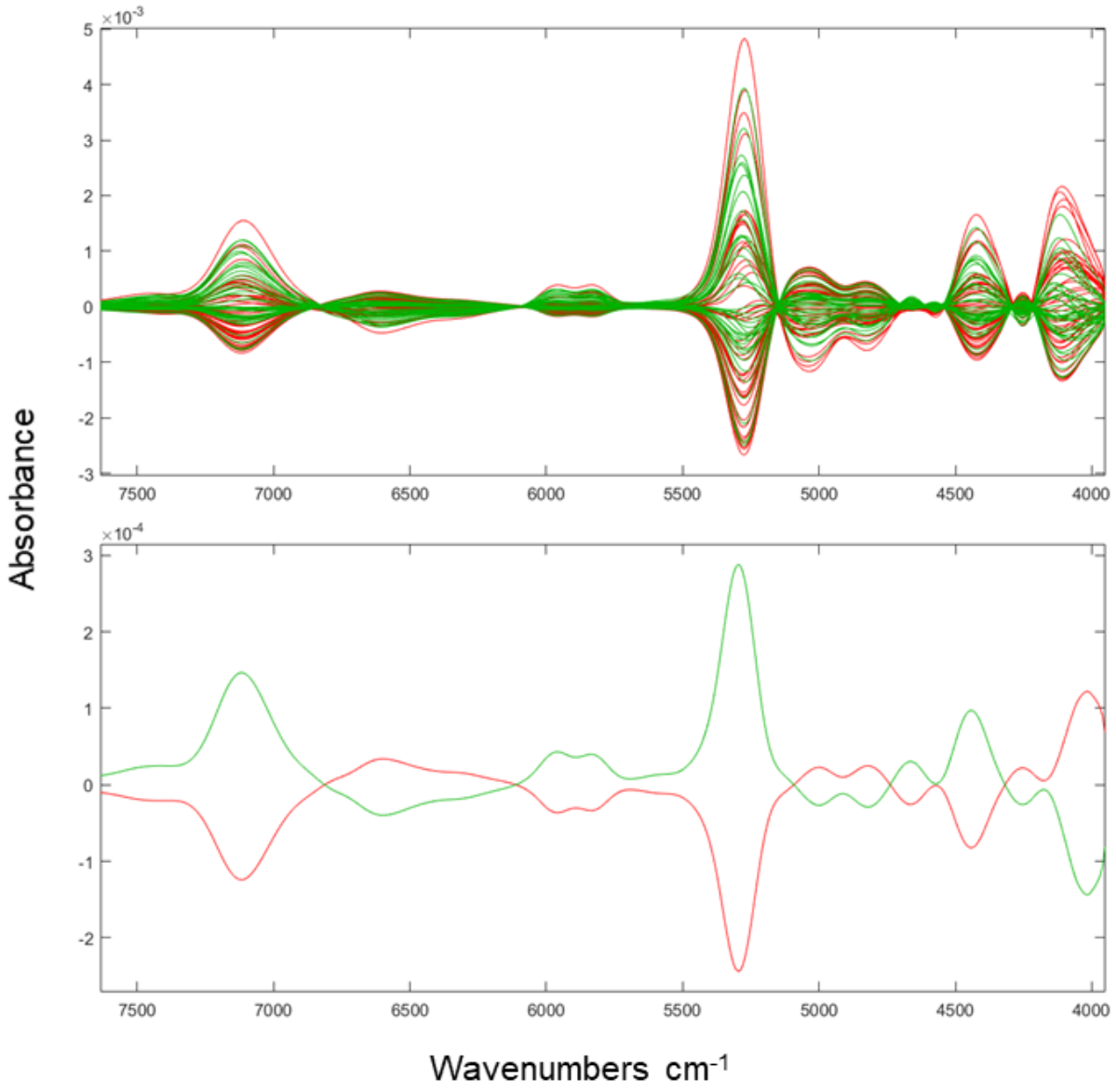


Figure 13.4. -- Raw (top) and average (bottom) Fourier transform near infrared spectra (10,000-4,000 cm^{-1}) of voucher *Etelis carbunculus* (green) and *Etelis boweni* (red) otoliths by species (n = 85), pre-processed by 13-point Savitzky-Golay smoothing (2nd order polynomial), 13-point Savitzky-Golay first derivative transformation, and mean-centering.

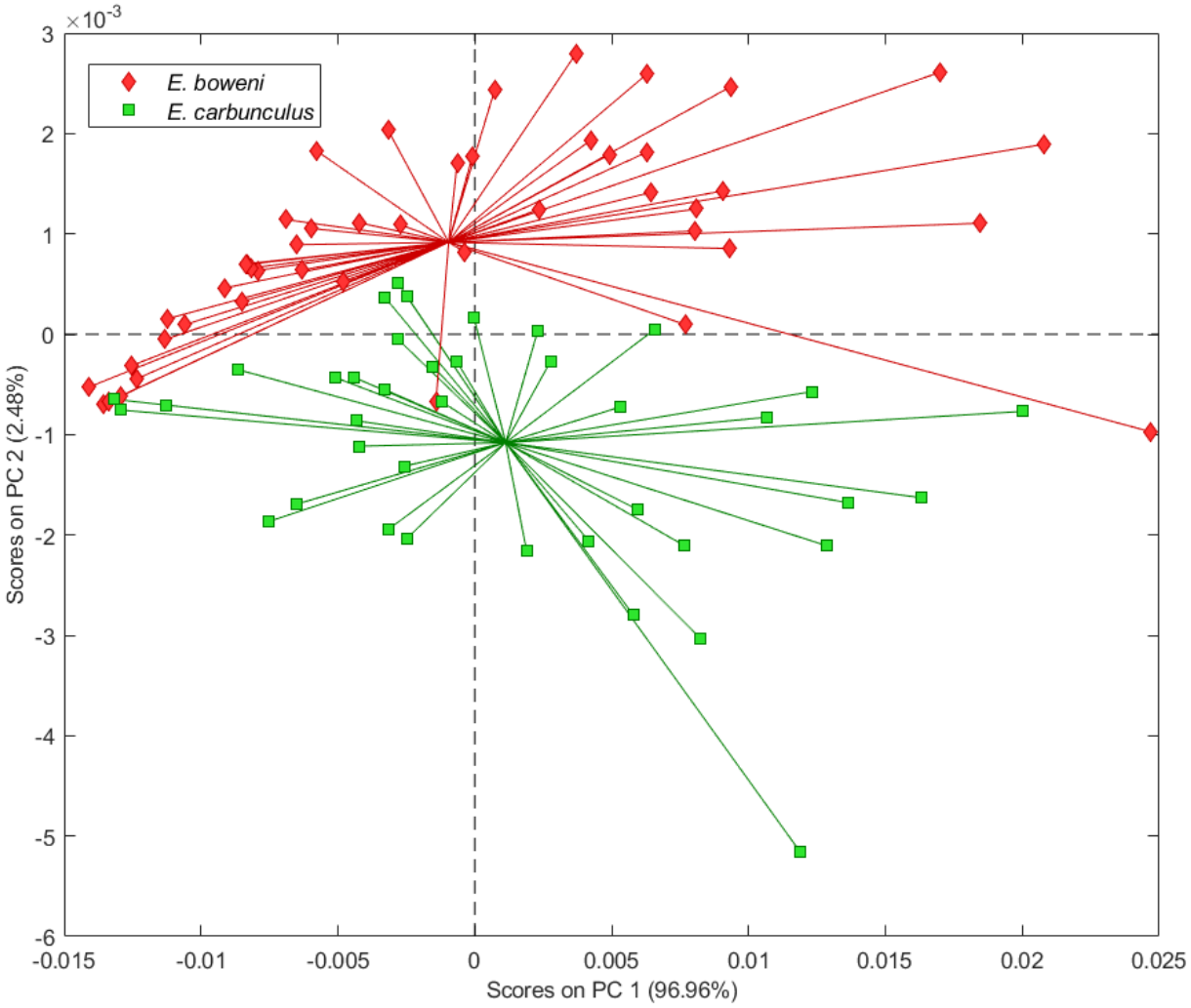


Figure 13.5. -- Scores plot of PC1 and PC2 from principal component analysis of Fourier transform near infrared spectral data from voucher otolith samples colored by *Etelis* species.

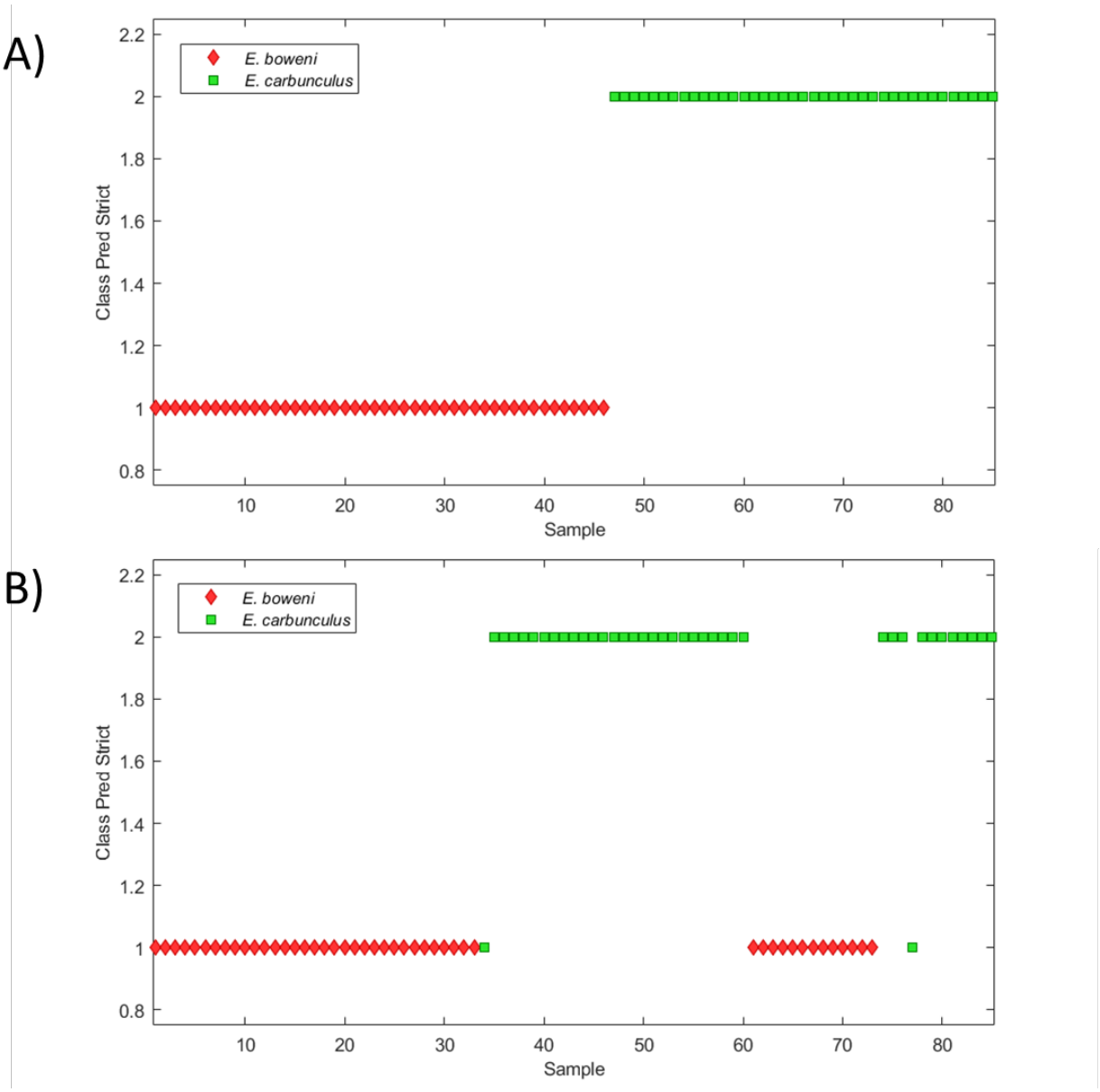


Figure 13.6. -- Partial least squares discriminant analysis model validation results showing *Etelis* species classification predictions by calibration models built using voucher samples *(n* = 85). Models were developed and validated on, A) all data (full random cross-validation) or B) data split into calibration (70%) and validation (30%) datasets. For the bottom panel, samples 1-60 represent the calibration (i.e., training) dataset and samples 61-85 represent the validation dataset. For each plot, class 1 = true *Etelis boweni* and class 2 = true *Etelis carbunculus*.

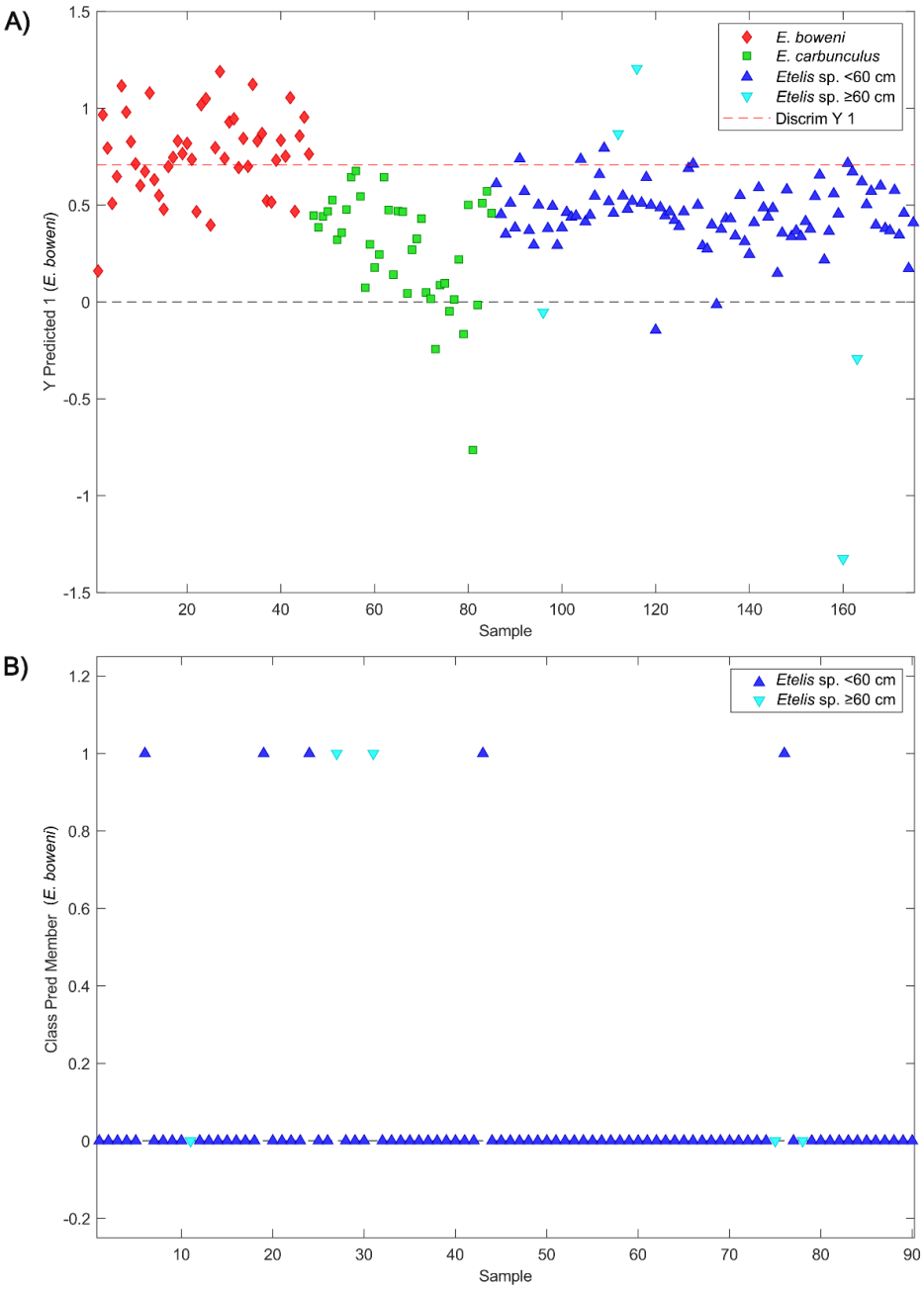


Figure 13.7. -- Partial least squares discriminant analysis model testing results showing the model trained on voucher Fourier transform near infrared spectra used to predict the species of Guam *Etelis* sp. samples. For A) samples 1-85 represent the calibration (i.e., training) dataset and samples 86-176 represent the Guam (i.e., unknown) dataset. The discrimination line (red dash) indicates the point at which samples above are predicted as *Etelis boweni*, and below are predicted as *Etelis carbunculus*. Guam data are displayed by a size range cut-off, given that only *Etelis boweni* are expected to grow larger than 60 cm FL. B) Model predictions for *E. boweni* are shown.

14. Trials and Tribulations of Using FT-NIR Spectroscopy on Coastal Pelagic Species: Method Development for Scanning Pacific Sardine Otoliths

Emma M. Saas[1], Brittany D. Schwartzkopf[2], Emmanis Dorval[3], and Dianna L. Porzio[4]

[1] Saltwater Inc. (under contract with Southwest Fisheries Science Center)
733 N Street
Anchorage, AK USA

[2] Southwest Fisheries Science Center
NOAA, National Marine Fisheries Service
8901 La Jolla Shores Drive
La Jolla, CA USA

[3] Lynker (under contract with Southwest Fisheries Science Center)
202 Church Street SE, Suite 536
Leesburg, VA USA

[4] Marine Region
California Department of Fish and Wildlife
2451 Signal Street
AltaSea Berth 59
San Pedro, CA USA

INTRODUCTION

One innovative approach that is being evaluated to rapidly age fish is Fourier transform near infrared (FT-NIR) spectroscopy. FT-NIR spectroscopy is a non-destructive analytical tool that collects information about a sample's light absorption within the near infrared electromagnetic spectrum (800-2,500 nm). The technology itself has been well established in the dairy and agricultural industry for decades (Rasco et al. 1991, Sorvaniemi et al. 1993); however, its use in fisheries research is still relatively novel. The research and development for implementation of FT-NIR spectroscopy at the six National Oceanic and Atmospheric (NOAA) Fisheries Science Centers is part of a NOAA strategic initiative (Helser et al. 2019a). The initiative funds research to develop methodology for respective species of interest at each center and potentially makes FT-NIR spectroscopy ready for practical use on a regular basis.

The application of FT-NIR spectroscopy has proved to be a useful tool for fisheries scientists despite the limited understanding of what is being measured within an otolith. FT-NIR spectroscopy has been used to scan the otoliths and derive ages of individual fish of several species, including walleye pollock (*Gadus chalcogrammus*) (Helser et al. 2019b), red snapper (*Lutjanus campechanus*) (Passerotti et al. 2020a, b), saddletail snapper (*Lutjanus malabaricus*) (Wedding et al. 2014, Robins et al. 2015), Pacific cod (*Gadus macrocephalus*) (Healy et al. 2021), and barramundi (*Lates calcarifer*) (Wright et al. 2021). The methods for analyzing otoliths of the aforementioned species include scanning the otoliths convex side down at a consistent orientation but vary slightly in the accessories used. Helser et al. (2019b) and Passerotti et al. (2020a, b) used a gold-coated reflector (i.e., stamp) over the otolith, though the latter found the additional use of a Teflon disc improved the prediction models due to the small

size of otoliths. Wedding et al. (2014) placed their samples in quartz glass vials, and Robins et al. (2015) used no accessories at all. The use of scanning accessories likely needs to be examined for each new species to be aged using FT-NIR spectroscopy.

The Southwest Fisheries Science Center in La Jolla, California, is tasked with assessing coastal pelagic species (CPS), which include performing stock assessments. Pacific sardine (*Sardinops sagax*) are an important CPS whose populations are assessed annually, and thus ages are generated frequently. Pacific sardine range from southeast Alaska down to Baja California, can grow up to about 41 cm standard length (SL), though individuals caught in the fishery are most frequently under 30 cm SL, and are short lived with a maximum reported age of 15 years old, though most individuals collected from trawl surveys and port sampling are less than 5 years old (Eschmeyer et al. 1983, Kuriyama et al. 2020).

The use of FT-NIR spectroscopy technology to estimate age must be evaluated on a species-to-species basis, as some otolith morphology and increment deposition patterns might be a better fit for this type of analysis than others. Pacific sardine were chosen to be evaluated due to their importance in CPS stock assessments, management, and the California Current Ecosystem. Pacific sardine present a particular challenge to FT-NIR spectroscopy due to their small size and short lifespan; their otoliths are extremely small (<6 mm, Javor et al. 2011) compared to long-lived species whose otoliths grow to much larger sizes. In this study, we developed a methodology that best captures the spectral information from Pacific sardine otoliths. Otolith condition, placement, and spectrometer accessories were assessed for optimal spectral output. Otoliths were scanned using 10 accessories, such as a chrome ring and a Teflon disc, to investigate what tools were needed to accommodate the small otolith size.

METHODS

Pacific sardine were collected along the west coast from CPS spring and summer trawl surveys conducted by the NOAA's Southwest Fisheries Science Center (NOAA-SWFSC) from 2004 to 2021, and port sampling conducted by California Department of Fish and Wildlife (CDFW), Washington Department of Fish and Wildlife (WDFW), and Oregon Department of Fish and Wildlife (ODFW) from 2004 to 2016 (see Dorval et al. 2022 for detailed sampling methods). Samples collected for Pacific sardine range from Vancouver Island, British Columbia, Canada (49.65°N, 125.44°W) to around San Diego, California (33.25°N, -117.56°W). All Pacific sardine samples that were analyzed were considered to be the northern stock, using the most up-to-date population structure assessment (Zwolinski and Demer 2023). After sample collection, sagittal otoliths were extracted, cleaned, dried, and stored in labeled plastic tubes. A set of Pacific sardine otoliths from an age validation study were included as 'known age' samples (K. C. James et al., SWFSC, in review). Juvenile fish were collected in 2014 and 2015, were chemically marked with oxytetracycline (OTC), and maintained in captivity for up to a year. Sagittal otoliths were extracted after mortality and stored dried in black, labeled plastic tubes.

Traditional Ageing

Pacific sardine otoliths were traditionally aged using the methods described by Yaremko (1996), which include submerging whole otoliths in distilled water with a black background. The number of annuli on the distal side of the otolith was counted using a dissecting microscope at 25x magnification with reflected light. Only otoliths whose ages were agreed upon by all three

age readers were chosen for the reference sets to create calibration models for each species. The NOAA-SWFSC and CDFW data sets only comprised otoliths whose ages were previously included in stock assessments. To increase the sample size for older age classes of Pacific sardine, we included otoliths collected off Oregon and Washington and aged by WDFW, as older Pacific sardine are generally found in the Pacific Northwest. Among all sources, we aimed to have at least 25 samples per age class for the reference set, but this was hard to achieve for older ages that were not frequently caught (Table 14.1).

Scanning Specifications

For all experiments and model predictions described below, sagittal otoliths were scanned using near infrared spectroscopy on a Bruker TANGO-R FT-NIR single-channel spectrometer (Bruker Optics, Ettlingen, Germany). Spectral data were collected using the OPUS software v8.5(SP1). Scanning specifications for the TANGO were set for 64 scans (averaged into a final raw absorbance value), with a frequency of 16 cm^{-1} at 8 wavenumber intervals. All spectral data were plotted as absorbance values by wavenumbers ranging from 11,528 to 3,952 cm^{-1}. All otoliths were placed convex side down and were weighed to the nearest 0.0001g after being scanned. Primary scanning and analysis were performed at the NOAA-SWFSC in La Jolla, California, with some additional scans for experiments performed at the NOAA-SWFSC in Santa Cruz, California.

Experimental Scans

A series of experimental scans determined the best placement, otolith condition, and accessories for the otoliths in this study. Unless otherwise stated, all experimental scans were done only on the otoliths collected from the NOAA-SWFSC trawl survey, hereafter referred to as the Pacific sardine trawl set (Table 14.1). Otoliths were scanned using a chrome ring and a large gold stamp for the initial tests to determine best placement and otolith condition.

The first sets of experimental scans were completed to determine the best orientation and position on the scanning window. Results were judged by visually assessing the raw spectra in OPUS to see if it met the criteria for a 'usable' scan. A usable scan of an otolith has minimal noise and a near-flat ~11,528 to 7,192 (or >8,000) wavenumber range, as well as a scan that falls below 2 AU (absorbance unit) on the y-axis (Jason Erikson, Bruker, pers. comm). We first scanned three pairs of Pacific sardine otoliths in four different orientations in the center of the window: 0°, 90°, 180°, 270° (Fig. 14.1a). We then scanned those same three Pacific sardine otoliths in nine different positions on the window with the 0° orientation (Fig. 14.1b). Otoliths were placed in the 0° orientation in the center of the window (position 5; Fig. 14.1b) for all subsequent experimental scans.

Otoliths are sometimes broken or slightly dirty (Fig. 14.2). Breakage and excess tissue aren't uncommon, so we tested these conditions to determine if those otoliths provided usable scans. Left and right otoliths, where one was broken and the other was whole, were scanned from 42 individuals from the Pacific sardine trawl sample. Age predictions from a calibration model from only broken otoliths, and only whole otoliths were then compared. This same analysis was

then done for left and right otoliths collected from 40 individual fish where one otolith was dirty and the other one was clean.

To determine the accessory that produced the best scan, we tried several combinations of accessories to determine what tools, if any, were needed to give the most reliable, precise, and repeatable results. The nine combinations of scanning accessories that were tested on five Pacific sardine otoliths included 1) no accessory, 2) a small chrome stamp, 3) a small gold stamp, 4) a large gold stamp, 5) a chrome ring, 6) a chrome ring plus a flat chrome stamp, 7) a chrome ring plus a small gold stamp, 8) a chrome ring plus a large gold stamp, 9) and a Teflon ring (Fig. 14.3). The flat chrome stamp (Fig. 14.3b), rather than the chrome stamp (Fig. 14.3a), was used in combination with the chrome ring to prevent light from escaping through any gaps. Accessories that resulted in scans that visually lacked the above criteria for a 'usable' scan were excluded from any future testing.

To test for fine-scale variability between large gold stamps of the same manufacturer and model, two otoliths were scanned by one individual on Santa Cruz SWFSC's TANGO with the chrome ring and the La Jolla SWFSC's large gold stamp, then without changing any other variables, scanned again with the Santa Cruz SWFSC's large gold stamp. The resulting raw spectra were judged qualitatively.

RESULTS

For both rotation and otolith placement, consistency was most important. Raw spectra appeared different among orientations (Fig. 14.4). The position on the TANGO window was very important, and the resulting raw spectra differed greatly among some positions (Fig. 14.5).

Having the otolith in position seven, eight, or nine (bottom left, bottom center, bottom right, respectively) produced the worst scans, with almost no prominent peaks, whereas position five (center) appeared to produce the best scans with very prominent peaks (Fig. 14.5).

Raw spectra from broken and whole otoliths of the same individual showed some differences (Fig. 14.6), whereas raw spectra from dirty and clean otoliths of the same individual showed little difference (Fig. 14.7). It is important to note that otoliths that were dirty were not completely covered in tissue, but rather had random spots of tissue occurring at different locations on the surface of the otolith.

The choice of accessory greatly influences the shape and absorbance values observed for raw spectra for the Pacific sardine otoliths (Fig. 14.8). The accessories that were completely ruled out for further use due to a lack of a near-flat 11,528 to 8,000 wavenumber range (recommended by Jason Erikson, Bruker, pers. comm.) included the small chrome stamp and small gold stamp, with or without the chrome ring, the large gold stamp without a chrome ring, and Teflon disc (Fig. 14.9). The three accessories set-ups that were chosen for future evaluation on the trawl set for Pacific sardine were 1) no accessory, 2) chrome ring only, and 3) chrome ring and large gold stamp.

There appeared to be a difference in absorbance values between the raw spectra from the same otoliths scanned with La Jolla's gold stamp and Santa Cruz's gold stamp, where La Jolla's gold stamp produced higher absorbance values (Fig. 14.10).

CONCLUSIONS

Overall, consistency is the key to obtaining reliable FT-NIR spectroscopy scans for small CPS. Otoliths should be scanned in a manner that is easily reproduced from scan-to-scan and technician-to-technician. Left and right otoliths that were both clean and whole showed no significant difference, which is consistent with Robins et al.'s (2015) findings for red snapper otoliths. Whole, clean, and dry otoliths must be in the same orientation and position on the window for every scan. Researchers have a higher risk for error when placing small otoliths on the scanning window than larger otoliths, so extra care needs to be taken. Small amounts of dried tissue on an otolith seems to be acceptable, though clean otoliths are always preferred if available. Markings to indicate correct otolith placement, without compromising the integrity of the TANGO's quartz scanning window, are helpful. Minimizing the number of accessories means fewer changing variables over time, such as scratches and replacement parts, yet using no accessory yields unusable scans for the smallest otoliths. We used our best judgment to decide that the gold stamp potentially added unnecessary variability for Pacific sardine and we concluded using only the chrome ring is the best accessory, but this may not be true for other species with larger otoliths. In conclusion, the method developed in this study may not be the ideal solution for all species with small otoliths, but hopefully provides a starting point for future species-specific method development.

Table 14.1. -- Sample sizes per age class for the reference set from all data sources for Pacific sardine (n = 537) that were used to create a calibration model. CDFW = California Department of Fish and Wildlife, WDFW = Washington Department of Fish and Wildlife, OTC = oxytetracycline, SWFSC = Southwest Fisheries Science Center.

	Age (years)									
Source	0	1	2	3	4	5	6	7	8	9
CDFW - Fishery	9	13	19	19	9	9	4	3	2	0
WDFW - Fishery	0	0	0	0	0	0	14	22	10	10
SWFSC – OTC	92	131	3	0	0	0	0	0	0	0
SWFSC - Trawl	6	15	20	18	18	17	21	25	28	0

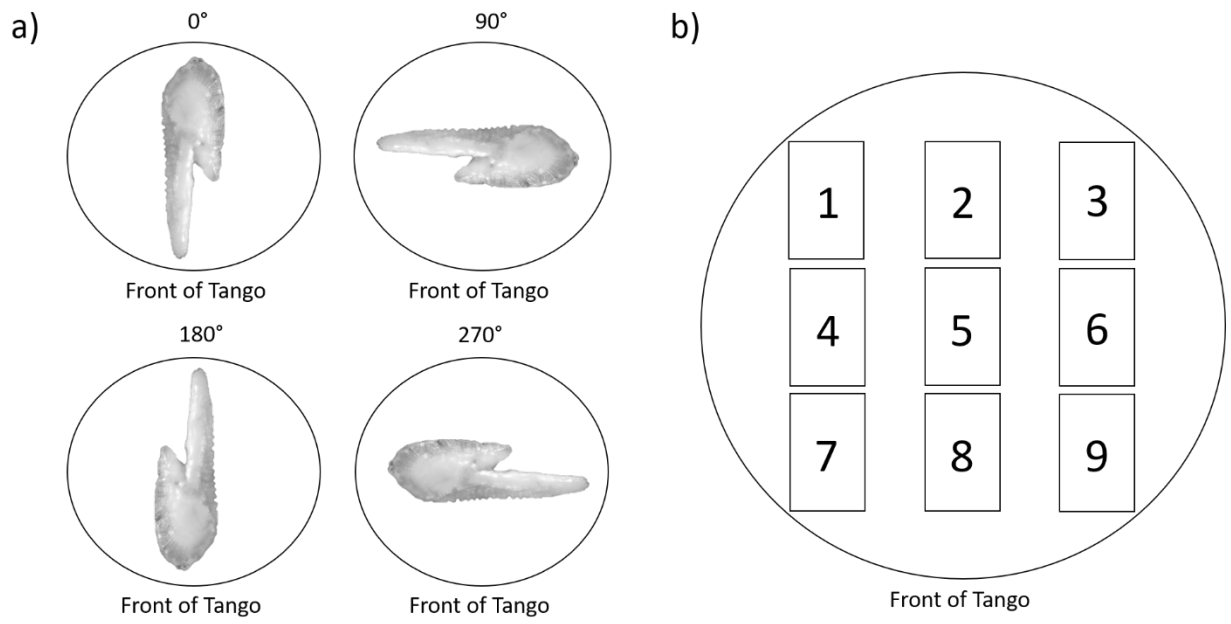


Figure 14.1. -- a) Otoliths were scanned in four orientations, 0°, 90°, 180°, and 270°. b) Otoliths were scanned in nine different positions on the scanning window. The otoliths shown here are from Pacific sardine (*Sardinops sagax*) viewed as if you were directly above the scanning window looking down.

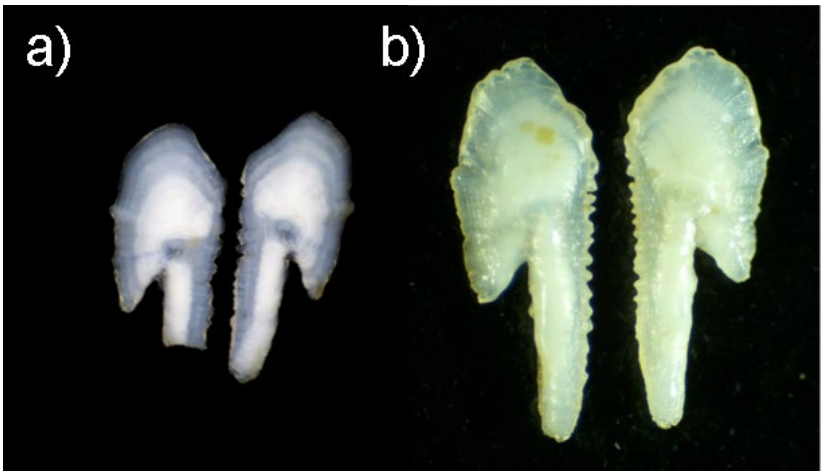


Figure 14.2. -- Examples of a) whole and broken and b) dirty (left) and clean (right) sagittal otolith pairs that were used to examine scan difference between otoliths of different condition. The otoliths pictured in a) were submerged in water.

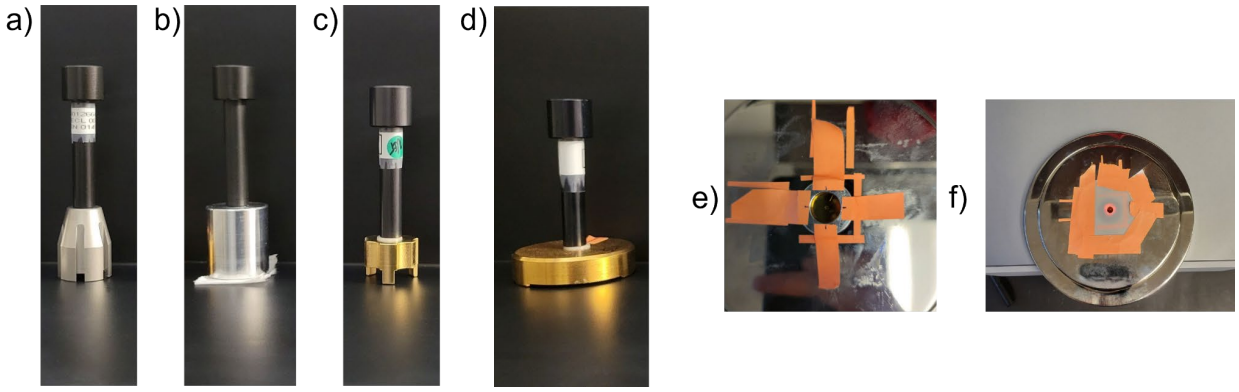


Figure 14.3. -- The accessories used to test scanning methods, sometimes in conjunction with one another, were a) small chrome stamp, b) flat chrome stamp, c) small gold stamp, d) large gold stamp, e) chrome ring taped to the TANGO, and f) Teflon disc taped to the TANGO.

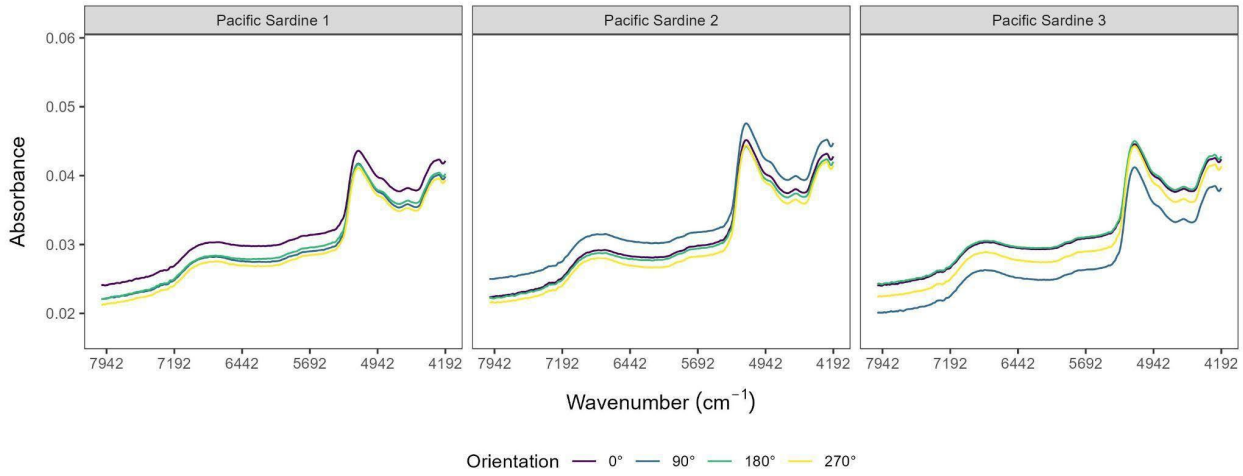


Figure 14.4. -- Raw spectra of three Pacific sardine (*Sardinops sagax*) otoliths in four different orientations. See Figure 14.1a for what each orientation represents.

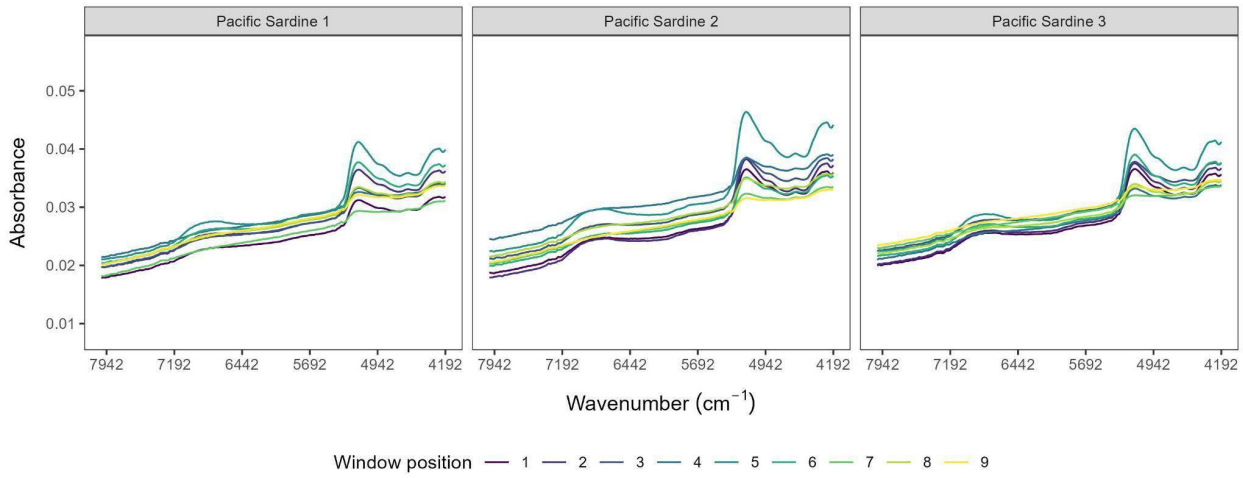


Figure 14.5. -- Raw spectra of three Pacific sardine (*Sardinops sagax*) otoliths in nine different positions on the TANGO window. See Figure 14.1b for what each window position represents.

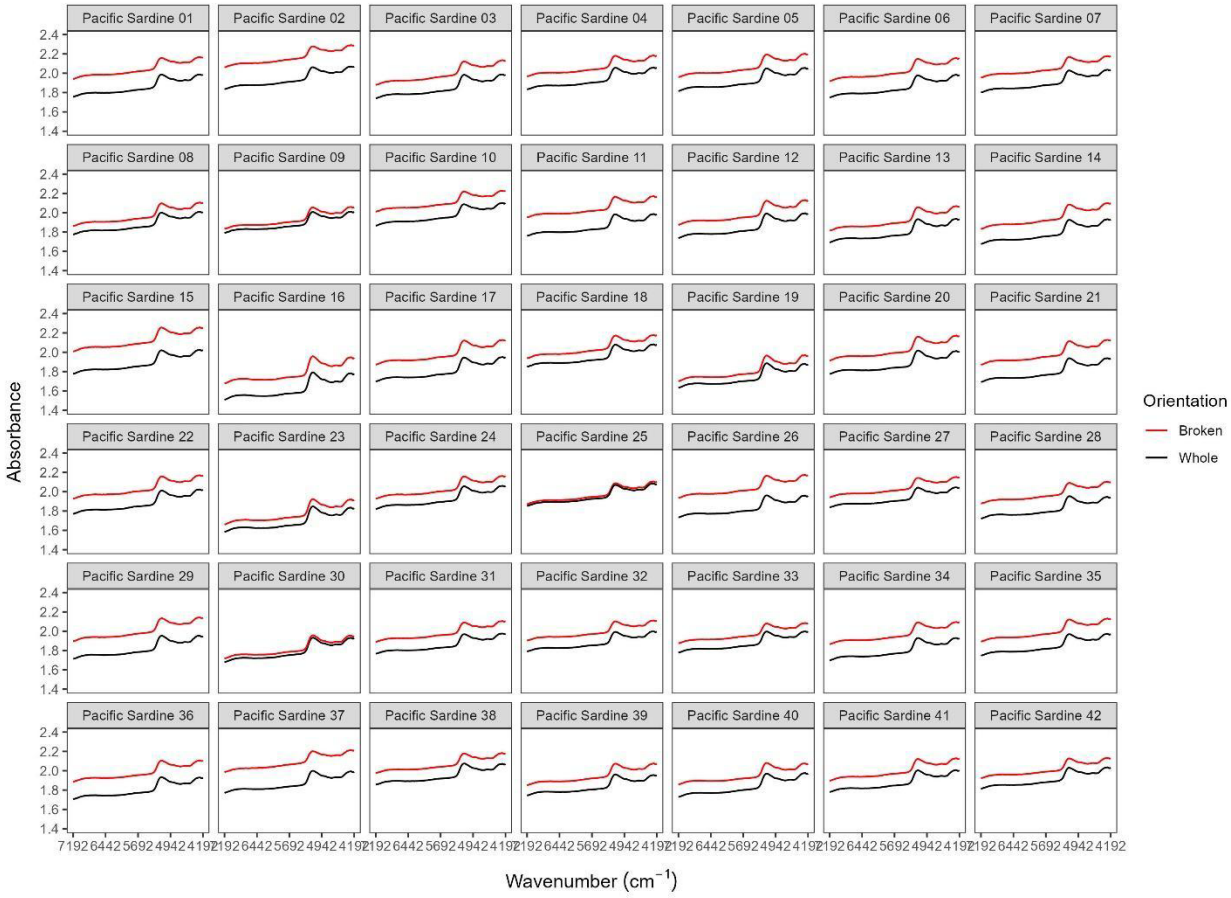


Figure 14.6. -- Raw spectra of Pacific sardine (*Sardinops sagax*) otoliths pairs from 42 individuals with one otolith being broken (red) and the other being whole (black) from each individual.

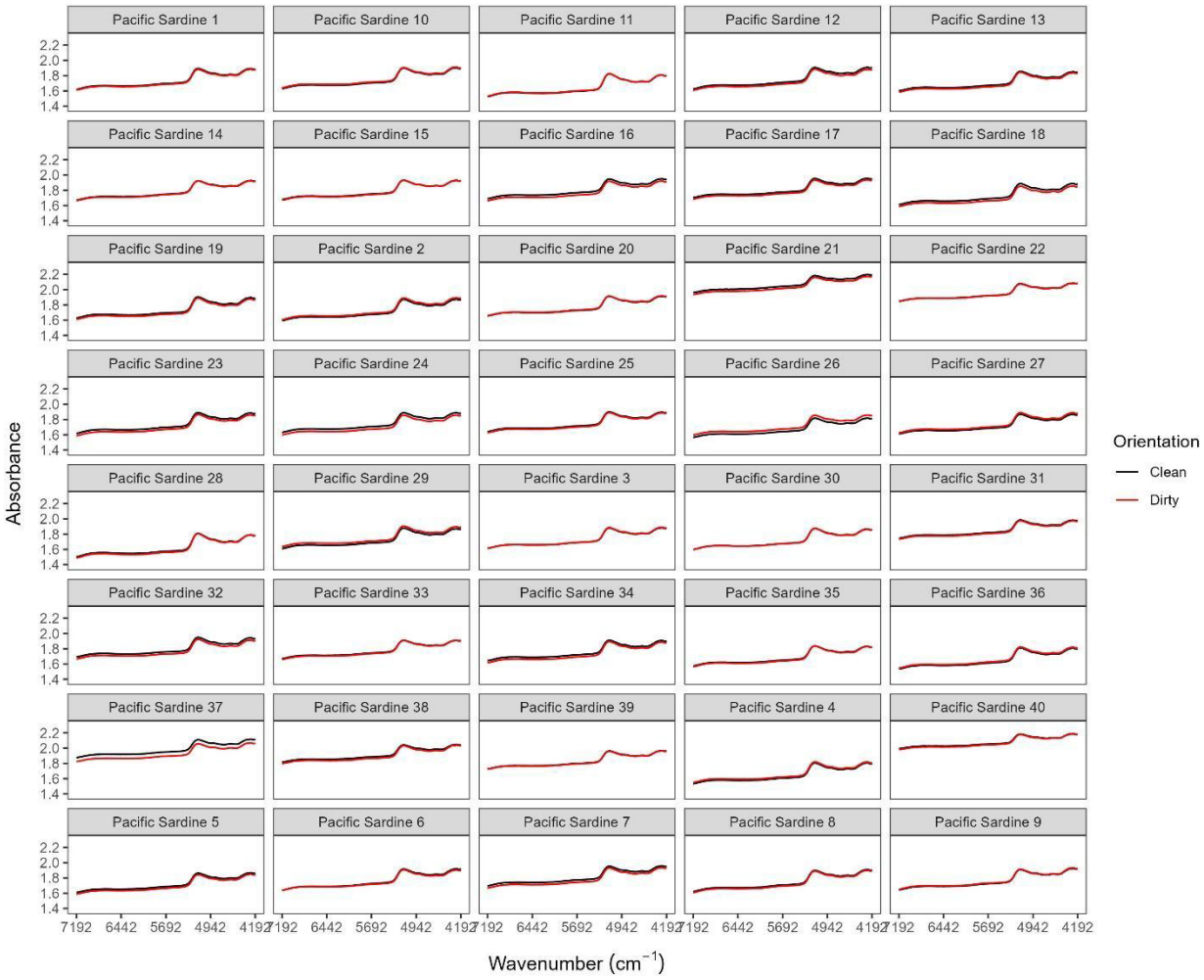


Figure 14.7. -- Raw spectra of Pacific sardine (*Sardinops sagax*) otoliths pairs from 40 individuals with one otolith being dirty (red) and the other being clean (black) from each individual.

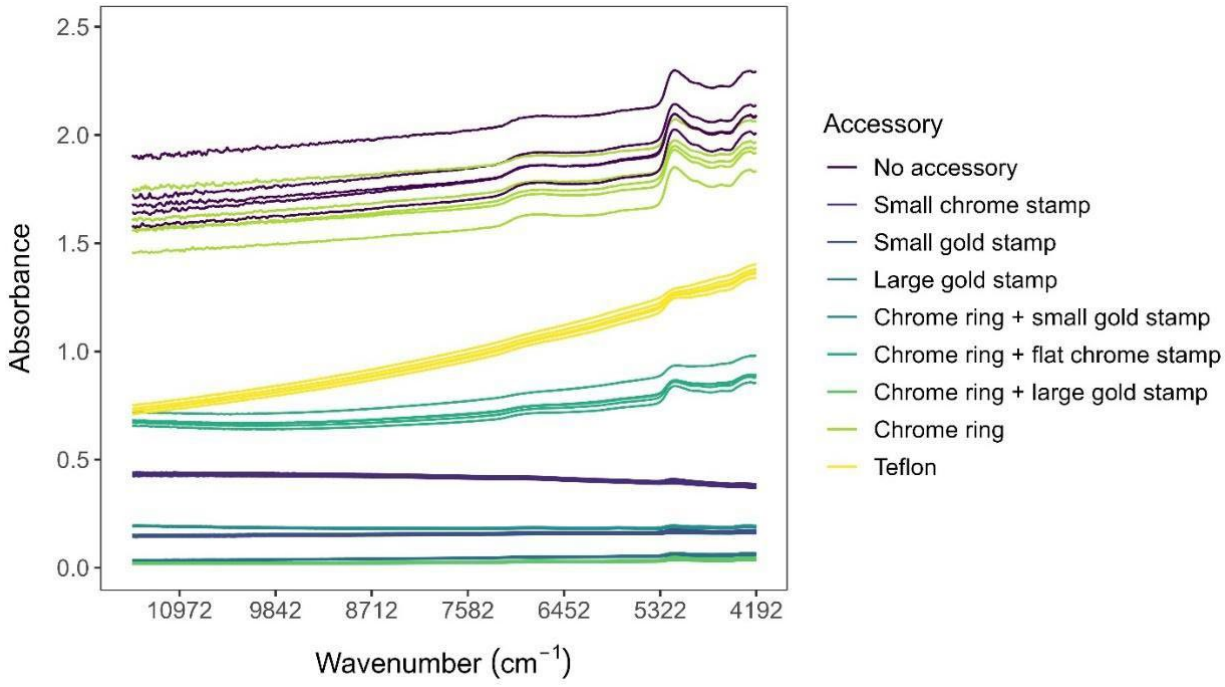


Figure 14.8. -- Raw spectra of the nine different accessory combinations for Pacific sardine (*Sardinops sagax*). This figure is meant to showcase the large differences in absorbance values as the spectral shape of certain accessories is masked due to the differences in y-axis values.

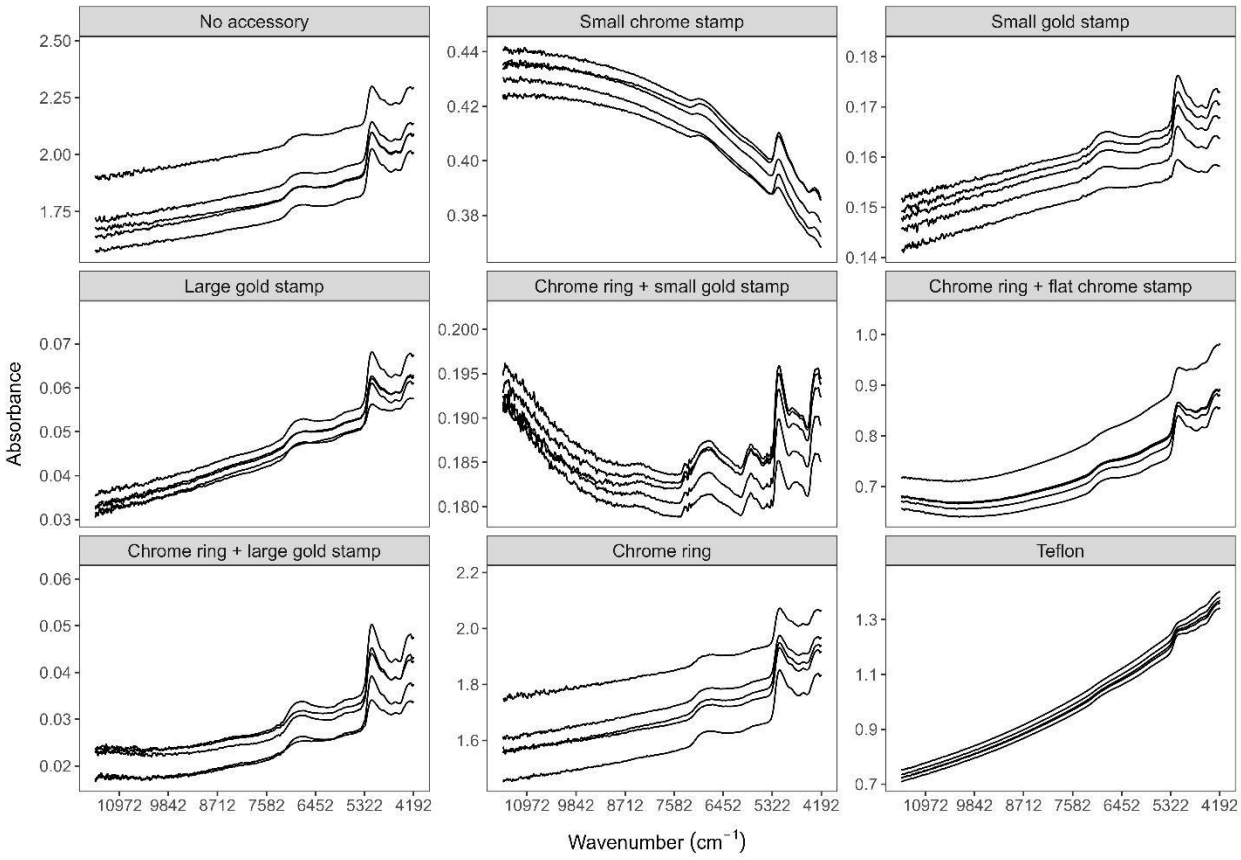


Figure 14.9. -- Raw spectra of five Pacific sardine (*Sardinops sagax*) otoliths for the nine different accessory combinations. This figure is meant to showcase the different shapes of the raw spectra, so each accessory has a different y-axis.

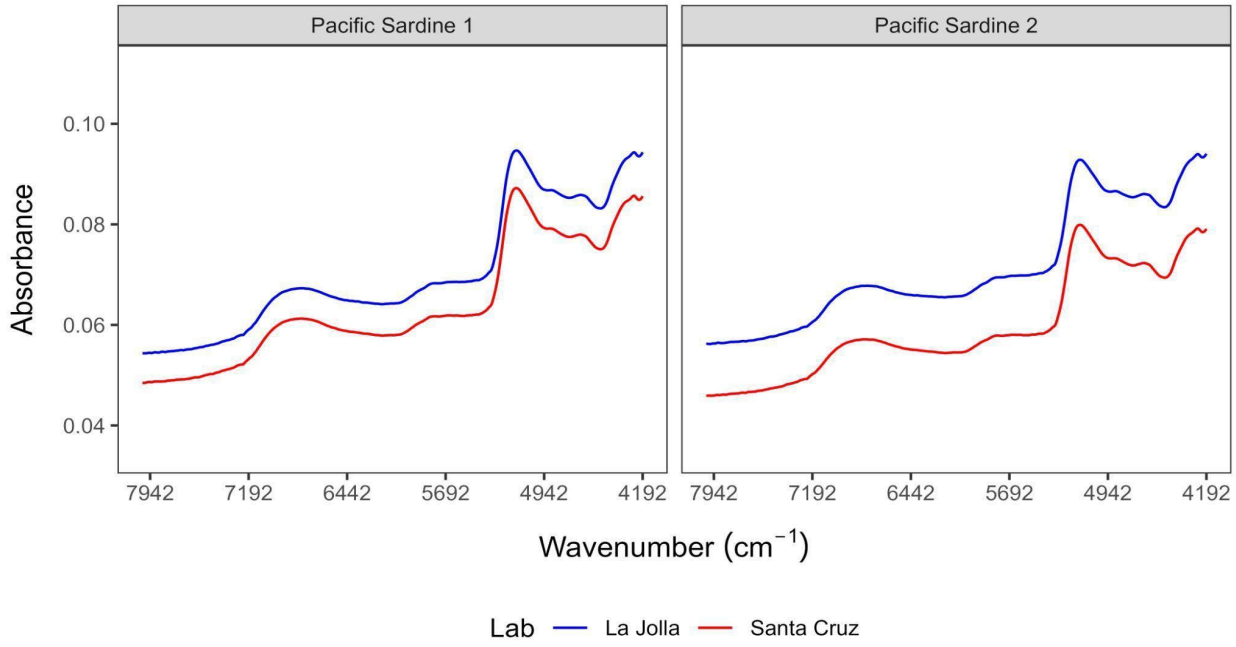


Figure 14.10. -- Raw spectra of two Pacific sardine (*Sardinops sagax*) scanned on Santa Cruz SWFSC's TANGO with the chrome ring and the La Jolla SWFSC's large gold stamp (blue), then without changing any other variables, scanned again with the Santa Cruz SWFSC's large gold stamp (red).

15. Benefits and Challenges of Using FT-NIR Spectroscopy for Age Estimation at the Northeast Fisheries Science Center

Eric M. Robillard[1], Michelle S. Passerotti[2], and Alexander M. Rubin[2,3]

[1] Northeast Fisheries Science Center
NOAA, National Marine Fisheries Service
166 Water Street
Woods Hole, MA USA

[2] Northeast Fisheries Science Center
NOAA, National Marine Fisheries Service
28 Tarzwell Drive
Narragansett, RI USA

[3] IBSS Corporation
1110 Bonifant Street
Silver Spring, MD USA

INTRODUCTION

Fourier transform near infrared (FT-NIR) spectroscopy is an emerging method to estimate age of marine fishes. FT-NIR spectroscopy has been previously shown to be an effective method at predicting ages of individuals using partial least squares (PLS) regression models trained on known-age samples (Arrington et al. 2022, Healy et al. 2021, Passerotti et al. 2020b, Rigby et al. 2016). However, the efficiency and cost-savings of FT-NIR spectroscopy relative to traditional ageing techniques is species- and laboratory-specific and has yet to be evaluated directly within the Northeast Fisheries Science Center (NEFSC) Age and Growth Laboratory.

Within the framework of high-volume production ageing, we compare the feasibility of estimating ages using FT-NIR spectroscopy for Acadian redfish *Sebastes fasciatus* to that of traditional ageing methods that require processing of otoliths and visual counts of annual growth rings under a microscope. Production ageing of otoliths at the NEFSC uses a unique high-throughput processing setup that employs the use of barcoding, data entry applications and a Labcut 250B saw (Extec, Enfield, CT) to maximize efficiency. This setup allows multiple otoliths to be processed simultaneously and represents a different perspective on efficiency that has not been directly compared to FT-NIR spectroscopy ageing methods to date. Herein, we use three metrics to compare these methods: efficiency, cost, and ageing precision.

METHODS

Processing

Processing otoliths for ageing with the Labcut 250B saw consists of four steps: 1) otoliths are mounted on polystyrene resin in trays, 2) once the resin is dry, the blocks with otoliths are removed from the trays, 3) the blocks are cut using the Labcut 250B saw, and 4) the sections are mounted on clear plastic slides (NEFSC in review). Each block holds on average 75 samples, which can be cut simultaneously with the saw.

For FT-NIR spectroscopy, all samples were scanned on a Bruker TANGO-R near infrared spectrometer (Bruker, Billerica, MA) per the methods outlined in Rubin et al. (this volume). All otoliths were cleaned with 70% ethanol and blotted dry using laboratory tissue before scanning. Otoliths then were scanned sulcus side down with the rostrum oriented horizontally on the sample window while covered with a rubber-rimmed gold reflective stamp while scanning to reduce stray light. All absorbance spectra were collected at a resolution of 64 scans every 16 wavenumbers.

Efficiency

To estimate efficiency of the traditional Labcut 250B method, processing times were recorded for 4,000 otoliths across all steps including embedding, sectioning, and mounting sections on slides and averaged to calculate the number of otoliths that could be processed per hour (Table 15.1). For FT-NIR spectroscopy, efficiency was estimated using scans of Acadian

redfish otoliths from the fall 2015 (n = 984) and fall 2016 (n = 925) Northeast Trawl Survey. Processing time for FT-NIR spectroscopy included cleaning the otolith, placing it on the sample window, initiating and completing a scan, and placing the otolith back in the envelope. Times were averaged per otolith (Rubin et al. this volume).

Cost

Costs were estimated using metrics of initial and long-term cost. Initial cost for each method includes the purchase of two Labcut 250B saws and two Tango-R spectrometers (one as a backup, which would be necessary to ensure no stoppage of work should one malfunction) and all supplies needed for one year of processing. The long-term cost includes supplies needed to produce about 60,000 samples and maintenance costs.

Precision

All otolith samples were aged by an expert age reader using a Leica MZ6 microscope under reflected light at 25X to 50X magnification. To test for precision, a random subsample of 345 otoliths was re-aged by the same reader with no knowledge of prior ages. Percent agreement (PA) and coefficient of variation (CV) were calculated and symmetry tests conducted (Campana et al. 1995, Chang 1982, Evans and Hoenig 1998).

FT-NIR spectroscopy age prediction models were created per the methods outlined in Rubin et al. (this volume). The models were assessed for accuracy using the coefficient of determination (r^2), root mean square error (RMSE) of cross validation and of prediction, residual

prediction deviation (RPD), model bias, and slope. To facilitate direct comparison of precision and bias metrics between FT-NIR spectroscopy and traditional ages (PA, CV, and the Evans-Hoenig symmetry test), raw FT-NIR spectroscopy (i.e., fractional) ages were rounded to the nearest integer using a standard rounding routine, meaning any fractional age ending in ≥ 0.5 was rounded up, and anything ending in ≤ 0.4 was rounded down.

RESULTS

Efficiency

Using the Labcut 250B method, it took 56 minutes to process 70 otoliths per hour to the point of being ready to age. Using FT-NIR spectroscopy, it took 130 minutes to clean and scan 70 samples (Table 15.1).

Cost

For the Labcut 250B saw, initial cost estimates are around $38,000, while for FT-NIR spectroscopy, the initial cost is around $150,000. For the long term, the Labcut 250B saw costs about $3,000, while the FT-NIR spectroscopy long-term cost ranges from $6,000 to $17,000 depending on software needed for data analysis and whether or not a service contract is needed for maintenance and repairs (Table 15.1).

Precision

Ageing precision for traditional ageing of Acadian redfish otoliths produced a 69.6% agreement and a CV of 1.81% between reads. FT-NIR spectroscopy ages produced a 29.5% agreement and a CV of 15.88% relative to traditional ages. The Evans-Hoenig test of symmetry for traditional age estimates did not find any bias across ages ($\chi 2 = 1.14$, d.f. = 4, $p = 0.889$). FT-NIR spectroscopy age estimates were not biased overall ($\chi 2 = 7.84$, d.f. = 8, $p = 0.480$) but did however show bias for some individual age classes up to age 5 (Fig. 15.1).

DISCUSSION

Our results show that, with minimal investments in sample processing technology for traditional ageing, the traditional method utilized by the NEFSC for otolith processing is twice as fast as the FT-NIR spectroscopy method. We did not include in our analysis the time required for producing ages because the process for producing FT-NIR spectroscopy ages requires model building and data optimization, which is a significant time investment on a much different scale that is difficult to compare to traditional ageing. On average, 15 samples can be aged per hour using traditional methods, while model building and data optimization to produce the FT-NIR spectroscopy model for Acadian redfish took several weeks and is still considered preliminary. This time estimate could be reduced for future FT-NIR spectroscopy models, but at this time it is still less efficient than producing a traditional age estimate.

FT-NIR spectroscopy is three times more expensive in the short-term and five times more expensive in the long-term when compared to the normal operating costs of the Labcut 250B

processing method. In terms of precision, FT-NIR spectroscopy predicted ages had lower agreement and higher variation relative to traditional ages than was demonstrated by traditional inter-reader variability. Given the lower efficiency in processing time, increased cost, and lower precision of FT-NIR spectroscopy age prediction relative to traditional methods, we caution potential adopters of FT-NIR spectroscopy methodology to do a cost analysis of the two methods before implementing this new technology. While some species require less efficient otolith processing methods (e.g., those cut individually with low-speed saws), roughly 80% of the work done at the NEFSC utilizes the high-throughput Labcut 250B saw. Unique methodologies specific to each laboratory will define whether FT-NIR spectroscopy conveys time and cost savings in the long term.

Table 15.1. -- Comparison of cost and efficiency of Traditional (Labcut 250B saw) ageing method versus Fourier transform near infrared spectroscopy (FT-NIRS). Initial cost includes the purchase of two Labcut 250B saws and two Tango-R spectrometers and supplies needed for one year of processing. The long-term cost includes supplies needed to produce about 60,000 samples and maintenance costs. Efficiency is the amount of time it takes to process 70 samples.

		Traditional	FT-NIRS
Short-Term Cost	Equipment	$36,000	$130,000
	Supplies	$2,300	$1,000
	Total	**$38,300**	**$131,000**
Long-Term Cost	Supplies	$2,300	$1,000
	Replacement parts	$500	$5,600
	Service contract	0	$9,000
	Software	0	500
	Total	**$2,800**	**$16,100**
Efficiency	Mount/Clean	42	70
	Section/Scan	14	60
	Total	**56**	**130**

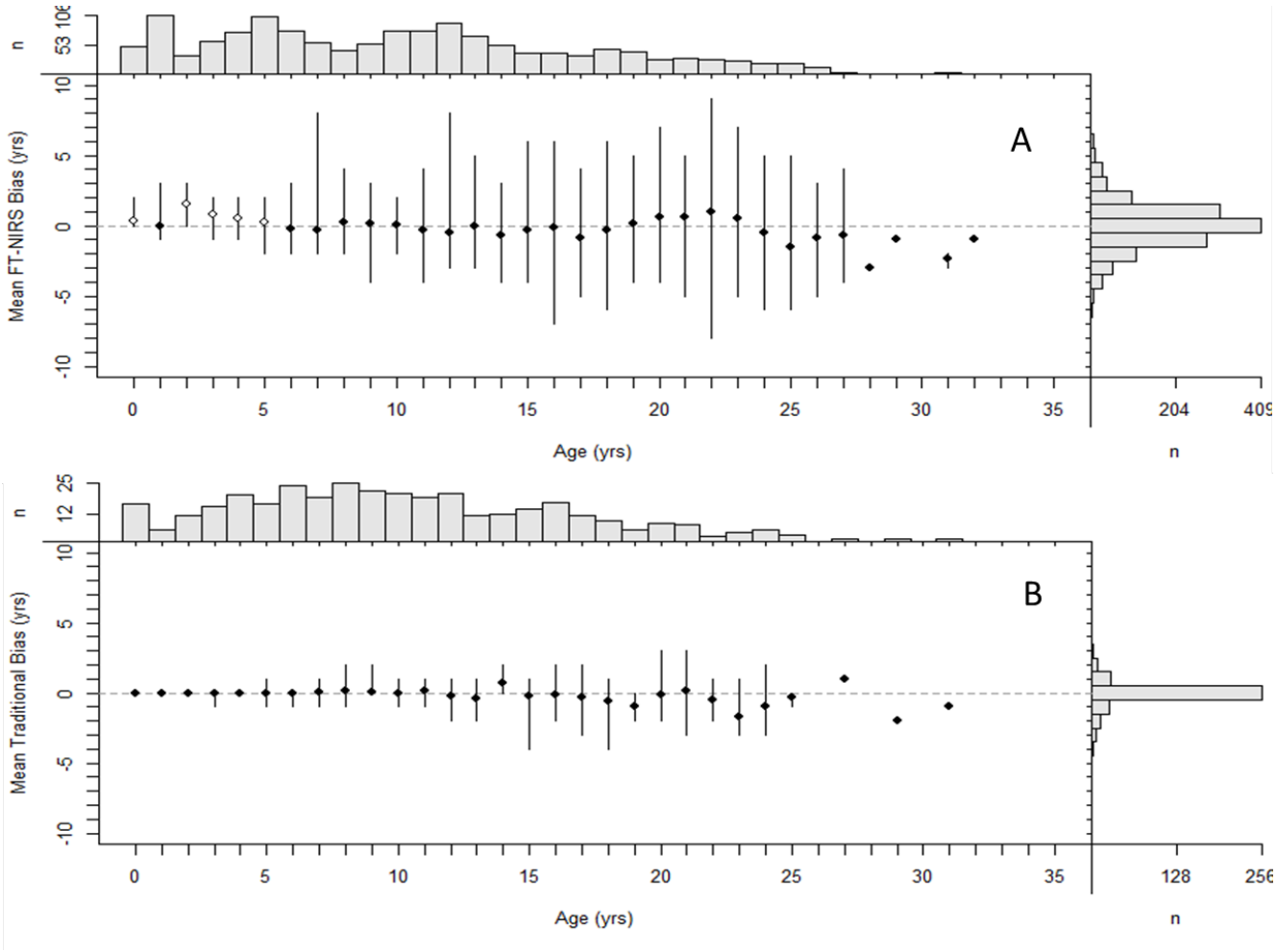


Figure 15.1. -- Mean bias (black circles) ± standard deviation for Acadian redfish *Sebastes fasciatus*. A) Fourier transform near infrared spectroscopy (FT-NIRS) and B) traditional ages by age class for all test sets combined, as predicted by the All Regions Combined Truncated Calibration Model.

16. Deep Learning Coupled with Fourier Transform Near Infrared Spectroscopy of Otoliths Improves Age Prediction for Long-Lived Fish

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

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

[2] Southeast Fisheries Science Center
NOAA, National Marine Fisheries Service
3500 Delwood Beach Road
Panama City, FL USA

ABSTRACT

Fish age estimation is crucial for effective stock management, but traditional methods are labor-intensive and subject to poor repeatability for longer-lived species. This study investigated an alternative approach utilizing Fourier transform near infrared (FT-NIR) spectroscopy of otoliths coupled with multimodal convolutional neural networks (MMCNN). Multiple previous studies applied partial least squares regression (PLS) to otolith spectroscopic data for age prediction. One recent study applied the FT-NIR/MMCNN approach to shorter-lived walleye pollock (*Gadus chalcogrammus*). In this study, we extended the approach to two longer-lived economically-valuable species: northern rockfish (NRF, *Sebastes polyspinis*) from the Gulf of Alaska (GOA) and red snapper (RS, *Lutjanus campechanus*) from the U.S. Gulf of Mexico (GOM). We trained MMCNN to integrate spectral data with biological and geospatial data. MMCNN performed well, yielding high r^2 values (0.92 for NRF and 0.91 for RS) and low RMSE values (3.4 years for NRF and 1.0 years for RS) for training data. The models demonstrated high accuracy for a holdout data set yielding high r^2 values (0.92 for NRF and 0.89 for RS) and low RMSE values (3.4 years for NRF and 1.1 years for RS) for test data. The Bland-Altman plots confirmed the agreement with low systemic bias between MMCNN predictions and traditional microscope-based ages. The SHAP (SHapley Additive exPlanations) feature importance analysis revealed the significant influence of otolith spectra, otolith weight, fish length, and geospatial data on age predictions. This study demonstrated the potential of the FT-NIR/MMCNN approach for efficient age estimation in long-lived fish species, contributing to enhanced fisheries management and conservation efforts.

INTRODUCTION

Fish age estimation is a vital aspect of fisheries management, providing crucial data for understanding fish populations and informing sustainable harvesting practices. Traditional age estimation methods, particularly those based on microscopic examination of otoliths, have proven to be labor-intensive and prone to reader variability, which can hinder accurate and timely assessments. To overcome these limitations, researchers have explored innovative techniques like Fourier transform near infrared (FT-NIR) spectroscopy for more efficient and reliable age predictions. A few recent studies implemented a traditional chemometric approach of using partial least squares regression (PLS) to build a model using otolith FT-NIR spectroscopic data of various fish species (Wedding et al. 2014, Helser et al. 2019b, Healy et al. 2021, Passerotti et al. 2020b). A more recent study of eastern Bering Sea (EBS) walleye pollock (*Gadus chalcogrammus*) otoliths used FT-NIR spectroscopy coupled with multimodal convolutional neural networks (MMCNN) to rapidly predict fish age with comparable performance to traditional microscope-based methods (Benson et al. 2023). Benson et al. (2023) suggested the MMCNN approach provides improved age predictions compared to traditional chemometric models such as PLS. While previous studies have demonstrated the success of this approach for certain fish species, its applicability to economically valuable long-lived species, such as northern rockfish (*Sebastes polyspinis*) from the Gulf of Alaska (GOA) and red snapper (*Lutjanus campechanus*) from the U.S. Gulf of Mexico (GOM), remains to be explored. Northern rockfish (NRF) are commercially important rockfishes that live 80+ years and are of particular interest due to their high abundance in the GOA and EBS waters (Kastelle et al., 2016). Red snapper (RS) are an ecologically and economically important reef fish that live 50+

years and are estimated to be rebuilding in GOM waters (SEDAR 2018). By extending the application of FT-NIR spectroscopy coupled with MMCNN to these species, our study aims to shed light on the potential advantages and limitations of this method, offering valuable insights into its effectiveness and potential contributions to fisheries management and conservation efforts.

METHODS

Data Collection

Otoliths from NRF were collected during biennial bottom trawl surveys conducted by the Alaska Fisheries Science Center (https://www.fisheries.noaa.gov/alaska/ecosystems/alaska-fish-research-surveys) in the GOA in the years 2013, 2015, 2017, and 2019. The data set comprised 1,460 otoliths, along with corresponding fish length in addition to geographic coordinates, depth, and temperature data (Table 16.1A, Fig. 16.1A-B). The Age and Growth Program at the Alaska Fisheries Science Center collected otolith weights and fish ages using traditional microscope-based methods, following established protocols described in Matta et al. (2012) and Helser et al. (2019b). For RS, otoliths (n = 6,053) were collected from GOM fishery dependent and fishery independent sources between the years 2017 and 2019. The Biology and Life History Branch at Panama City weighed, sectioned, and aged RS otoliths using traditional, microscope-based methods. The data included ages, substock information assigned as East GOM or West GOM using the Mississippi River as the separating line (SEDAR 2018), fork length, and otolith weight (Table 16.1B, Fig. 16.2A-B).

Spectroscopic Measurements

Diffuse reflectance measurements from all otoliths were obtained using Bruker Optics Tango R or MPA II spectrometers. The OPUS software version 7.5 (Bruker Optics, Ettlingen, Germany) was used for spectral acquisition. Before scanning, NRF otoliths were removed from glycerin-thymol solution and patted dry with a laboratory tissue. At the time of otolith collection, all RS otoliths were stored dry in coin envelopes; therefore, no preparation was needed prior to scanning. Each otolith was placed in a concave up position on the spectrometer's integrated sphere and covered with a gold-coated reflector stamp (see Benson et al. (2020) for a detailed description and pictures of scanning setup). Final absorbance spectra were acquired between 11,500 and 4,000 cm^{-1}, averaging 64 scans at a resolution of 16 cm^{-1}. Spectral data for each species were preprocessed with first derivative, second polynomial, 17-point Savitzky-Golay smoothing (Figs. 16.1C-F, 16.2C-D).

Data Analysis

The data set for each species was divided into training and test groups using the "onion" method in the chemometrics software Solo 8.7 (Eigenvector Research, Inc., Manson, WA, USA). The training group was used to train a neural network, while the test group was kept aside to evaluate a model on unseen data. For NRF, 90% of the samples (n = 1,418) were assigned to the training set, and the remaining 10% (n = 142) were designated as the test set. The 90/10 split was selected to maximize the amount of data in the training set due to the total low sample size. Integer age was assigned as the target variable for NRF. For RS, specimens with higher-

confidence ages (n = 1,406) were placed in the training set, and the remaining data were randomly split, with 74% assigned to the same training set. The specimen with the oldest microscope-based age was also included in the training set. The final training (n = 4,845) and test (n = 1,208) sets for red snapper comprised an 80/20 data split, with fractional age as the target variable.

The MMCNN model, as described in Benson et al. (2023), was applied to both species (Fig. 16.3). The model consisted of two primary branches, one associated with spectral data and the other with biological and geospatial data. The spectral data branch included a convolutional layer and the output was flattened and concatenated with the biological/geospatial data branch. A fully connected dense layer activated by a rectified linear unit (ReLU) introduced nonlinearity to the model and dropout was used for regularization to prevent overfitting. The final fully connected dense layer had a single output node with a linear activation function. Python 3.7.0 (Van Rossum and Drake 2009) with TensorFlow v. 2.5.0 (Abadi et al. 2016) and Keras API (Chollet 2015) implementation were used to train the models. The hyperband optimization (HB) was applied to improve the models' generalizing abilities (Li et al. 2017).

Performance Evaluation

The performance and robustness of the models were assessed using coefficients of determination (r^2) for training and test data. Root mean square error (RMSE), which represents the average magnitude of the errors between predicted and traditional ages, was calculated for both data sets to measure the accuracy of the predictions. Bland-Altman (BA) plots were employed to compare MMCNN predictions with traditional microscope-based ages, enabling the

identification of any systematic differences or biases between the two methods (Bland and Altman 1999). The plot allowed examination of the level of agreement in terms of the measurement variability. It consisted of a scatter plot where the x-axis represented the average of the two measurements being compared and the y-axis showed the difference between the two measurements for each data point. Each data point represented a single measurement. A center line, which represented the overall bias between the two measurements, was drawn at the mean difference between the two methods. Two additional lines, which represented the range within which approximately 99% of the differences between the two measurements were expected to lie, calculated as the mean difference ±2.58 times the standard deviation of the differences. Any data points that fell outside the limits of agreement suggested discrepancies between the two ageing methods for those particular measurements.

Comparison of accuracy between ages generated by the traditional and FT-NIR age estimation method was done by plotting the frequency of the mean bias by age. Traditional bias was calculated as a difference between two trained age readers' microscope-based age estimates for the same specimen. FT-NIR bias was calculated as a difference between age predicted by the MMCNN models and traditional microscope-based age estimation (Helser et al. 2019b). Fractional numbers were rounded to the nearest integer. We examined precision for both traditional methods and FT-NIR method by calculating percent agreement (PA), the percent agreement within one year (PA ± 1), coefficient of variation (CV), and average percent error (APE) (Campana 2001, Ogle 2017b, Ogle 2018). The metrics for each species were calculated on the subset of specimens that had two traditional ages in addition to the age predicted by the MMCNN model.

Model Interpretability

To interpret the MMCNN predictions, SHAP (SHapley Additive exPlanations) was used to compute a contribution of each feature to the model's output (Shapley 1953, Lundberg and Lee 2017). The SHAP values, based on cooperative game theory, provided a fair and consistent measure of feature importance. The most influential features driving the model's predictions were identified and visualized using SHAP feature importance plots. Each bar in the plot represented a feature with its length corresponding to the magnitude of its effect on the model's output. The plot allowed us to identify the most influential features driving the model's predictions. The spectral wavenumber contributions were aggregated by thousands and all features were ordered according to their importance.

RESULTS AND DISCUSSION

The application of FT-NIR spectroscopy combined with MMCNN proved to be a promising approach for accurate and efficient fish age estimation. The MMCNN models exhibited strong performance, demonstrating a high degree of correlation between FT-NIR ages and traditional microscope-based ages for both NRF and RS. The scatter plots of the MMCNN age predictions showcased a close alignment with the reference microscopic ages, indicating accuracy and reliability of the models for both species (Fig. 16.4). For NRF, the MMCNN model achieved an r^2 of 0.92 and an RMSE of 3.4 years for both the training and test data sets. At least 67% of the estimated NRF ages were predicted within 3.4 years of traditional microscopic ages, while 95% of the estimated ages were within 6.8 years of the reference ages. Similarly, the RS

MMCNN model demonstrated strong predictive capabilities, with an r^2 of 0.91 and an RMSE of 1.0 year for the training data set, and an r^2 of 0.89 and an RMSE of 1.1 years for the test data set. At least 67% of the estimated RS ages were accurately predicted within 1.1 years of traditional microscopic ages, and 95% of the estimated ages fell within 2.2 years of the reference ages. A higher mean age and maximum age for NRF (max age = 58 years, mean age = 25 years; Table 16.1A) compared to RS (max age = 50 years, mean age = 6 years; Table 16.1B) could potentially explain higher RMSE for NRF. In addition, the majority of the RS specimens were aged 20 years old or less (Fig. 16.2A-B) with just a few older specimens being included in the training data set and not available for the test set (Fig. 16.4C-D). For NRF, the training and test data set included both young and old fish, which led to a wider spread of ages in the dataset (Figs. 16.1A-B, 16.4A-B). Predicting ages accurately for both young and old fish could have been more challenging, resulting in higher RMSE. Another factor that could have contributed to the NRF having higher RMSE is a smaller sample size that could lead to less representative data for training the model. Another factor that could contribute is inherent biological variability, leading to age-related growth variations even within the same age group. The growth patterns of RS otoliths in this study led to a linear relationship of otolith weight with age (Fig. 16.2B), while a non-linear growth pattern of NRF where growth slows with age led to a non-linear relationship of otolith weight with age (Fig. 16.1B), which can make age prediction more challenging and contribute to higher RMSE.

The performance metrics of the MMCNN models were comparable to previous studies that utilized FT-NIR spectroscopy and PLS for longer-lived fish age estimation. For instance, Passerotti et al. (2020b) reported an r^2 of 0.92 and an RMSE of prediction of 0.99 to 1.02 years for red snapper. Wedding et al. (2014) achieved an r^2 of 0.94 and an RMSE of prediction of 1.54

years for shorter-lived saddletail snapper (*Lutjanus malabaricus*). Our results were also consistent with Benson et al. (2023), who applied MMCNN to FT-NIR spectral, biological, and geospatial data of shorter-lived walleye pollock. Benson et al. (2023) reported an r^2 of 0.92 and an RMSE of prediction of 0.91 years for the test set.

The BA plots provided additional evidence of the agreement between MMCNN-predicted ages and traditional microscope-based ages (Figs. 16.5A and 16.6A). Low systematic bias for either species was indicated by the center line not deviating significantly from zero (0.47 years for NRF and -0.06 years for RS). The narrow limits of agreement for RS (+2.86/-2.97 years) indicated a better level of agreement compared to NRF (+9.06/-8.13 years). The wider limits of agreement indicated greater variability and less agreement between the two ageing methods for NRF. The presence of multiple data points falling outside of the limits of agreement in the RS plot might be attributed to the large range of RS otolith sizes. Passerotti et al. (2020b) noted that RS otoliths' FT-NIR spectra might be affected by otolith size and otolith thickness in addition to physical and chemical changes underlying otolith growth. The mean bias by age was equivalent between the traditional and FT-NIR methods up to around 40 years of age for NRF (Fig. 16.5B) and up to around 18 years of age for RS (Fig. 16.6B).

For the NRF model predictions, the PA of 13.9% and PA ± 1 year of 41.9% were lower than the PA of 28.7% and PA ± 1 year of 64.9% for the traditional microscopic-based between-reader agreement (Table 16.2A). Between-reader APE and CV were higher for the FT-NIR method, with an APE of 5.4% and CV of 7.6% in comparison to an APE of 3.3% and CV of 4.7% for the traditional method (Table 16.2A). The typical inter-reader precision for NRF is usually higher (PA = 44.4%, APE = 2.9%, CV = 4.1%; n = 2852; Goetz et al. 2012) than in our study, while the mean age is usually lower (~16 years) than in our study (25 years; Table 16.1A).

Since the FT-NIR method relies on traditional age data, difficulties with generating ages by the traditional method most likely affected accuracy and precision of the FT-NIR method. For the RS model predictions, the PA of 52.1% and PA ± 1 year of 88.79% were lower than the PA of 68.7% and PA ± 1 year of 93.6% for the traditional microscopic-based between-reader agreement (Table 16.2B). Between-reader APE and CV were higher for the FT-NIR method, with an APE of 5.5% and CV of 7.8% in comparison to an APE of 3.3% and CV of 4.7% for the traditional method (Table 16.2B). Our results were slightly better than the previous study that utilized a PLS model for GOM RS spectral data and reported a PA of 41-50%, PA ± 1 year of 87.1%, and APE of 7.3-13.2% (Passerotti et al. 2020b). The FT-NIR method's performance metrics for either species could also be affected by the fact that we had to round fractional ages. Further studies are needed to establish the best protocol for establishing rounding convention for the FT-NIR ages.

The SHAP feature importance plots provided valuable insights into the significant factors influencing the MMCNN predictions (Fig. 16.7). Otolith spectra in the 4,000 cm^{-1} wavenumber region emerged as the most influential feature for both species, confirming the significance of spectral information in age estimation. Additionally, otolith spectra in the 7,000-5,000 cm^{-1} wavenumber region played a key role in age predictions, particularly for RS. Otolith weight emerged as the most influential among non-spectral features for both species, highlighting the impact of otolith size on age determination. Specifically, for the NRF model, otolith weight and spectra in the 5,000 cm^{-1} region exhibited considerable influence. For the RS model, spectra in the 5,000 cm^{-1} and 7,000 cm^{-1} regions held greater importance than otolith weight. Fish length, along with latitude for NRF and substock for RS, were also influential features, aligning with existing knowledge of these factors' influence on fish growth and age.

These results are comparable to the recent study using the MMCNN on walleye pollock data with otolith spectra in the 7,000-4,000 cm^{-1} wavenumber region having the highest impact on predicting fish age (Benson et al. 2023). Otolith weights were not available in that study; however, fish length and latitude were the most important features among biological and geospatial data ingested by the model. While the MMCNN model demonstrated excellent performance, enhancing its interpretability remains an area of improvement. Exploring visualization techniques or attention mechanisms could provide deeper insights into the model's decision-making process.

Despite the promising results, several limitations and future considerations should be acknowledged. The relatively small sample size for NRF could potentially impact the model's generalizability, calling for further exploration with larger and more diverse datasets. Updating the models with known-age fish or external validation against alternative age estimation methods would strengthen the model's reliability. Temporal validation across different time periods could also account for potential variations in fish growth rates and age estimation. Extending the study to include data from different geographic regions and a wider range of fish species would provide a more comprehensive assessment of the FT-NIR/MMCNN approach under varying environmental conditions and life histories. Addressing the discussed limitations and considering future model updating methods will further advance the FT-NIR/MMCNN method as a valuable tool.

Our results demonstrated the ability of the FT-NIR/MMCNN approach to closely align its age predictions with traditional microscope-based ages for both species, indicating its accuracy and reliability. The FT-NIR spectroscopy of otoliths coupled with MMCNN demonstrated robust and accurate fish age predictions for long-lived fish species. The study's

findings contribute valuable insights into the potential use of this method for diverse fish species, supporting sustainable fisheries practices and effective population monitoring. The model's performance metrics, agreement with traditional methods, and identification of influential features underscore the efficacy of the FT-NIR spectroscopy of otoliths coupled with MMCNN as a valuable tool for enhancing fish age estimation accuracy and efficiency.

Table 16.1. -- Summary of available data for (A) northern rockfish *Sebastes polyspinis* and (B) red snapper *Lutjanus campechanus*.

(A)

	Age (years)	Total length (mm)	Otolith weight (g)	Depth (m)
Min	3	140	0.02	57
Max*	58	480	0.44	255
Mean	25	391	0.22	116
SD	12	42	0.07	30
Median	22	400	0.23	110

* The oldest northern rockfish aged by the Age and Growth Program between 1990 and 2021 was 88 years and 410 mm (Goetz et al. 2012).

(B)

	Fractional age (years)	Fork length (mm)	Otolith weight (g)
Min	0.70	215	0.14
Max	49.64	892	5.56
Mean	5.9	513	0.87
SD	3.48	114	0.46
Median	4.95	493	0.73

*The maximum calendar age used in the most recent assessment was 57 years (SEDAR 2022).

Table 16.2. -- Precision statistics for the (A) northern rockfish *Sebastes polyspinis* (*n* = 296) and (B) red snapper *Lutjanus campechanus* (*n* = 1,090) collections. Specimens that weren't assigned a second traditional age have been excluded. PA = percent agreement, PA ± 1 yr = percent agreement within 1 year, APE = average percent error, and CV = coefficient of variation.

(A)

Method	PA	PA ± 1 yr	APE	CV
Traditional	28.7%	64.9%	3.3%	4.7%
FT-NIR	13.9%	41.9%	5.4%	7.6%

(B)

Method	PA	PA ± 1 yr	APE	CV
Traditional	68.7%	93.6%	3.3%	4.7%
FT-NIR	52.1%	88.79%	5.5%	7.8%

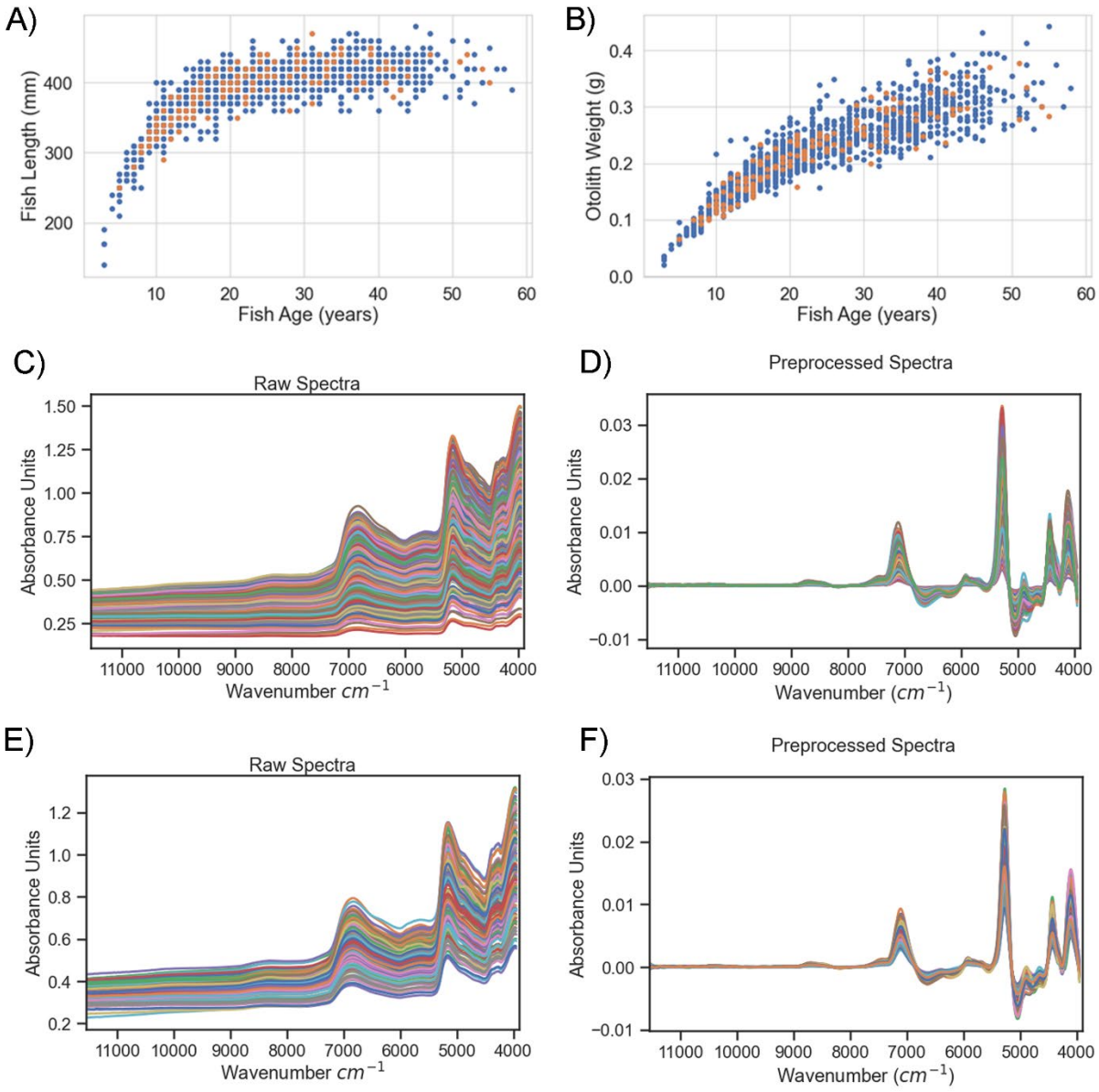


Figure 16.1. -- Northern rockfish *Sebastes polyspinis* (NRF) (A) fish length-for-age and (B) otolith weight-for-age scatter plots with blue dots indicating training data set and orange dots indicating test set. (C) Raw and (D) first-derivative-transformed otolith Fourier transform near infrared (FT-NIR) spectra for NRF training set, (E) raw and (F) first-derivative-transformed otolith Fourier transform near infrared (FT-NIR) spectra for NRF test set.

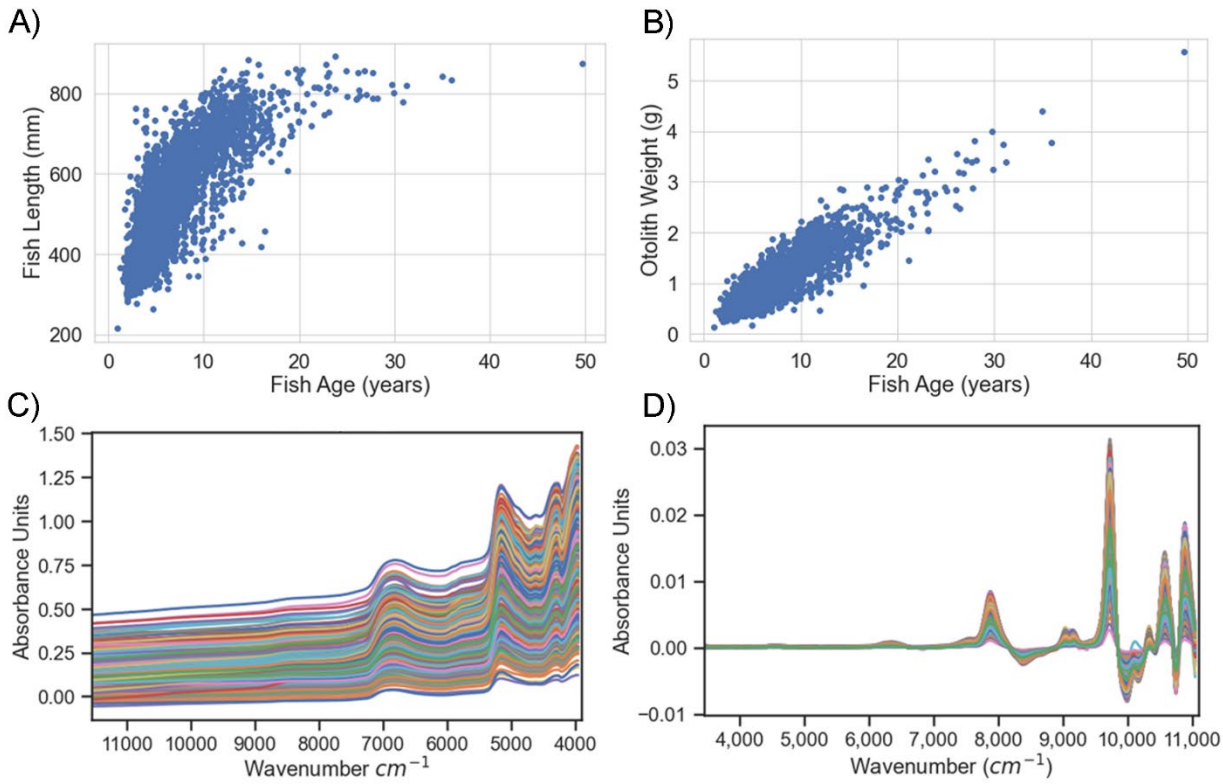


Figure 16.2. -- Red snapper *Lutjanus campechanus* (RS) (A) fish length-for-age and (B) otolith weight-for-age scatter plots, and (C) raw and (D) first-derivative-transformed otolith Fourier transform near infrared (FT-NIR) spectra for RS combined data set.

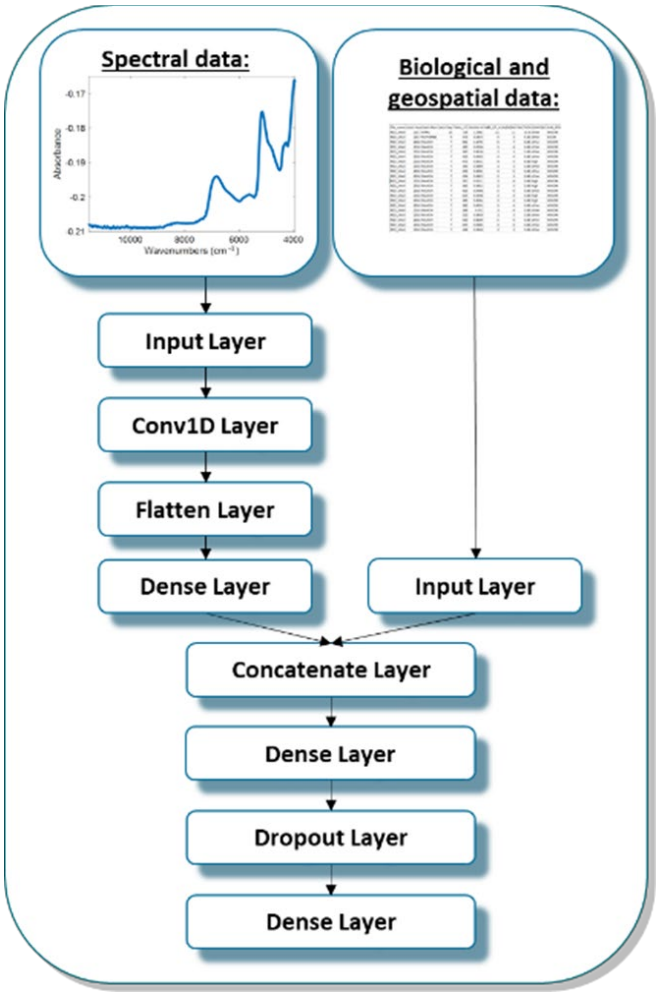


Figure 16.3. -- Schematic diagram of the structure of the multimodal convolutional neural network (MMCNN).

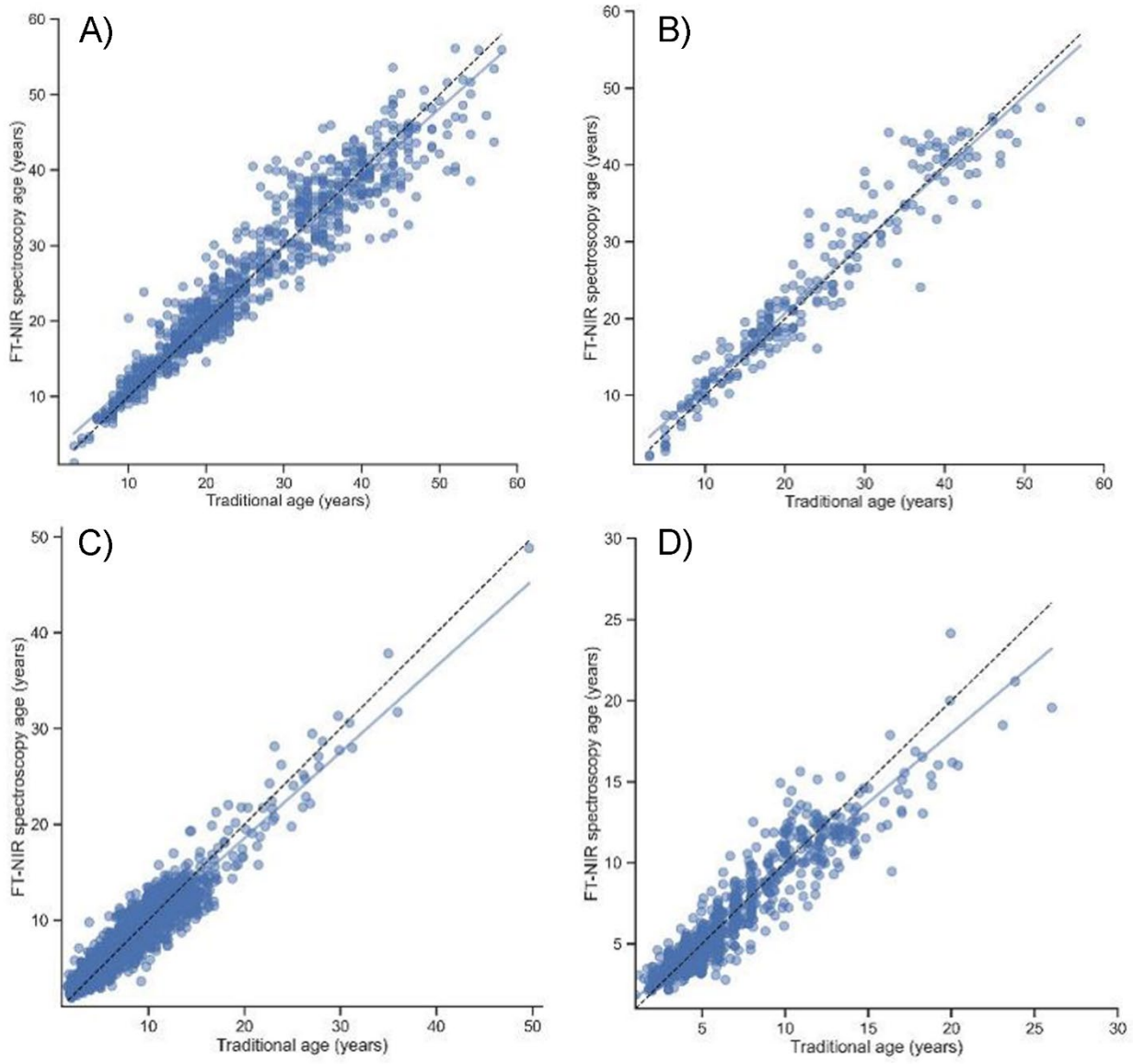


Figure 16.4. -- Scatterplots showing MMCNN fish age predictions versus traditional age estimates for northern rockfish *Sebastes polyspinis* (A) training and (B) test sets and red snapper *Lutjanus campechanus* (C) training and (D) test sets. The dashed line is 1:1 agreement and the solid line is linear regression fit. Point density is illustrated by overlapping semi-transparent data points.

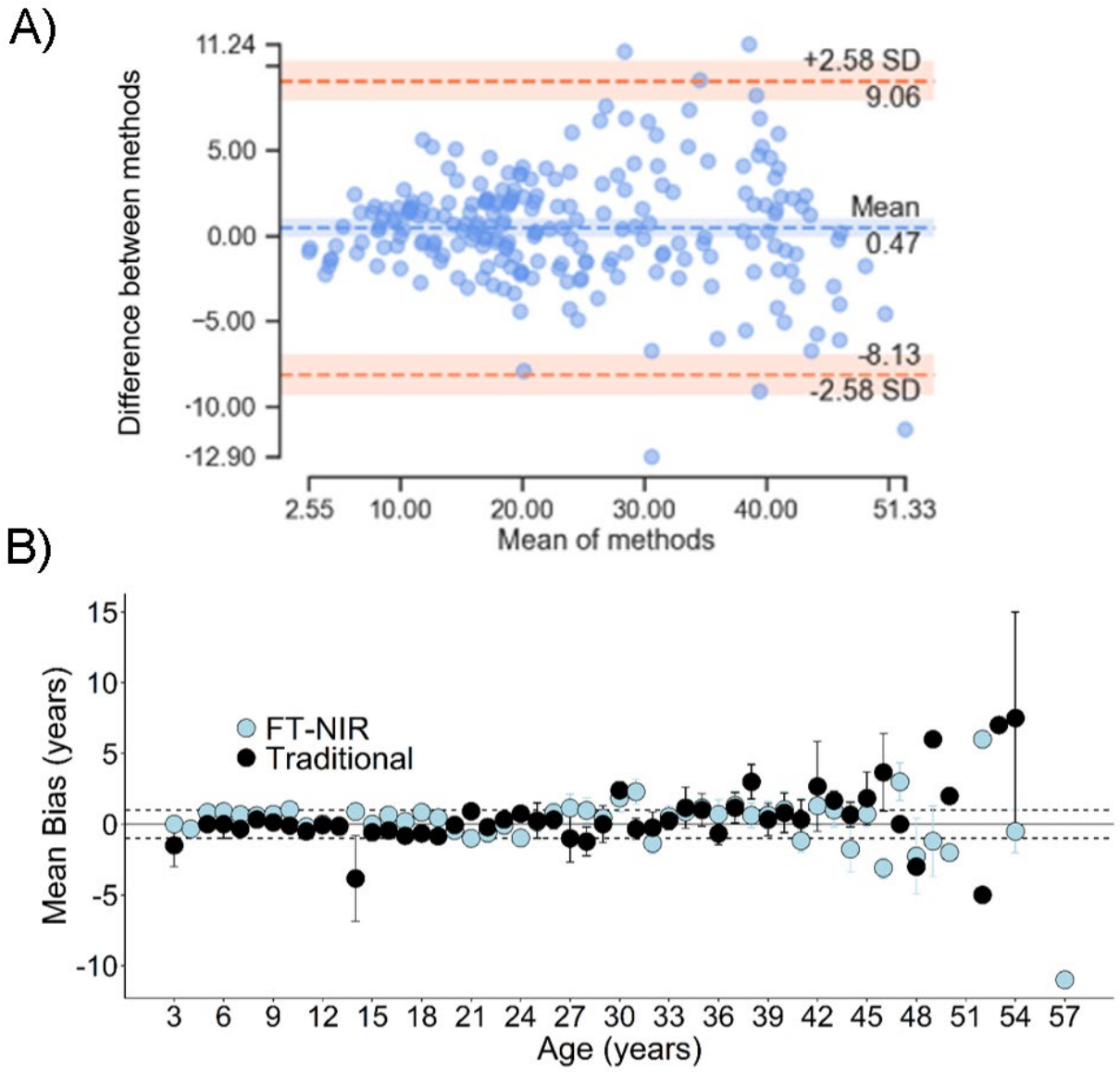


Figure 16.5. -- (A) Bland-Altman plot showing difference in traditional microscope-based age estimates and age estimates predicted by the model versus the mean of the reference and predicted age estimates for northern rockfish *Sebastes polyspinis* (NRF). Point density is illustrated by overlapping semi-transparent data points. (B) Frequency of mean bias by age class for NRF. The error bars show one standard error intervals around the mean. The solid line represents a 100% agreement line. The dashed lines represent ±1 year.

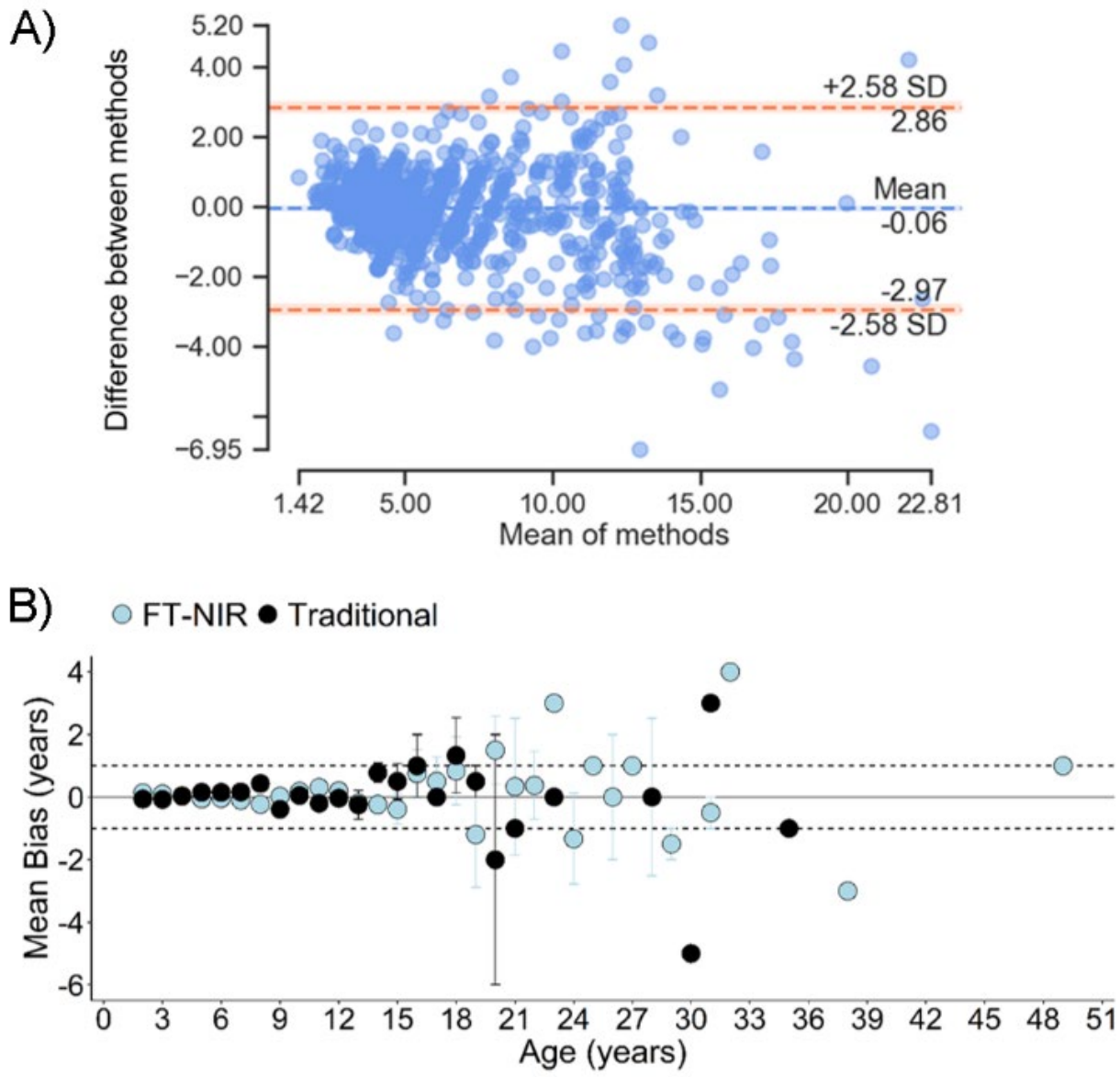


Figure 16.6. -- (A) Bland-Altman plot showing difference in traditional microscope-based age estimates and age estimates predicted by the model versus the mean of the reference and predicted age estimates for red snapper *Lutjanus campechanus* (RS). Point density is illustrated by overlapping semi-transparent data points. (B) Frequency of mean bias by age classes for RS. The error bars show one standard error intervals around the mean. The solid line represents a 100% agreement line. The dashed lines represent ±1 year.

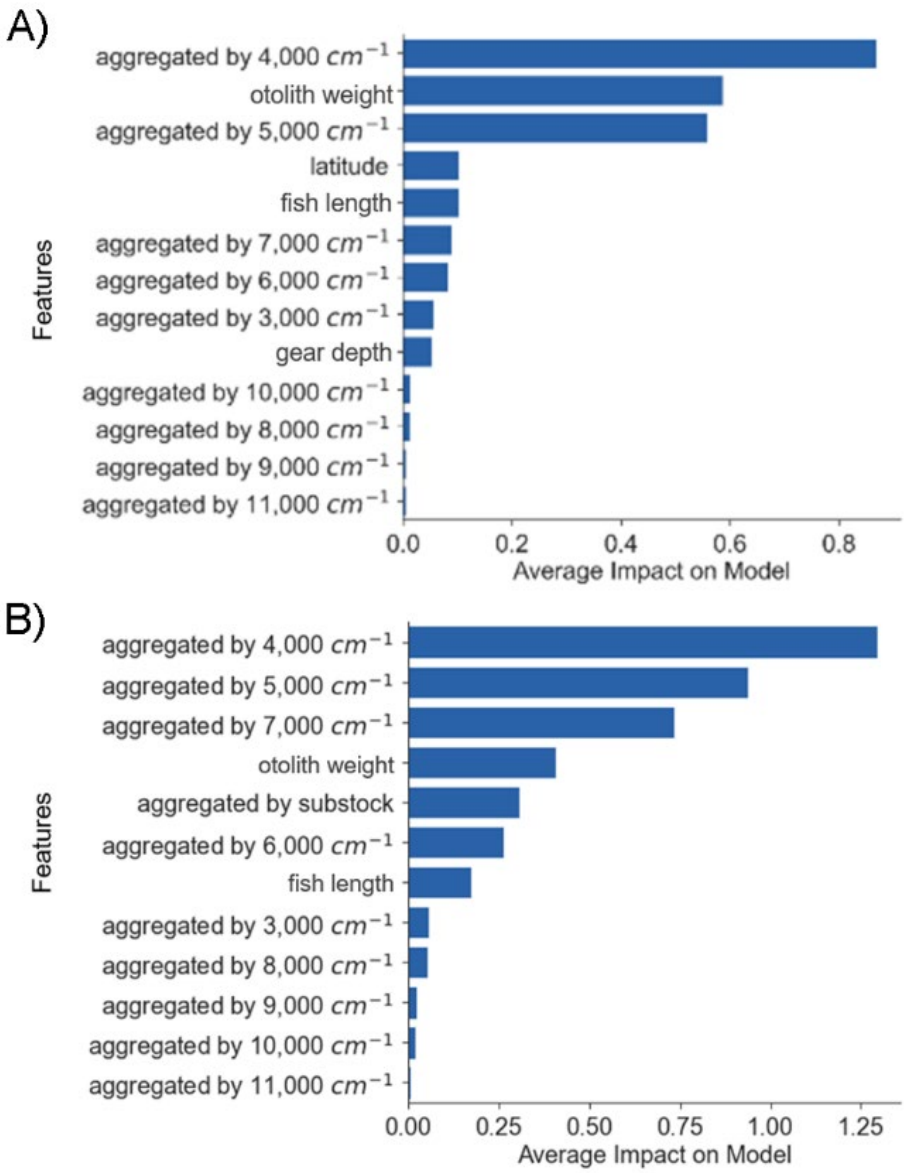


Figure 16.7. -- The relative importance of the input features for the (A) northern rockfish *Sebastes polyspinis* and (B) red snapper *Lutjanus campechanus* models. Wavenumbers are aggregated by 1,000 cm^{-1}. The y-axis indicates the feature name in order of importance from top to bottom. The x-axis indicates the average of the absolute Shapley value of each feature.

17. Automatic Fish Age Prediction Using Deep Machine Learning – Combining Otolith Images, NIR Spectra, and Metadata Features

Aotian Zheng[1], Yudong Li[1], Ara Vardanyan[1], Stefan Arsov[1], Jenq-Neng Hwang[1], Irina M. Benson[2], Thomas E. Helser[2], Jonathan A. Short[2], Kathrin A. Bayer[2], Craig R. Kastelle[2], Beverly K. Barnett[3], Amanda R. Rezek[3], Farron R. Wallace[3], Ronald L. Hill[3], William Kline[3], Laura A. Thornton[3], Nancy Evou[3], Robert J. Allman[3], and Naeem Willett[3]

[1] University of Washington
1410 NE Campus Parkway
Seattle, WA USA

[2] Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

[3] Southeast Fisheries Science Center
NOAA, National Marine Fisheries Service
75 Virginia Beach Drive
Miami, FL USA

ABSTRACT

Fish ages, which are critical for managing stocks, are used to estimate growth rates, mortality, age at maturity, and population trends. Unfortunately, traditional otolith age estimation methods are labor-intensive and subject to poor repeatability. While Fourier transform near infrared (FT-NIR) spectroscopy of otoliths has recently emerged as a highly efficient fish age estimation method, our project investigates another possible data modality for rapid and effective fish age estimation, red-green-blue (RGB) imagery of whole fish otoliths. Our focus is on how effective the three different data modalities (otolith images, FT-NIR spectra, and associated biological and geospatial data) are for deep learning-based fish age prediction, both individually and when combined. We use data collected from three different species in our model training and evaluation: walleye pollock, red snapper, and Atlantic menhaden.

INTRODUCTION

Traditional fish ageing methods generally involve manual counting of growth patterns on either otoliths or scales by trained experts. These methods are not only time consuming and labor intensive, but they also contain some degree of uncertainty, as even expert readers can often arrive at different ages for the same otolith sample. Recent research has explored using Fourier transform near infrared (FT-NIR) spectroscopy (Helser et al. 2019b) for rapid and effective ageing of otoliths. Besides FT-NIR spectroscopy, focus has also been put on using otolith red-green-blue (RGB) imagery in combination with deep learning (Moen et al. 2018, Politikos et al. 2021) for ageing as well. We analyze the effectiveness of the three data modalities: FT-NIR spectra, otolith image, and associated metadata (biological and geospatial), in fish age prediction and the effectiveness of combining all three modalities in one neural network model. Most of our experiments are conducted on walleye pollock (*Gadus chalcogrammus*) data, but we also show results on red snapper (*Lutjanus campechanus*) and Atlantic menhaden (*Brevoortia tyrannus*).

METHODS

We designed our model to utilize all three available data modalities by having three different input branches, one for each modality. The corresponding embedding features for each data modality were then concatenated together before being fed to a final, unified prediction layer. Following Politikos et al. (2021), we used an inceptionV3 (Szegedy et al. 2016) convolutional neural network (CNN) architecture for the image branch, while both the spectral data and metadata used a simple multilayer perceptron (MLP). Since we formulated ageing as a

classification task, we trained our network using cross-entropy loss. Rather than training with a one-hot age label, we instead used a soft label to indicate the uncertainty in our ages obtained from human experts. Our methodology is illustrated in Figure 17.1.

Data

We evaluated our method on data from three different species: walleye pollock, red snapper, and Atlantic menhaden. Details of the three datasets can be seen in Figures 17.2, 17.3, and 17.4. For walleye pollock metadata, we used fish length, otolith weight, latitude, longitude, sex, gear depth, bottom depth, and catch month. The samples in the walleye pollock test set were also aged by a second reader so we were able to compare our model versus reader accuracy to a reader versus reader agreement. For red snapper metadata, we used fish length, otolith weight, sex, and catch location. For Atlantic menhaden metadata, we used fish weight, fish length, and catch month. Note that for Atlantic menhaden, we did not have spectra data, only images and metadata.

RESULTS AND DISCUSSION

We used classification accuracy to evaluate our model performance compared to a given reader age. Due to the difficulty associated with age determination, even for trained human readers, we used top-2 accuracy for walleye pollock and red snapper ageing (Figs. 17.5, 17.6). In top-2 accuracy, a prediction is considered correct if the groundtruthed age is one of the two ages

with the highest scores based on our model output. Regular top-1 accuracy was used for Atlantic menhaden ageing (Fig. 17.7).

Top-2 accuracy for walleye pollock ageing was very high at low ages but gradually dropped as age increases (Fig. 17.5). This trend can likely be attributed to two factors. First, the difficulty of determining an exact age from an otolith increases as the fish gets older, even for expert human readers. Second, older ages (>10 years) are what we call minority classes, meaning they are rarer than younger ages and have significantly less training data. This lack of training data also affects the model's ability to accurately age older otoliths. One interesting thing to note is that for our walleye pollock data, there was an anomalous dip in accuracy at ages 11 and 12, which was not explained by the two factors we identified.

Comparison to Humans

To evaluate whether our deep learning model can replace humans in otolith ageing, we compared results with expert human readers on the walleye pollock test set (Fig. 17.8). We found that humans achieved a 93.5% reader agreement with a 1-year margin, meaning that 93.5% of the time, two expert human readers estimated ages that were within 1 year of each other. Our model achieved a slightly lower accuracy of 88.1% with a 1 year margin, meaning that for 88.1% of the samples, our model's top age prediction was within 1 year of the human expert-provided ground truth age. While our current deep learning model still fell a bit short of human experts, it was only at higher ages (>10) and can potentially be remedied through long-tail training methods.

Ablation Study

We conducted an ablation study on the walleye pollock dataset to determine how effective each data modality was individually for age prediction and how well they complemented each other when combined together (Table 17.1). To better understand the contribution of each data modality to age discrimination, we removed one or more input branches from our model in each experiment of our ablation study. We found that when only using one data modality, image was the most informative, providing a top-2 accuracy of 76.4% compared to 73.0% of metadata and 70.6% of spectra data. From combining data modalities together, we found that each data modality offered unique information for age prediction, as models using multiple modalities always produced better results than only using a single input modality. Our best result was achieved from using all three data modalities, giving us an 81.1% top-2 classification accuracy.

Table 17.1. -- Ablation study of model performance with different combinations of input modalities.

Input Data			Method	Test Accuracy	
Image	**Meta**	**Spectral**		**Top 1**	**Top 2**
x			CNN	50.6%	76.4%
	x		MLP	42.0%	73.0%
		x	MLP	43.2%	70.6%
	x	**x**	MLP	48.0%	75.3%
x	**x**		CNN	51.5%	78.1%
x		**x**	CNN	50.3%	77.8%
x	**x**	**x**	CNN	52.8%	**81.1%**
			Reader vs Reader	**67.0%**	

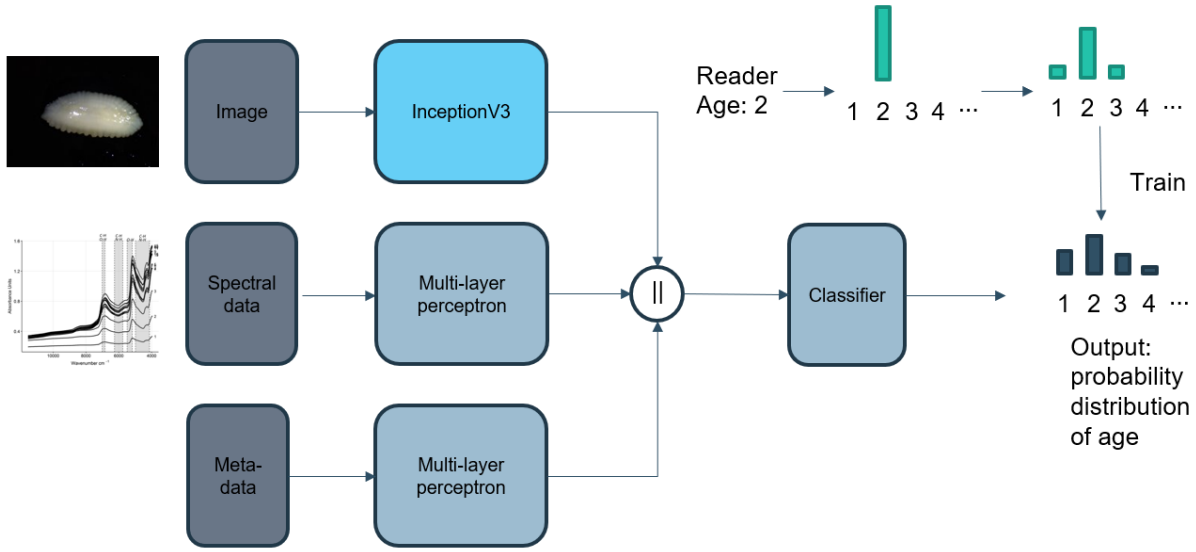
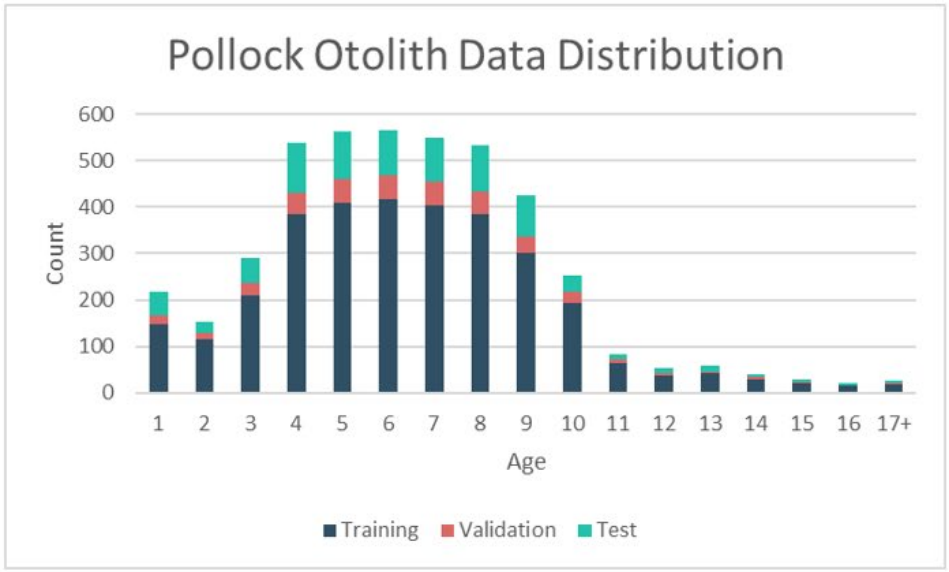


Figure 17.1. -- Our model architecture for combining otolith imagery, Fourier transform near infrared spectra, and associated meta-data for age classification.

- 4379 samples
- 17 classes



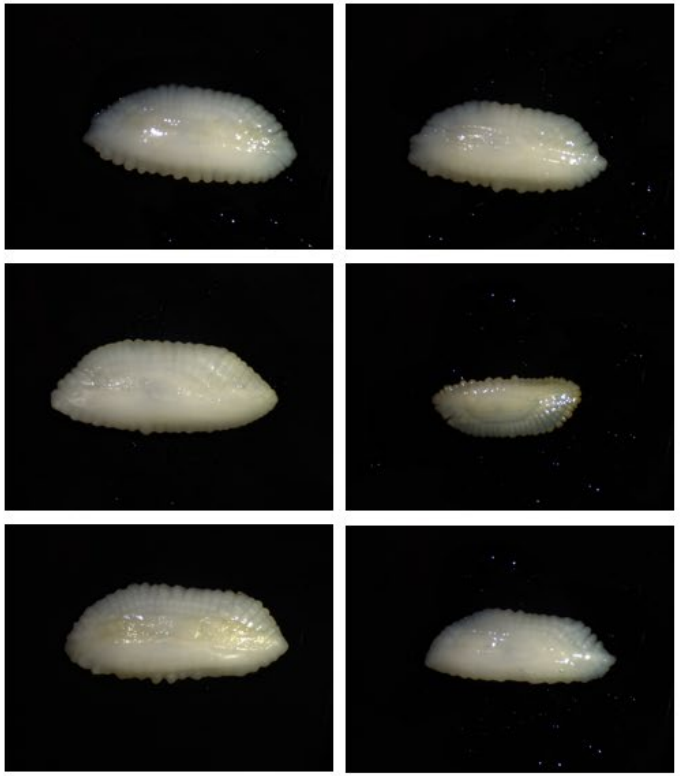


Figure 17.2. -- Age distribution of walleye pollock (*Gadus chalcogrammus*) data. Data were split roughly into 70% training, 10% validation, and 20% test sets. Example walleye pollock otolith images are on the right.

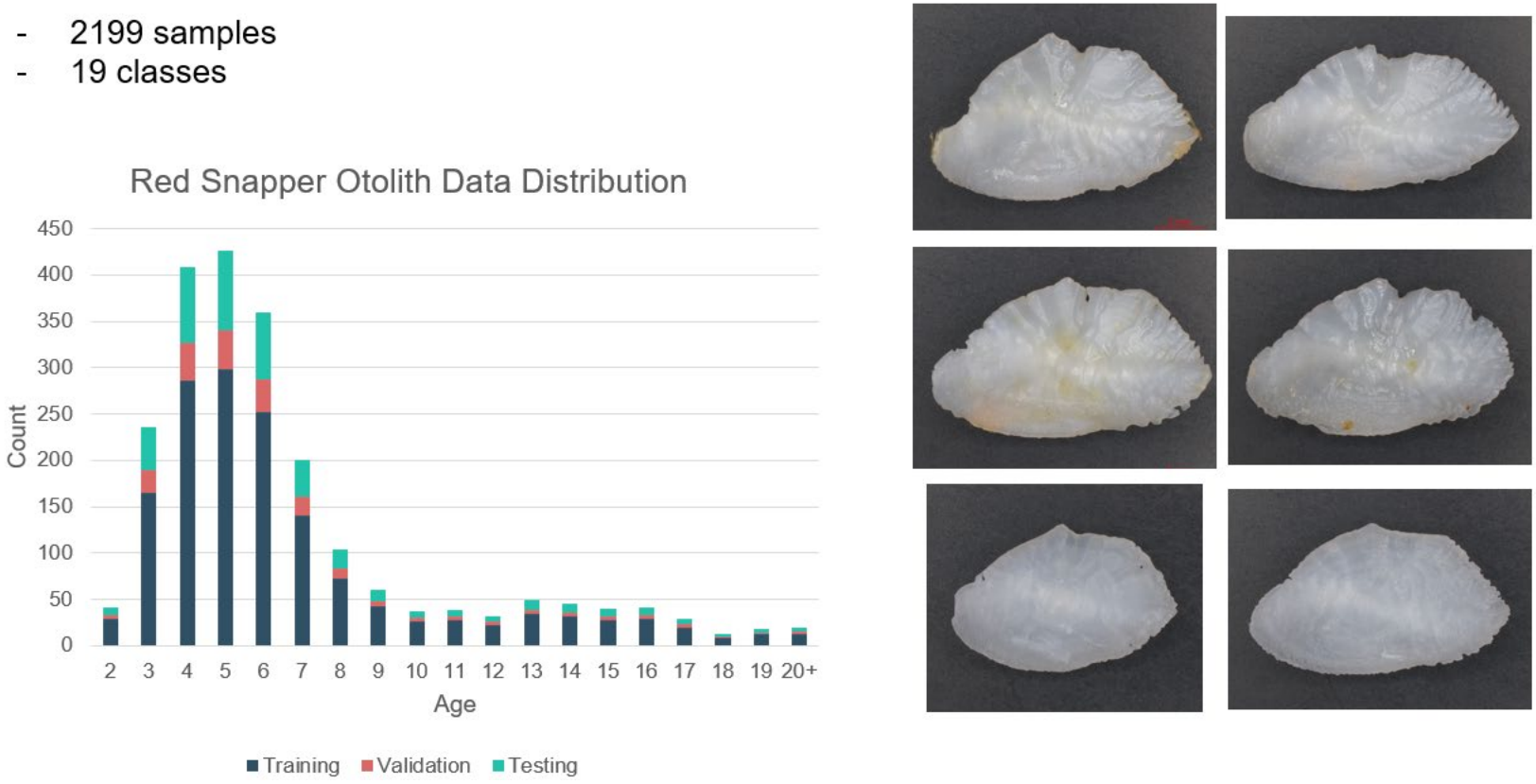


Figure 17.3. -- Age distribution of red snapper (*Lutjanus campechanus*) data. Data were roughly split into 70% training, 10% validation, and 20% test sets. Example red snapper otolith images are on the right.

- 4199 samples
- 5 classes

- Image
- Metadata: weight, length, month

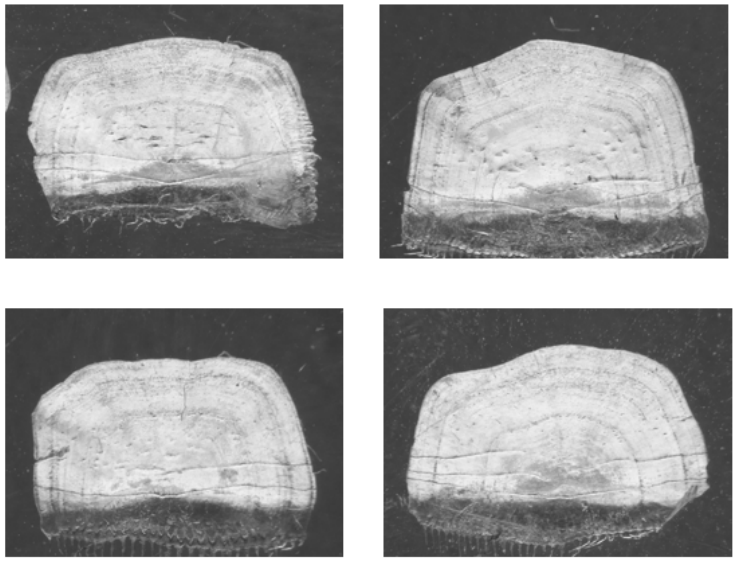


Figure 17.4. -- Age distribution of Atlantic menhaden (*Brevoortia tyrannus*) data. Data were roughly split into 70% training, 10% validation, and 20% test sets. Example Atlantic menhaden scale images are on the left.

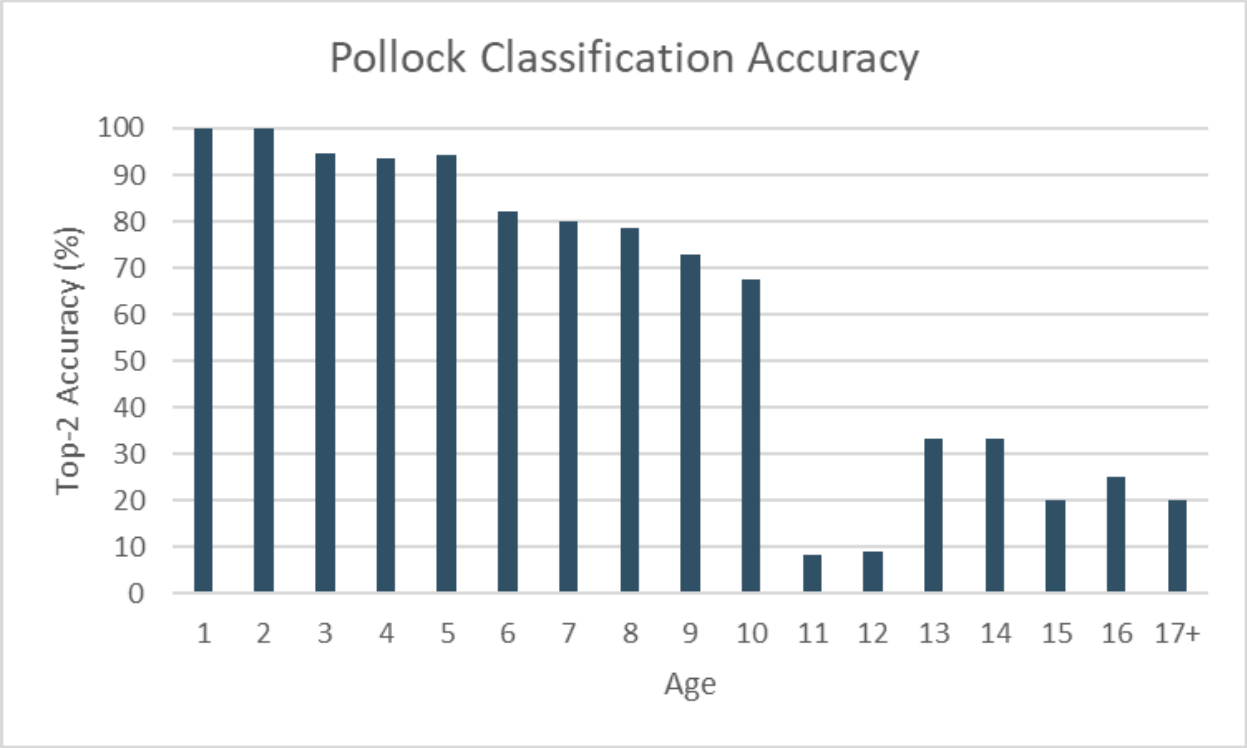


Figure 17.5. -- Walleye pollock (*Gadus chalcogrammus*) classification accuracy on test set using our best model combining all three data modalities. Top-2 accuracy was used in the evaluation, with overall accuracy being 81.1%.

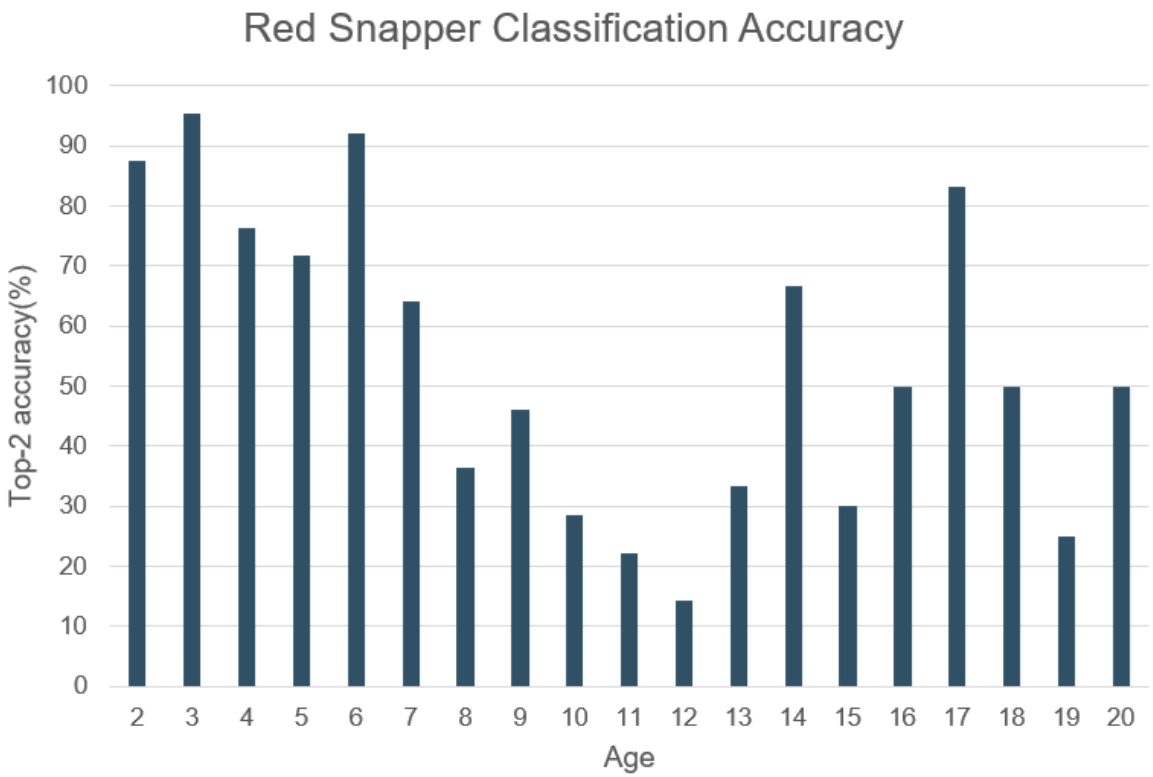


Figure 17.6 -- Red snapper (*Lutjanus campechanus*) classification accuracy on test set using our best model combining all three data modalities. Top-2 accuracy was used in the evaluation, with overall accuracy being 70.1%. Note: Groundtruth labels for red snapper data are being reevaluated. Results in this figure may not be reliable.

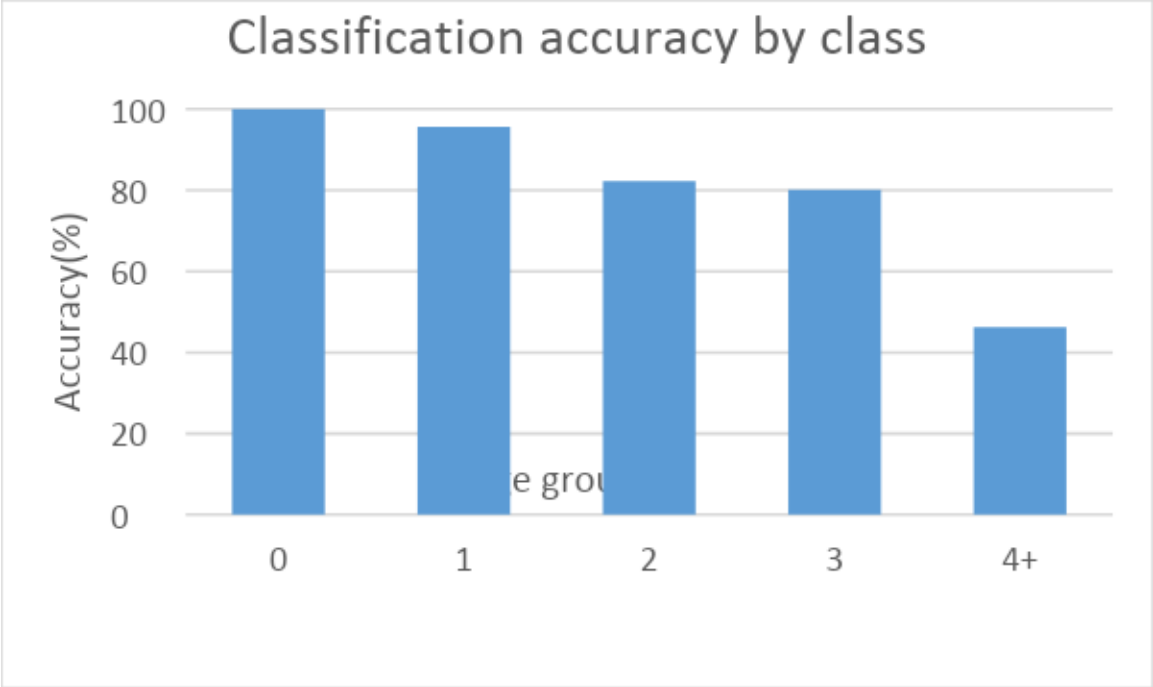


Figure 17.7. -- Atlantic menhaden (*Brevoortia tyrannus*) classification accuracy on test set using our best model combining image and metadata. Top-1 accuracy was used in the evaluation, with overall accuracy being 87.3%.

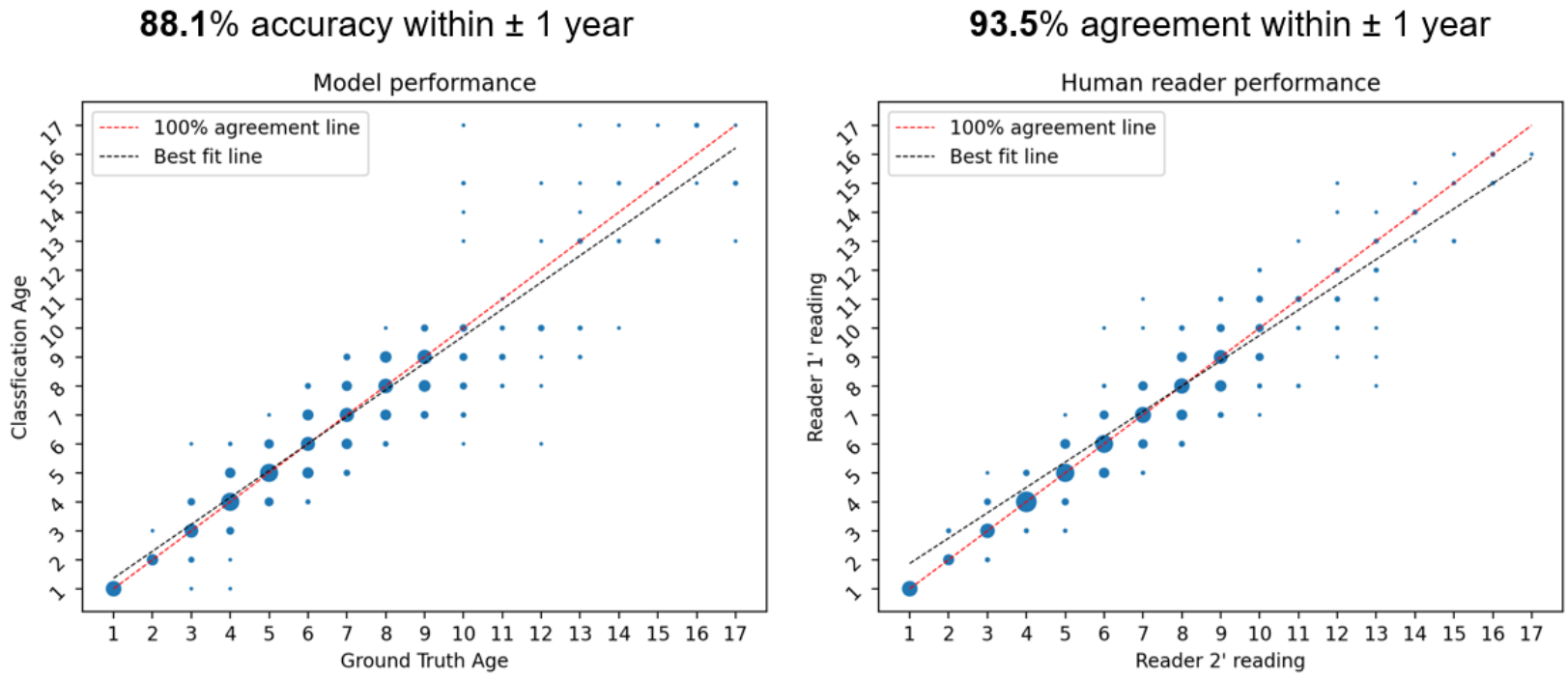


Figure 17.8. -- Comparison of model performance versus human reader for walleye pollock (*Gadus chalcogrammus*). Model prediction versus ground truth age is shown on the left, while reader versus reader age is shown on the right.

18. Calibration and Variation of FT-NIR Otolith Spectra Among NIR Spectrometers and Species

Andy D. Ostrowski[1], Beverly K. Barnett[2], Jamie M. Clark[1,3], Jennifer C. Potts[1], Alexander M. Rubin[4,5], Michelle S. Passerotti[4], Melissa H. Monk[6], Jessica J. M. Choi[6,7], Brittany D. Schwartzkopf[8], Emma M. Saas[8,9], and Emmanis Dorval[8,10]

[1] Southeast Fisheries Science Center
NOAA, National Marine Fisheries Service
101 Pivers Island Road
Beaufort, NC USA

[2] Southeast Fisheries Science Center
NOAA, National Marine Fisheries Service
3500 Delwood Beach Road
Panama City, FL USA

[3] Cooperative Institute for Marine and Atmospheric Studies
University of Miami
4600 Rickenbacker Causeway
Miami, FL USA

[4] Northeast Fisheries Science Center
NOAA, National Marine Fisheries Service
28 Tarzwell Drive
Narragansett, RI USA

[5] IBSS Corporation
1110 Bonifant Street
Silver Spring, MD USA

[6] Southwest Fisheries Science Center
NOAA, National Marine Fisheries Service
110 McAllister Way
Santa Cruz, CA USA

[7] Fisheries Collaborative Program
Institute for Marine Ecosystems and Climate
University of California Santa Cruz
110 McAllister Way
Santa Cruz, CA USA

[8] Southwest Fisheries Science Center
NOAA, National Marine Fisheries Service
8901 La Jolla Shores Drive
La Jolla, CA USA

[9] Saltwater Inc.
733 N Street
Anchorage, AK USA

[10] Lynker
202 Church Street SE, Suite 536
Leesburg, VA USA

INTRODUCTION

Traditional ageing techniques have long been standardized using reference collections between age readers and laboratories to ensure consistency in the ages produced. This allows confidence when combining multiple datasets for analysis in stock assessments. Fourier transform near infrared (FT-NIR) spectroscopy is currently being researched as a potential technology to replace traditional processing and ageing techniques. Using spectroscopy, a beam of light is transmitted through a sample, in this case a whole otolith, and the amount of light absorbed is correlated to traditional age estimates yielding a predictive model to estimate fish age from absorbance data alone. As with traditional ages, FT-NIR spectroscopy ages must also be standardized across spectrometers to ensure consistency across multiple machines, laboratories, and users. To date, this has not been researched, and with the increase of machines across multiple NOAA laboratories throughout the country, it is a unique opportunity to investigate scanning consistency and to quantify ageing precision for FT-NIR spectroscopy ages across spectrometers and laboratories.

METHODS

Otoliths from Beaufort's laboratory collection for gag (*Mycteroperca microlepis*, n =116) and vermilion snapper (*Rhomboplites aurorubens*, n = 154) from the U.S. South Atlantic were chosen as calibration samples to be scanned using either Bruker Tango (Beaufort, Narragansett, Santa Cruz, and La Jolla laboratories) or Bruker MPA II (Panama City [PC] laboratory) spectrometers (Bruker Corp., Billerica, MA). All laboratories followed the same scanning

protocol that the samples were originally scanned with in Seattle 2019 on PC's MPA II. This included the use of a standard gold stamp with a rubber gasket to reduce stray light and identical scanning protocols (16 cm^{-1} resolution, 64 sample scan time and background scan time, and wavelength from 11,550 to 3,950 cm^{-1}). Once all the samples were scanned by each laboratory, spectra were analyzed using Bruker's OPUS software to compare spectral data and resulting calibration models. Next, we analyzed the new spectral scans and input them into an established calibration model created on PC's MPA II while meeting in Seattle during 2019. The same preprocessing techniques were used and models were reran, then compared using model rank, the coefficient of determination (r^2), root mean square error, bias, and residual prediction deviation (Table 18.1). Lastly, the best predictive model was chosen by species, and the other laboratory location spectra were used as a test set to that model.

RESULTS AND DISCUSSION

When comparing the raw spectral files between laboratories and machine types, two things were clear: raw spectra differed between spectrometer types (TANGO vs. MPA II) and differed according to use of the gold stamp (Figs. 18.1 and 18.2). Both of these factors caused problems for standardizing spectra and calibration models across laboratories. The MPA II spectrometer had lower raw absorbance values than the Tangos, while Tango raw spectra overlapped across all laboratories (Figs. 18.1 and 18.2). This seemed to be due to different preamplifier gain settings within the two spectrometer models, where the amplitude shift in spectra was corrected for in the model build of the Tango but was not in that of the MPA II. Regarding the effect of stamp use on spectral data, one laboratory (Santa Cruz) scanned half of

the sample sets with a standard gold stamp and the other half without. When no stamp was used, higher absorbance amplitudes were observed than were with a standard stamp, regardless of the species (Figs. 18.1 and 18.2). Although preprocessing can remove many artifacts of ambient scanning conditions, these differences may nonetheless affect model results and should be avoided if possible.

These results are preliminary and sample sizes were small; hence, it is difficult to definitively state conclusions without further investigations and statistical testing. Further, the acceptable range of variability between instruments needs further investigation. Third, since the laboratories involved in this project are new to optimizing models, we investigated creating new models from each location's scans without any preprocessing designated. Instead, each model was run, had outliers removed, reran as needed, optimized, and the best model was chosen. Model output was compared between laboratory locations and the best tested model was graphed (Figs. 18.3 and 18.4). Regarding differences in model building between users, preprocessing choices varied by user and led to differences in overall model results between laboratories, indicating that preprocessing techniques need to be standardized by species. For example, Narragansett and Beaufort laboratory users optimized Narragansett spectra without any preprocessing designated, whereas other users did not. While some variation in raw spectra was observed across laboratories, each calibration model generally identified the same samples as spectral outliers during model testing (i.e, youngest ages for gag and chipped otoliths samples for vermilion snappers).

This was the first attempt to evaluate cross-machine and cross-user precision in spectral data collection and model building. Initial results indicate that scanning between machine types when utilizing the same scanning protocol provides consistent spectral data; however, it's

important to establish calibration sets similar to reference collections that are used to maintain consistency in traditional ageing. Another important takeaway from this study is that consistent scanning methods need to be established between machines prior to the start of scanning to ensure that results can be combined. Finally, model building should be well thought out by all users for individual species, and decisions should be made jointly on which parameters and preprocessing should be utilized.

Future discussions should include but not be limited to the following questions:

1. How should spectra be analyzed if scanning methods are changed? Is there a correction factor?
2. What is the threshold of acceptable variability between models?
3. How do we handle new spectral scans while utilizing existing models?
4. What is the minimum number of samples needed to create a model (calibration set vs. test set)?
5. How does the physical size of the sample affect models (i.e, some otoliths are much smaller or larger than the aperture window)?
6. How can we handle user variability in model optimization? How do we ensure we are using the same criteria for removing outlier samples?
7. Should we work with other data providers to create the same preprocessing protocol prior to scanning?
8. Should spectra be combined across data providers to create a single model?

Table 18.1. -- Model comparison output by location using established preprocessing protocols from previous scans (using the Panama City laboratory's spectrometer in Seattle) for vermilion snapper *Rhomboplites aurorubens* (top) and gag *Mycteroperca microlepis* (bottom). Preprocessing methods are outlined for each species. PC = Panama City, LJ = La Jolla, BFT = Beaufort, NAR = Narragansett, SC = Santa Cruz, RMSECV = root mean square error of the cross validation, RPD = residual prediction deviation.

Vermilion snapper
Preprocessing method: Multiplicative Scattering Correction
Regions of interest: 7456-4944, 4504-4000

	PC Seattle	LJ	BFT	NAR	PC	SC-all	SC-Gold	SC-No gold
Rank	3	2	6	3	4	3	3	2
R2	80.22	79.55	76.39	73.93	81.24	77.14	61.50	74.07
RMSECV	1.06	1.05	1.12	1.18	1.00	1.10	1.03	1.21
Bias	0.00286	0.00053	-0.34200	0.00769	0.00816		0.01130	0.02440
RPD	2.25	2.21	2.06	1.96	2.31	2.09	1.61	1.96

Gag grouper
Preprocessing method: 1st Derivative + Vector Normalization
Regions of interest: 7456-6400, 5504-4944, 4776-4000

	PC Seattle	LJ	BFT	NAR	PC	SC-all	SC-Gold	SC-No gold
Rank	5	4	4	5	6	4	2	2
R2	83.06	76.49	80.49	73.49	78.49	65.54	63.16	56.30
RMSECV	0.746	0.879	0.801	0.933	0.848	1.060	1.110	1.050
Bias	0.00689	0.01760	0.00747	0.00821	0.03450	-0.00347	0.03750	-0.05110
RPD	2.43	2.06	2.26	1.94	2.16	1.70	1.65	1.51

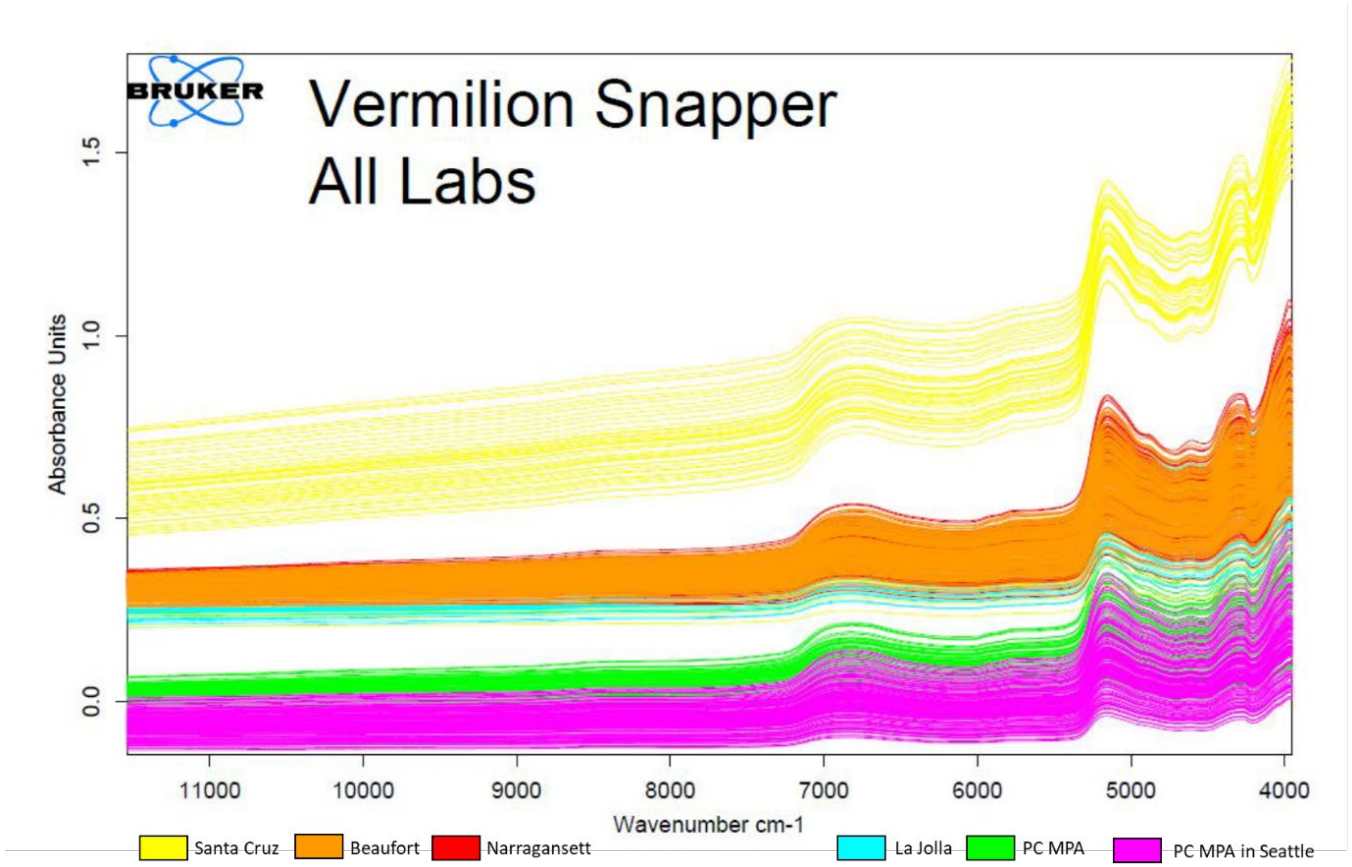


Figure 18.1. -- Raw spectral data of vermilion snapper (*Rhomboplites aurorubens*) otoliths scanned at each location's laboratory using two types of spectrometer: Tango (Santa Cruz, Beaufort, Narragansett, and La Jolla laboratories) and MPA II (Panama City [PC] laboratory). Note: Santa Cruz measured half of the samples without a gold stamp (upper level yellow spectra), while the other half were measured with a gold stamp (which line up with the rest of the Tango scans).

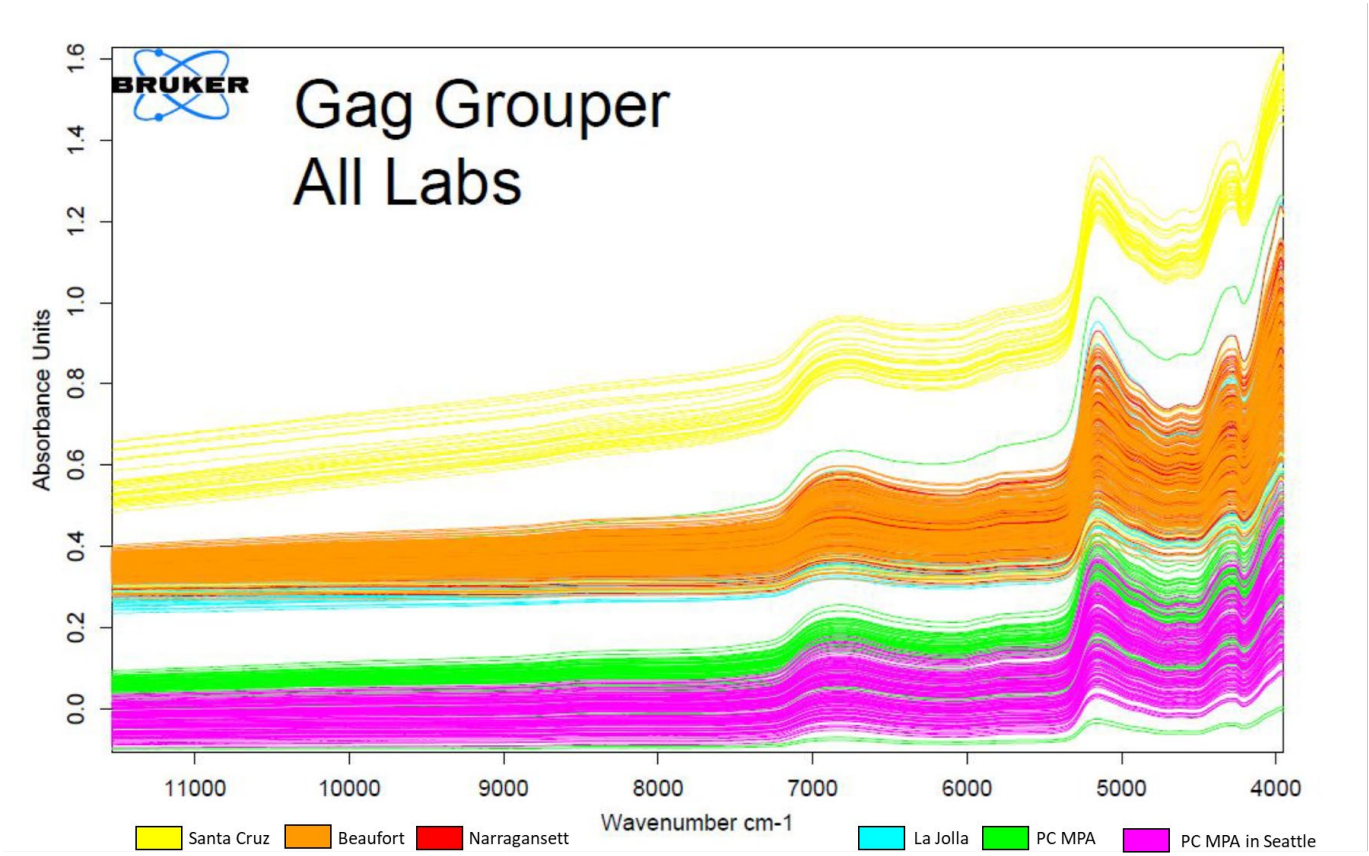


Figure 18.2. -- Raw spectral data of gag grouper (*Mycteroperca microlepis*) otoliths scanned at each location's laboratory using two types of spectrometer: Tango (Santa Cruz, Beaufort, Narragansett, and La Jolla laboratories) and MPA II (Panama City [PC] laboratory). Note: Santa Cruz measured half of the samples without a gold stamp (upper level yellow spectra), while the other half were measured with a gold stamp (which line up with the rest of the Tango scans).

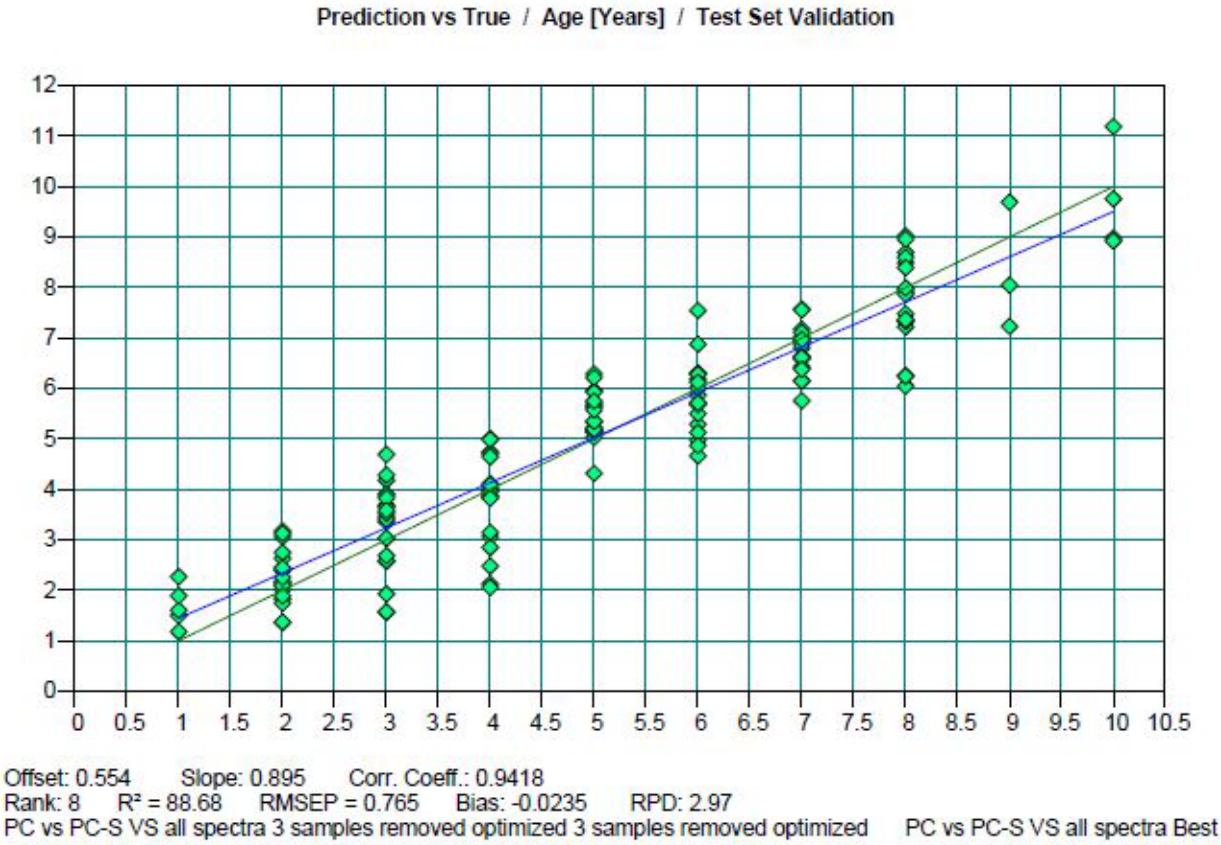


Figure 18.3. -- Vermilion snapper (*Rhomboplites aurorubens*) best model comparison output using other location scans as test sets. The best predictive model was from the Panama City (PC) laboratory, and the best test set was also from PC scans from Seattle in 2019. Model output is displayed under the graph.

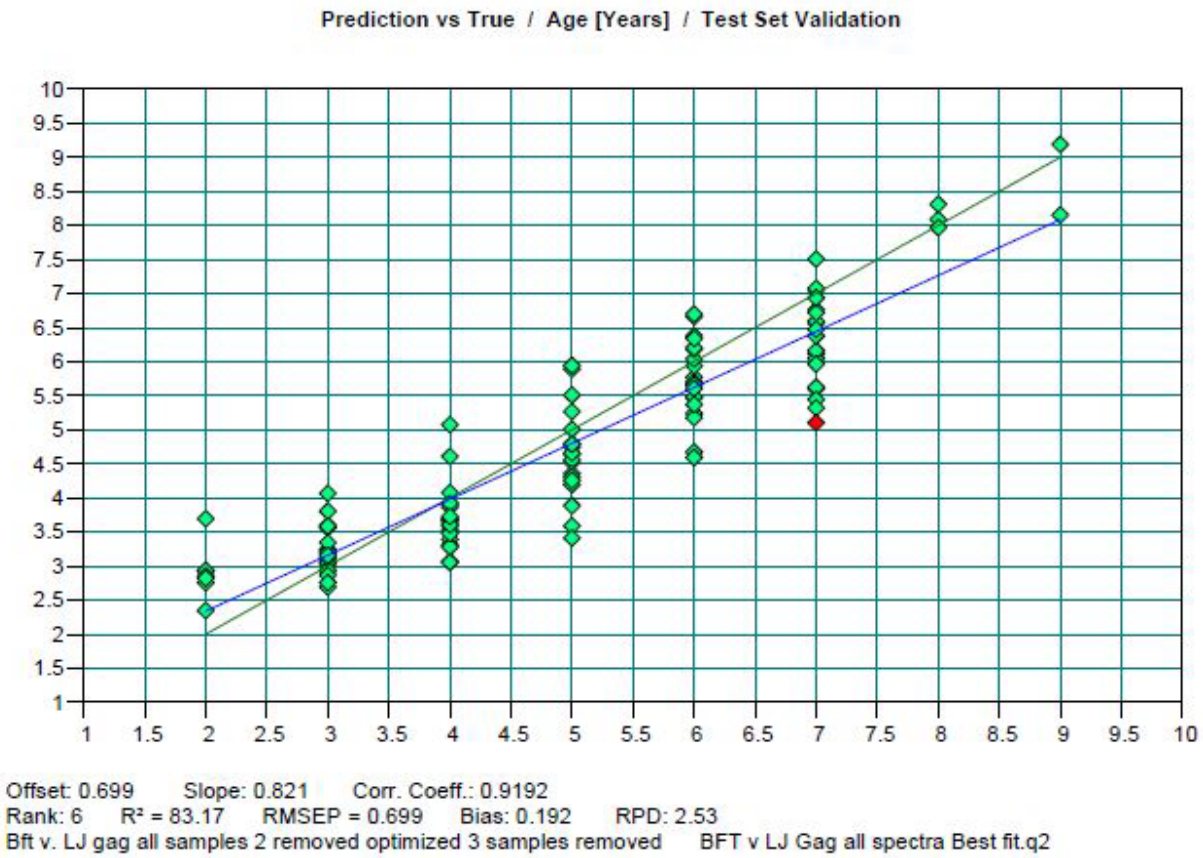


Figure 18.4. -- Gag grouper (*Mycteroperca microlepis*) best model comparison output using other location scans as test sets. The best predictive model was from Beaufort and the best test set was from La Jolla. Model output is displayed under the graph.

19. A Simulation-Based Approach to Evaluate Best Practices for Estimating Fish Age Using FT-NIR Spectroscopy

Morgan B. Arrington[1,3], Jordan T. Healy[2], Thomas E. Helser[3], Esther D. Goldstein[3], Irina M. Benson[3], Brenna C. Hsieh[3], and André E. Punt[2]

[1] Cooperative Institute for Climate, Ocean, and Ecosystem Studies (CICOES)
University of Washington
3737 Brooklyn Ave NE
Seattle, WA USA

[2] School of Aquatic and Fishery Sciences
University of Washington
1122 NE Boat Street
Seattle, WA USA

[3] Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

ABSTRACT

Fourier transform near infrared (FT-NIR) spectroscopy applied to otoliths has been used to predict the age of fish with comparable precision and improved efficiency relative to traditional, microscope-based methods. This approach requires the development of predictive calibration models that relate otolith spectra with fish age. However, before FT-NIR spectroscopy can be operationalized for production fish ageing and fisheries management, factors that may impact predictive skill of calibration models while maintaining efficiency must be investigated. The influences of calibration data set sample size (the number of otoliths that require traditional ageing) and sample selection methods on model predictive skill have not been thoroughly evaluated. Using eastern Bering Sea (EBS) walleye pollock as a case study, we conducted a simulation study of model predictive skill across 39 scenarios differing in selection methods and sample sizes. Specifically, we resampled with replacement from a subset of walleye pollock otoliths collected between the years of 2014-2021 (n = 2,055) and compared three selection methods: 1) random selection from samples where two age readers agreed on age based on the assumption that these ages were the most "accurate"; 2) random selection from the full subset of samples; and 3) selection using the Kennard-Stone algorithm to sample from the full domain of spectral variation. We evaluated each selection method at a range of sample sizes, each with 200 iterations, to determine the approach that maximizes efficiency but best represents the larger data set. Partial least squares regression calibration models were fitted to each data set and predictive skill on future, unseen data was evaluated by applying each model to a hold-out data set (n = 7,272) and estimating the root mean square error (RMSE) between predicted age and reference age for each scenario. We also evaluated two model validation approaches: cross-

validation and external validation, and their relative ability to accurately estimate model performance on new data (represented by model performance on the hold-out data set). For EBS walleye pollock collected between 2014 and 2021, calibration sample selection using the Kennard-Stone algorithm resulted in the most consistent model predictive skill at sample sizes greater than 500 (~7% of the hold-out data set), with negligible gains over 900 (~12% of the hold-out data set). The random selection approach was nearly comparable in performance to Kennard-Stone. Our results did not support any benefit of restricting calibration samples to those for which two readers agreed on fish age and also suggest that model cross-validation is a good indicator of future performance on unseen data when the Kennard-Stone selection approach is utilized but tended to overpredict performance when the random selection approach was employed. External validation tended to over-predict future predictive performance on unseen data for both selection approaches, especially when the external validation data set was small. Future work could explore the optimal sample size for an external validation data set that would better quantify model performance on new data. This case study outlines calibration model optimization techniques and approaches that can be universally applied to new species and data sets.

INTRODUCTION

Information on the age of fish is an essential component of fisheries research and hence management decision-making. Accurate age data allow for a more robust understanding of population dynamics and contribute to our ability to conserve and manage species effectively (Campana 2001, Lai and Gunderson 1987). Fish age data are essential when applying age-structured methods of stock assessment, as they inform estimates of the size of population cohorts as well as important life-history parameters related to growth, mortality, maturity, and longevity.

At the Alaska Fisheries Science Center alone, over 30,000 age estimates from over 22 species of fish are generated annually for inclusion in stock assessment (Lambert et al. unpublished report, Helser et al. this volume-a), representing an enormous time and monetary investment for processing and traditional age estimation. Fourier transform near infrared (FT-NIR) spectroscopy has been shown to be an effective method for predicting the age of numerous fish taxa with improved efficiency relative to traditional methods (Rigby et al. 2014, Wedding et al. 2014, Rigby et al. 2016, Helser et al. 2019a, Helser et al. 2019b, Arrington et al. 2022, Healy et al. 2021, Passerotti et al. 2020a, Passerotti et al. 2020b, Wright et al. 2021, Passerotti et al. 2022, Benson et al. 2023).

The eastern Bering Sea (EBS) walleye pollock (*Gadus chalcogrammus*) fishery is one of the largest fisheries in the Alaska region (1.2 million t, on average, since the late 1970s) and accounts for roughly a third of all production age estimates. The FT-NIR spectroscopy approach has predicted the age of EBS walleye pollock with comparable precision to traditional methods and improved efficiency (Helser et al. 2019b, Benson et al. 2023). Helser et al. (2019b) found

that 75% of FT-NIR age estimates for EBS walleye pollock collected during 2016 and 2017 from partial least squares regression (PLS) models were equivalent to traditional production integer ages. In contrast, the traditional age reading method results in only 68% agreement between reader and tester ages. Benson et al. (2023) found that for EBS walleye pollock collected between 2014 and 2018, 81% of FT-NIR age estimates from multimodal convolutional neural network models (MMCNN) were the same as traditional production ages, in contrast to only 73% agreement between reader and tester in the traditional method. Both PLS and MMCNN resulted in predicted ages that had less bias for most age classes than age estimates compared between two age readers in traditional methods.

However, despite promising proof-of-concept studies, the FT-NIR method of fish ageing has yet to be formally used for life history studies or for production ages included in stock assessments. Best practices still need to be developed to understand methodological impacts on the predictive skill of FT-NIR models when applied to future data. Here, we develop a simulation framework to explore the influence of calibration data set sample size and selection method on PLS model consistency and predictive accuracy. We define consistency as the range in predictive performance of calibration models on future data. We define predictive accuracy as how well a calibration model predicts future data. Both consistency and accuracy are evaluated based on root mean square error (RMSE) between model predictions and traditionally estimated age. We address the following three questions:

- Question 1: What are the best methods for selecting a calibration data set to optimize performance on new data?

- Question 2: What is the minimum sample size required for a consistent calibration data set?
- Question 3: What are the best methods for validating a calibration model to determine how it will perform on new data?

METHODS

Walleye pollock otoliths included in this study were collected on fishery-independent EBS bottom trawl surveys conducted by the Alaska Fisheries Science Center between the years 2014 and 2021. Otoliths were stored in glycerin-thymol solution after extraction at sea. Traditional age estimates for all samples included in this study (n = 9,327) were determined microscopically by counting pairs of annual bands corresponding to the age of the fish using one paired sagittal otolith (Matta and Kimura 2012). Two age readers independently read a randomly selected 20% subset of specimens to evaluate the precision and bias of traditional methods. Traditional age reading methods for walleye pollock are destructive, so the other otolith was used for spectroscopic evaluation.

Spectroscopy

Spectral absorbance data were collected on either a Bruker TANGO-R FT-NIR spectrometer (Bruker Optics, Ettlingen, Germany) or a Bruker MPA II FT-NIR spectrometer with an integrating sphere. Whole, unaltered otoliths were removed from their vials and excess glycerin-thymol was blotted off using Kimwipes (Kimtech Science). The otolith was then placed

on the sampling window distal side up in a 0° orientation and covered with a gold transflectance cap. Absorbance data were acquired at 16 cm^{-1} resolution with 64 averaged scans between the wavenumbers 11,500 and 4,000 cm^{-1}.

Analysis

We conducted data analysis in R (version 4.3.0, R Foundation for Statistical Computing, Vienna, Austria, see https://www.Rproject.org), with chemometric package *mdatools* (version 0.14.0, Kucheryavskiy (2020a), https://github.com/svkucheryavski/mdatools). Raw spectral data were pre-processed with mean centering and Savitzky-Golay smoothing (first derivative, second order polynomial, 17-points; Savitzky and Golay 1964), to remove the offset between instruments while preserving the chemical signal of interest (Benson et al. 2023, Luthria et al. 2011).

We removed outliers based on orthogonal distances (Q) and score distances (Hotelling's T^2). This was done by calculating critical limits using a data-driven approach assuming a joint distance that follows a chi-squared distribution (Rodionova and Pomeranstev 2020). Values that fell outside a significance level of 0.01 were considered outliers and removed from the data set (n = 100). We split the remaining data into two data sets. One data set (hereinafter referred to as the double-read data set) had all specimens that were read by two age readers (n = 2,055), and the other data set had just a single read age (hereinafter referred to as the hold-out data set) (n = 7,272) (Fig. 19.1).

We then evaluated three methods for selecting calibration models from the double-read data set at a range of sample sizes: 1) the "agree ages" approach – we used just the spectra from otoliths where both the reader and tester agreed on the age (n = 1,386) – this was based on the

assumption that these otoliths had the least error in reference ages, 2) the "random selection" approach - we randomly selected a subset of data for the calibration model from the full double-read data set; and 3) the "Kennard-Stone algorithm" approach - we used the Kennard-Stone algorithm (Ferreira et al. 2021, Kennard and Stone 1969) to select representative samples that encompassed the full range of spectral variation in the double-read data set. When selecting a subset of data for a calibration sample, there are many possible combinations depending on the total sample size. Some combinations may, based on random chance, be better or worse at predicting future samples than others. By utilizing a simulation-based approach with n = 200 repetitions, we were able to simulate sampling from a larger population and generate a distribution of possible results.

Agree ages approach

To evaluate the agree ages approach, we resampled with replacement from the specimens where the reader and tester agreed on age (1,386 samples out of the 2,055 samples in the double-read data set) at sample sizes from n = 100 to the maximum possible, n = 1,386, at intervals of 100. We resampled 200 data sets per calibration sample size. Each had a paired external validation set that included any samples from the double-read data set not included in each calibration data set.

Random selection approach

To evaluate the random selection approach for selecting calibration samples, we resampled with replacement from the full double-read data set (n = 2,055) at sample sizes from n = 100 to n = 1,386 to be equivalent to the agree ages approach. This was to eliminate variation in

model performance due to different calibration sample sizes, and resulted in 200 simulated data sets per calibration sample size, each with a paired external validation data set that included any samples from the double-read data set not included in each calibration data set (Fig. 19.1).

Kennard-stone algorithm approach

To evaluate the Kennard-Stone algorithm for selecting calibration samples, we resampled with replacement from the full double-read data set (n = 2,055) to generate 200 simulated data sets. We then applied the Kennard-Stone algorithm to each simulated data set to select calibration sets with sample sizes from n = 100 to n = 1,386 to be equivalent to the agree ages and random selection approaches. This resulted in 200 simulated calibration sets per calibration sample size, each with a paired external validation set that included any samples from the double-read data set not included in each calibration data set (Fig. 19.1).

Evaluating model performance

To evaluate model performance in each scenario, we then fitted PLS models (Wold et al. 1984, Helser et al. 2019b) to each calibration set selected via each method at each sample size. To evaluate and compare model performance on new data, we applied each model to predict ages for the hold-out data set (n = 7,272) (Fig. 19.1). Model predictive skill was summarized by calculating the root mean square error (RMSE) of predicted ages relative to reference ages for each iteration in each scenario. This resulted in 200 RMSE values per method, per sample size.

In most proof-of-concept studies, a hold-out data set with such a large number of samples (n = 7,272 for EBS walleye pollock) with traditionally estimated reference ages is not typical. To simulate model evaluation methods more typical in a proof-of-concept study, we evaluated the accuracy of external validation and cross-validation methods for estimating future performance

on new data (represented by model performance on the hold-out data set) (Fig. 19.1). To evaluate external validation methods, we applied the same PLS models to each paired external validation set and calculated the RMSE of the predicted ages versus traditionally estimated reference ages. This resulted in 200 RMSE values per method, per sample size. PLS models were also tested through cross-validation by fitting models using each data set iteratively. In each iteration, 20 samples were systematically left out and parameter estimates were used to estimate the age of the left-out samples. The mean error of all predictions versus reference ages was calculated as the RMSE of cross-validation. This resulted in 200 RMSE values per method, per sample size.

RESULTS AND DISCUSSION

The results of this simulation suggest that model predictive accuracy (represented by RMSE) can vary depending on what samples are included in the model calibration set, the selection approach utilized, and the sample size. Additionally, it is important to give model validation approaches careful consideration as they can over-estimate model predictive performance on new data.

When all calibration sets were used to predict fish age for the hold-out data set (n = 7,272) to represent predictive skill on new data, the best and most consistent model performance was achieved using the Kennard-Stone algorithm at the highest sample size tested (n = 1,386) (Fig. 19.2). The most variability in model performance was for sample sizes less than 500, suggesting that more traditional ages are required to reliably select a representative sample of the population. However, there were negligible gains over 900 samples, suggesting that this may be a good target number for traditionally estimated reference ages for EBS walleye pollock.

Although there was near full overlap among the Kennard-Stone and the other two selection methods' RMSE values, the random and agree ages approaches had larger ranges in RMSE that encompassed models with poorer predictive performance than the Kennard-Stone approach (Fig. 19.2). Of the three methods, the agree ages approach resulted in a range that encompassed the poorest models at calibration sample sizes over 500 (Fig. 19.2). These results do not support a benefit to restricting calibration samples to those with agreement between the primary age reader and the test reader. Though random selection resulted in some models with slightly poorer performance than the Kennard-Stone algorithm, they had nearly comparable RMSE values and both methods may be useful in different contexts (Fig. 19.2). When selecting a subset of samples for a spectroscopy study, random selection of samples for traditional ageing may be most appropriate if spectra are not yet collected, since the Kennard-Stone algorithm requires information about spectral variation. However, in a scenario where spectra are available that do not yet have reference ages, the Kennard-Stone selection method can be used to optimize the samples to process for traditional age estimation.

Finally, model validation techniques are often used to estimate the future performance of a model on new data in a proof-of-concept study where data are limited. We found that cross-validation over-estimated predictive skill for the random selection approach, showing better RMSE than when the same calibration models were applied to the hold-out data set (Fig. 19.3a). However, cross-validation adequately represented the predictive skill for the Kennard-Stone selection approach and resulted in a conservative estimate of model predictive skill, except for the highest two sample sizes (Fig. 19.3b). External validation (using paired validation data sets) tended to overestimate predictive skill for the random approach at sample sizes greater than 300 (Fig. 19.4a). It also tended to overestimate predictive skill for the Kennard-Stone selection

approach at sample sizes greater than 200 (Fig. 19.4b). In a data-limited scenario, if using the Kennard-Stone algorithm for selecting calibration samples, our recommendation would be to use cross-validation to estimate predictive performance on future, unseen data. If using random selection, future work could explore the adequate sample size of an external validation data set that would result in better estimates of performance on new data. Anecdotally for these results, when ~1,700-1,900 samples are included in the external validation relative to 100-300 samples in the calibration, the external validation better estimated "actual" performance as represented by predictive skill on the hold-out data set (n = 7,272) (Fig. 19.4a,b). However, calibration models fitted on just 100-300 samples were also highly variable and inconsistent in their predictive performance on future, unseen data (Fig. 19.2). To more accurately estimate the future predictive skill of calibration models fitted on an optimal number of 500-900 samples, a larger paired external validation data set of spectra with traditional ages would likely be required than was available in our double-read data set (n = 2,055) (Fig. 19.1).

This simulation utilized EBS walleye pollock for a case study, yet the simulation framework developed in the study is flexible and can be applied in other contexts. Findings from this study may be generalizable across species. Recommendations to use the Kennard-Stone approach or random selection approach for calibration sample selection are likely broadly applicable. Similarly, our results highlight important considerations for validation methods that accurately estimate a model's future performance on new data. We recommend building from this framework to explore best practices on an application-specific basis since optimal calibration sample sizes may vary by species and modeling approach (e.g., MMCNN). A better understanding of these best practices, such as the impact of calibration sample size and selection

methods on PLS model consistency and predictive accuracy for any candidate species, is important for the successful use of FT-NIR spectroscopy for production fish ageing.

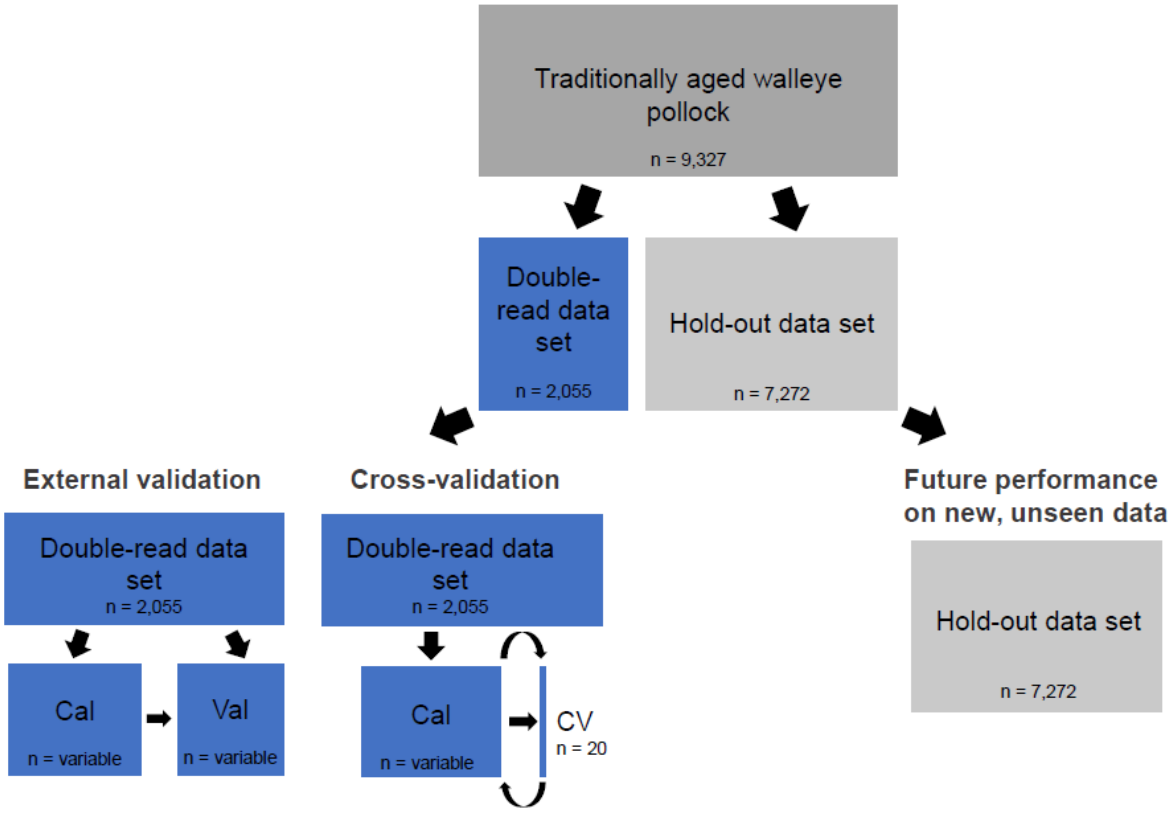


Figure 19.1. -- Schematic of study design. The full data set including all traditionally aged eastern Bering Sea walleye pollock *Gadus chalcogrammus* (n = 9,327) from 2014 to 2021 was split into a "double-read data set" with all samples that had a reader and tester age (n = 2,055), and a "hold-out" data set with all other samples that had just one traditionally estimated age (n = 7,272). The double-read data set was used for comparing calibration data set sample size and selection methods so that the agree ages selection approach, which required that two age readers agreed on age, could be compared to random selection and the Kennard-Stone algorithm. The hold-out data set was used to represent model performance on future, unseen data. The hold-out data set was kept unchanged so that calibration models could be compared against a comparable benchmark. Two typical validation approaches - external validation and cross-validation - were compared in their ability to estimate future model performance on new, unseen data (represented by hold-out data set). Above, "cal" represents calibration and "val" validation.

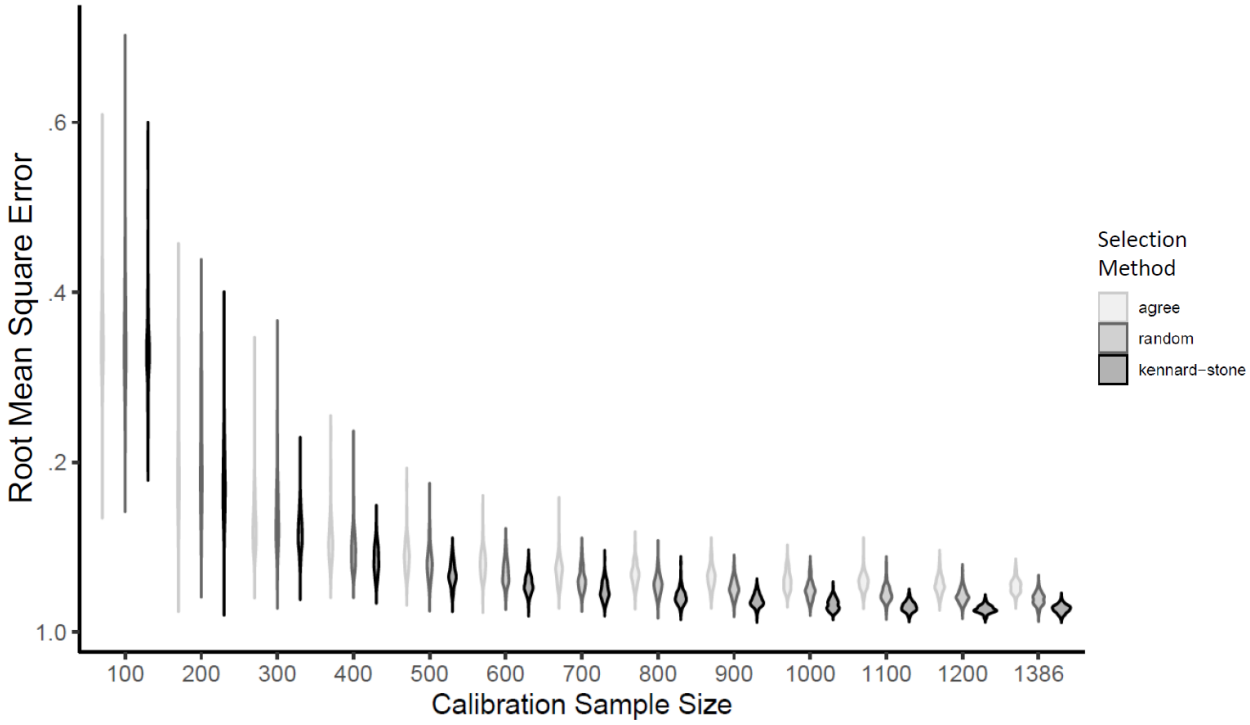


Figure 19.2. -- Violin plots (kernel density) comparing root mean square errors among the three calibration selection approaches to predict ages for the hold-out data set (n = 7,272). This is shown for calibration sample sizes 100-1,386 at intervals of 100. Each has 200 iterations.

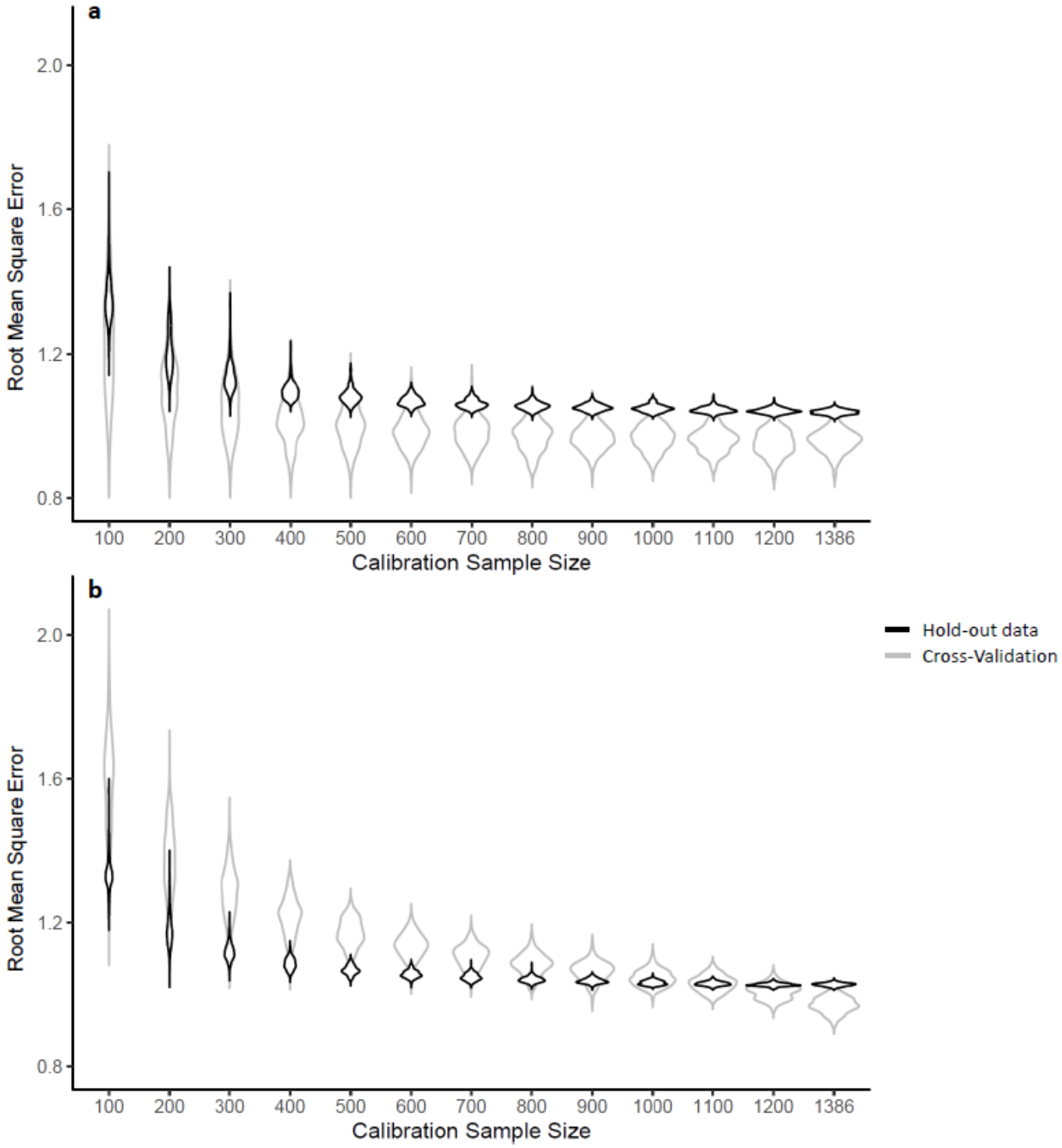


Figure 19.3. -- Violin plots (kernel density) comparing estimated model predictive skill (root mean square error) from model cross-validation versus actual performance on future, unseen data represented by the hold-out data set for a) the random selection approach, and b) the Kennard-Stone algorithm selection approach.

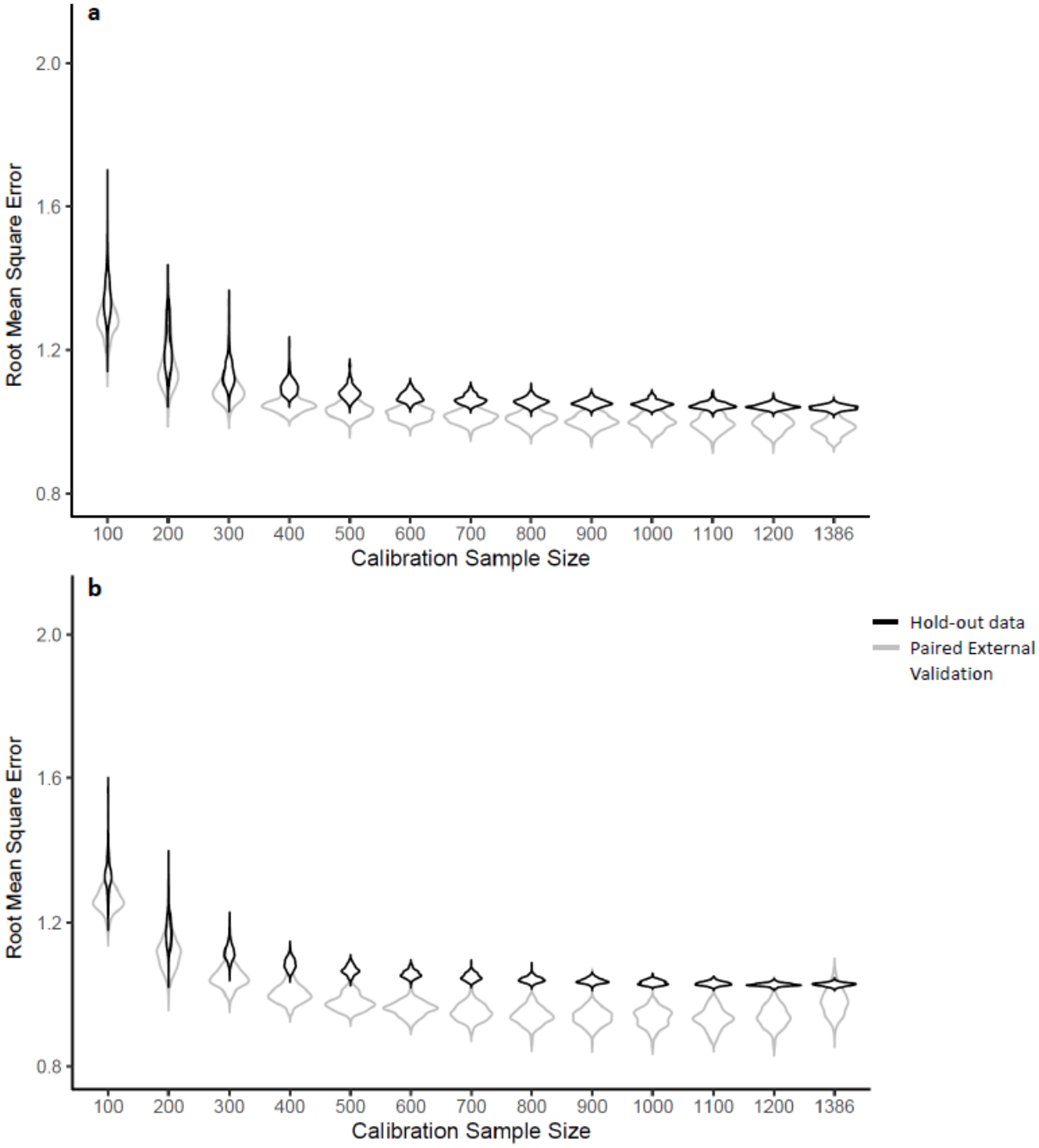


Figure 19.4. -- Violin plots (kernel density) comparing estimated model predictive skill (root mean square error) from external validation (paired validation sets) versus actual performance on future, unseen data represented by the hold-out data set for a) the random selection approach, and b) the Kennard-Stone algorithm selection approach.

20. Database Design and Considerations with FT-NIR Spectra Data Collection and Management

Jonathan A. Short

Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

INTRODUCTION

Fourier transform near infrared (FT-NIR) spectra have been recently discovered to be informative in age estimation of many species of fishes (e.g., Helser et al. 2019a, Robins et al. 2015, Wedding et al. 2014). To take full advantage of this new data stream, a well-designed and scalable database is useful for managing the data collection process and for creating a searchable clearinghouse of spectral data. The Alaska Fisheries Science Center (AFSC) Age and Growth Program (AGP) uses a relational SQL Server database called AGE3 to house specimen and age data and it has been expanded to include FT-NIR data. As of this writing, the AGE3 database houses specimen data for over two million fish, over one million traditional ages, and over 100,000 FT-NIR spectra.

Fish specimens come to the AGP from survey and fishery collections that have already been through a quality control process and have unique fish identifiers. Survey collections have a unique combination of vessel code, cruise number, species, and specimen number. Fishery collections are given unique barcodes. With either combination of identifiers, we can relate to catch data, such as the date of collection and latitude and longitude.

DATABASE DESIGN AND ARCHITECTURE

The AGE3 database is a relational database that has data objects that reflect physical objects. For example, the "Fish" data object or table contains attributes directly relating to fish specimens, such as fork length, sex, and body weight (Fig. 20.1). Another data object called "Age structures" is related to Fish via a primary key-foreign key relationship and contains

attributes related to the structure such as structure type (otolith, vertebra, scales, etc.), structure label information, and structure storage location. FT-NIR data are collected on parts of fish and not the entire fish, so the "NIR scans" object is related to the Age structures table. In this way we can spectrometrically interrogate many aspects of the fish without duplicating attributes such as fish length and weight and while maintaining data integrity. The AGE3 database stores the paths to the spectral files on a network share and metadata describing the conditions of the structure scanned. The database also stores information relating to instrument settings and quality control notes and changes.

The database consists of a SQL Server 2016 back end housed in the AFSC server space. It is accessible with Windows Authentication, so users do not need to go through an additional login procedure after logging in to their computer workstation. The database front ends are a custom Python/Qt interface, several Microsoft Access front ends, and may be queried directly with code and an ODBC connection. The database structure and user access are limited to set processes so it is not possible for users to inadvertently change data that are outside their purview.

SPECTRUM COLLECTION

An important aspect of the collection process is creating unique spectrum file names. This is accomplished in the AGE3 database by beginning with fish specimen data already in the database and tracking the number of times a fish structure has been scanned. Data sessions are created and the structures to be scanned are added to it by a data manager. Scripts in the database interface check the number of times a structure has been scanned to ensure data integrity. Once

the session is set up and samples are available, then it is ready for scanning. The data manager writes a description of the sample selection process and the treatment involved with each data session (Fig. 20.2).

Once a session is ready, a scan form is opened by the NIR operator (Fig. 20.3). The scan form is pre-filled with specimen numbers or barcodes of the samples ready to be scanned. The NIR operator records sample conditions such as broken, crystallized, structure side, percent affected, sample weight and any other comments pertaining to the sample, and then pastes the database-generated spectrum file name into the name field in the Bruker Opus software. Opus scans the structure and generates the spectrum file in a temporary network location to which all AGP users have read/write access. The AGE3 scan form locks completed records one by one as the operator clicks the save button and checks that the spectrum file is generated in the expected file location.

Once all spectra in a session are collected, the NIR operator checks their data and marks the session as complete. This automatically generates an email notification to the data manager that a session is ready for QA/QC. After quality control, the data manager executes a script that moves all spectrum files in a session to a permanent archival location with read-only access.

DATA CLEARINGHOUSE

Archived NIR data can be retrieved by querying all available data in the AGE3 database. Data analysts can search with many criteria, since all related metadata are stored in the database (Fig. 20.4). A Microsoft Access front end has common criteria preset into a data selection form. This can build a query that can be further modified to the specific needs of the analyst. A script

in the front end copies all selected spectra and related metadata from the archive library to the destination of choice.

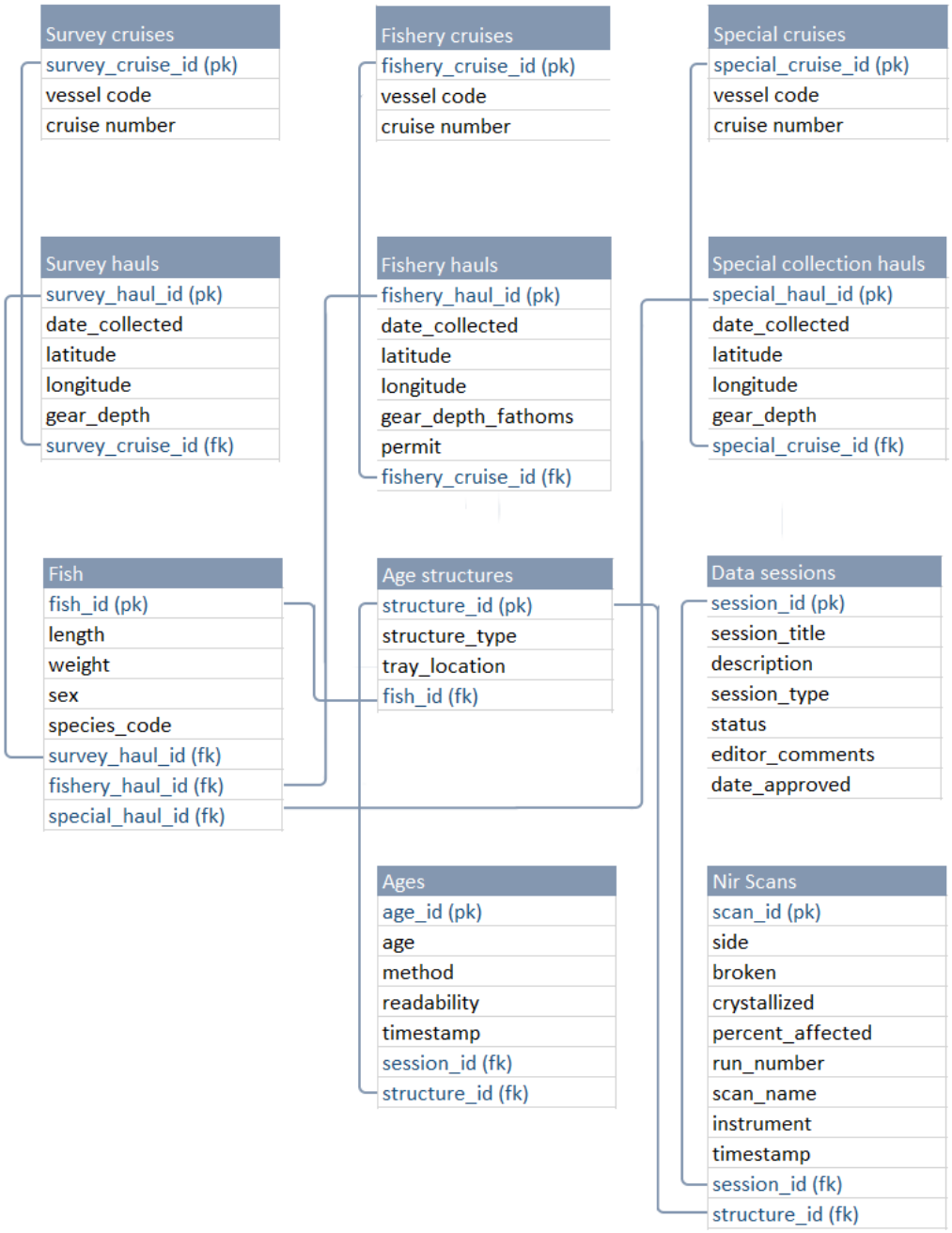


Figure 20.1. -- A simplified table structure of the Alaska Fisheries Science Center AGE3 database showing entities and attributes, primary keys and foreign keys for collection data, fish data, and near infrared (NIR) spectra.

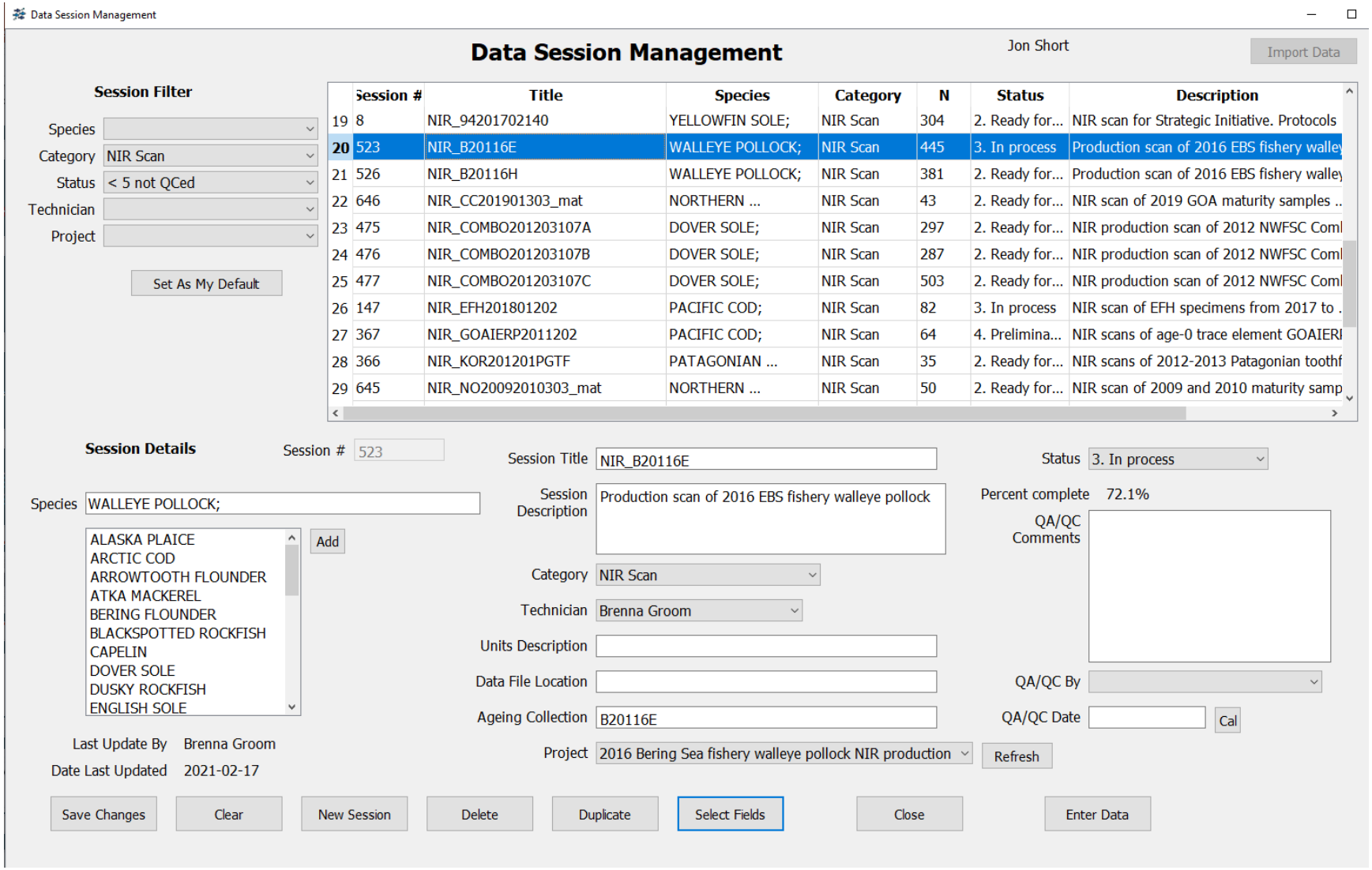


Figure 20.2. -- The AGE3 data session management form. This form displays metadata relating to near infrared (NIR) spectra in the active collection process and gives NIR operators access to in-process data sessions.

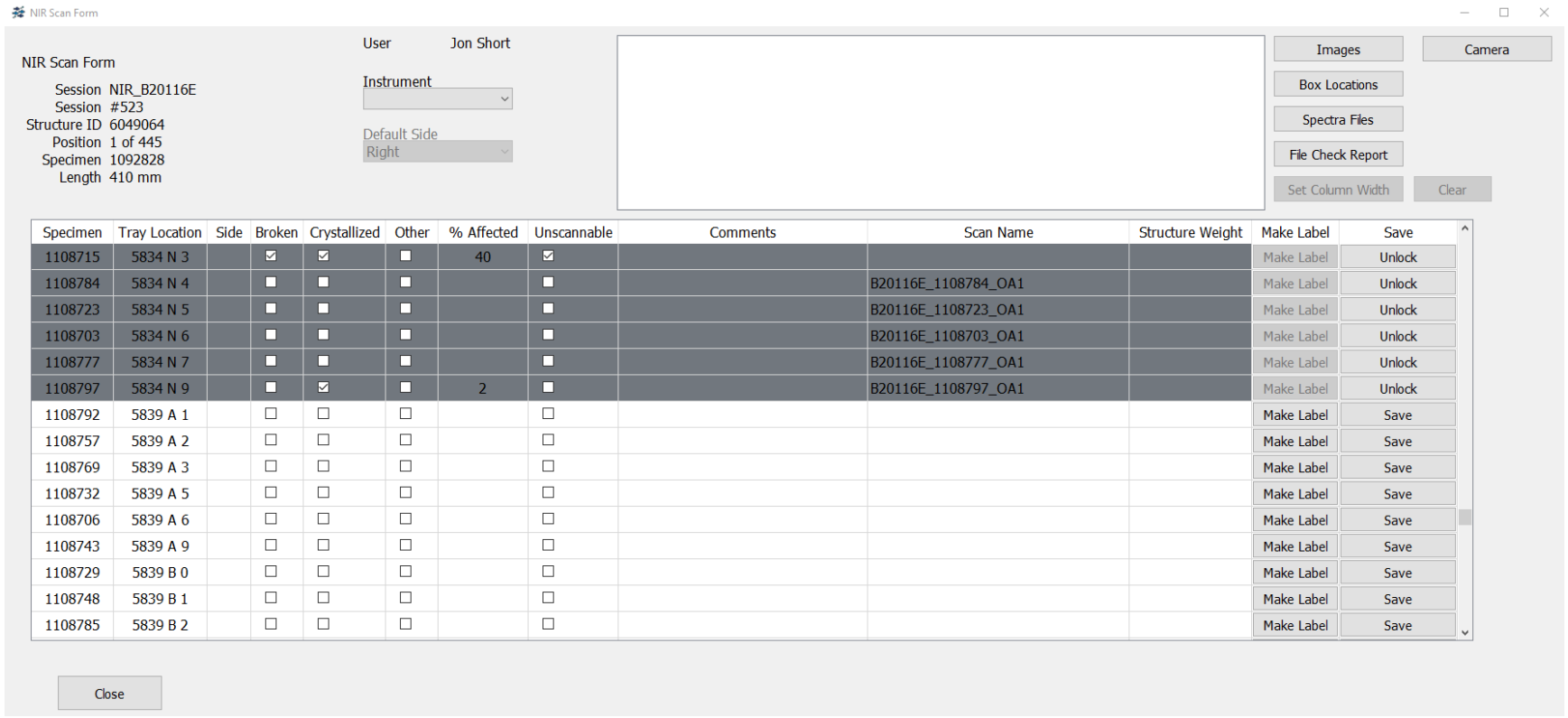


Figure 20.3. -- The near infrared (NIR) scan form for a fishery collection in AGE3. NIR operators enter otolith characteristics during the scanning process and this form creates the Scan Name, which is copied to Opus to create the spectrum file name.

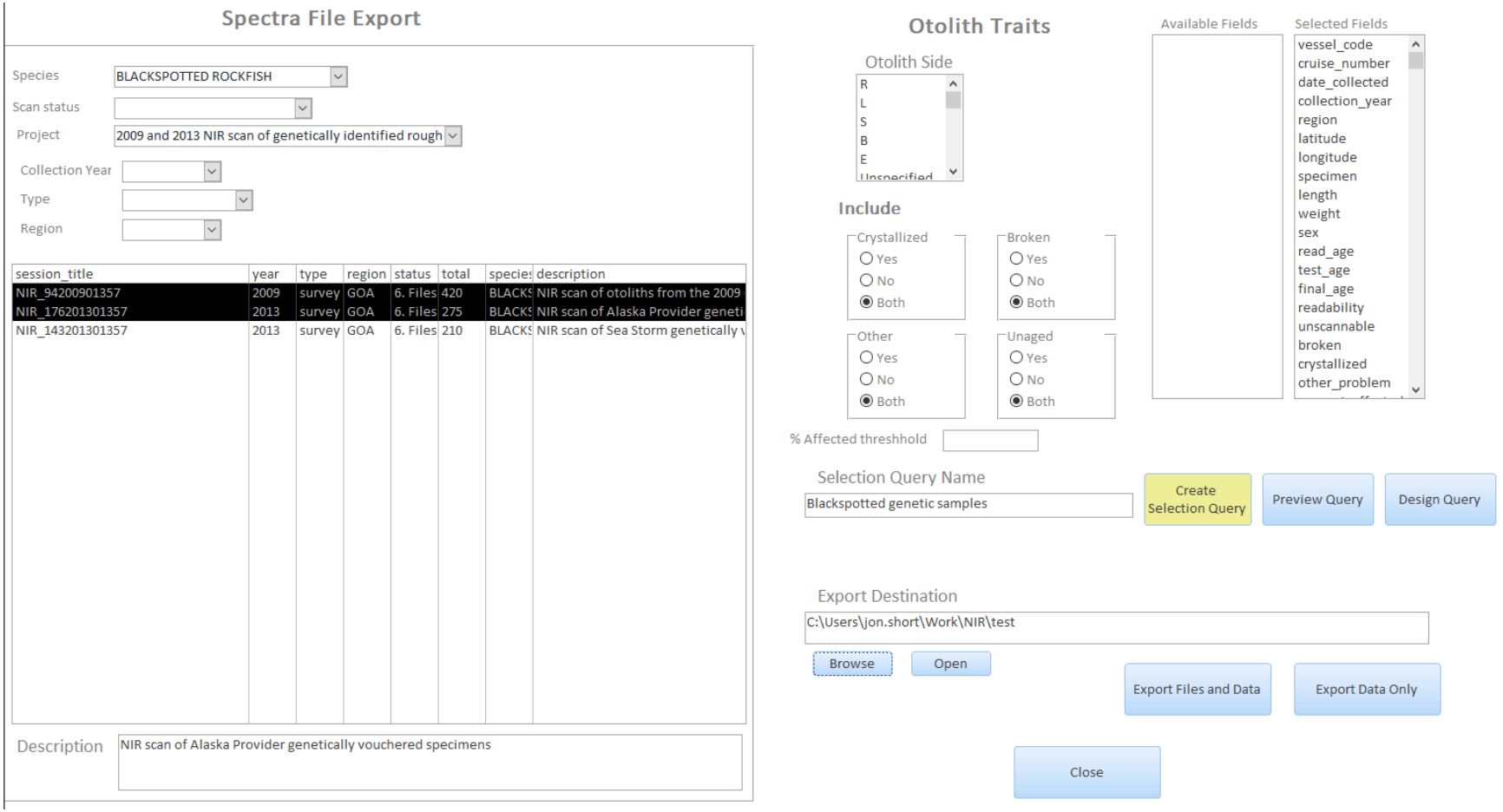


Figure 20.4. -- The spectra file export form is a clearinghouse of spectral data. Analysts can filter data sessions by species, status, project, year, type, and region, then select or exclude specimens based on otolith characteristics. With the script in this form, spectra and metadata will be copied from the archived library to the analyst's chosen file destination.

21. Envisioning the Future of Production Fish Ageing: End-to-end Integration of the FT-NIR Spectroscopy Age Estimation Enterprise at the Alaska Fisheries Science Center

Thomas E. Helser[1], Irina M. Benson[1], Mary Elizabeth Matta[1], Esther D. Goldstein[1], Brenna C. Hsieh[1], Morgan B. Arrington[1,2], and Jonathan A. Short[1]

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

[2] Cooperative Institute for Climate, Ocean, and Ecosystem Studies (CICOES)
University of Washington
3737 Brooklyn Ave NE
Seattle, WA USA

ABSTRACT

Fourier transform near infrared (FT-NIR) spectroscopy of otoliths is emerging as a highly efficient fish age estimation alternative to the labor intensive, time-consuming traditional microscope-based approach (TMA). Within the National Oceanic and Atmospheric Administration (NOAA) National Marine Fisheries Service (Fisheries), seven biological laboratories are engaged in production ageing that provides critical data for stock assessments in support of managing the nation's fishery resources. Research by Alaska Fisheries Science Center (AFSC) scientists demonstrated the potential utility of FT-NIR spectroscopy for ageing eastern Bering Sea (EBS) walleye pollock, which led to NOAA Fisheries funding a 5-year strategic initiative in 2020 entitled "A revolutionary approach for improving age determination efficiency in fish using Fourier transform near infrared (FT-NIR) spectroscopy." The roadmap toward operationalizing the FT-NIR spectroscopy ageing technology across science centers was envisioned as encompassing three major related tasks: (1) application development, (2) application implementation, and (3) stock assessment integration. Within this framework, examples of successful fish ageing applications of FT-NIR spectroscopy have been documented for numerous species and regions over the last several years by the Strategic Initiative Development Team (SIDT). In a few regions, headway has also been made in application implementation through simulation studies for best practices and even stock assessment integration by evaluating stock assessment model sensitivity to data products generated using FT-NIR spectroscopy. However, before such innovative methods attain a technical readiness level (TRL) for transition, we must begin to focus on adaptation of FT-NIR spectroscopy methods into existing TMA operations that result in a smooth delivery of data streams with

essential quality control processes that generate consistent and reliable data products. This chapter describes in detail the end-to-end integration of FT-NIR spectroscopy methods for age estimation envisioned for the production ageing process at the AFSC. In this case study, we employ a predictive model for EBS walleye pollock, leveraging over 9,000 observations spanning from 2014 to 2018, to investigate the envisioned process for generating future age predictions.

INTRODUCTION

The central principle governing fish age determination based on the growth patterns visible in otoliths has changed little over the last century. While ageing fish using scales can be documented as far back as the early 1700s, Johannes Reibisch was the first to describe a protocol for ageing fish using otoliths in 1899 (Jackson 2005). Microscopic counting of the annual growth zones in otoliths was the foundation of the ageing method then and is still the basis of age estimation today. With the exception of advancements in microscopy and otolith preprocessing techniques, the standard practice today remains the visual examination and enumeration of annual banding patterns. Ageing laboratories worldwide have improved age estimation methods through various otolith processing techniques to enhance the microstructure for visual interpretation. These techniques involve clearing whole otoliths in various solutions or sectioning them and then applying burning or staining processes to improve contrast between growth zones. Another common practice is “double reading”, wherein the same otolith is evaluated by a second (or even third) analyst to estimate precision or repeatability of age estimates. Hence, the age reading process aligns with the scientific method and allows for the application of statistical methods to estimate age reading precision. McBride (2015), Campana (2001), and others have discussed statistical techniques that can be used to assess age reading precision. Generating ages for stock assessments follows the scientific method provided that ages are validated, age determination criteria are consistently applied, and independent double readings are conducted. Campana (2001) provides a thorough treatise of age validation methods along with measures of statistical precision. The entire process of generating ages can be expensive, time consuming,

and labor intensive, especially for species with long lifespans or complex otolith processing methods.

In recent decades, efforts have been made to increase the efficiency and improve repeatability of age determination with machine-based technologies. These technologies have utilized otolith morphometrics (Pilling et al. 2007, Fablet et al. 2009, Mahe et al. 2016), image analysis (Nasreddine et al. 2013), or both (Troadec and Benzinou 2002, Fablet and Le Josse 2005), with variable levels of success. When a large number of age estimates is needed for stock assessments, the goal is to increase efficiency without hindering age accuracy or precision (Fablet and Le Josse 2005). Although a number of alternative approaches have been explored for use in age estimation, none have been widely adopted as a large-scale production tool for age determination. Following the lead of several promising studies in Australia (Wedding et al. 2014, Robins et al. 2015), scientists at the AFSC have shown that Fourier transform near infrared (FT-NIR) spectroscopy applied to walleye pollock (*Gadus chalcogrammus*) otoliths (Helser et al. 2019b) holds promise of high-throughput fish ageing. FT-NIR spectroscopy is used in a wide variety of industries including pharmaceutical, chemical, petrochemical, agricultural, and dairy manufacturing. FT-NIR spectroscopy is a vibrational spectroscopy technique that functions by exciting covalent bonds (O-H, C=O, C-H, C-N, and N-H) at the molecular level using NIR light and measuring absorbance in the wavenumber range 4,000 to 12,500 cm^{-1}. The light's interaction with the sample results in measurable vibrational frequencies represented by spectral signatures associated with molecular combinations and overtones that make up compounds in the sample (Conzen 2014, Siesler et al. 2002). In otoliths, these spectral signatures correlate well with age, enabling the establishment of a mathematical relationship between absorbance and fish age.

With the promise for improving efficiency and repeatability of generating age data for stock assessments, the National Oceanic and Atmospheric Administration (NOAA) National Marine Fisheries Service (Fisheries) has undertaken a nationally coordinated research-and-development effort across seven biological laboratories to integrate FT-NIR spectroscopy technology into the fish age determination enterprise (Helser et al. 2019a). NOAA Fisheries funded a 5-year strategic initiative entitled "A revolutionary approach for improving age determination efficiency in fish using FT-NIR spectroscopy" in 2020. With the participation of national and international scientists and industry experts in FT-NIR spectroscopy, a framework toward operationalizing the FT-NIR spectroscopy ageing technology across fisheries science centers emerged from a three-day planning workshop in April 2019 (Helser et al. 2019a). The Strategic Initiative Development Team (SIDT) envisioned a "roadmap" consisting of three related and overlapping tasks: (1) application development, (2) application implementation, and (3) stock assessment integration (Helser et al. 2019a). Application development focused on the exploration of the technology to generate reliable NIR spectra (good signal-to-noise) from otoliths and determine whether a strong significant mathematical relationship between otolith spectra and fish ages of different species and in different regions of the country could be established. In application implementation, the goals were to investigate the delivery of new data generated by predictive models, define and evaluate model performance, establish protocols for data processing and quality control, and develop a manual of best practices to guide decisions related to sample size, sampling distributions, ageing precision, and performance measures. While not the end point, the stock assessment integration objective sought to evaluate the sensitivity of population model output of critical management outcomes and benchmarks from incorporating FT-NIR spectroscopy data. Taken together, each element of the hierarchy builds

upon the previous tier with "discovery switch-backs" helping to refine and improve the technical knowledge of the system.

Collectively, the NOAA Fisheries ageing laboratories have made reasonably good progress on application development, despite significant delays and setbacks due to the 2020-2022 global pandemic. The science underpinning the use of FT-NIR spectroscopy for age prediction is demonstrated by a number of published studies across fish taxa and regions in the U.S. (Helser et al. 2019a, Helser et al. 2019b, Healy et al. 2021, Arrington et al. 2022, Passerotti et al. 2020b, Passerotti et al. 2022, Benson et al. 2023), with others in Australia (Wedding et al. 2014, Rigby et al. 2016, Wright et al. 2021), and more are yet to be published (including a number of studies described in these proceedings). Although work is ongoing or still needed across all science centers, these efforts hold great promise for improving efficiency and repeatability of generating age data for stock assessments using FT-NIR spectroscopy. At the AFSC, simulation studies are underway to evaluate best practices, and two case studies for eastern Bering Sea (EBS) walleye pollock and Pacific cod (*G. macrocephalus*) demonstrate the sensitivity of the stock assessment models to using FT-NIR spectroscopy-generated age compositions compared to the traditional method of ageing (TMA). The fourth element, which was not identified in the planning document, relates to the technical readiness level (TRL) and full transition of the new method into existing operations that result in a smooth end-to-end pipeline with essential quality control processes that generate consistent and reliable data. This chapter describes the end-to-end integration of FT-NIR spectroscopy methods for age estimation envisioned within the production ageing framework at the AFSC's Age and Growth Program (AGP).

THE UNIFIED SYSTEM

The AFSC conducts production ageing on over 22 species of fish in Alaskan waters, generating on average between 30,000 and 40,000 ages per year for inclusion in stock assessments[1]. Technical readiness and transition of FT-NIR spectroscopy-based age estimation may occur a few species at a time and may not be applicable for all species. In the FT-NIR spectroscopy method, ages are generated from a mathematical relationship between otolith spectra and associated TMA ages using statistical methods such as principal component regression (PCR) or partial least squares regression (PLS) to predict age using a small set of intermediate linear latent variables (Chen and Wang 2001). For species in which the TMA precision is very low, FT-NIR spectroscopy may simply not be possible without propagating excess error into the model age predictions. The impact of ageing imprecision and bias on FT-NIR-based age predictions is currently being investigated using simulation analysis (See Arrington et al. this volume-b). Walleye pollock and Pacific cod account for approximately 50% of the AGP annual production demand. FT-NIR spectroscopy models for EBS walleye pollock and Pacific cod, which have well developed age-structured stock assessment models and reasonable TMA ageing precision, have shown promising results (Helser et al. 2019b, Healy et al. 2021, Benson et al. 2023) and most likely will represent the first stocks to reach a TRL for transition. Moreover, we have advanced from using typical chemometric modeling techniques such as PCR or PLS to deep machine learning, specifically multimodal convolutional neural networks (MMCNN), to improve age prediction for these species (Benson et al. 2023). Deep

[1] G. Lambert et al., unpublished report. Importance of age data collection for stock assessments: A U.S. national perspective. Report from the Otolith Sample Size Working Group. NOAA Fisheries.

machine learning provides a number of advantages over classical linear methods by dealing with system nonlinearity, automatic feature extraction from spectra and metadata, and error quantification when used within a Bayesian framework. Regardless of which analytical tool is used to predict ages, what is clear is that both TMA and FT-NIR spectroscopy age estimation approaches have to operate in parallel in a unified process. The current TMA production process needs to adapt FT-NIR spectroscopy for species that have reached a sufficient TRL for transition because both, to a greater or lesser extent, rely on reference ages (TMA) to evaluate proper system functioning. For FT-NIR spectroscopy, reference ages are needed for evaluating predictive model performance and for model updating, if necessary. Model updating is another area of research that is being investigated by scientists at the AFSC. Model updating or recalibration may be required in situations where there is unseen variability in the spectral or reference data or when the system response function changes due to instrumentation (Wise and Roginski 2015).

Below we describe an augmented process for integrating the FT-NIR spectroscopy ageing method into the existing AGP workflow system. This chapter uses EBS walleye pollock as an illustrative example and draws upon a number of other chapters in this volume to illustrate an emerging vision of transitioning this species into an end-to-end parallel process within an existing fish ageing enterprise. The scenario presented in this paper is based on an already validated age prediction model for walleye pollock using over 9,000 otolith spectra and TMA ages from 2014-2018 EBS bottom trawl surveys. This MMCNN model was published by Benson et al. (2023) and will not be described in detail here. Rather, our intent is to predict new unseen ages from this base model and otolith scans collected during the 2019 and 2021 EBS bottom

trawl surveys and highlight the process flow, data pipelines as well as data quality control measures.

We illustrate a simplified schematic of the process (Fig. 21.1) through which the 2019 and 2021 pollock sample otoliths and data streams flow. The large rectangle on the left side of the schematic, identified as microscopic ageing in which the goal is to maintain consistency in the reference data, represents the established TMA process in the AGP (more details of the AGP's database design, and how FT-NIR data will be incorporated into the workflow, can be found in Short this volume). The TMA process is discussed in Section 1 below, with further details found in Matta et al. (this volume-a). The entirely new parallel process of FT-NIR spectroscopy age prediction is represented by diamonds and the workflow represented by arrows on the right side of Figure 21.1. The associated QC tools applied at each stage to evaluate the reliability of spectral data collection are described in Section 2 below, with further details in Goldstein et al. (this volume-a). Following the solid arrows, the data stream enters the FT-NIR spectroscopy-based model performance evaluation component of the process. This is described in Section 3 below. Under acceptable performance (solid arrow), age predictions move to final reporting, library storage, and sharing with end users. The dotted arrows illustrate the alternative flow when new variability in the data does not conform to expected model performance and there is a need for model updating. Some model updating strategies are presented and discussed in Section 4 below.

1. TMA Age Estimation and Data Quality Control

The traditional age determination process within the AGP, including quality control (QC) procedures, is well established (Kimura and Anderl 2005, Matta and Kimura 2012, Matta et al.

this volume-a). Walleye pollock age determination is based on mixed methods, primarily consisting of surface ageing (generally younger fish with clearer annuli) and sectioning-and-toasting (Matta and Kimura 2012). TMA QC metrics used by the AGP are illustrated for the 2014 to 2018 walleye pollock age data from which the MMCNN model was developed (Fig. 21.2). Total collection size was over 9,000 analyzed otoliths, with slightly more than 2,100 double reads to give a percent agreement (±0 years) of about 67%, average percent error (APE) of 3.05%, and Chang's coefficient of variation (CV) of 4.32%. Relative bias between age readers was low (not significant based on tests of symmetry) and a robust regression of reader on tester age gave $r^2 = 0.9$ and RMSE = 1.0 years. This indicates that 67% or 95% of ages from the read age were within 1 or 2 years of the test age, respectively.

When applying the published age determination criteria (Matta and Kimura 2012) for walleye pollock and conducting the usual 20% random subsample for re-analysis, QC metrics of precision for double-read TMA data from the 2019 and 2021 EBS BTS showed similar results, except that the samples consisted of fewer older fish than compared to the 2014-2018 sample (Matta et al. this volume-a). TMA double reads in 2021 showed a slightly higher CV than expected, with a slight relative bias. It should be noted that while existing TMA procedures use a 20% random subsample, efforts are underway to evaluate whether a lower percentage of double reads, such as 10%, could be used to evaluate reference age consistency in the parallel FT-NIR spectroscopy-based ageing process. Crude estimates of walleye pollock TMA ageing efficiency derived from empirical data found that approximately 35 ages could be produced per person per day[1], although these are likely underestimates of productivity. Enhancements in the AGP AGE3 database will allow better tracking of time-flow statistics for future evaluation. With better time-flow statistics for both the TMA and FT-NIR spectroscopy ageing processes, we hope to conduct

a thorough cost-benefit analysis to evaluate new technology utilization, efficiency, and its effect on program resources.

2. FT-NIR Spectral Data Acquisition and Quality Control

Quality of spectral data collection from otoliths begins with a well-designed relational database that has a user-friendly interface and an architecture that collects and stores unique spectral files, executes QA/QC tools to evaluate spectral reliability, and archives spectra to a permanent location. The AGP AGE3 database has been enhanced to accommodate the new spectral data stream and details can be found in Short (this volume). This section will describe the flow of spectral data acquisition and the QC measures that have been implemented to ensure reliability, with additional details described in Goldstein et al. (this volume-a). As shown in Figure 21.1, key personnel at this stage of the process include instrument operators (who are also production agers) and a lab manager who is responsible for quality checking each scan session and warehousing the data in a permanent library. The operator shares responsibility for the quality of spectral data because standardized sample presentation methods for each species have been developed and documented in a user manual.

The AGP FT-NIR spectroscopy laboratory currently runs two Bruker MPA II FT-NIR instruments (illustrated in Fig. 21.1). The production ageing needs of survey (RACE) and observer (FMA) and special collections are entered in a web-based application (AGPS) and are prioritized on the basis of demand and capacity of AGP resources and stock assessment priorities. Similarly to the age reading sheets created during the TMA ageing process, a data session for otolith scanning is created by the database manager to identify specimens to be scanned during that session. From a typical annual RACE EBS bottom trawl survey collection

for walleye pollock, there may be approximately 1,500 to 1,600 otoliths, so the data manager will create six scan sessions with approximately 280 otoliths (two boxes) each. An operator will select a session, physically locate the associated otolith boxes, open the session form on the AGE3 interface, and begin the scanning process (Short this volume). The scan session interface is pre-populated with specimen number or barcodes as well as metadata associated with the particular biological structure (otolith, etc.), including vessel, region, fish length, and other fields related to the specimen. Several QC tools have been built into the complete data pipeline process. The first, real-time and integral to the AGE3 process, is that the operator can compare the recorded fish length with that predicted from a large-sample PLS model that relates otolith spectra to fish length after each scan. Large departures between observed and predicted fish length can aid in identifying scan anomalies, including misplacement on the instrument sample window, incorrect sample identification number, or other mistakes that might otherwise go undetected. Additional QC tools have been developed to monitor scan data reliability, including *i*) generating absorbance profiles for each otolith, *ii*) generating absorbance profiles at the end of the session to check for anomalies, and *iii*) running a script that displays PCA plots, Hotelling's T^2, and Q-residuals to identify gross outliers (Fig. 21.3). In Figure 21.3, multi-panel plots associated with QC tools of the scanning process are illustrated for EBS walleye pollock otolith scans using data from 2021, with details described in Goldstein et al. (this volume-a). Wise and Roginski (2015) provide a thorough discussion of these diagnostic tools for spectral data. Once final data checking is complete, the lab manager executes a script to move all spectral files to a permanent location. While we are still evaluating the time-flow statistics for the complete scan process, spectral data acquisition yields an efficiency of about 40-60 scans per hour per instrument, but this can vary by operator and species.

3. FT-NIR Spectroscopy Model Development and Performance Evaluation

The next step in the process is where the data analyst employs a predictive model that has been trained and tested (sometimes called calibrated and validated) to predict fish ages from future otolith spectra. The process for developing an FT-NIR spectroscopy predictive model is highlighted in numerous chapters in this volume; however, the specific methodological approach using MMCNN pertaining to this discussion can be found in Benson et al. (2023). The data analyst examines the quality of the spectral data yet again, often looking for outliers that may exert undue leverage. Codes related to the condition of the otolith (whether broken, cracked, crystalized, etc.) during the scanning process can sometimes be identified as the reason for extreme predictions. As previously mentioned, our scenario is based on a model already developed by Benson et al. (2023) using MMCNN from otolith spectral data, reference TMA ages, and meta-data from 2014-2018. Results from their analysis yielded r^2 values of 0.92 and 0.91 and RMSE values of 0.83 and 0.91 years for training and test data, respectively (Fig. 21.4). Good model performance can be seen showing minimal differences between predicted ages and TMA ages, with comparable precision as illustrated in agreement plots between the two methods. Age predictions for 2019 and 2021 are shown in Figure 21.5, including some diagnostic plots that illustrate relative differences in FT-NIR and TMA age predictions. Quantities of relative bias for TMA ages were calculated as (read age – test age) and for FT-NIR as (NIR age – final age). For most ages, the relative bias between methods is consistently near the zero-reference line for the expected predictions of the base model. The model performance measures using FT-NIR spectroscopy are on par with those expected from independent double-read data from the TMA approach (r^2 = 0.90 and RMSE = 1.0). Interpretation of this

performance indicates that 67% or 95% of ages from the FT-NIR spectroscopy age are within 1 or 2 years of the final TMA age, respectively.

In order to evaluate the performance of this MMCNN model for pollock age predictions from future samples, in particular for otolith scans collected in 2019 and 2021, we proposed using Bayesian uncertainty quantification or Bayesian dropout for deep machine learning (Gal and Ghahramani 2016, Gal et al. 2022, Herlau et al. 2022). In concept, it operates by stochastically omitting neurons within the network and can be interpreted as a form of Bayesian model averaging (Herlau et al. 2022). This approach has emerged as a powerful tool within the machine learning field to provide optimal inference under model misspecification. Specifically, we developed 95% marginal posterior predictive internals about the base MMCNN model as the criteria to evaluate future model performance. Figure 21.4 (bottom panel) illustrates the Bayesian dropout approximation for 95% marginal posterior intervals along with the predicted walleye pollock ages from the same 2014-2018 data set. The jagged nature of the posterior interval along the traditional age domain axis is what would be expected from averaging all posterior predictive densities for each data point. Of course, posterior predictive intervals would appear much smoother with age under infinite domain intervals but instead are constrained by integers. For evaluation of model performance, we simply calculated the probability that the observations (age predictions) from future samples in 2019 and 2021 were outside the predictive interval at the 5% level, which would be expected from random chance, or the *prob* $(2.5\% < \hat{a} > 97.5\%) \leq 5\%$. The calculated probability based on the criteria suggests that predicted ages for both 2019 and 2021 are less than 5% (Fig. 21.6). Under this scenario, the model is performing as expected from the null hypothesis and no action to consider modeling updating would be needed. We have not explored other predictive model performance procedures, but consistent with our current

MMCNN modeling approach, the Bayesian dropout technique is a natural extension. Moreover, the flexibility of MMCNN models may provide a natural means of updating existing predictive models under unseen variability and poor model performance, which is discussed in the next section.

4. Model Updating Strategies

When introducing an innovative technology into an existing process control system, the ultimate goals are data stream reliability and efficiency. In this context, we have described a scenario where the TRL of FT-NIR spectroscopy fish ageing has advanced to a point that warrants a serious consideration of its integration within the existing TMA production ageing enterprise, along with potential modification for optimal coadaptation. Within this system, any action or intervention in the process creates an associated cost to efficiency, aimed at maintaining data reliability. As discussed in the previous section, model performance for predicting 2019 and 2021 walleye pollock ages was deemed satisfactory, obviating the need for model updating. However, should model updating be required, we will outline general steps that could be taken, along with the reasons for such actions. Additionally, we will address outstanding questions that can guide future research to establish best practices. Barring the need for recalibration due to instrument standardization issues, which could be addressed using more prescriptive measures like transfer models (Wise and Roginski 2015), the central questions revolve around the nature of the dataset used for model updating. In essence, determining the number of samples needed and the selection process for expanding the calibration space to encompass new variability within the system become crucial. Although there are currently no

definitive answers for applications in fish age estimation, ongoing simulation studies that describe the general framework are currently underway (Arrington et al. this volume-a).

Considering the schematic in Figure 21.1, the dotted arrows indicate situations where modeling updating becomes necessary due to new unseen data variability. In such cases, both otolith spectra and associated TMA (reference) ages will be needed to augment the calibration data set. While otolith spectra will routinely and automatically be collected within the outlined framework, reference ages associated introduces an additional cost. Several key considerations are discussed below:

1. TMA methods prescribe a 20% subsample from a collection for double reading. Can this be reduced to 10% to maintain consistency and measure precision in the TMA and be applied to FT-NIR model updating when needed?
2. Sample size requirements for the model updating dataset may be affected by the inherent ageing precision level of the species. For instance, if model predictions are poor and ageing precision is low, how do we disentangle prediction error from new spectral variation versus poor reference age estimates?
3. The sample selection method could necessarily be intertwined with sample size (10% up to 20%). Methods that select samples based on maximizing spectral variability, such as Kennard-Stone, may require fewer samples. However, this approach may not represent a "random" or "representative" subsample of the population.
4. The sample size and sample selection method of the updating dataset may also be dictated by the predictive algorithm or model used to calibrate otolith spectra to fish age.

In deep machine learning, techniques called “transfer learning” can be used to fine-tune the original model based on a new dataset (Alzubaidi et al. 2021).

CONCLUSIONS

In this paper, and volume as a whole, we have described an innovative machine-based technology that could perhaps transform the fish ageing enterprise within NOAA Fisheries. Taken together, FT-NIR spectroscopy coupled with MMCNN represents a machine learning (ML) system. Deployment of ML systems must follow a well-defined, principled process that embodies data processing, training, testing, and model deployment within the production process that it is to replace, or adapt into. Lavin et al. (2022) described the framework for TRL for machine learning systems that is far more detailed than provided here. Nevertheless, several key ideas for a ML TRL framework relevant to our work include: 1) ML system generality, 2) nonlinearity, 3) ML-specific failure modes, and 4) probabilistic ML systems.

Robust research and development that leads to broad scale applicability of technological system development for the domain of interest satisfies the proof of concept stage in the ML system lifecycle framework. FT-NIR spectroscopy in other domains, as mentioned earlier, is a well-developed technology that obviates the need for the earliest stages of life cycle development, such as greenfield research or proof of principle development. In fact, early research in Australia (Robins et al. 2015) and the United States (Helser et al. 2019a) have already shown promising results that to some extent satisfy the proof of concept development stage for applications in fish ageing. According to Lavin et al. (2022), the proof of concept stage is the entry point in the ML system R&D framework of most organizations, and this is what our SI

framework refers to more broadly as application development. Within these proceedings, we have shown broad-scale applicability of the technology and ML system in real-world scenarios with robust, verifiable data. In general, sufficient confidence in the technology must be achieved with real-world data of the case-use within the application context to move forward in TRL.

Most ML system workflows use prescriptive compartmentalized, linear processes through the various stages of ML lifecycle development. Instead, a robust workflow should be thought of as nonlinear, allowing "discovery switchbacks" that iterate over previous, more basic, developmental stages as knowledge of the technology improves. This has been our experience moving from FT-NIR spectroscopy application development to application implementation and back again. We found "out of the box" PLS models to be successful in modeling otolith spectra to TMA ages for a variety of species within different regions, yet some systemic issues in model performance were rectified by revisiting basic model concepts in deep machine learning. In another instance, engineering modifications were needed to augment FT-NIR spectrometer sample presentation to achieve improved performance (signal-to-noise) for a number of species and life stages where otolith size was too small for the sample window diameter. These are just a few of the many discovery switchbacks from our experience.

Prior to technology deployment, failure modes must carefully be addressed both in the productization of the system and ML-specific algorithms for future predictions. Adhering to strict QC standards will be key to deploying the ML system in the context of the larger process. Here we have shown tools and metrics to measure the quality of data streams from both reference and spectral data, even before being ingested downstream into the ML system. Moreover, ML-specific failure modes need to be addressed as early on as possible prior to deployment and on a continuing basis once deployment occurs. Failure points may include

nonlinearity in predictions from otolith spectra and reference data, unseen variability in reference data or spectra from biological/environmental interactions, or domain shifts that might mimic population changes. The key is to have a tool embedded in the system to monitor systemic or dynamic changes and continually evaluate model performance and intervene when needed to make corrections or updates.

The probabilistic ML system should quantify the uncertainties of components and propagate them through the system. We are implementing simulation studies to quantify the broadest axis of uncertainty associated with the various elements within the system. Probabilistic ML methods, rooted in Bayesian probability theory, provide a principled approach to representing and manipulating uncertainty about models and predictions. As discussed, we have implemented the Bayesian dropout method for the EBS walleye pollock case study, which addresses uncertainty around model structure. However, methods to account and quantify uncertainty in data, system fault points, and unseen future variability is still an active research area. Component uncertainties must be assembled in a principled way to yield a meaningful measure of overall system uncertainty, based on safe decisions and best practices.

Reaching operational readiness of a system comprising innovative technological hardware and machine learning algorithms has been discussed as we have envisioned FT-NIR spectroscopy age prediction at the AFSC. The simple framework we started with did not follow the detailed steps that are found in other application domains such as engineering and aerospace, but in concept follows the R&D procession needed in application development, data training, data testing, quality control monitoring, and system integration, if not using different terminology. Nevertheless, the key ideas and objectives are the same for robust technologies and

ML systems that strive to integrate or co-adapt innovations into existing systems of data pipelines, software, hardware, database architecture, and human interactions.

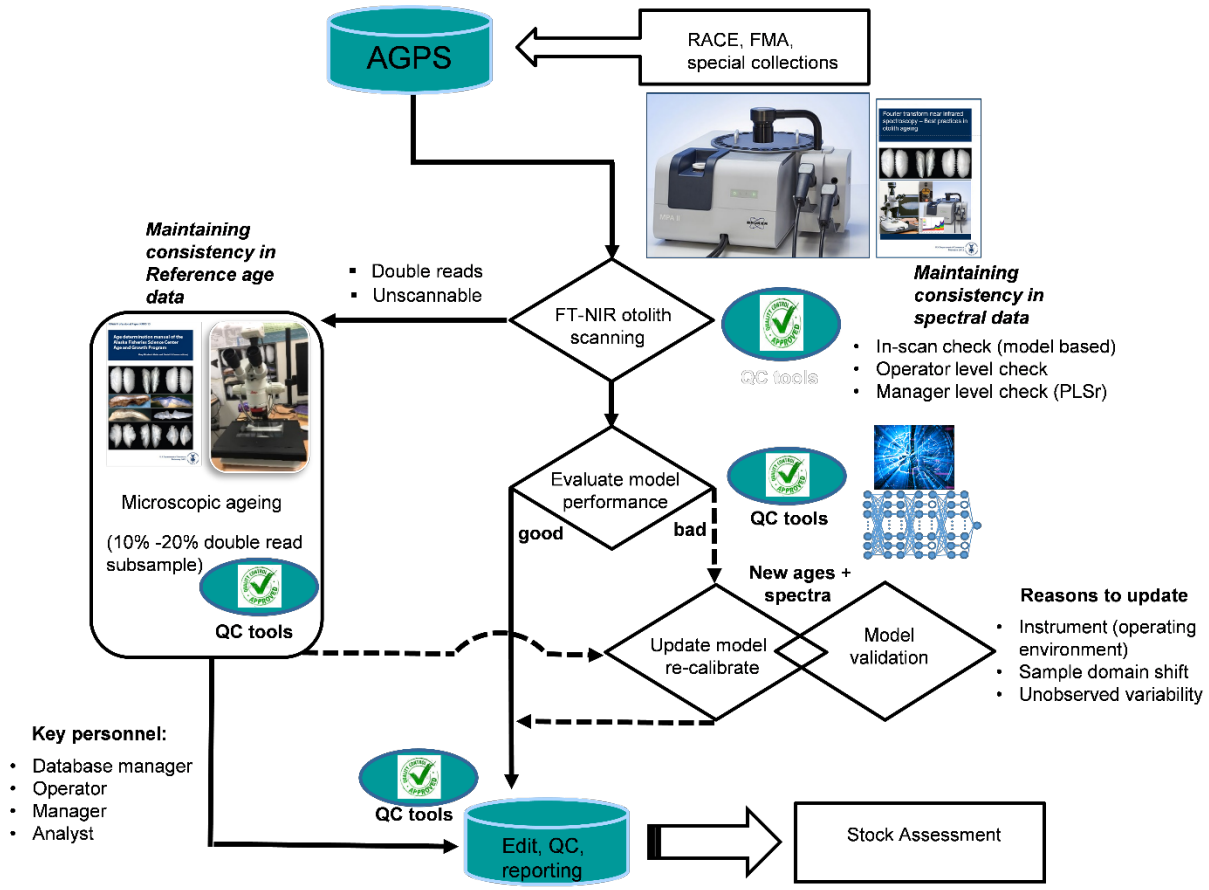
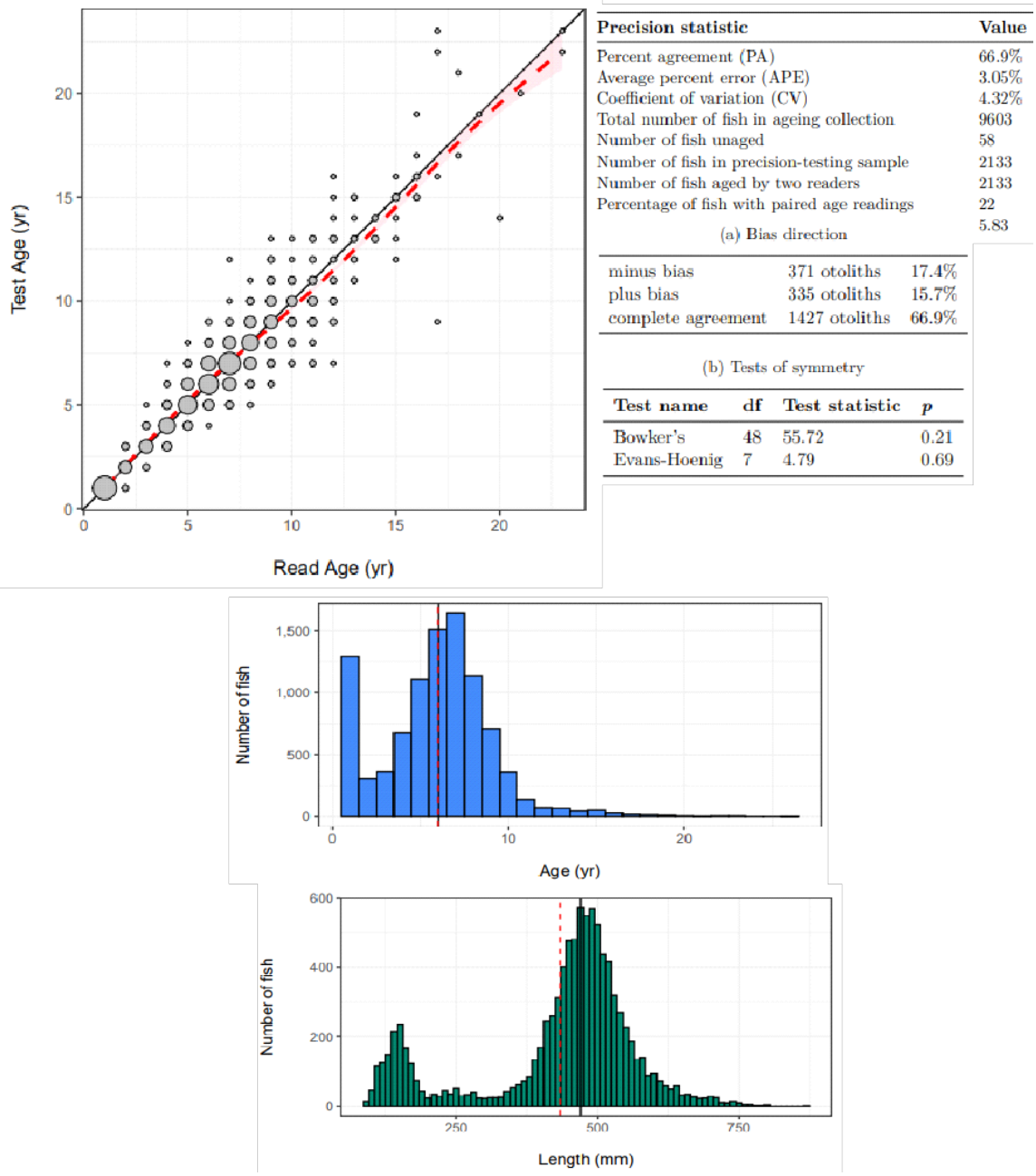


Figure 21.1. -- Stylized schematic illustrating the proposed adaptation of Fourier transform near infrared (FT-NIR) spectroscopy fish ageing process at the Alaska Fisheries Science Center's Age and Growth Laboratory.



Precision statistic	Value
Percent agreement (PA)	66.9%
Average percent error (APE)	3.05%
Coefficient of variation (CV)	4.32%
Total number of fish in ageing collection	9603
Number of fish unaged	58
Number of fish in precision-testing sample	2133
Number of fish aged by two readers	2133
Percentage of fish with paired age readings	22
	5.83

(a) Bias direction

minus bias	371 otoliths	17.4%
plus bias	335 otoliths	15.7%
complete agreement	1427 otoliths	66.9%

(b) Tests of symmetry

Test name	df	Test statistic	*p*
Bowker's	48	55.72	0.21
Evans-Hoenig	7	4.79	0.69

Figure 21.2. -- Selected tables and graphs from the Alaska Fisheries Science Center Age and Growth Laboratory's traditional method of ageing data quality control report.

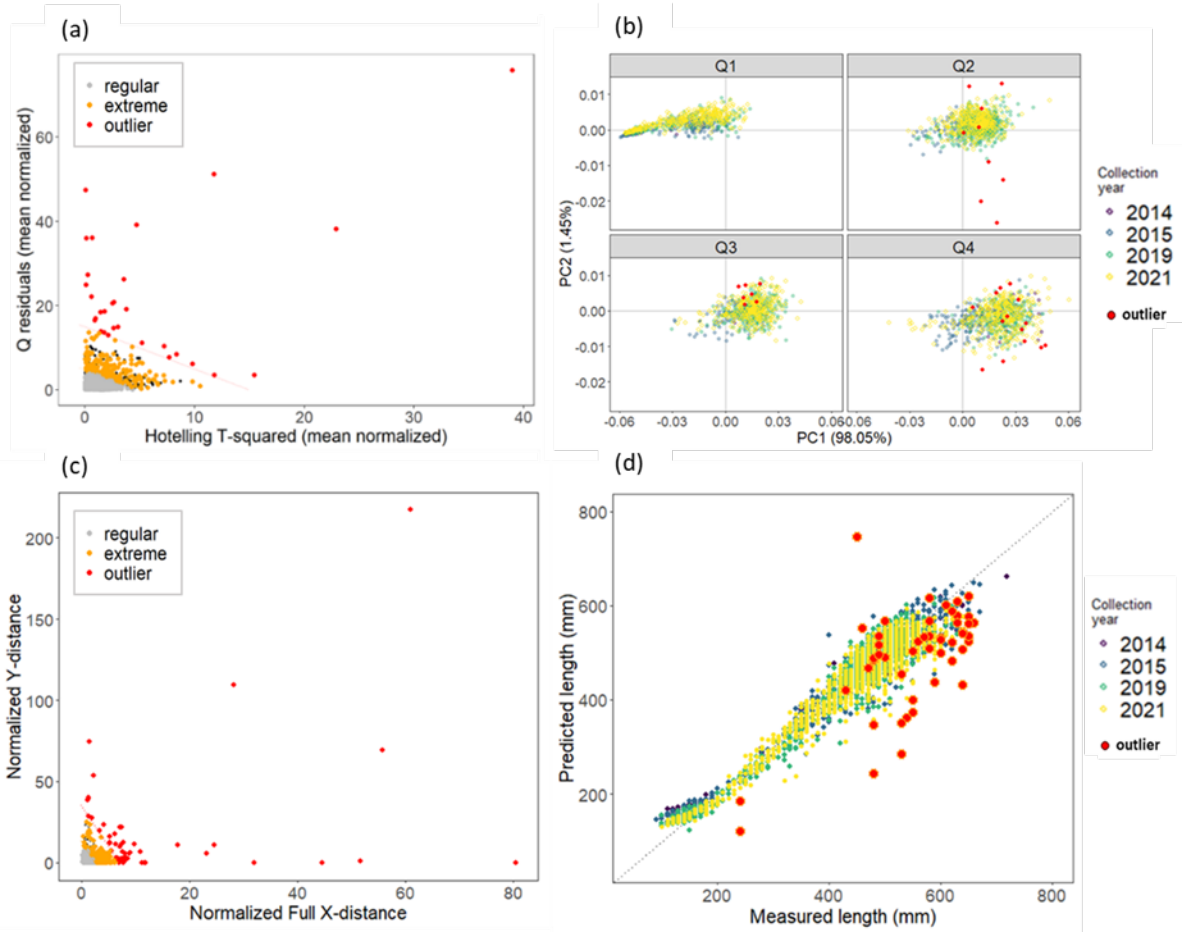


Figure 21.3. -- Principal component analysis extreme values and outliers for the 2021 case study using pre-processed data showing thresholds based on (a,b) Hotelling T^2 (T-squared) and Q residuals. (c, d) Extreme and outlier spectra identified by partial least squares models with fish length prediction based on Hotelling T^2, Q residuals, and Y-distance (fish body length observations compared to predictions. See Goldstein et al. (this volume-a) for details.

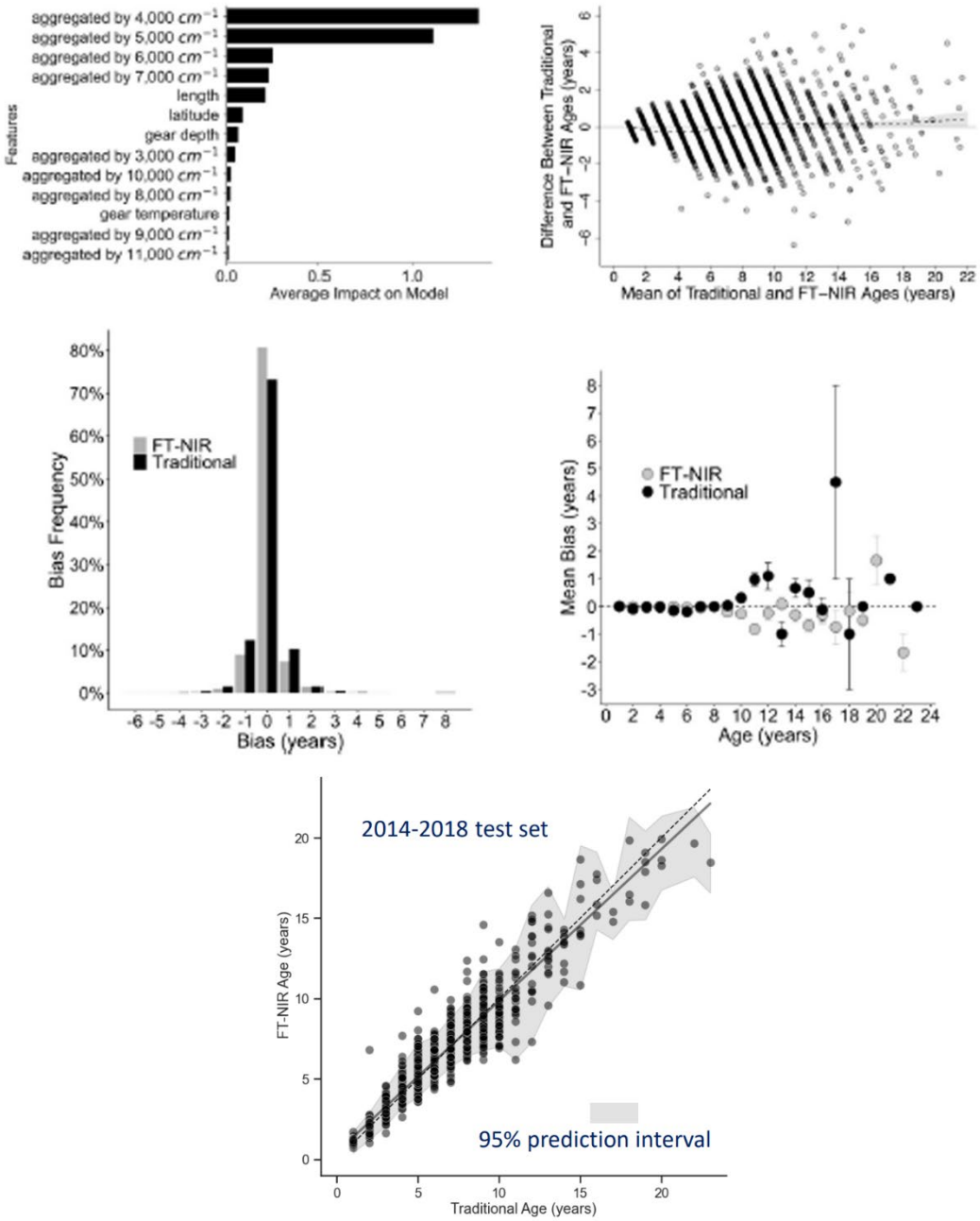


Figure 21.4. -- Results from multimodal convolutional neural network model applied to 2014-2018 eastern Bering Sea shelf bottom trawl survey walleye pollock (*Gadus chalcogrammus*) otolith Fourier transform near infrared (FT-NIR) spectra and associated biological and geospatial data. Over 9,000 walleye pollock otoliths were analyzed using FT-NIR spectrometers (Bruker Tango-R and Multipurpose Analyzer II) to generate this base model.

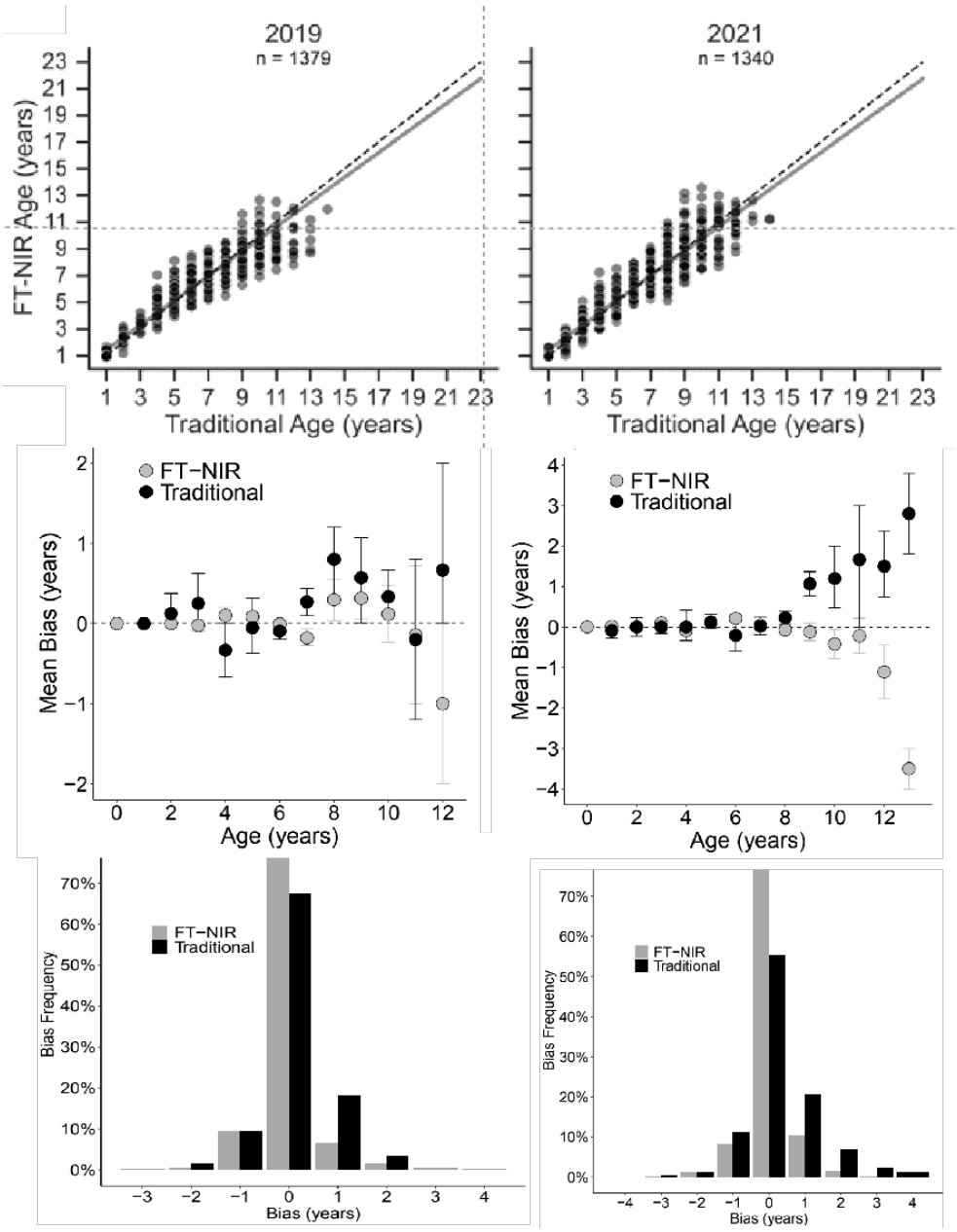


Figure 21.5. -- Results of applying the multimodal convolutional neural network model (Base Model) to walleye pollock (*Gadus chalcogrammus*) otolith spectra and associated metadata collected during the 2019 and 2021 eastern Bering Sea shelf surveys.

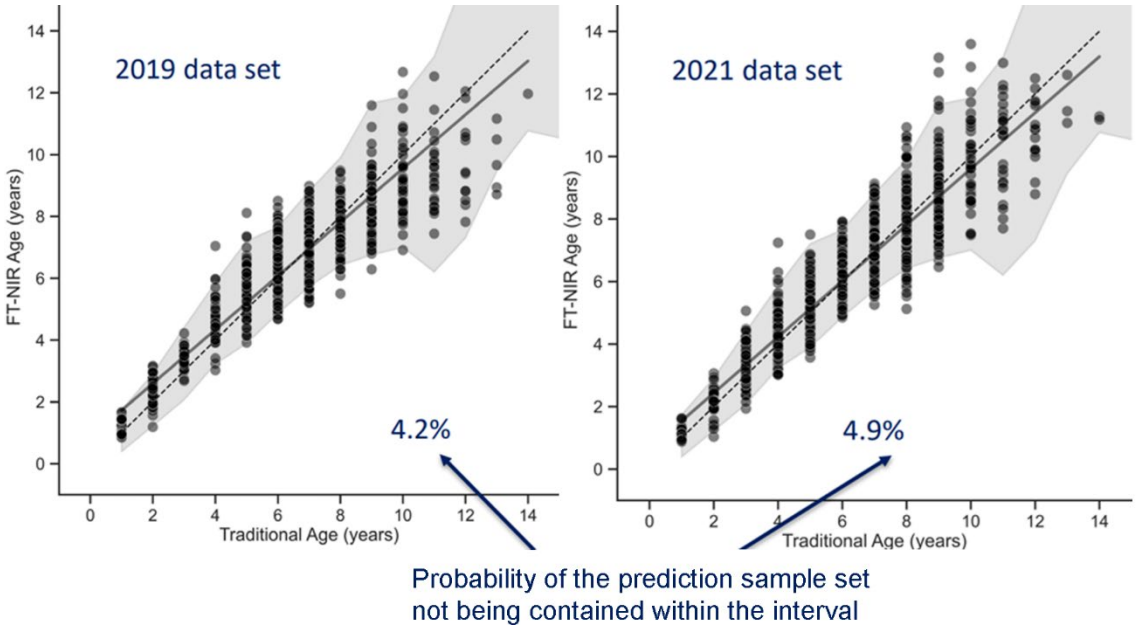


Figure 21.6. -- Age prediction of walleye pollock (*Gadus chalcogrammus*) collected during the 2019 and 2021 eastern Bering Sea shelf surveys. Predictions are from the multimodal convolutional neural network (MMCNN) Base Model using otolith Fourier transform near infrared (FT-NIR) spectra and other biological and geospatial data and plotted against traditional age estimates. Shadowed region is the MMCNN "Base Model" 95% credible interval from the Bayesian dropout method, to which the 2019 and 2021 predictions are compared to evaluate predictive model performance.

22. A Simulation Framework to Examine the Effect of Ageing Error on FT-NIR Model-Based Age Predictions

Morgan B. Arrington[1,3], Thomas E. Helser[3], Irina M. Benson[3], Mary Elizabeth Matta[3], Esther D. Goldstein[3], and André E. Punt[2]

[1] Cooperative Institute for Climate, Ocean, and Ecosystem Studies (CICOES)
University of Washington
3737 Brooklyn Ave NE
Seattle, WA USA

[2] School of Aquatic and Fishery Sciences
University of Washington
1122 NE Boat Street
Seattle, WA USA

[3] Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

INTRODUCTION

Fourier transform near infrared (FT-NIR) spectroscopy is an emerging technology in the field of fish ageing. It has shown promise for estimating fish ages from biological structures including otoliths, vertebrae, and spines from a variety of species more rapidly than traditional, microscope-based methods (Rigby et al. 2014, Wedding et al. 2014, Robins et al. 2015, Rigby et al. 2016; Helser et al. 2019b, Passerotti et al. 2020a, Passerotti et al. 2020b, Arrington et al. 2022, Healy et al. 2021, Benson et al. 2023). Fish ages, an important data type used to better understand fish ecology and population dynamics, are a critical component of integrated stock assessments to inform fisheries management (Maunder and Punt 2013, Methot and Wetzel 2013). Age estimation using FT-NIR spectroscopy may allow for efficiencies in time and monetary expenditure, improved spatial and temporal data collection for some species, and improved reproducibility relative to traditional methods.

The ability to estimate fish age using FT-NIR spectroscopy depends upon calibrating a predictive model between fish age and the near infrared (NIR) spectrum of an ageing structure (e.g., otolith). Each NIR spectrum captures a unique absorbance signature across a range of wavenumbers when an otolith is irradiated with NIR light. A spectrum is determined by the type and quantity of organic bonds present in a material and can also be affected by the material's structure (Siesler et al. 2008). The molecular or structural basis for NIR otolith spectra correlating with fish age is still not fully understood. Passerotti et al. (2022) found that it is likely a complex combination of physical and chemical properties in otoliths that vary with age that are detectable by NIR light, including otolith size and thickness, structure, and the main constituents of calcium carbonate and protein.

There are currently two sources of uncertainty in predictive models when utilizing the FT-NIR spectroscopy approach for age prediction: spectral variability at age and observation error in reference values. Variability in otolith spectra-at-age is ubiquitous among species. This may be, in part, because the composition and structure of otoliths are influenced by numerous other biological and environmental factors in addition to age (Chang and Geffen 2013). Indeed, otolith spectra have also been used to discriminate among species and ecosystems (Benson et al. 2020). Other factors such as storage medium and storage time may also contribute variability to spectra (Robins et al. 2015, B. Hsieh, AFSC, pers. comm.). Additionally, variability in spectra-at-age may be driven by some otolith spectra getting misassigned to inaccurate age classes. When the FT-NIR method is used for age prediction, traditional estimates of animal age are used as the response variable (henceforth referred to as the reference value) for model calibration, since the true age of fishes is often not known. Traditional age estimation is done by an analyst counting growth increments on a biological structure. There is often ambiguity in the interpretation of species' ageing structures that results in inaccurate age estimates, henceforth referred to as observation error.

Statistical models assume variance between the model and the data (here, the modeled spectra-at-age and observed spectra-at-age). However, they do not often account for measurement error (here, observation error in the age estimate) (Buonaccorsi 1996). There are established methods to account for observation error when age estimates are used as a regressor (Punt et al. 2008, Cope and Punt 2007), but to our knowledge, no methods yet exist to account for observation error when age is used as a reference value, as is the case in the FT-NIR spectroscopy approach.

The FT-NIR ageing method shows promise for age prediction, but the application of the approach is still limited by our lack of understanding of how sources of uncertainty in predictive models impact age predictions. Specifically, it is unknown whether variability in spectra-at-age may lead to greater prediction uncertainty (Helser et al. 2019b), or whether prediction error is due to observation error in reference ages. This information is important for understanding where to focus efforts to improve the predictive accuracy of this technology to achieve technical readiness (Helser et al. this volume-a). Here, we present a preliminary study to investigate these two sources of uncertainty influencing model-based age predictions for eastern Bering Sea (EBS) Pacific cod (*Gadus macrocephalus*). A simulation approach was used to disentangle observation error from spectral variability to address the following questions:

1. How does observation error in the reference age impact model predictions?
2. How does spectral variability at age impact model predictions?

METHODS

We characterized the impact of two sources of uncertainty on age predictions from partial least squares regression (PLS) models fitted between Pacific cod age and otolith spectra: observation error in the reference value and spectral variability at age. A simulation approach was used in R version 4.3.0 (R Core Team 2023). The data simulated in this study were based on Pacific cod collected on EBS summer bottom trawl fishery-independent surveys conducted by the Alaska Fisheries Science Center between the years 2010 and 2019.

Broadly, in this approach, we simulated datasets with "true ages" (henceforth referred to as "known ages") and paired "age estimates" for four different spectral variability scenarios: low, medium, high, and empirical. Each dataset had a known age vector (n x 1), an estimated age vector (n x 1), and a matrix of spectra (n x p), where n = sample size and p = number of channels/ wavenumbers.

$Y_{i \times 1}$ = known age

$Y^*_{i \times 1}$ = estimated age

$X_{i \times p}$ = spectra

p = number of channels

i = number of individuals

We simulated 100 datasets for each spectral variability scenario (400 datasets total). A model was fitted between spectra and known age, and spectra and estimated age for each of these datasets. Predictive accuracy was compared among models to evaluate the effect of observation error in the reference values. More details on each step of this approach are included below.

Age Data Simulation

The known age vector, Y, was generated using a uniform distribution to represent theoretical fish with integer age a of 1 through 14 years. Each age class had 100 individuals (n = 1,400). This was a simplifying assumption used for this preliminary study, but future work

should explore using more biologically realistic age distributions. We simulated an age estimate for each known age "fish" by first modeling empirical ageing error for EBS Pacific cod using the method described in Punt et al. (2008). This method allows the estimation of an ageing error matrix that estimates the relationship between traditionally estimated fish age and known age. The empirical age estimates used to estimate the ageing error matrix were determined microscopically by at least two analysts counting pairs of annual bands corresponding to the age of the fish using sagittal otoliths (Matta and Kimura 2012). In the resulting ageing error matrix, each known age, *a,* has a unique probability distribution F*a* of being estimated age 0 to 14 years. This was used to simulate an estimated age Y_i* by resampling with replacement from set {0,1,2…14} for each individual Y_i based on the F*a.* The resulting matrix had a vector of known ages (Y_i) and a vector of age estimates (Y^*_i) that were based on the modeled relationship between known age and estimated age for Pacific cod (Fig. 22.1).

Spectral Data Simulation

Spectra were then simulated to allow us to introduce different levels of simulated variability with respect to age. Spectra were simulated based on empirical Pacific cod otolith spectra to retain empirical shapes and profiles. Empirical spectra were available from a subset of the EBS Pacific cod otoliths (collection years 2013-2018, n = 10,255). See Benson et al. (2020) for more information on spectral data acquisition methods. We simulated spectra based on the principle that a spectral matrix $X_{i \times p}$ can be decomposed via principal component analysis (PCA) into a score matrix $T_{i \times a}$, and loading matrix $P^T_{a \times p}$, and a residual error matrix $E_{i \times p}$ where i is the number of individuals, a is the number of components, and p is the number of wavenumber

channels (Beattie and Esmonde-White 2021) (Equation 1). These separate matrices can then also be reconstructed into spectra by multiplying the score vector by the transposed PC loading matrix (Fig. 22.2). The more PCs included in the reconstruction, the smaller the residual error and the closer the spectra will be to the original (Beattie and Esmonde-White 2021).

(1) $$X_{i \times p} = T_{i \times a} * P^{T}_{a \times p} + E_{i \times p}$$

$T_{i \times p}$ = score matrix

$P_{p \times a}$ = loadings matrix

p = number of wavenumber channels

a = number of components

i = number of individuals

First, we decomposed empirical Pacific cod otolith spectra into scores and loadings using PCA (Equation 1). Spectra had been preprocessed using Savitzky-Golay smoothing based on standard methods at the Alaska Fisheries Science Center to remove baseline shifts and prepare data for modeling (Helser et al. 2019b, Arrington et al. 2022, Healy et al. 2021, Benson et al. 2023). We standardized the scores from each principal component (PC) 1-5 to have unit variance (Fig. 22.3). The first 5 PCs were selected because they explained a prior threshold variance of greater than 99.95%. We then averaged scores among specimens of the same age to obtain an average score by age for each PC 1-5. Variability was introduced around each average age-specific score to simulate the four variability scenarios. In each scenario, we included 100 random draws per age and per PC to match the simulated age data set. Our four scenarios were:

the “real world” scenario with scores sampled from the real, empirical distributions; a low variability scenario that was informed by observations of the low end of standard deviation (SD) in the score distributions by age; a medium variability scenario; and a high variability scenario, also informed by the higher end of SD in score distributions by age. For the low, medium, and high variation scenarios, we randomly sampled from normal distributions with a mean equal to the mean score by age and a SD of 0.01 for low variability, 0.05 for medium variability, and 0.10 for high variability to simulate scores (Fig. 22.4). Principal components are uncorrelated, so we sampled from individual distributions. However, sampling from normal distributions was a simplifying assumption for this preliminary approach. True distributions were sometimes closer to bimodal (Fig. 22.3), and future work could explore using different distributions to simulate variability in scores. The scores with simulated variability were then multiplied by the original transposed loadings matrices and back-transformed to spectra (Equation 1, Fig. 22.5). This approach allowed us to avoid making any assumptions about the empirical shape of spectra that is captured and retained by using empirical loadings matrices. We then bootstrapped this step with 100 datasets per scenario, each with 100 individuals per known age 1 to 14 years (n = 1,400). Due to small sample sizes of empirical spectra over age 10 that may not have been representative, we truncated our simulated dataset at a known age of 10 years (some simulated estimated ages exceeded this).

Evaluation of Uncertainty

Finally, we fitted PLS models between simulated spectra and known age and between simulated spectra and estimated age for each spectral variability scenario. This allowed us to

evaluate the effect of observation error in reference age on model predictions at different levels of spectral variability. Predictions were rounded to the nearest integer and evaluated relative to known fish age using root mean square error (RMSE), the coefficient of determination (r^2), and age-bias plots (Campana et al. 1995) for all scenarios. Future work could explore additional methods and metrics to quantify and compare predictive accuracy.

RESULTS AND DISCUSSION

FT-NIR spectroscopy has shown great potential for predicting the age of fish from otoliths with improved efficiency, repeatability, and near-similar accuracy to traditional methods for many species including EBS Pacific cod. However, to date, age estimates have been relied upon for model calibration and evaluation, since true fish ages were unknown. The effect that observation error in reference ages may have on model predictions could not be evaluated. This study is the first to evaluate the impact of ageing error on FT-NIR spectroscopy-based age predictions from fish otoliths. The preliminary results of this study suggest that uncertainty in model-based age predictions is driven by spectral variability at age for EBS Pacific cod more so than by observation error in traditional ages, but that bias in reference ages may still have an impact. At lower levels of spectral variability, we observed a greater impact of ageing bias on the models' predictive skill. At higher levels of spectral variability, we observed a reduced impact of ageing bias on the models' predictive skill, including in the "real world", empirical scenario. At higher levels of spectral variability, nonlinearity emerged with a tendency for models to overpredict young ages and underpredict older ages.

Spectral variability at age was the greater driver of uncertainty in model-based age predictions in this simulation. At the lowest level of simulated spectral variability, predicted ages were within just one year of the known age regardless of which reference value (known age or estimated age) was used except for age-10 fish (Figs. 22.6, 22.7, 22.8, Table 22.1). At empirical levels of spectral variability, there was greater uncertainty in model-based age predictions regardless of whether known ages or estimated ages were used as the models' reference values (Figs. 22.6, 22.7, 22.8, Table 22.1). This suggests that the variability at age in empirical spectra cannot be attributed to ageing error alone. It is still unknown whether variability in spectra-at-age is largely due to heterogeneity in the molecular or structural composition of otoliths at age or affected by other factors such as otolith storage. Effects of storage time on otolith spectra have been observed for saddletail snapper *(Lutjanus malabaricus)* and barramundi *(Lates calcarifer)* (Robins et al. 2015), which may contribute to spectral variability at age, but more work is needed to explore this for Pacific cod. Storage media may also contribute to storage effects, and while snapper and barramundi otoliths were stored dry, Pacific cod otoliths included in this study were stored in glycerin thymol.

Bias in traditionally estimated reference ages also impacted age predictions in this study. The age reading error model best supported by the Akaike information criterion was a curvilinear coefficient of variation (CV) and a curvilinear bias between estimated age and known age. Pacific cod are among the most difficult Alaska groundfish species to age and commonly have strongly defined "checks" in early life that make growth patterns hard to interpret (Matta and Kimura 2012). The modeled bias for Pacific cod included in this study was most prominent for younger (under age 2 years) and older (over age 6 years) fish, which was reflected in the simulated age estimates that tended to be overestimates of known age (Fig. 22.1). At lower levels

of spectral variability, we observed a greater impact of ageing bias on the models' predictive skill with larger differences in the RMSE, r^2, and bias plots between age estimate models and known age models (Figs. 22.6, 22.7, 22.8, Table 22.1). At higher levels of spectral variation, we observed a reduced impact of ageing bias on the model's predictive skill, including in the "real world", empirical scenario (Figs. 22.6, 22.7, 22.8, Table 22.1). We also observed slightly improved r^2 values for predictions from models fitted to age estimates compared to models fitted to known ages (Fig. 22.7, Table 22.1). This was an unexpected result but was likely due to the positive bias in age estimates counteracting nonlinearity in model predictions at older ages.

At higher levels of spectral variability, we observed increased nonlinearity with a tendency for models to overpredict young ages and underpredict older ages (Figs. 22.6, 22.8). Nonlinearity was also observed in the case study applying FT-NIR spectroscopy to age Pacific cod, with a tendency to underpredict the age of older fish (Healy et al. 2021). Since the true ages of fish were not known in that study, predictions were compared only relative to age estimates. It was unknown whether the cause of the nonlinearity was age estimation error in reference values, nonlinearity inherent in the spectra, or a combination.

The presence of nonlinearity in predictions for both known age models and age estimate models in this simulation study supports that the nonlinearity in predictions is driven, at least in part, by the spectra as opposed to solely by observation error in traditionally estimated reference ages. The increase in nonlinearity with spectral variability also suggests that when more spectral variability at age is present, the effect of the nonlinearity on predictions may be exacerbated. This is to be expected given that there would likely be more overlap between spectra of subsequent age classes. This can be somewhat visualized in the distributions of scores-at-age for the four spectral variability scenarios (Fig. 22.4).

Several possible causes for nonlinearity in spectra with fish age have been suggested. There is a decline in otolith accretion with age for many fish species (Helser et al. 2012, Matta et al. 2020), and therefore the relationship between fish age and analyte concentration in otoliths may not stay consistent. There may also be a nonlinear relationship between spectra and the analyte being measured. This can be caused by a decline in light attenuation with sample thickness (Padalkar and Pleshko 2015, Passerotti et al. 2020b) and violations to the Beer-Lambert Law, where light is either scattered from the sample or absorbed by the sample (as can be the case with thick samples) instead of being transmitted (Miller et al. 1993, Lui et al. 2018, Benson et al. 2023).

An alternative modeling approach may help to resolve nonlinearity in EBS Pacific cod age predictions. Linear multivariate methods, such as the PLS models used in this study, can often deal with nonlinearity through wavenumber selection and dimensionality reduction (Miller at al. 1996). However, they are not always successful and other nonlinear modeling approaches may be better suited (Miller et al. 1993, Benson et al. 2023). Only spectra with the lowest variability at age were able to produce linear age predictions in this study using PLS models. Benson et al. (2023) found that a multimodal convolutional neural network model outperformed PLS models for predicting the age of walleye pollock (*Gadus chalcogrammus*), especially for older fish. This approach may be successful at resolving nonlinearity in predictions for Pacific cod.

It is important to note that model predictive accuracy is reduced in the empirical scenario in the present study compared to Healy et al. (2021); however, this is likely due to differing goals between studies. The goal of Healy et al. (2021) was to optimize age predictions, while the goal of the present study was to evaluate the impact of error in reference ages. Models in this present

study were not optimized for wavenumber region or number of latent variables. Instead, predictive accuracy was compared among scenarios to evaluate them relative to one another. When RMSE and r^2 are calculated relative to the age estimate instead of known age, they are further worsened (Table 22.2). This indicates that in real applications of this technology, model predictions may actually be closer to true fish ages than the age estimates being used to evaluate them.

We make several assumptions in this study that may impact our findings. We assumed a uniform distribution of known ages 1 to 10 to build our models, but future work should explore the effect of a more realistic age structure, which is an important consideration when characterizing ageing error (Nesslage et al. 2022). Since bias in age estimates relative to known age was highest for older fish and there are typically fewer older fish in real-world populations, we would expect that our RMSE values may reflect a greater impact of ageing error on model predictions than in real-world scenarios. Additionally, though we did our best to disentangle two confounded sources of error, we used empirical spectra that we averaged based on traditional age estimates (since true fish age is unknown) as the foundation of our simulation. This approach assumes an even distribution of ageing error around true age, which it would appear is not entirely true for Pacific cod (Fig. 22.1). However, the results of the low spectral variability scenario indicate that linear age predictions with good predictive accuracy are possible using our simulation approach. Finally, we added random variability to average PC scores by age for the low, medium, and high scenarios by drawing from normal distributions. This was a simplifying assumption. However, when compared to the empirical scenario for which variability was drawn from empirical distributions, the low, medium, and high scenarios seem reasonable (Fig. 22.4). The largest difference between the empirical and non-empirical scenarios is that the empirical

scenario had varying levels of spectral variability at age, while the non-empirical scenarios had constant variability across the full age range (Figs. 22.4, 22.6, 22.8). It is still unknown why spectral variability changes with age but may reflect the individual growth variability in fish populations that can potentially increase with age (Quinn and Deriso 1999, Ogle et al. 2017).

The results of this study address two sources of uncertainty in PLS models fitted between Pacific cod otolith spectra and fish age. This is the first study to separately evaluate the effects of these two sources of uncertainty on PLS model predictions for otolith spectra, and preliminary results suggest that ageing error in reference values may not impact model predictions to the same degree as spectral variability and nonlinearity with age for EBS Pacific cod. Some of the findings of this study may be applicable to other species, such as the impact of bias on model-based age predictions. But, ageing error and otolith spectra can vary by region (Robins et al. 2015, Benson et al. 2020, Passerotti et al. 2020b) and species (Benson et al. 2020), and we recommend exploring the effect of ageing error on a region- and species-specific basis. This simulation framework is flexible and can be applied to explore the effects of age estimation error and spectral variability on model-based age predictions for a range of scenarios and species. This is important for understanding the utility of the FT-NIR spectroscopy approach and where to focus efforts to improve predictive accuracy to achieve technical readiness.

ACKNOWLEDGMENTS

We thank Brenna Hsieh, Jonathan Short, and the AFSC Age and Growth Program for spectral data collection, traditional age estimation, quality control and assurance, and support with data access, and without whom this study would not have been possible. We thank the AFSC Resource Assessment and Conservation Engineering division for survey data collection. We also thank Matthew Siskey for assistance in getting started with the "Puntilizer" code, and are grateful for feedback from attendees at the FT-NIR Spectroscopy 2023 Workshop that improved this abstract.

Table 22.1. -- The root mean square error (RMSE) and r^2 values between model predictions and known age. Each spectral variability scenario is represented by 100 iterations of models fitted to age estimate reference values.

Spectral variability	Model reference age type	RMSE (years) known age	r^2 known age
empirical	known age	1.39-1.54	0.62-0.70
	age estimate	1.43-1.58	0.66-0.73
low	known age	0.20-0.27	0.99-0.99
	age estimate	0.50-0.56	0.97-0.97
medium	known age	0.94-1.06	0.84-0.88
	age estimate	1.01-1.14	0.85-0.89
high	known age	1.56-1.74	0.45-0.59
	age estimate	1.61-1.77	0.52-0.64

Table 22.2. -- The root mean square error (RMSE) and r^2 values between model predictions and estimated age. Each spectral variability scenario is represented by 100 iterations of models fitted to age estimate reference values. This represents the status quo method of model calibration and evaluation when true fish age is unknown.

Spectral variability	Model reference age type	RMSE (years) age estimate	r^2 age estimate
empirical	age estimate	1.62-1.78	0.56-0.66
low	age estimate	0.58-0.64	0.96-0.97
medium	age estimate	1.12-1.28	0.81-0.86
high	age estimate	1.80-1.98	0.40-0.54

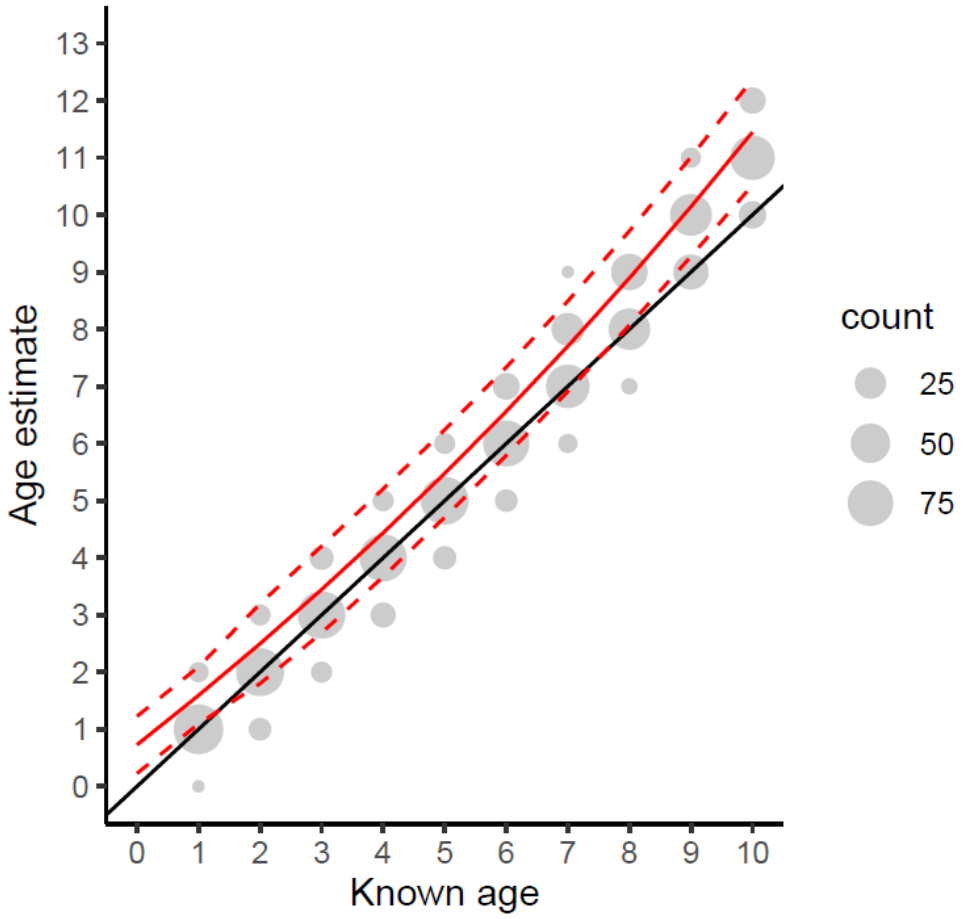


Figure 22.1. -- Simulated known age and corresponding age estimates (in gray) relative to the modeled relationship between known age and estimated age for eastern Bering Sea Pacific cod (*Gadus macrocephalus*) collected from 2010-2019 (solid red line) with 95% confidence intervals (dashed red lines). The size of the gray point reflects the number of simulated observations. The one-to-one line is shown in black.

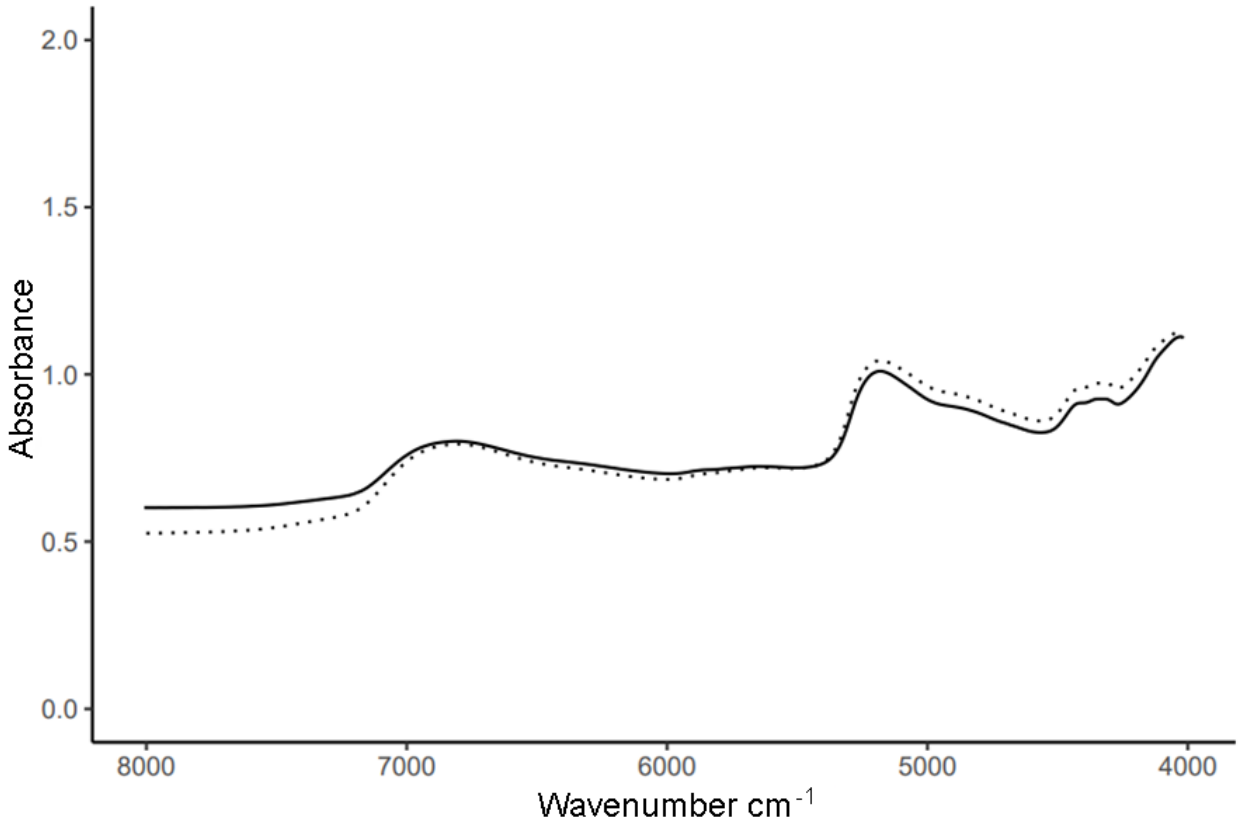


Figure 22.2. -- One empirical, raw spectrum (solid line), and the same spectrum recomposed using principal component (PC)1 scores multiplied by the transposed PC1 loading (dotted line).

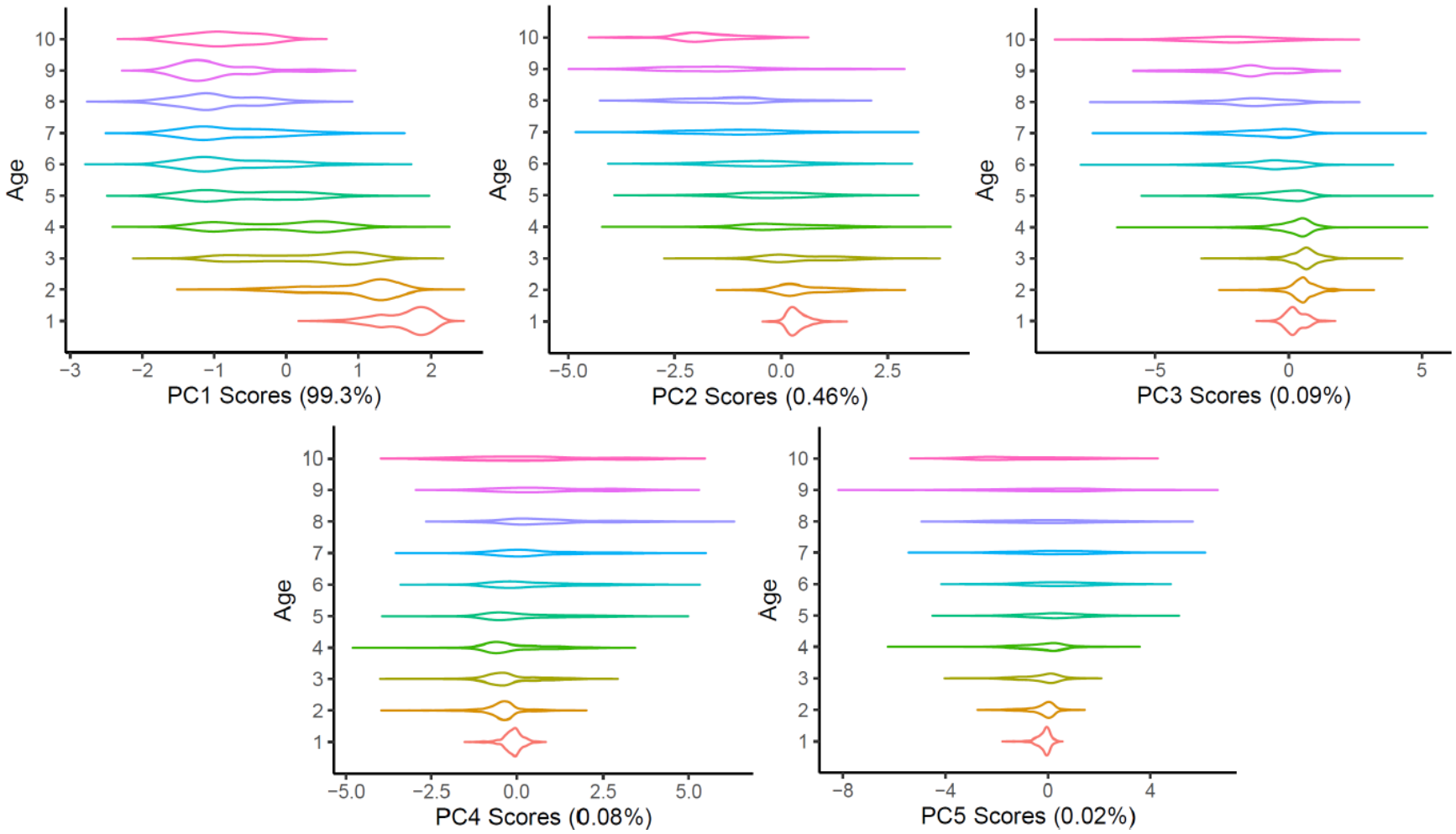


Figure 22.3. -- Kernel density plots showing empirical variability in standardized scores for principal components (PCs) 1 through 5 by empirically estimated age. This is the foundation for calculating the average score by age for each PC 1 to 5 to simulate spectra.

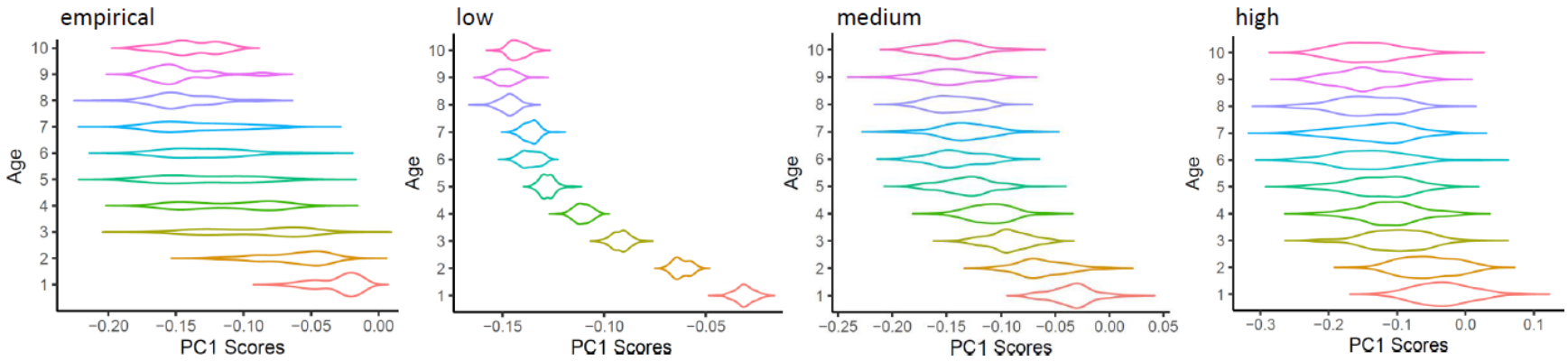


Figure 22.4. -- Simulated scores for just principal component (PC) 1 for empirical or "real world" variability, low variability (standard deviation, SD = 0.01), medium variability (SD = 0.05), and high variability (SD = 0.10). Scores were simulated for PCs 1 to 5.

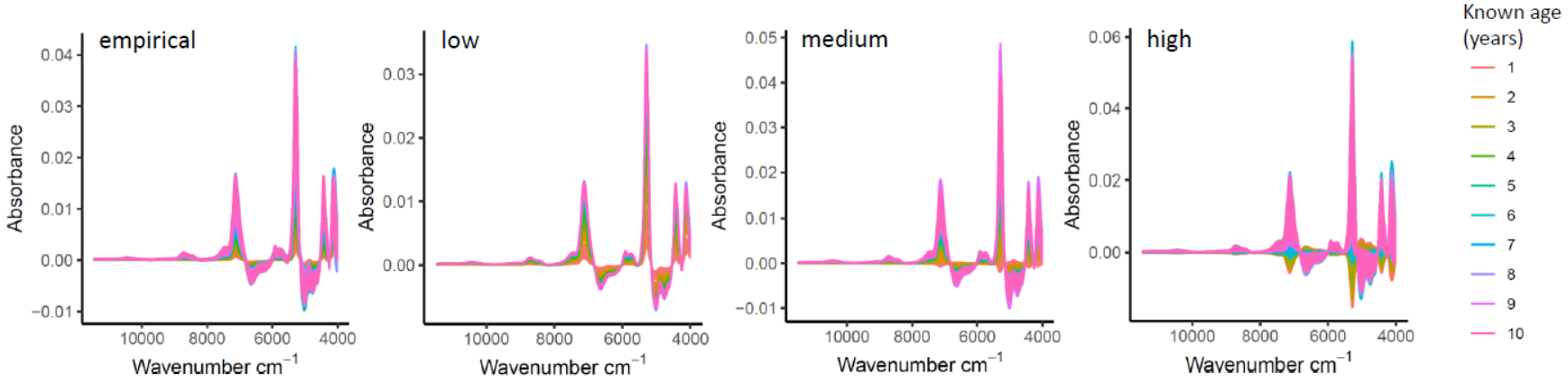


Figure 22.5. -- Simulated spectra for four spectral variation scenarios: empirical, low, medium, and high. Example of one data set per scenario; each had 100 resampled data sets.

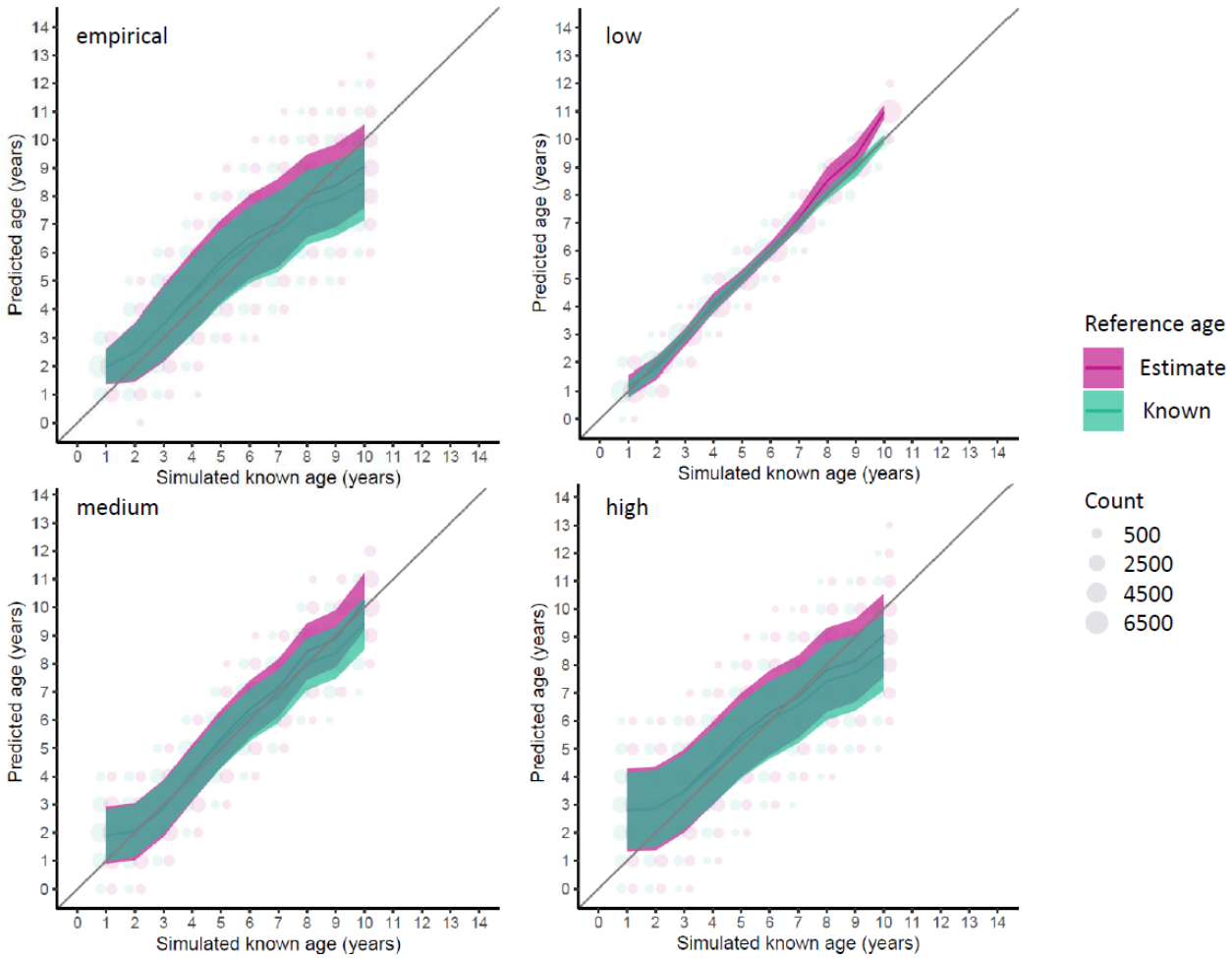


Figure 22.6. -- Predicted ages from models fitted between simulated spectra and known age (green) or age estimate (pink) plotted relative to known age. Predictions are shown as dots that scale with count and the average prediction with standard error is shown as a ribbon. This illustrates the impact of ageing error in the reference value on model predictions at four different levels of spectral variation: empirical, low, medium, and high.

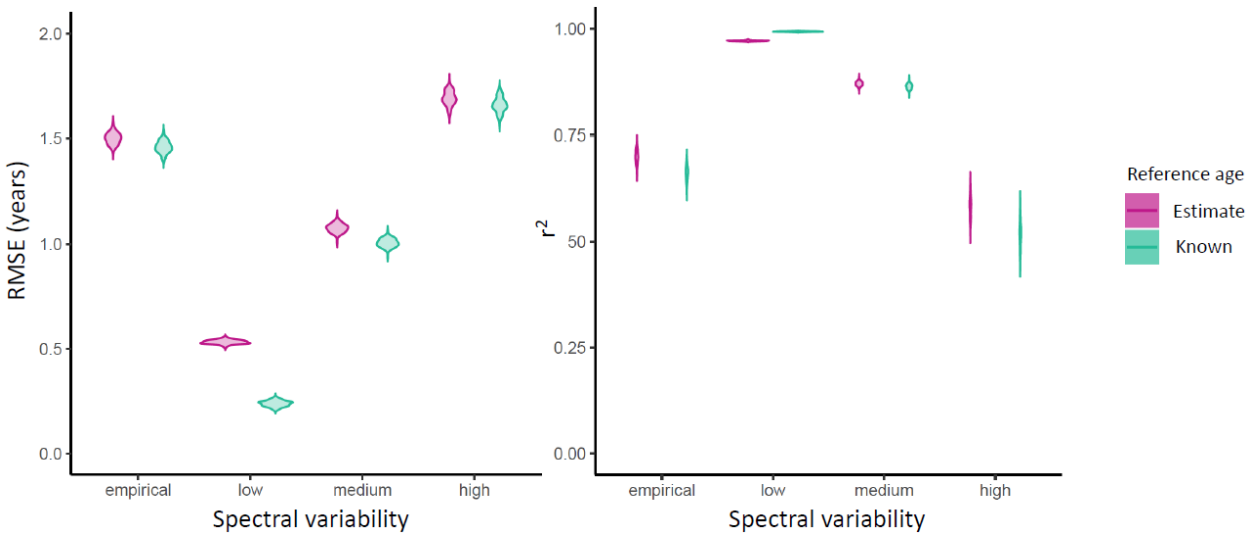


Figure 22.7. -- The root mean square error (RMSE) and r^2 values for model predictions relative to known age. Kernel density plot comparisons between RMSE and r^2 for age estimate models and known age models are shown for the four spectral variability scenarios. Ranges in values are based on 100 iterations per reference age type per scenario.

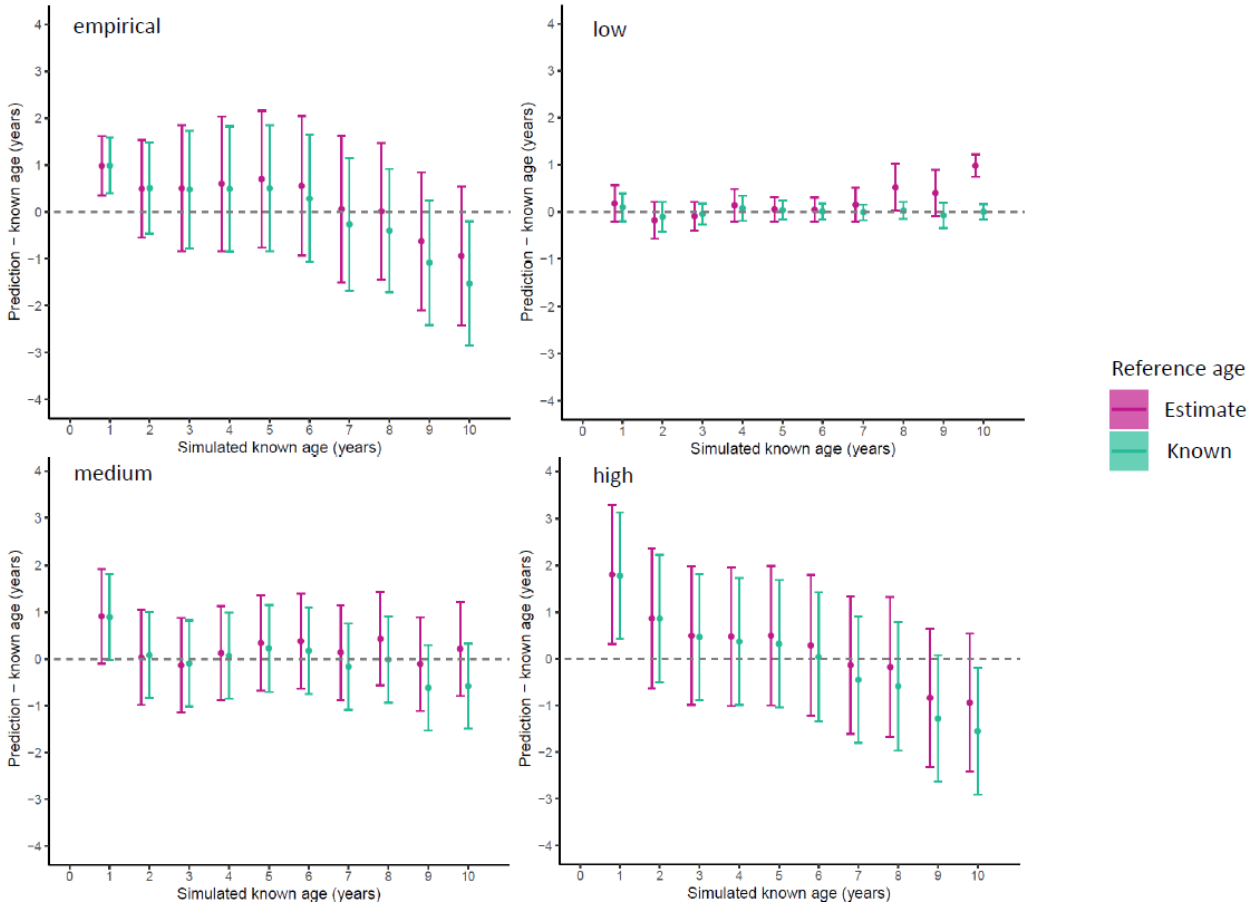


Figure 22.8. -- Prediction bias relative to known age compared among levels of spectral variation: empirical, low, medium, and high for models fitted to age estimate (pink) and known age (green).

23. Fourier Transform Near Infrared Spectroscopy Data and Application Within the Eastern Bering Sea Pollock Assessment Model

James N. Ianelli

Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

ABSTRACT

Adopting age determination error within stock assessments has been a feature applied and available in stock assessments since the early 1980s (Fournier and Archibald 1982). Findings from applying preliminary walleye pollock (*Gadus chalcogrammus*) age data using Fourier transform near infrared (FT-NIR) spectroscopy method are summarized below.

1. Modification of assessment model a step forward: The study concludes that the modification of the assessment model used in the research represents a positive step forward in accurately evaluating the fishery or population under investigation. The specific modifications made to the model may be highlighted, along with their impact on improving the accuracy or reliability of the assessment results.

2. Flexibility to use data from any source/year/gear type: The study emphasizes the importance of flexibility in utilizing data from various sources, years, and gear types. By incorporating data from diverse sources, the assessment becomes more comprehensive and robust. This flexibility allows for a broader understanding of the fishery or population dynamics and facilitates better-informed management decisions.

3. More detail and random effects application: The study suggests that incorporating more detail and applying random effects in the assessment model can lead to more accurate and nuanced results. By accounting for additional variables or factors that

influence the fishery or population, the assessment becomes more comprehensive and better represents the complex dynamics at play.

4. Preliminary results seem promising: The preliminary results obtained from the assessment are described as promising. This indicates that the initial findings suggest positive outcomes or trends regarding the health and sustainability of the fishery or population. However, further analysis and validation may be necessary to confirm the reliability of these results.

5. Age error "shape" relatively limited, application needs more careful evaluation: The study finds that the shape of age errors, which refers to the pattern or distribution of errors associated with age determination, is relatively limited. However, the summary cautions that the application of age error estimation requires more careful evaluation. This implies that while the estimation method shows some promise, its practical application needs to be thoroughly examined to ensure its accuracy and suitability for future assessments.

FT-NIR DATA AND APPLICATION COMPARISONS

For the data made available from 2014-2018 using the FT-NIR determined ages, we see that both the proportions-at-age and body mass-at-age have differences (Figs. 23.1 and 23.2). Interpretation of these patterns suggest that there may be some important age-specific patterns between the traditional age determination methods and the FT-NIR age determination approach. Some standard assessment model diagnostics (root mean square error and effective sample size) suggest that the model fit to the traditional age data is better than the fit to the alternative FT-NIR age data (Table 23.1). Preliminary application of the new data compared with the conventional age determination approach does have some impact on the assessment results (Fig. 23.3). However, more work is needed to estimate the observation-error matrices used in stock assessments. Specifically, this should entail a careful study of conventional reader/tester age estimates with those from FT-NIR estimates.

Table 23.1. -- Indicative eastern Bering sea walleye pollock (*Gadus chalcogrammus*) model diagnostics using the "conventional" age-determination error matrix and data compared with the Fourier transform near infrared (FT-NIR) matrix and data, from various sources (BTS = bottom trawl survey, ATS = acoustic trawl survey, AVO = acoustic vessels of opportunity). Lower values of root mean square error (RMSE) and higher values of effective sample size "effective N" indicate better fits.

Component	Conventional	FT-NIR
RMSE BTS	0.16	0.16
RMSE ATS	0.21	0.20
RMSE AVO	0.22	0.21
Effective N Fishery	1,022	899
Effective N BTS	213	210
Effective N ATS	253	256

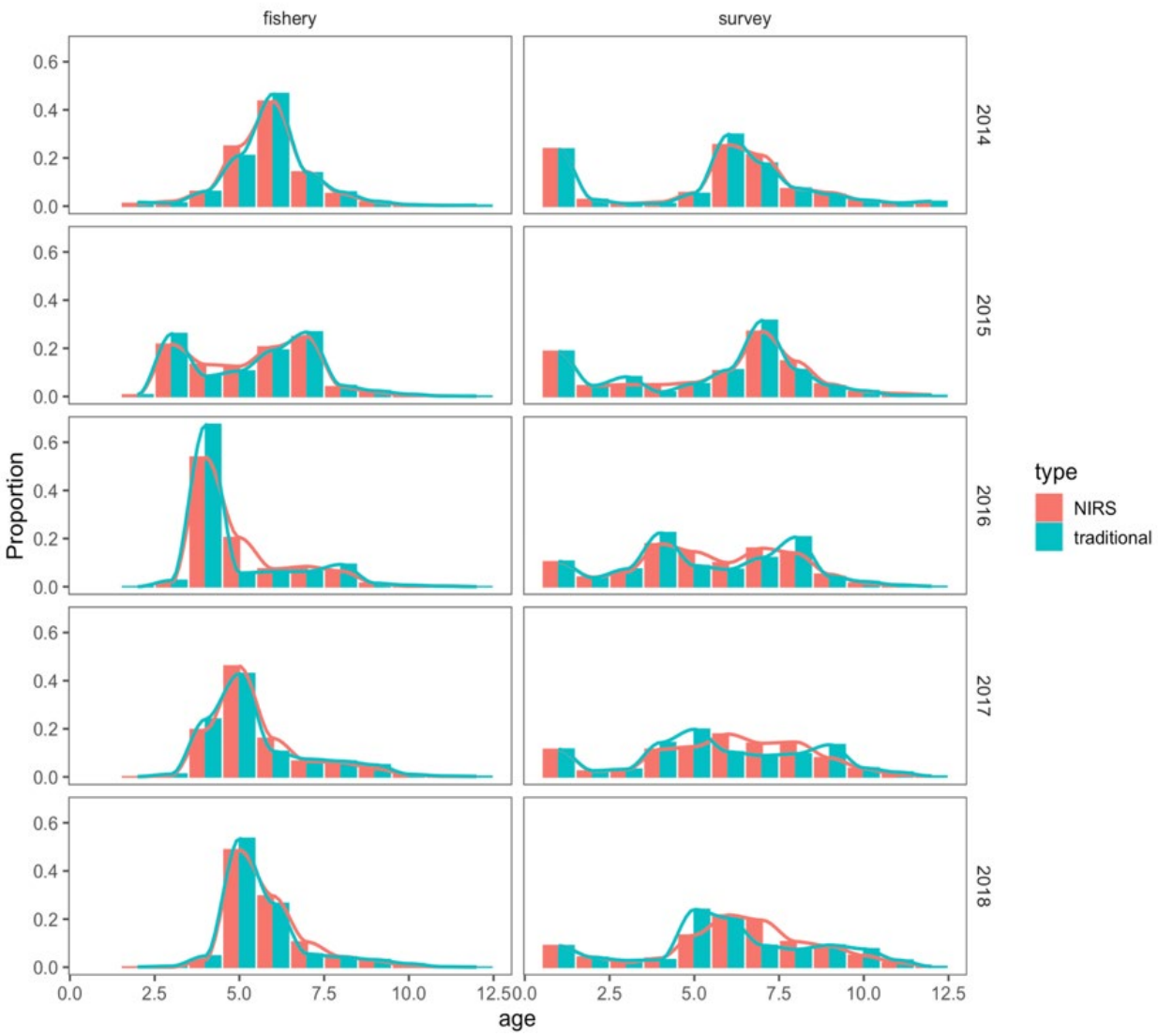


Figure 23.1. -- Comparison of the data in terms of proportions-at-age over time for the Fourier transform near infrared spectroscopy (NIRS) and traditional age determination data. Fishery data is shown on the left and survey on the right.

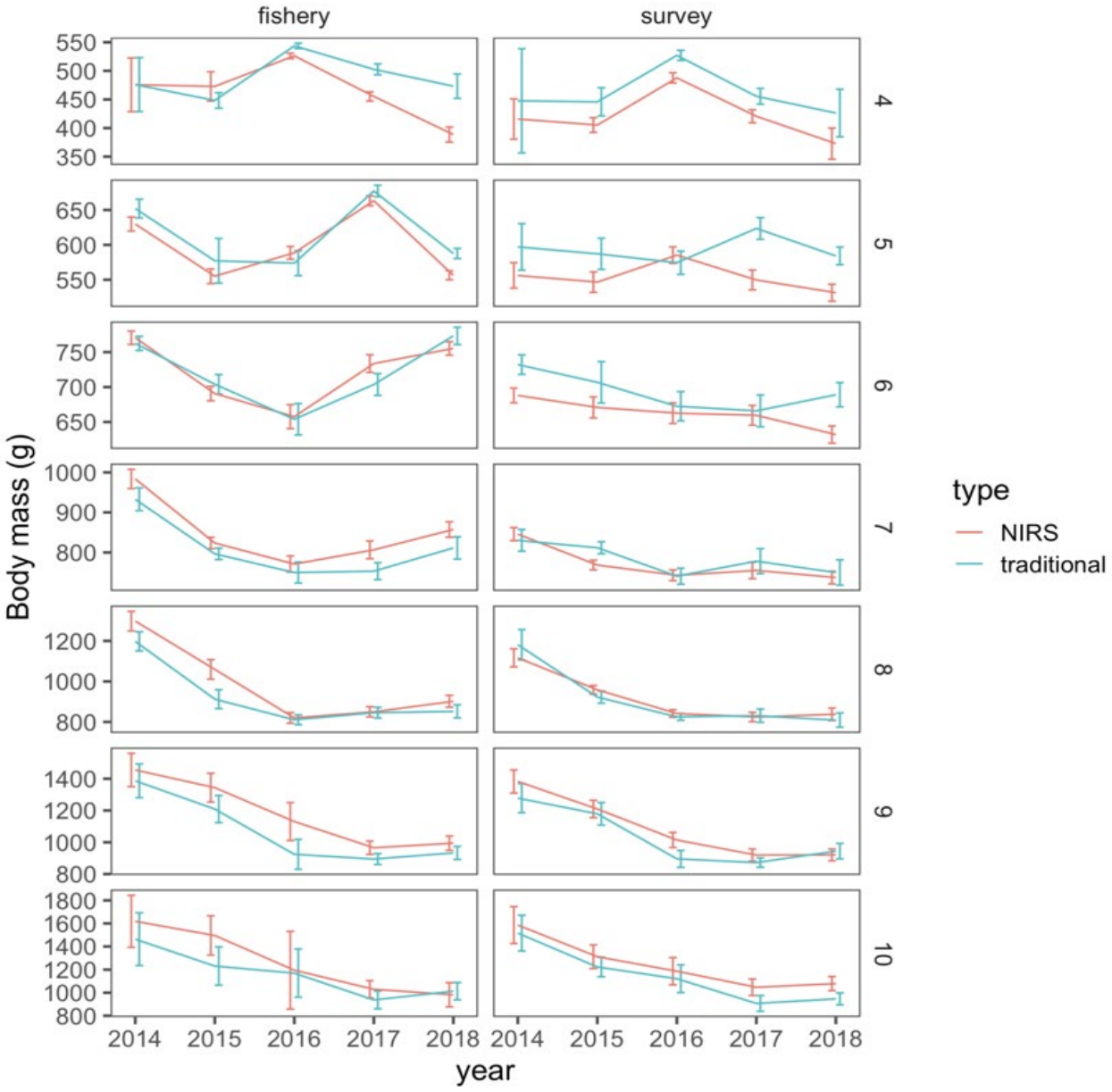


Figure 23.2. -- Comparison of the data in terms of body mass-at-age over time for the Fourier transform near infrared spectroscopy (NIRS) and traditional age-determination data. Fishery data is shown on the left and survey on the right. Rows represent age classes.

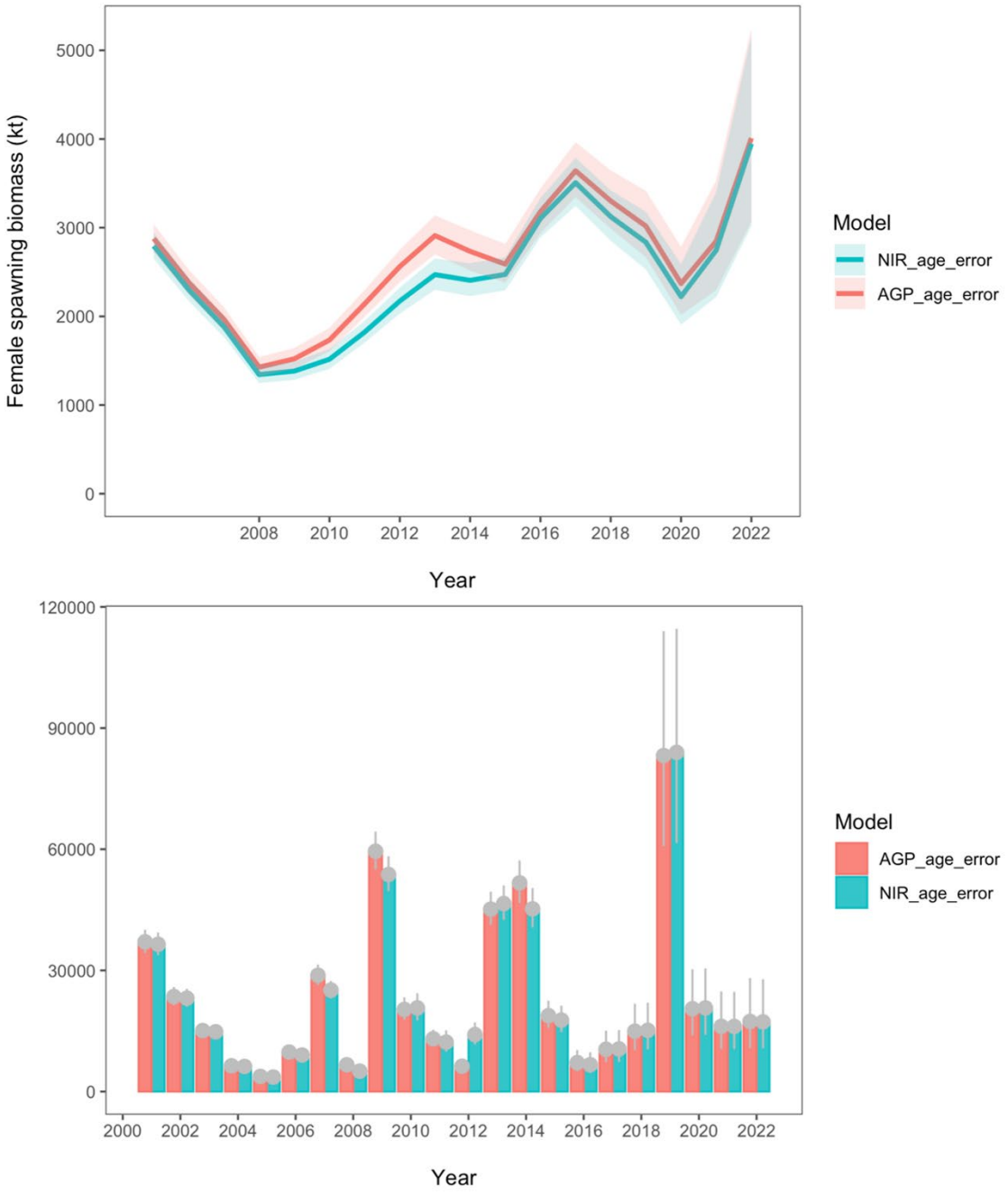


Figure 23.3. -- Preliminary indication of comparisons of a stock assessment model run with the conventional Alaska Fisheries Science Center Age and Growth Program's conversion matrix (AGP_age_error) with that of the Fourier transform near infrared estimate (NIR_age_error). Spawning biomass estimates are shown on the top panel and recruitment on bottom.

24. Sensitivity of Pacific Cod Stock Assessment to Alternative Age Composition Data Derived from Fourier Transform Near Infrared Spectroscopy of Otoliths

Thomas E. Helser[1], Matthew R. Siskey[2], Irina M. Benson[1], and Steven J. Barbeaux[1]

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

[2] Washington Department of Fish and Wildlife
1111 Washington Street SE
Olympia, WA USA

ABSTRACT

Fourier transform near infrared (FT-NIR) spectroscopy of otoliths is emerging as a highly efficient fish age estimation alternative to the labor intensive, time consuming traditional microscope-based approach (TMA). However, before such innovative methods can be operationalized in the production fish ageing and management process, FT-NIR spectroscopy generated data must be vetted within the stock assessment model. Using southeastern Bering Sea Pacific cod as a case study, we present the first investigation of the stock assessment model sensitivity to alternative age data predicted from FT-NIR spectroscopy and neural networks (NIR-NN). Specifically, we employed Markov Chain Monte Carlo simulation of the Pacific cod stock assessment model to evaluate the sensitivity of key internal parameters and externally derived management quantities by substituting a 6-year span of survey age compositions (2013-2018) that were based on more than 8,000 generated NIR-NN ages. Our results showed that assessment model outcomes were highly insensitive to the substitution of TMA age compositions for age compositions from the NIR-NN ageing approach, such that the probability of observing differences for any parameter or derived management quantity was no more than 4%.

INTRODUCTION

The eastern Bering Sea (EBS) Pacific cod (*Gadus macrocephalus*) fishery is one of the largest fish stocks managed in the Alaska region, bringing in a yearly average of 185,896 metric tons from 1991 through 2018 (Thompson 2018). Pacific cod stocks are primarily managed with an age-structured, length-at-age-structured, or stage-structured stock assessment model (Thompson 2018). The annual EBS shelf bottom trawl surveys provide age compositions that along with abundance trends represent crucial data inputs for the assessment, shaping essential model parameters estimated internally, such as natural mortality, recruitment, and growth. These quantities then offer insights into population abundance, productivity, and female spawning biomass level relative to fishery management benchmarks. Subsequently, this information serves as the scientific advice that guides resource managers in establishing harvest goals (Hilborn and Walters 1992, Maunder and Punt 2013). Integrated stock assessments, like those used for Pacific cod (Methot and Wetzel 2013, Thompson 2018), heavily rely on age and length data (Maunder and Punt 2013, Ono 2015). However, traditional ageing methods, which involve some form of otolith preparation (such as sectioning and roasting) followed by microscopic evaluation of annual growth rings, can be subjective, prone to error, and labor intensive. Campana (2001) has suggested that age data obtained through traditional methods are considered one of the most expensive data sources for stock assessments.

Ageing error is one of the major sources of measurement errors on composition data most often used in age-structured assessments (Deriso et al. 1985). Ageing error can lead to profound uncertainty on estimates of growth, weight- or maturity-at-age, natural mortality, and productivity (i.e., fewer older fish means higher natural mortality and higher productivity) that

will in turn affect management reference points (Reeves 2003, De Pontual et al. 2006, Bertignac and De Pontual 2007). Two properties of ageing error are precision and accuracy (Kimura and Lyons 1991, Beamish and McFarlane 1995). Imprecision occurs when age reading errors occur at random between independent trials of age assignment (e.g., by two age readers). Accuracy, or bias, occurs when there is a systematic difference between the age assigned and the true age. In practice, both types of error occur during the age determination process, but generally precision is most often routinely quantified for use in stock assessments by production ageing laboratories.

Recent research has explored a novel approach of using Fourier transform near infrared (FT-NIR) spectroscopy of otoliths as a rapid, nondestructive, and more efficient method of generating fish ages (Helser et al. 2019a). Near infrared (NIR) spectroscopy has been widely used in agriculture, pharmaceutical, and petrochemical industries, and most recently in terrestrial ecology (Vance et al. 2014) and fishery science (Helser et al. 2019a, Helser et al. 2019b). It functions by exciting covalent bonds (O-H, C=O, C-H, C-N, and N-H) at the molecular level using NIR light and measuring absorbance in the wavenumber range 4,000 to 12,500 cm^{-1}. The light interaction with the sample results in measurable vibrational frequencies represented by spectral signatures associated with molecular overtones and combinations that make up compounds in the sample (Conzen 2014, Siesler et al. 2002). In otoliths, these spectral signatures are a good proxy for fish age. The use of FT-NIR spectroscopy of otoliths to estimate fish age have been reported for a number of species (Wedding et al. 2014, Helser et al. 2019b, Passerotti et al. 2022), including EBS Pacific cod in particular (Healy et al. 2021). Healy et al. (2021) found the Pacific cod multiyear model yielded high precision metrics (r^2 = 0.869, root mean square error RMSE = 0.614 years, percent agreement PA = 63%, coefficient of variation CV = 7.412%), which were highly comparable to precision metrics from the traditional microscopic

approach (r^2 = 0.763-0.869, RMSE = 0.639-0.737 years, PA = 63%-70%, CV = 5.671-6.698%). Moreover, recent work by Benson et al. (2023) has combined FT-NIR spectroscopy of otoliths and neural networks (NIR-NN) for fish ageing that directly ingests the whole range of collected spectral information and corresponding geospatial and biological data that improve predictive skill.

With the promise for improving efficiency and repeatability of generating age data for stock assessments, the National Oceanic and Atmospheric Administration, National Marine Fisheries Service has undertaken a nationally coordinated R&D effort across seven biological laboratories with the goal of operationalizing this new technology into the fish age determination enterprise (Helser et al. 2019a). These efforts hold great promise for improving efficiency and repeatability of generating age data for stock assessments; however, no study to date has evaluated the impact of these new data products on model outputs used for management systems. Here, we present a first case study that uses more than 8,000 otolith spectra and biological data across 6 years to generate ages using the NIR-NN model, which are then incorporated as age composition data in the EBS Pacific cod stock assessment. Specifically, this evaluation was performed as a sensitivity analysis by replacing ages generated by the traditional microscope-based approach (TMA) with NIR-NN predictions. We evaluate a number of scenarios in which age compositions from each ageing method are incorporated into the assessment model with different assumptions about ageing error using Markov Chain Monte Carlo simulation and compare key assessment model internally-estimated parameters and derived management quantities.

METHODS

Pacific cod otoliths were taken from recent collections for this analysis. Pacific cod otoliths and corresponding biological data (fish length and sex) were collected during 2013-2018 EBS bottom trawl surveys (BTS) conducted by Alaska Fisheries Science Center (AFSC) Groundfish Assessment Program (https://www.fisheries.noaa.gov/alaska/science-data/groundfish-assessment-program-bottom-trawl-surveys). The BTS, which covers the southeast EBS continental shelf, occurs annually and employs a 20 × 20 nautical mile fixed grid design at 0-50 m (inner shelf), 50-100 m (middle shelf), and 100-200 m (outer shelf) depths. The survey area averages 493,000 km^2 over approximately 300-356 hauls annually (Thompson 2018). Otoliths are taken from fish subsampled within each haul; fish are dissected on deck and ages are later determined in the laboratory. BTS estimates of abundance and corresponding age and length compositions are generated using a depth-area stratified design-based approach (Lauth et al. 2019).

In statistical analysis involving otolith spectra, interest is in predicting the target variable (often referred to and used interchangeably as reference data), which in our case is the TMA age. For the reader not familiar with chemometrics, in simple terms, it involves reducing the dimensionality of otolith spectra (the *X* data matrix) while at the same time maximizing the correlation with reference values (*Y* data vector) through a small set of intermediate linear latent variables (Chen and Wang 2001). Standard classical methods of generating this relationship involve techniques such as principal component regression or partial least squares (PLS) regression. Our approach, however, employs deep machine learning, but the reference data preparation and spectral data acquisition are essentially the same, described below.

Reference Age Data

Pacific cod ages for all samples (n = 8,261) were estimated by analysts enumerating annual growth rings of otoliths, following the ageing guidelines and procedures of the AFSC's Age and Growth Manual (Matta and Kimura 2012). This was done for only one of the paired sagittal otoliths belonging to a given fish, leaving the other otolith for spectroscopic analysis. Age determination is performed using a dissecting microscope under reflected light, which is preceded by preparation techniques such as cutting to expose the otolith core and roasting to enhance annual bands. The AFSC has historically employed a random 20% double read between two independent analysts (referred to as a reader and tester) to monitor precision and repeatability of the TMA process (Matta and Kimura 2012). The metrics of precision are expressed as percent agreement, average percent error, and the coefficient of variation (CV) (Matta and Kimura 2012). Since several of our assessment model scenarios included TMA age reading error, this was quantified through double-read data from the 1993-2019 EBS BTS.

Otolith Spectral Acquisition and Data

For spectroscopic analysis, whole otoliths were removed from their vials and gently blotted dry with Kimwipes to remove excess glycerol-thymol (storage medium) before placement on the sample window of a Bruker Tango-R or MPA II FT-NIR spectrometer. Otoliths were covered with a gold stamp and analyzed using diffuse reflectance on an integrated sphere, and spectra were collected at step sizes of 8 cm^{-1} for MPA II and 8.3 cm^{-1} for TANGO R, with 64 co-scans, measuring absorbance values from 12,500 to 4,000 cm^{-1}. Helser et al.

(2019b) reported that scan time to acquire otolith spectra ranges between 40 and 60 seconds. The entire paired data set of TMA age, otolith spectra, and associated biological data was split into training (80%) and testing (20%) sets using the "onion" algorithm (Gallagher et al. 2020). The "onion" algorithm was applied to each year in the split and then recombined to capture annual variability in the spectra in both train and test data sets. Spectral data were pre-processed using 1st derivative and a Savitsky-Golay (2nd order polynomial, 17 points) smoothing filter. Spectral variability across collection years and detection of outliers was evaluated using principal component analysis (PCA) and measures such as Hotelling's T^2 and Q-statistic (Mujica et al. 2011, Rodionova and Pomerantsev 2020). A more comprehensive description of sample presentation, spectral data preprocessing and outlier detection can be found in Benson et al. (2020, 2023).

Multimodal Convolutional Neural Network Model

The proposed multimodal convolutional neural network (MMCNN) was built by incorporating otolith spectra and biological datasets as multimodal input within different branches. Benson et al. (2023) provide a more detailed description of the MMCNN model, although the general model architecture is described here. The model is trained to extract information out of the entire spectral data range before integrating it with information extracted from additional data blocks that includes corresponding biological data. Biological information routinely collected at sea include fish length and fish sex, both of which were proposed in the model since they bear an allometric relationship with fish somatic growth and otolith accretion rates.

We employed Python 3.7.0 using TensorFlow and Keras to train and test a neural network model. The model consisted of two primary branches and 14 layers. The first branch was associated with spectral data input (layer 1) and represented 5 layers with the convolutional layer (layer 2) activated by the rectified Linear Unit (ReLU) activation function for nonlinear activation. The convolutional neural network is capable of extracting the abstract features of FT-NIR spectra and selecting most featured wavenumbers (Chen and Wang 2019). Output was flattened (layer 3) before passing to the fully connected dense layer (layer 4) and regularized through dropout (layer 5). For this model, all fully connected dense layers were activated with the exponential linear unit (eLU) activation function to capture the non-linearity in the data. The dropout layers were used for regularization to prevent overfitting. A number of nodes for fully connected dense layers and a dropout rate for dropout layers were tuned using the Hyperband optimization tuning algorithm.

The second branch was associated with corresponding biological data that have an effect on fish growth and provide information that improves the ageing model performance. Fish sex was preprocessed with one-hot encoding, which represented categorical variables as numerical values in a machine learning model. Fish length was scaled with a zero mean and unit variance. The concatenated preprocessed biological data input layer (layer 6) was processed with dense layer (layer 7) and regularized with dropout layer (layer 8) and then concatenated to the second branch (layer 9). Within the combined, two fully connected dense layers (layers 10 and 12) were regularized through dropout layers (layers 11 and 13). Then the output of the previous layers was fed to the final fully connected dense layer (layer 14) with a single output node of predicted ages and linear activation function.

Model performance was evaluated through plotting NIR-NN predicted ages on TMA ages and calculating r^2 and RMSE of test data. The SHapley Additive exPlanations (SHAP) method was used to interpret the importance of features for predictions (Lundberg and Lee 2017). NIR-NN model performance was also evaluated with the TMA approach using a modified Bland-Altman plot (Ogle 2017a, 2018) and also a measure of relative bias with and without respect to age using:

$$average\ bias = \frac{\sum_{i=1}^{n}(A1 - A2)_i}{n}.$$

In the above equation, *A1* and *A2* represent TMA age estimates of the reader and tester, respectively. For the NIR-NN model predictions, the same equation was used for the difference between the NIR-NN model predicted age and the average of the *A1* and *A2* TMA estimates.

Stock Assessment Model

The EBS Pacific cod stock assessment is an age-structured model based on Stock Synthesis V3.30.16.02 (Methot and Wetzel 2013), which is an integrated statistical catch-at-age modeling framework coded using Automatic Differentiation Model Builder (ADMB; Fournier et al. 2012). The base model specification used in this analysis most closely resembles Model 16.6 of the 2020 stock assessment report (Thompson et al. 2018). Data inputs included: 1) BTS swept area biomass 1982-2019, 2) BTS age compositions from 1994-2019, 3) BTS size compositions 1982-2019, 4) fishery catches 1977-2020, and 5) fishery size compositions 1977-2020. The base model included BTS shelf age compositions, as the goal of this analysis was to evaluate the

sensitivity of model parameter estimation (e.g., natural mortality, recruitment, and growth) and derived-quantity outputs most dependent on those data to changes to this data source (e.g., missing data, alternate age data from the same specimens generated via FT-NIR spectroscopy, and alternate ageing error estimates via FT-NIR spectroscopy ageing). Model parameters estimated internally include: 1) survey catchability, q, 2) natural mortality rate, M, 3) natural log of initial recruitment, $\ln R_0$ and time-varying recruitment deviations, 4) growth parameters (L_{Min}, L_{Max}, von Bertalanffy k, Richards k, and variability for size at minimum and maximum age), 5) fishery and survey selectivity parameters, and 6) Dirichlet-multinomial log(theta) parameters (Table 24.1).

One difference between our base model (i.e., the TMA comps/TMA error model) and the model employed for stock assessment (i.e., the original model) was that the otolith ages used for generating age compositions in this study included a subset of the original specimens. A single whole sagittal otolith is needed for instrument scanning to acquire spectra, and in some cases both otoliths have already been processed for TMA age estimates. All age compositions used in this analysis were calculated using a design-based estimator rather than model-based estimation (e.g., via VAST). Another difference between our base model and the original model was that ageing error was only estimated for the years 1977-2007, while the years 2008-2019 used an ageing error vector calculated externally to the model via the nwfscAgeingError package (Punt et al. 2008) which estimates age imprecision and age reading bias-at-age.

For the ageing error vector estimated within Stock Synthesis (years 1977-2007), there are seven parameters that are used in estimation of ageing error-at-age. The second and third ageing error parameters are attributed to the bias at the starting age (age-0) and bias at the maximum age (age-20) in the form of an additive offset from unbiased age. These two parameters were

estimated within the time-varying module of Stock Synthesis in order to incorporate the block structure that was specified in the original assessment model (i.e., estimated for the 1977-2007 block and not estimated for 2008-2019). Two ageing error vectors were calculated externally to Stock Synthesis using the Punt et al. (2008) method. The first ageing error vector, which is representative exclusively of the TMA method, was calculated using 8,998 double reads from 1993-2020 (i.e., reader age vs. tester age). This error vector was referred to as TMA error. The second vector, which represented error arising from both the TMA and the NIR methods, consisted of triplicate age outcomes for the same otolith. Here the error vector was calculated using 1,946 double reads from 2013-2018 (reader and tester) and the FT-NIR spectroscopy predicted age (from NIR-NN). This vector is referred to as *TMA-FTNIR* error. These ageing error vectors were used as inputs for the ageing error module within SS for the second time block of 2008-2019, thus ageing error parameters were not estimated for these years. A fourth and final vector of error was calculated from a sample of 600 otoliths that were predicted with the NIR-CNN model from otoliths collected from 2016 and 2017 EBS BTS that were scanned on two different instruments and by two different operators. This was not used in SS but rather simply used to compare measurement error arising from spectral data collection.

Model scenarios developed for this analysis included age compositions generated from the TMA and NIR-NN approaches and two alternative error vector assumptions (Table 24.1). The default or base scenario used the 2013-2018 TMA age compositions and the TMA ageing error (identified as *TMA comps/TMA ageing error*). An alternative scenario employed the same specification and inputs as the base model scenario but used 2013-2018 age compositions generated from the NIR-NN approach and assumed the same TMA age reading error. This scenario is identified as *Mixed comps/TMA ageing error.* Another considers the same model

specification and inputs as the previous scenario using NIR-NN age compositions but here augments ageing error to account for both the TMA ages (e.g., reader and tester) and the difference between TMA and NIR-NN ages (involving triplicate outcomes). This scenario was identified as *Mixed comps/Mixed ageing error*. A final scenario was used to examine the effect or sensitivity of the choice of the 2013-2018 time period over which the TMA versus NIR-NN age composition substitutions were made (*Missing comps/TMA error model*). Here, the base model was simply run by removing age composition data for years 2013-2018.

We employed Markov Chain Monte Carlo (MCMC) simulation to quantify and compare sensitivity of assessment outcomes to the use of NIR-NN versus TMA data sources for age composition and ageing error. We specified non-informative uniform priors on all parameters except for M, which was empirically derived, approximating a log-normal density (Thompson et al. 2020). We performed MCMC on the four different scenarios (Table 24.1) to compare model outcomes between ageing methods and assumed ageing errors. Marginal posterior densities (MPD) of the key parameters from 900 MCMC iterations (1,000,000 with 1/1,000 thinning and a 100-iteration burn interval) were compared to evaluate stock assessment model sensitivity.

RESULTS

Ageing error of Pacific cod using the TMA method for the 2013-2018 period (percent agreement ± 0 years = 65.3%; CV = 6.96%) was typical of historic data throughout the time series. Pacific cod is considered a moderately difficult species to age given the relatively low mean age of the sample (~3.9 years) and relatively short lifespan of the stock (90% of ages were below 8 years old). There was no indication of relative bias between TMA double read data

based on the Evan-Hoenig test of symmetry ($p > 0.05$). Robust regression of age estimates from the two analysts resulted in an r^2 of 0.86 and an RMSE of 0.52, indicating that age estimates of a second age reader are expected to be within 0.52 years of the first reader 67% of the time. All error types estimated from the model resulted in a best fit model (AIC_c) which was represented by a non-linear standard deviation (SD) of the observed error with "true age" (Fig. 24.1). The TMA ageing method showed the greatest magnitude error (TMA error; Fig. 24.1) while the mixed error was intermediate in magnitude to TMA ageing error or measurement error (FT-NIR spectroscopy). Not surprising was that the FT-NIR spectroscopy error was the smallest, suggesting that repeatability of age prediction is high even when otolith NIR scanning is performed by different operators and on different instruments.

Spectral data collection from Pacific cod otoliths showed consistently good signal to noise, with absorbance increasing within the 7,000 to 4,000 cm^{-1} wavenumber regions (Fig. 24.2A). Two distinct groups of raw spectra (Fig. 24.2A) can be seen as a baseline shift that is associated with otolith scanning using two different FT-NIR spectrometers (Tango-R vs. MPA II). Instrument differences in baseline absorbance are not unusual and otolith spectra were harmonized by preprocessing techniques employed prior to modeling (Fig. 24.2B, 24.2D). Preprocessed spectra plotted in bivariate space indicate that the vast majority of spectral variability is accounted for by the 1st and 2nd principal components. Further, the 'onion' algorithm extracts roughly the same extent of spectral variability for both the train and test data sets (Fig. 24.2C).

The NIR-NN was trained to extract information out of spectral data, including all wavenumbers, before integrating the output with the scaled data block that includes corresponding biological data. The data types were ordered according to their importance for

prediction in the NIR-NN model using the mean of absolute Shapley values (Fig. 24.3). The collective contributions of otolith spectra in the range of 7,000-4,000 cm^{-1} wavenumbers had the highest impact on predicting fish age, followed by fish length and sex. The NIR-NN model results show good performance resulting in a high degree of correlation between NIR-NN predicted and TMA estimated ages (Fig. 24.3). The NIR-NN model using training data resulted in good performance with $r^2 = 0.93$ and RMSE = 0.52 years. Likewise, on the test data set (which represents new unseen variability by the model, $n = 1{,}652$) the NIR-NN model yielded similarly good performance, with $r^2 = 0.92$ and only a slight increase in the RMSE (0.56 years).

Predicting Pacific cod age using otolith spectra and associated biological data is expected to result in age predictions within 0.6 years of the TMA age 67% of the time and within 1.2 years 98% of the time. Precision associated with NIR-NN predictions (RMSE = 0.56 years) was clearly consistent with double reads of the TMA age determination approach, which had an RMSE of 0.52 years. Comparing NIR-NN model age predictions with TMA estimated ages, Bland-Altman plots showed a high level of consistency of ages from the different methods over the range of ages, where the majority of deviations fell within 1.96 SD (Fig. 24.3D). Age predictions were highly consistent along a 1:1 line, although there was a very slight under-prediction for fish greater than 10 years. Very few older Pacific cod ages were observed, and therefore predictions may be improved by inclusion of older age fish in the training model.

A measure of relative bias or differences between ages estimated using TMA and NIR-NN predictions also showed a very high degree of consistency between ageing methods, except for a departure after 9 years of age (Fig. 24.4A). Since so few ages beyond age 9 were available, there was insufficient information to evaluate relative differences between ageing methods for older fish. Overall, up to 90% of the observations resulted in zero bias between NIR-NN

predicted ages and TMA estimated ages, compared to roughly 70% from TMA age estimates (Fig. 24.4B). Nearly all estimated or predicted ages were within 1 year of age for both methods (Fig. 24.4B).

Survey age compositions generated from the design-based estimator using data derived from TMA and NIR-NN ageing methods were comparable, with only slight variations in proportions-at-age (Fig. 24.5). The largest differences in proportions-at-age were for age-3 in 2018 (TMA = 0.213, NIR-NN = 0.166), age-4 in 2017 (TMA = 0.264, NIR-NN = 0.307), and age-4 in 2014 (TMA = 0.139, NIR-NN = 0.092). Similarly, model fits to the age composition data had little variation across scenarios, with the largest differences in expected proportion-at-age being for age-5 in 2018 (*TMA comps/TMA error* = 0.236, *Mixed comps/Mixed error* = 0.249, *Mixed comps/TMA error* = 0.235). Only very small differences between age composition likelihood were observed, suggesting model fits among these scenarios were equally good (or poor). Recent assessments have reported a relatively strong 2013 year-class, seen as age-1 fish in the 2014 age compositions. Age compositions for both ageing methods exhibit the movement of this year class through the age compositions in subsequent years (Fig. 24.5).

Marginal posterior densities generated from MCMC of key internal and externally derived management parameters among the different scenarios show little difference in the highest posterior density (HPD) and kernel densities (Figs. 24.6 and 24.7). Table 24.2 provides comparisons of key model parameters over all sensitivity scenarios. The two primary scenarios in which NIR-NN age compositions are substituted for TMA age compositions (*Mixed comps/TMA error* and *Mixed comps/Mixed error*) resulted in nearly identical median values and kernel densities for M and $\ln R_0$ (Fig. 24.6), but did shift slightly to higher values when NIR ageing error was assumed. The only estimated parameter with notable shifts in median value

across 1,000 MCMC iterations was Q, which shifted from the most negative values in the *TMA comps/TMA error* scenario to more positive values in other scenarios (Table 24.2). In terms of uncertainty, L_{max} displayed the highest degree of change in standard deviation despite this change being relatively small (*TMA comps/TMA error* = 3.206, *Missing comps/TMA error* = 3.765; Table 24.2). The years chosen (2013-2018) for data substitution did produce an effect on the sensitivity analysis (*Missing comps/TMA error*), such that kernel densities of key parameters shifted to lower median values. The relative change in kernel densities from the *Missing comps/TMA error* scenario was in fact greater than kernel density shifts across scenarios that used different age data products (*TMA comps/TMA error*, *Mixed comps/TMA error*, *Mixed comps/Mixed error*; Fig. 24.6).

Similar shifts in median values of the MPDs for derived management quantities such as relative spawning biomass ($SSB/SSB_{unfished}$) and unfished female spawning biomass were observed, but in all cases the sensitivity of this change resulted in less and a maximum of a 4% probability (Fig. 24.7). Differences in the probability of relative depletion between the NIR-NN and TMA age data scenarios and with either ageing error assumption were exceedingly low. All model run scenarios resulted in decreased median derived management quantity values and standard deviations in comparison to the base model scenario (*TMA comps/TMA error*; Table 24.2, Fig. 24.7). The *Missing comps/TMA error* scenario resulted in the largest reductions in the median value of derived management quantities (1.2% reduction for median of SSB_0 to 4.8% reduction for median of overfishing level OFL) as well as the standard deviation (3.2% reduction for SD of SSB_0 to 14.3% reduction for SD of 2020 recruitment). The *Mixed comps/Mixed error* scenario resulted in the smallest reductions in the median value of derived quantities (0.2% reduction for median of 2020 recruitment to 0.6% reduction for median of OFL) as well as

standard deviation (2.2% reduction for SD of OFL to 7.0% reduction for SD of 2020 recruitment).

DISCUSSION

Fish age determination based on growth patterns visible in otoliths has changed little over the last century. While ageing fish using scales can be documented as far back as the early 1700s, Johannes Reibisch was the first to describe a protocol for ageing fish using otoliths in 1899 (Jackson 2007). Microscopic counting of the annual growth zones in otoliths was the foundation of the ageing method then and is still the basis of age estimation today. FT-NIR spectroscopy of otoliths to predict fish age is an emerging technology that may have a transformative impact on traditional microscopic age determination. Studies suggest that FT-NIR spectroscopy of otoliths may produce efficiency gains by as much as 10x and substantially improve repeatability compared to traditional microscopic methods (Helser et al. 2019, Benson et al. 2023). Crude estimates of TMA age production rates suggest a throughput of 31 ages per person per day for Pacific cod[2]. In comparison, age predictions from FT-NIR spectroscopy with an already calibrated model can produce up to 35-50 ages *per hour*. Research is underway to better quantify time-flow statistics for cost-benefit analysis for both methods.

Fish otoliths largely comprise the same composition of molecular constituents dominated by $CaCO_3$ as aragonite and lesser amounts of organic material (Campana 1999), so it is not surprising that other studies applying FT-NIR spectroscopy found wavenumbers in the 7,000 to

[2] Lambert et al. 2017. Importance of age data collection for stock assessments: a U.S. national perspective. NOAA Fisheries, Alaska Fisheries Science Center, Unpublished report.

4,000 cm^{-1} range most informative for fish otoliths (Wedding et al. 2014, Helser et al. 2019b, Benson et al. 2020, Passerotti et al. 2020b). Many promising studies that have analyzed FT-NIR spectroscopy of otoliths using partial least squares for age prediction have been reported for teleost fish (Wedding et al. 2014, Helser et al. 2019a, Helser et al. 2019b, Wright et al. 2021, Passerotti et al. 2020b, Healy et al. 2021) as well as chondrichthyan species (Rigby et al. 2016, Arrington et al. 2022). For many of them, model performance has been quite good with coefficients of determination in the high 80% and even low 90%, and with precision on par with TMA precision estimates. Our study, and that of Benson et al. (2023), have coupled FT-NIR spectroscopy of fish otoliths and a deep machine learning approach that allowed the network to extract important spectral features automatically by utilizing a convolutional layer before connecting its output with additional biological or geospatial layers. Using NIR-NN for EBS walleye pollock *G. chalcogrammus*, Benson et al. (2023) reported superior performance, relative to classical partial least squares, that achieved not only better predictive skill but also rectified nonlinearity at older ages. While we did not compare our EBS Pacific cod NIR-NN model to the PLS methods, our model performance achieved similarly good results, with r^2 in low 90% and as good or slightly better precision compared to TMA.

Ageing error associated with TMA is well known in fish ageing, and statistical methods for quantifying it have been well documented (Beamish and Fournier 1981, Kimura and Lyons 1991, Campana 2001). The impact of ageing error on the performance of stock assessments has also been well documented in the literature, resulting in inappropriate yield projections and overfishing (Lai and Gunderson 1987), poorly defined stock-recruitment relationships (Richards et al. 1992), and in some cases overly optimistic yield projections (Reeves 2003). Punt et al. (2008) articulated the justification and necessity of including ageing error in stock assessment

models, widely used today (Methot and Wetzel 2013). Hence, alternative ageing methods used to generate age composition data, such as FT-NIR spectroscopy, must also seriously consider such sources of error and how to quantify them. Among the ageing error assumptions considered in this study, the *Mixed error* assumption is probably the most realistic for age compositions generated from the NIR-NN approach, although we examined an alternative of using TMA error. Statistically, the TMA approach gives rise to an estimate of age, whereas the NIR-NN approach gives rise to a prediction. The question arises of how best to quantify and propagate uncertainty from a predictive model that is trained on data that have inherent error among outcomes. Since the standard practice is to incorporate the error specific to the age composition data within the stock assessment model, it seems logical that error associated with both TMA ageing and FT-NIR spectroscopy predictions (no matter what modeling technique is used) should be used, as was done with the *Mixed comps/Mixed error* scenario. It should be noted that ageing bias for Pacific cod has yet to be quantified and analyzed within the assessment model. If ageing bias exists in TMA, it will also be the case for NIR-NN age predictions.

Finally, several considerations of this study are warranted. First, the time span of NIR data substitution was somewhat narrow relative to the entire range of years over which age composition data are used in the assessment. The over 8,000 ages generated from NIR otolith scanning from 2013-2018 was a practical consideration. Had resources been available to substitute NIR age composition data over the entire time series, might the results have been different, particularly during periods of more dynamic change in the population and fishery? Maybe not, because the EBS Pacific cod population experienced a rather dramatic change over the 2013-2018 period explored in this analysis (Thompson et al. 2020). Cod recruitment reached an all-time high in 2013, which was followed by a population size increase to a near record high

in 2014 and then to a near record low in 2018 (Thompson et al. 2020). The *Missing comps* scenario, which attempted to measure the "effect size" of the time period, showed the most substantive changes in model outputs, but yet the model results among the other scenarios in which NIR-NN age compositions were substituted for TMA resulted in minimal differences.

Our sensitivity analysis showed only minimal differences in model outputs when using FT-NIR spectroscopy generated age compositions. Under the $B_{40\%}$ minimum depletion threshold level, the model suggests the stock in 2020 was at about 52% of unfished female biomass and this depiction of stock status did not change under any sensitivity scenario. Key estimated parameters and externally derived management quantities were largely unchanged among the different scenarios and assumed ageing error structures based on the stock assessment model specification used for this sensitivity analysis. However, compared to some assessment models, the Pacific cod model only uses age data to a somewhat limited extent. Fishery age compositions are currently not used in the assessment. Time varying growth, which may be a biologically realistic feature of the population, is not specified. If future Pacific cod stock assessments evolve to incorporate more age data for other fleets (fisheries) or add complexity of their use, such as conditional age-at-length compositions, further sensitivity analyses of the use of FT-NIR spectroscopy generated age data such as this may be warranted.

ACKNOWLEDGEMENTS

We express sincere appreciation to NOAA/AFSC Age and Growth Program staff who scanned and aged Pacific cod otoliths. We also thank the NOAA Fisheries Science Board for funding this work under the NOAA Fisheries Strategic Initiative.

Table 24.1. -- Pacific cod *Gadus macrocephalus* Stock Synthesis (SS) model specifications and scenario descriptions used in the sensitivity analysis. TMA = traditional method of ageing; FTNIRS = Fourier transform near infrared spectroscopy.

SS Model Parameter	Description
Lmin	Length at minimum age
Lmax	Length at maximum age
Von Bert K	Von Bertalanffy growth coefficient
Richards	Richards growth coefficient
SD_young_Fem_GP_1	variability for size at age <= minimum age
SD_old_Fem_GP_1	variability for size at age >= maximum age
Ln_R0	log of virgin recruitment level
M	Natural mortality
Q	Catchability of survey
Size selectivity fishery params	5 of 6 double normal selectivity parameters
Size selectivity survey params	2 of 6 double normal selectivity parameters
ln(DM_theta)	3 log theta parameters

Sensitivity Scenario Name	Description
TMA comps/TMA ageing error	• TMA age compositions for all years • Estimate ageing error 1977-2007 • TMA ageing error (2-way; read & test ages) for all years 2008-2019
Missing comps/TMA ageing error	• TMA age compositions for all years • 2013-2018 missing • Estimate ageing error 1977-2007 • TMA ageing error (2-way; read & test ages) for all years 2008-2019
Mixed comps/TMA ageing error	• FTNIRS age compositions for 2013-2018 • TMA age compositions for other years • Estimate ageing error 1977-2007 • TMA ageing error (2-way; read & test ages) for all years 2008-2019
Mixed comps/Mixed ageing error	• FTNIRS age compositions for 2013-2018 • TMA age compositions for other years • Estimate ageing error 1977-2007 • Mixed ageing error (3-way; read, test, FTNIRS) for all years 2008-2019

Table 24.2. -- Summary statistics for estimated parameters and derived management quantities across 1,000 Markov Chain Monte Carlo (MCMC) runs. TMA = traditional method of ageing. See Table 24.1 for descriptions of model parameters and text for management quantities.

Parameter	Scenario	Median	SD	CV
Lmin	TMA comps/TMA error	14.6648	0.39341	0.02681
	Missing comps/TMA error	14.6454	0.39861	0.02722
	Mixed comps/TMA error	14.6699	0.41049	0.02796
	Mixed comps/Mixed error	14.6817	0.41323	0.02815
Lmax	TMA comps/TMA error	111.027	3.20622	0.02883
	Missing comps/TMA error	112.186	3.76545	0.03349
	Mixed comps/TMA error	111.813	3.69312	0.03289
	Mixed comps/Mixed error	111.391	3.55216	0.03179
Von Bert K	TMA comps/TMA error	0.11140	0.00928	0.08314
	Missing comps/TMA error	0.10779	0.01032	0.09556
	Mixed comps/TMA error	0.10928	0.01016	0.09313
	Mixed comps/Mixed error	0.11069	0.01005	0.09089
M	TMA comps/TMA error	0.36364	0.01259	0.03464
	Missing comps/TMA error	0.35787	0.01204	0.03364
	Mixed comps/TMA error	0.36248	0.01215	0.03351
	Mixed comps/Mixed error	0.36356	0.01207	0.03319
Richards K	TMA comps/TMA error	1.52657	0.04497	0.02943
	Missing comps/TMA error	1.54333	0.04837	0.03135
	Mixed comps/TMA error	1.53427	0.04802	0.03124
	Mixed comps/Mixed error	1.53252	0.04826	0.03151
Q	TMA comps/TMA error	-0.19012	0.07919	-0.41557
	Missing comps/TMA error	-0.14926	0.07148	-0.47303
	Mixed comps/TMA error	-0.17404	0.07436	-0.42821
	Mixed comps/Mixed error	-0.18258	0.07533	-0.40899
LnR0	TMA comps/TMA error	13.3122	0.11836	0.00889
	Missing comps/TMA error	13.2487	0.10959	0.00827
	Mixed comps/TMA error	13.2955	0.11218	0.00844
	Mixed comps/Mixed error	13.3096	0.11187	0.00840
Management quantity	Scenario	Median	SD	CV
OFL	TMA comps/TMA error	182,363	22,138	0.120216
	Missing comps/TMA error	173,475	19,771	0.113215
	Mixed comps/TMA error	178,285	20,808	0.115564
	Mixed comps/Mixed error	181,243	21,183	0.115999
SSB0	TMA comps/TMA error	1,274,570	33,655	0.026394
	Missing comps/TMA error	1,258,560	32,590	0.025894
	Mixed comps/TMA error	1,269,815	33,639	0.026521
	Mixed comps/Mixed error	1,271,135	32,915	0.025868
SSB2020	TMA comps/TMA error	647,083	58,205	0.089261
	Missing comps/TMA error	627,444	52,097	0.082814
	Mixed comps/TMA error	637,056	54,074	0.084511
	Mixed comps/Mixed error	644,436	54,938	0.084847
Recruitment	TMA comps/TMA error	476,590	58,100	0.120768
	Missing comps/TMA error	447,298	49,796	0.110445
	Mixed comps/TMA error	468,742	53,419	0.113134
	Mixed comps/Mixed error	475,393	54,000	0.112762

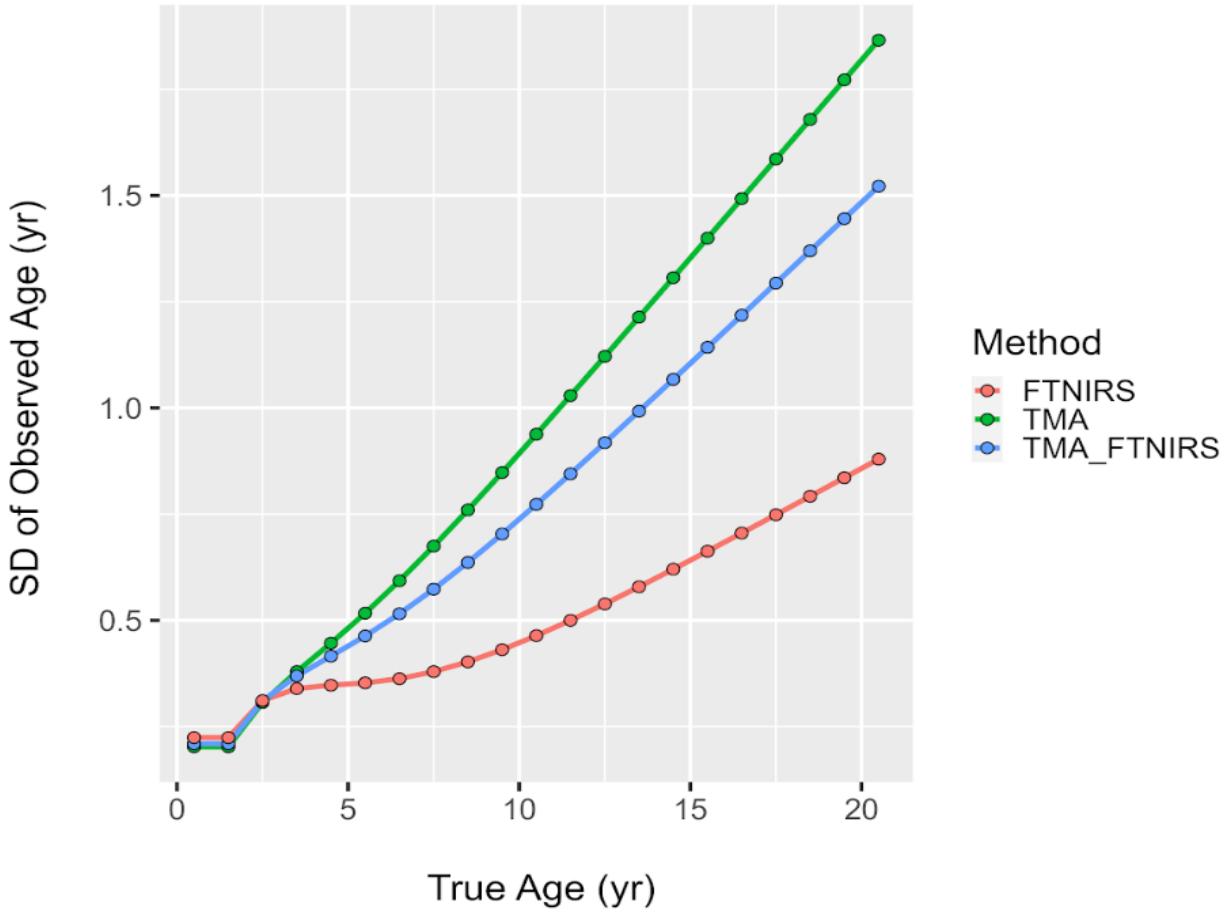


Figure 24.1. -- Ageing error vectors showing standard deviation (SD) of estimated age relative to "true age" from traditional method of ageing (TMA) and Fourier transform near infrared spectroscopy (FTNIRS) data inputs. TMA ages are represented by differences between reader age and tester age, FT-NIRS is the difference between multiple FTNIRS predicted ages from multiple instruments and analysts, and TMA-FTNIRS is the difference between tester age, reader age and FTNIRS predicted ages.

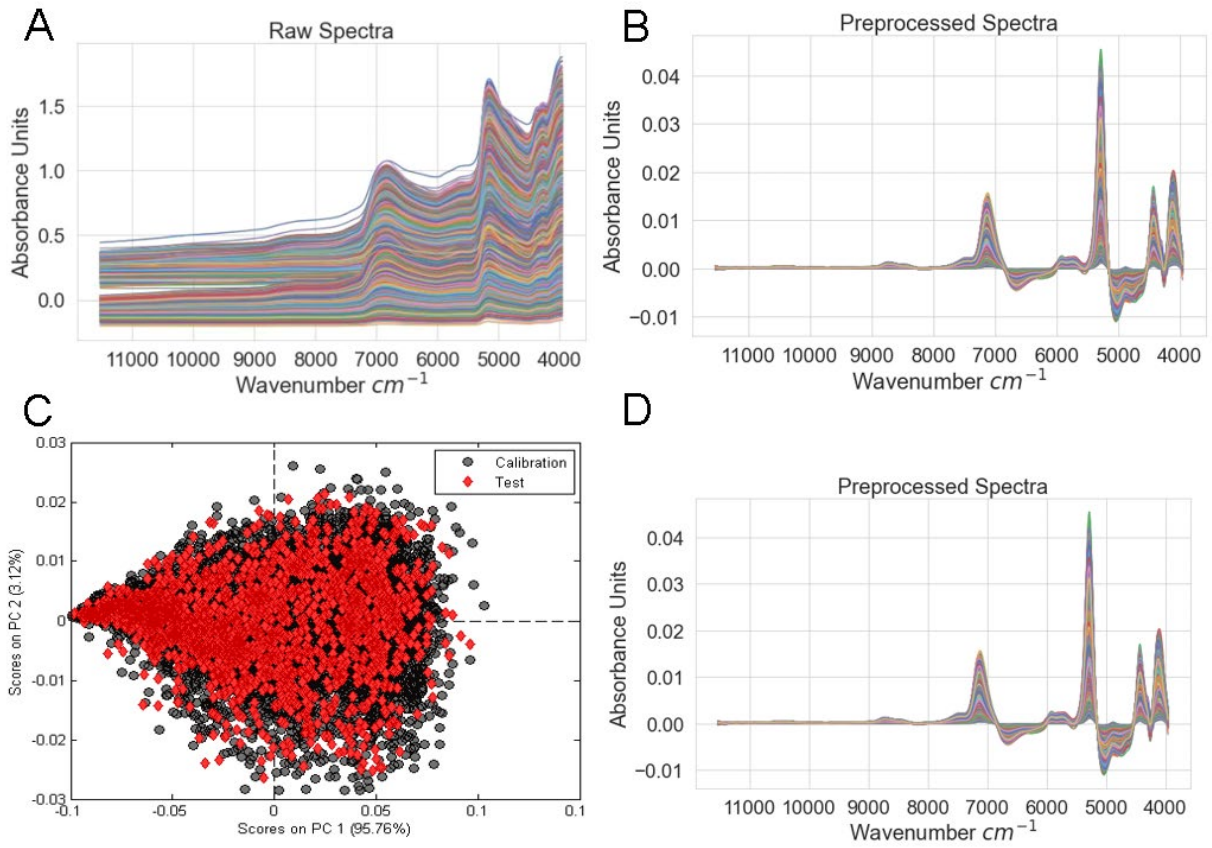


Figure 24.2. -- Pacific cod *Gadus macrocephalus* otolith Fourier transform near infrared spectra (measured as absorbance) from over 8,000 samples showing (A) raw spectra, preprocessed spectra (1st derivative, Savitsky-Golay smoothing) for (B) train and (D) test data sets, and (C) principal component analysis (PCA) of otolith spectra separated into training and testing data sets using the onion algorithm.

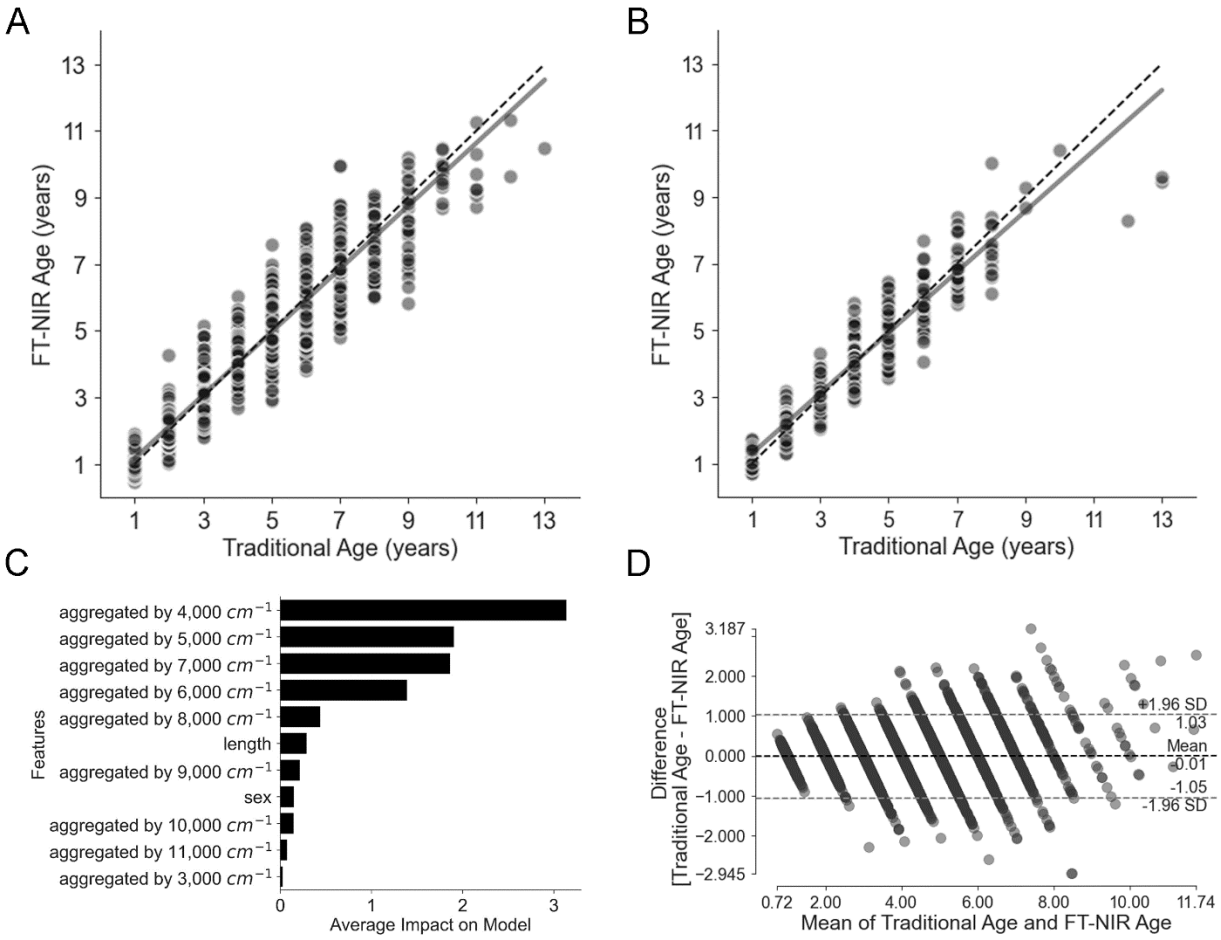


Figure 24.3. -- Age predictions from multimodal convolutional neural network (MMCNN) model for Pacific cod *Gadus macrocephalus* from otolith spectra and biological data as inputs. Model results are illustrated as the relationship between model predictions and traditional age for A) training set with $r^2 = 0.93$ and RMSE = 0.52 years and B) test data sets with $r^2 = 0.92$ and RMSE = 0.56 years. C) MMCNN model importance plot, and D) Bland-Altman plot of difference between methods.

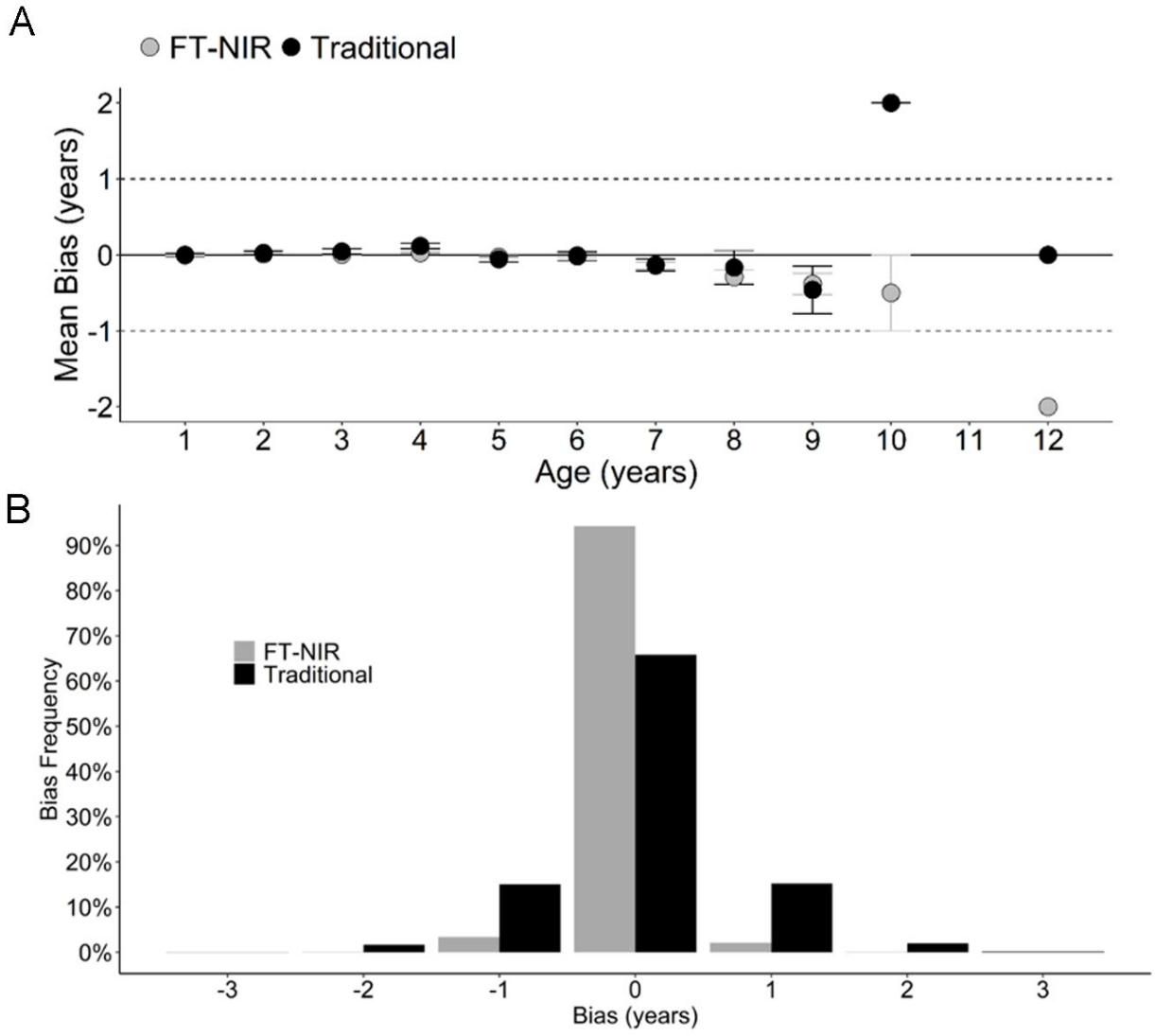


Figure 24.4. -- Relative bias or differences between estimated traditional (TMA) ages and Fourier transform near infrared (FT-NIR) spectroscopy predicted ages A) with respect to age and B) without respect to age. Relative mean bias was expressed as the difference between reader and tester age for the TMA method, and difference between multimodal convolutional neural network predicted age and final age for the FT-NIR method.

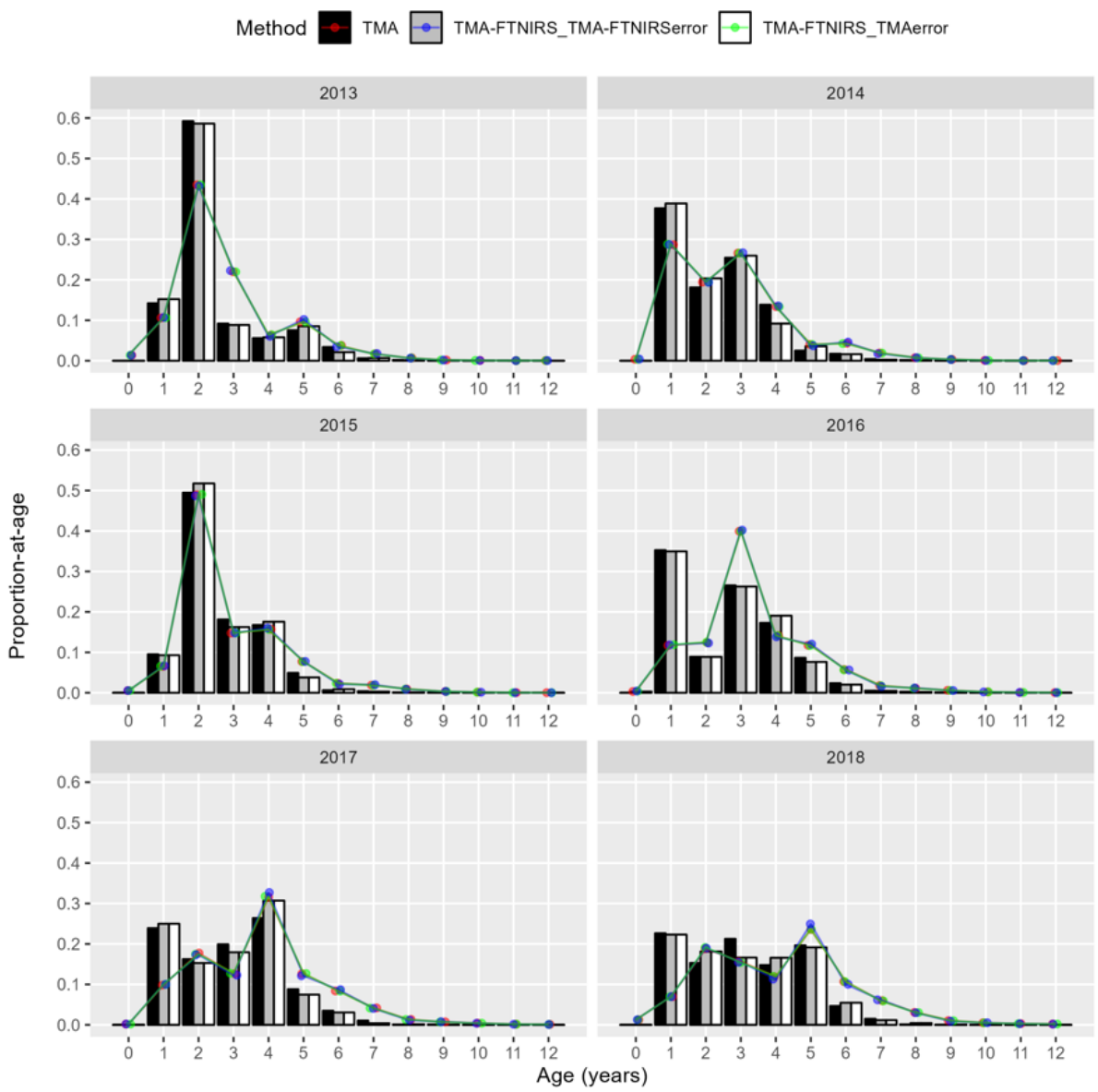


Figure 24.5. -- Pacific cod *Gadus macrocephalus* eastern Bering Sea shelf bottom trawl survey age compositions (expressed as proportions at age by year) generated using a design-based estimators and Stock Synthesis (SS) model fits from three scenarios based on traditional (TMA) and Fourier transform near infrared spectroscopy (FT-NIRS) ageing methods and associated ageing errors. Age compositions are shown as bars and SS model fits to age comps are shown as lines.

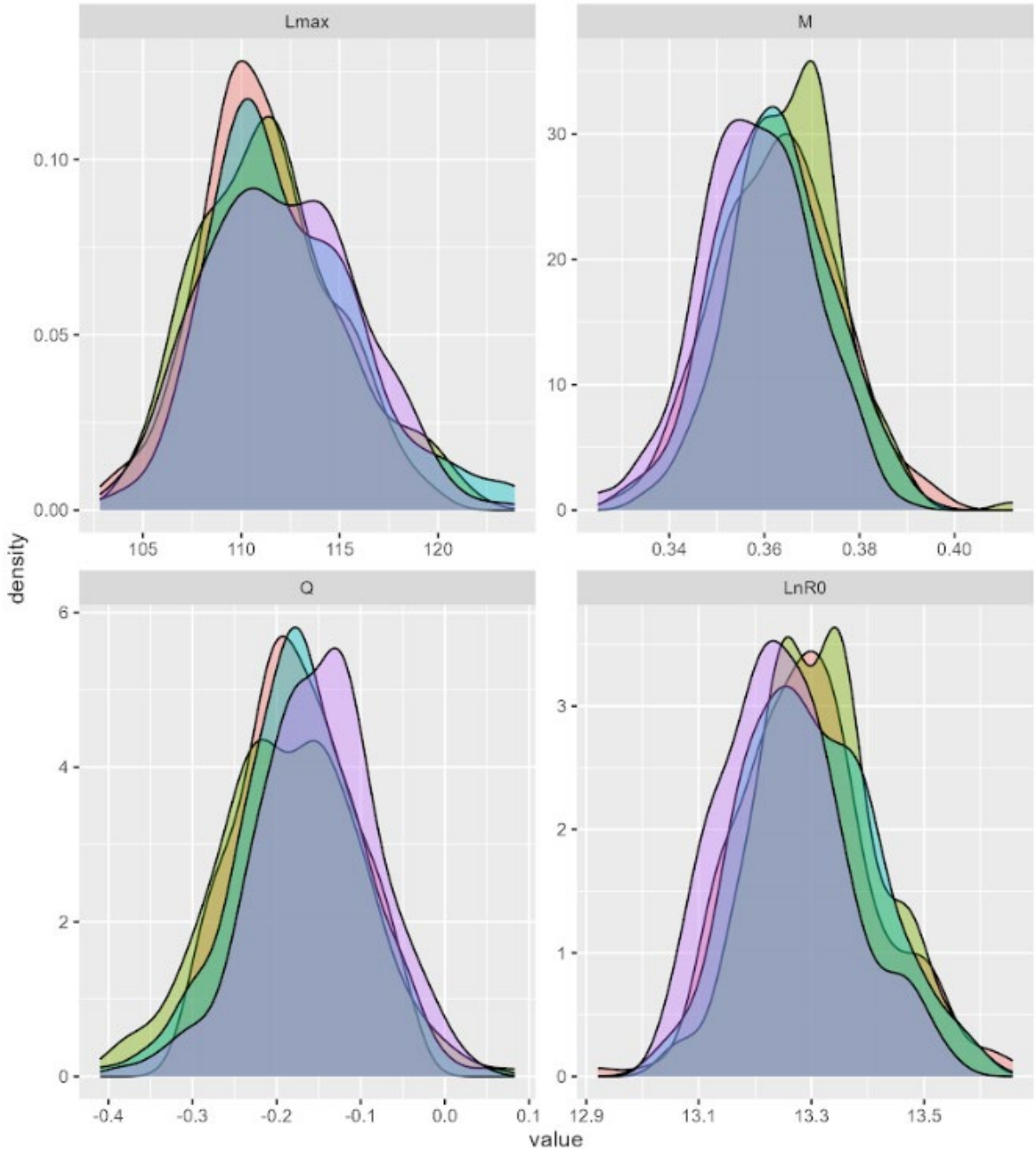


Figure 24.6. -- Marginal posterior kernel densities of internally estimated Pacific cod *Gadus macrocephalus* Stock Synthesis model parameters (*Lmax*, *M*, *Q*, and Ln*R0*) from 1,000 Markov Chain Monte Carlo simulations based on four different model scenarios comparing traditional and Fourier transform near infrared spectroscopy age composition and assumed ageing errors.

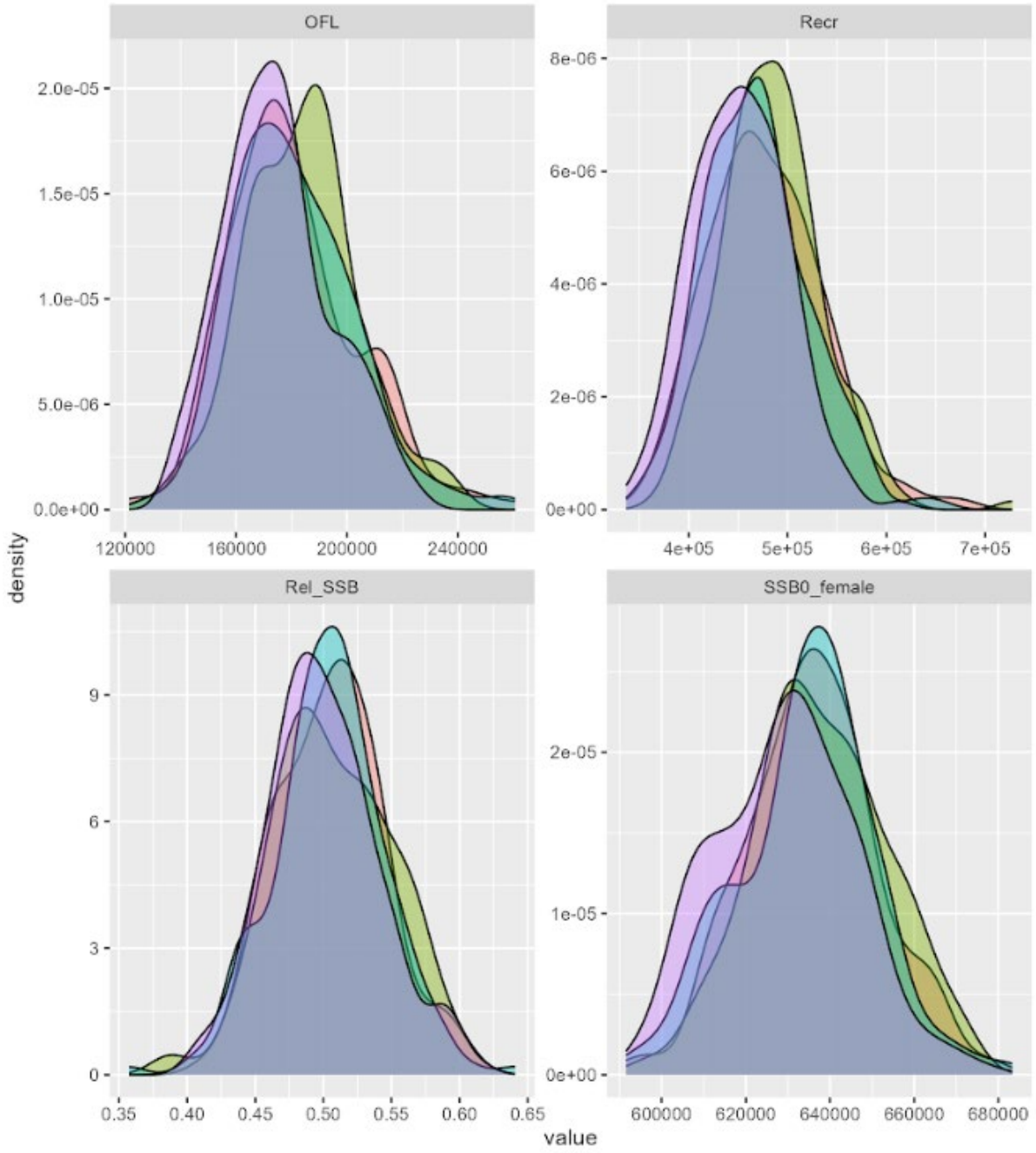


Figure 24.7. -- Marginal posterior kernel densities of externally derived management quantities (overfishing limit OFL, recruitment, relative spawning stock biomass ($SSB/SSB_{unfished}$), and unfished female SSB) in the Pacific cod *Gadus macrocephalus* Stock Synthesis model from 1,000 MCMC simulations based on four different model scenarios comparing traditional and Fourier transform near infrared spectroscopy age composition and assumed ageing errors.

25. Sensitivity of Gray Snapper (*Lutjanus griseus*) Stock Assessment Models to Age Inputs Estimated with Near Infrared Spectroscopy

Steven B. Garner[1,2], Beverly K. Barnett[2], Derek W. Chamberlin[3], Francesca C. Forrestal[4], Thomas E. Helser[5], Irina M. Benson[5], William Kline[2], and William F. Patterson III[3]

[1] Cooperative Institute for Marine and Atmospheric Studies
University of Miami
4600 Rickenbacker Causeway
Miami, FL USA

[2] Southeast Fisheries Science Center
NOAA, National Marine Fisheries Service
3500 Delwood Beach Road
Panama City, FL USA

[3] Department of Fisheries and Aquatic Sciences|
University of Florida
7922 NW 71st Street
Gainesville, FL USA

[4] Southeast Fisheries Science Center
NOAA, National Marine Fisheries Service
75 Virginia Beach Drive
Key Biscayne, FL USA

[5] Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

INTRODUCTION

Modern, age-structured stock assessment models require age data of sufficient quality (i.e., high precision and accuracy), quantity (i.e., sample size), and representativeness (i.e., composition). Standard protocols for developing age estimates from fish ageing structures (primarily otoliths) require substantial sample processing procedures that can take several minutes per sample over multiple days. Thus, the number of samples available for input into stock assessments is dependent upon the processing power of ageing laboratories, while the quality of age estimates is dependent upon the ageing expertise of fishery scientists and species-specific otolith section clarity. Fourier transform near infrared (FT-NIR) spectroscopy has the potential to greatly increase sample processing efficiency in production ageing laboratories. Used in the food and drug industry for decades, FT-NIR spectroscopy has only recently been applied to ageing fish (e.g., Robins et al. 2015, Helser et al. 2019b, Passerotti et al. 2020b), but with great promise. With FT-NIR spectroscopy, all otoliths of interest are irradiated with near infrared light to collect spectral absorbance data. A subset of the scanned otoliths are sectioned and aged with traditional methods. The observed ages are paired with their spectral absorbance data, and a model is developed to predict age for the remainder of otoliths from only the spectral data. Considering that a spectral scan requires approximately one minute to complete, there is enormous potential for increasing production ageing efficiency with this approach. However, the effects of utilizing age estimates generated from FT-NIR spectroscopy on stock assessment outputs need to be explored. Our objective was to evaluate the sensitivity of the Gulf of Mexico gray snapper (*Lutjanus griseus*) stock assessment to FT-NIR spectroscopy-predicted ages by

comparing assessment model outputs under different simulation scenarios in the Stock Synthesis (SS) framework.

METHODS

Sample Selection, Model Building, and Simulations

Gray snapper otoliths were selected from the archive at the National Marine Fisheries Service Panama City Laboratory from the years 2016-2020, with a target sample size of 300 per year. Each otolith was placed distal side up on a quartz sampling window of a Bruker Multi-Purpose Analyzer II FT-NIR spectrometer to collect spectral scans of diffuse reflectance (Robins et al. 2015, Helser et al. 2019b). Samples were irradiated with near infrared light at 16 cm^{-1} resolution with absorbance measured in wavenumbers (cm^{-1}). Following scanning, otoliths were processed using standard protocols and sectioned on a Hillquist high-speed saw to produce a thin-section (~0.5 mm) for traditional ageing. Only whole, clean, unbroken, left otoliths were sampled to maintain consistency among FT-NIR scans. Observed ages were input along with paired spectral data into chemometric software (OPUS, v8.5, Bruker Optics) to develop predictive models using partial least squares regression (Chen and Wang 2001). Individual samples were selected from the calibration set iteratively and cross-validated to develop the best possible predictive model based on root mean square error (RMSE) and the ratio of prediction deviance (RPD). The predictive model built from the calibration set was then applied to the validation set (i.e., the remaining 70% of scanned otoliths) to predict age from spectral scans

without including the observed age estimate. Ages predicted with FT-NIR spectroscopy were then compared to their observed age estimate to compare accuracy and precision.

Simulation Set 1

The FT-NIR spectroscopy-predicted age estimates from the calibration and validation sets were substituted for their corresponding observed ages in the dataset used to assess gray snapper in the most recent stock assessment model (Southeast Data Assessment and Review 75). Ages without an FT-NIR spectroscopy-predicted age were unchanged and retained in the data. The assessment model was rerun to estimate a new base model and then run in the R package ss3sim (Anderson et al. 2014) with n = 10,000 iterations. The ss3sim function fits relevant distributions to the data in the operating (base) model from which to resample the data to the specified number of iterations and reruns the model estimating procedure (estimation model). The SS output of interest was compared to the base model and summarized across all iterations as relative error. Model outputs including spawning stock biomass (SSB), fraction unfished, B/B_{MSY} (ratio of observed biomass to biomass that would provide maximum sustainable yield, MSY), F/F_{MSY} (ratio of observed fishing mortality to the fishing mortality rate that would result in MSY), depletion, SSB at MSY, and unfished biomass and were compared between the base assessment model and the model with substituted FT-NIR spectroscopy-predicted ages. For the base model, ageing error was estimated by comparing observed ages from the primary reader to an expert reader and assuming no bias in the expert reader's age estimates. For the model with FT-NIR spectroscopy-predicted ages substituted for observed ages, ageing error was estimated by comparing the FT-NIR spectroscopy-predicted ages to the expert reader's age estimates. Age

data for 2016-2020 contained both observed ages and FT-NIR spectroscopy-predicted ages; thus, applying FT-NIR spectroscopy ageing error was a conservative approach given that observed ages had less imprecision than FT-NIR spectroscopy-predicted ages.

Simulation Set 2

Due to the low number of FT-NIR spectroscopy-predicted ages (~3-5%) relative to the number of observed ages in each year from 2016-2020, a second set of simulations was conducted with resampled FT-NIR spectroscopy ages. The FT-NIR spectroscopy-predicted ages were resampled (random with replacement) so that year-specific sample sizes of FT-NIR spectroscopy-predicted ages were equal to the number of observed ages in each year (Table 25.1). For example, FT-NIR spectroscopy-predicted ages for 2016 (n = 481) were resampled to equal the number of observed ages from 2016 (n = 3,643). Samples from the calibration and validation set were included in the resampling pool to maximize the age range and sample size comprising the resampled dataset. Each FT-NIR spectroscopy-predicted age was resampled with all of its corresponding data (e.g., observed age). The observed ages that corresponded to the resampled FT-NIR spectroscopy-predicted ages were used to estimate a new base model. The FT-NIR spectroscopy-predicted ages were then substituted for all the observed ages, and the SS model was rerun to get new parameter estimates. The model was then run through the simulation framework for comparison to the base model output from the simulation framework. The model with resampled observed ages (primary vs. expert age comparisons) and the model with the resampled, fully substituted FT-NIR spectroscopy-predicted ages (FT-NIR spectroscopy vs. expert age comparisons) was specified with the same ageing error matrices used in simulation set

1. As with the first set of simulations, only data from 2016-2020 were altered, while data from years prior to 2016 remained unchanged. Individual SS model outputs as well as simulation outputs were compared with relative error. Output was not compared between simulation sets 1 and 2, only within each simulation set.

RESULTS

In total, 1,419 gray snapper otoliths were available for FT-NIR spectroscopy scanning with $n = 420$ samples comprising the calibration set and $n = 996$ samples comprising the validation set (Table 25.1). Observed ages ranged from 2 to 29 yr, while FT-NIR spectroscopy-predicted ages ranged from 0 to 26 yr. The prediction model tended to overestimate age for fish <10 yr and increasingly underestimate age fish >15 yr. The relationship between the observed ages and FT-NIR spectroscopy-predicted ages in the calibration set had an r^2 of 0.92, an RMSE of the cross validation of 1.5 yr, and an RPD of 3.5. The validation set had an r^2 of 0.897, an RMSE of prediction of 1.6 yr, an RPD of 3.1, an average percent error (APE) of 7.6, and an average coefficient of variation (ACV) of 10.7 (Fig. 25.1).

Outputs between assessment models with primary reader ages versus FT-NIR spectroscopy-predicted ages were very similar in single model runs (Fig. 25.2), indicating that utilizing FT-NIR spectroscopy-predicted ages had a negligible effect on reference point estimates of biomass or fishing mortality. For example, relative error (RE) in SSB ranged from 0.002 to 0.007 across all years. Greater differences in reference points were observed between single models that used the resampled datasets, but RE was still low (RE ranged from -0.072 to 0.093 in SSB across all years; Fig. 25.2). The RE in B/B_{MSY} (≤10%) was higher than in F/F_{MSY}

(<5%), but median estimates were very similar between models with FT-NIR spectroscopy-predicted ages either partially (e.g., Fig. 25.3A vs. 25.3B) or fully substituted (e.g., Fig. 25.3E vs. 25.3F) for observed ages. Similarly, summary data indicated RE in median estimates of depletion, SSB at MSY, and unfished biomass were very similar within and between simulation sets when FT-NIR spectroscopy-predicted ages were substituted, partially or fully, for observed ages (Fig. 25.4). The RE in depletion estimates was similarly positively biased in summary data for sets 1 and 2; RE in median estimates of SSB at MSY and unfished biomass were unbiased in both simulation sets (Fig. 25.4).

DISCUSSION

FT-NIR spectroscopy-predicted ages were generally similar to observed ages but with greater imprecision and relatively low bias in the younger and older age classes. FT-NIR spectroscopy-predicted ages were overestimated for ages 2-5 yr and underestimated for ages >15 yr. The APE for FT-NIR spectroscopy-predicted ages versus expert reader ages was nearly three times higher than error between primary and expert readers. Imprecision of FT-NIR spectroscopy estimates for young ages, which had extremely high between-reader agreement, has a disproportionate impact on APE. Assessment output can be sensitive to imprecision in younger age estimates due to cohort smoothing of infrequent strong year-classes. However, bias in older ages has minimal impact on assessment outputs because older fish represent relatively few individuals in the landings or population, and most are contained in the plus group in age composition data for this species (≥21 yr). While, greater imprecision in FT-NIR spectroscopy-predicted ages is a concern, the assessment model was insensitive to this source of error,

especially when the FT-NIR spectroscopy-predicted ages comprised a low proportion of age samples. Results of simulation set 2 indicated low sample sizes of FT-NIR spectroscopy ages did minimize their impact on assessment outputs, but relative error was still low (<10%) when FT-NIR spectroscopy-predicted ages were fully substituted for their corresponding observed age for the last five years of the dataset (2016-2020). Generally, greater imprecision in age estimates will increase uncertainty in and possibly shift assessment outputs, but the stock assessment model appears robust to the increased imprecision associated with FT-NIR spectroscopy-predicted ages, at least on the scale of imprecision examined with gray snapper. Scanning additional otoliths would allow us to fully evaluate effects of utilizing FT-NIR spectroscopy-predicted ages versus observed ages by better refining imprecision between traditional and FT-NIR spectroscopy methods. Furthermore, scanning a larger number of younger and older individuals may reduce or remove the bias observed for these age classes at the sample sizes used in this study.

Table 25.1. -- Number of primary reader versus Fourier transform near infrared (FT-NIR) spectroscopy-predicted ages per year from 2016 to 2020 in each dataset. The simulation 1 dataset contains the FT-NIR spectroscopy-predicted ages (n = 999) from the validation set of scanned otoliths, while the simulation 2 dataset contains the FT-NIR spectroscopy-predicted ages from the calibration and validation sets resampled to match the year-specific sample sizes aged by the primary reader. Ages from prior to 2016 were unchanged in both datasets. FT-NIR spectroscopy-predicted ages were resampled with all corresponding data so that each had a paired observed age for re-estimating the base model that utilized resampled the data.

Simulation 1 dataset	Year	2016	2017	2018	2019	2020
Observed ages	n	3,643	2,773	3,795	4,109	1,861
NIRS-predicted ages	n	347	215	136	220	78
Simulation 2 dataset						
Observed ages	n	3,643	2,773	3,795	4,109	1,861
NIRS-predicted ages	n	3,643	2,773	3,795	4,109	1,861

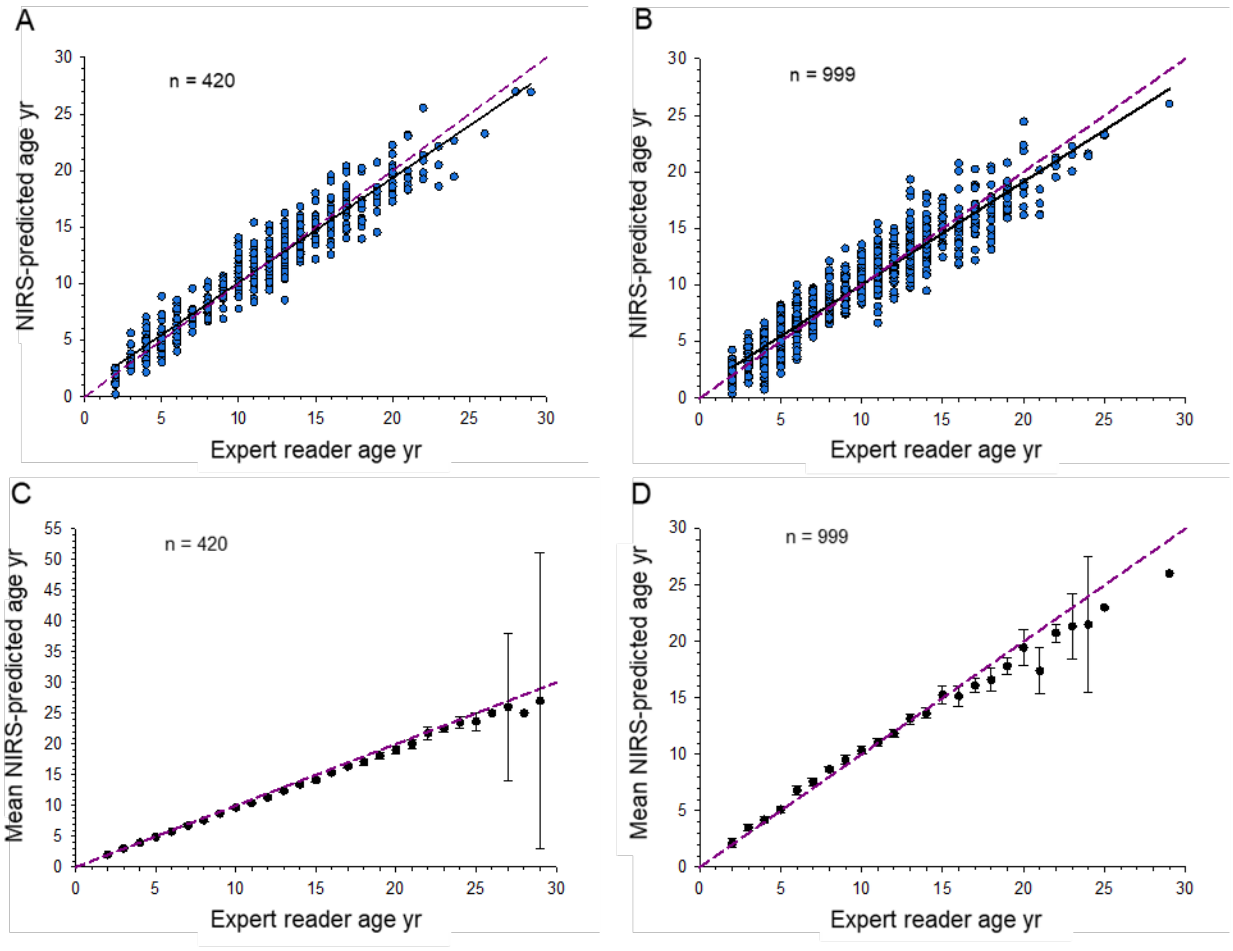


Figure 25.1. -- Observed (expert reader) age (yr) versus Fourier transform near infrared spectroscopy (NIRS) predicted age (yr) for the A) calibration and B) validation sets of gray snapper (*Lutjanus griseus*) otoliths. Panels C and D indicate age-specific means (with 95% CIs) for the calibration and validation sets, respectively. NIRS-predicted ages were rounded to the nearest integer age. Sample sizes are shown at the top left of each panel.

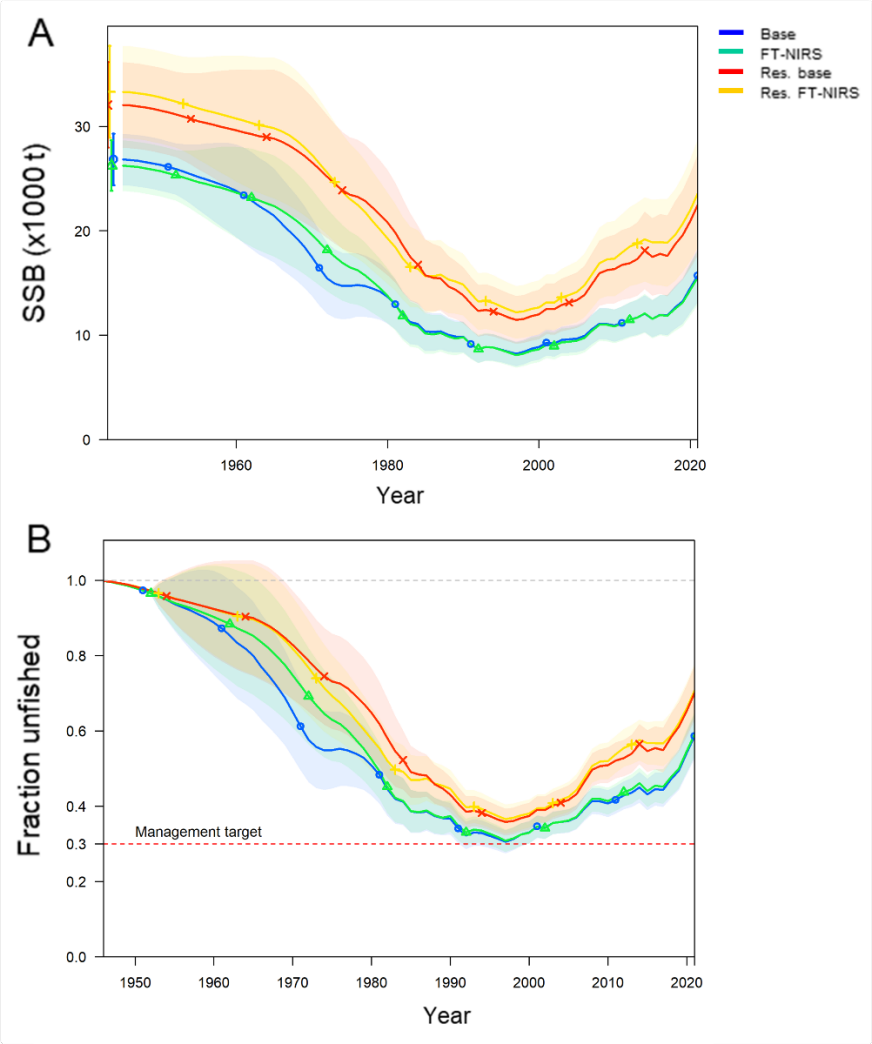


Figure 25.2. -- Single model output for A) spawning stock biomass (SSB, tons) and B) fraction unfished by year from each gray snapper (*Lutjanus griseus*) stock assessment model used in simulation sets 1 or 2. Models include the 1) primary reader ages (Base), 2) Fourier transform near infrared spectroscopy, FT-NIRS (FT-NIRS substituted ages), 3) resampled primary reader ages (Res. base), or 4) fully substituted, resampled FT-NIRS-predicted ages (Res. FT-NIRS). Models 1 and 2 should not be compared with models 3 and 4 as they utilize different datasets but are shown together for simplicity. Color-corresponding shaded regions indicate model uncertainty (95% CI) in the estimate of SSB or fraction unfished.

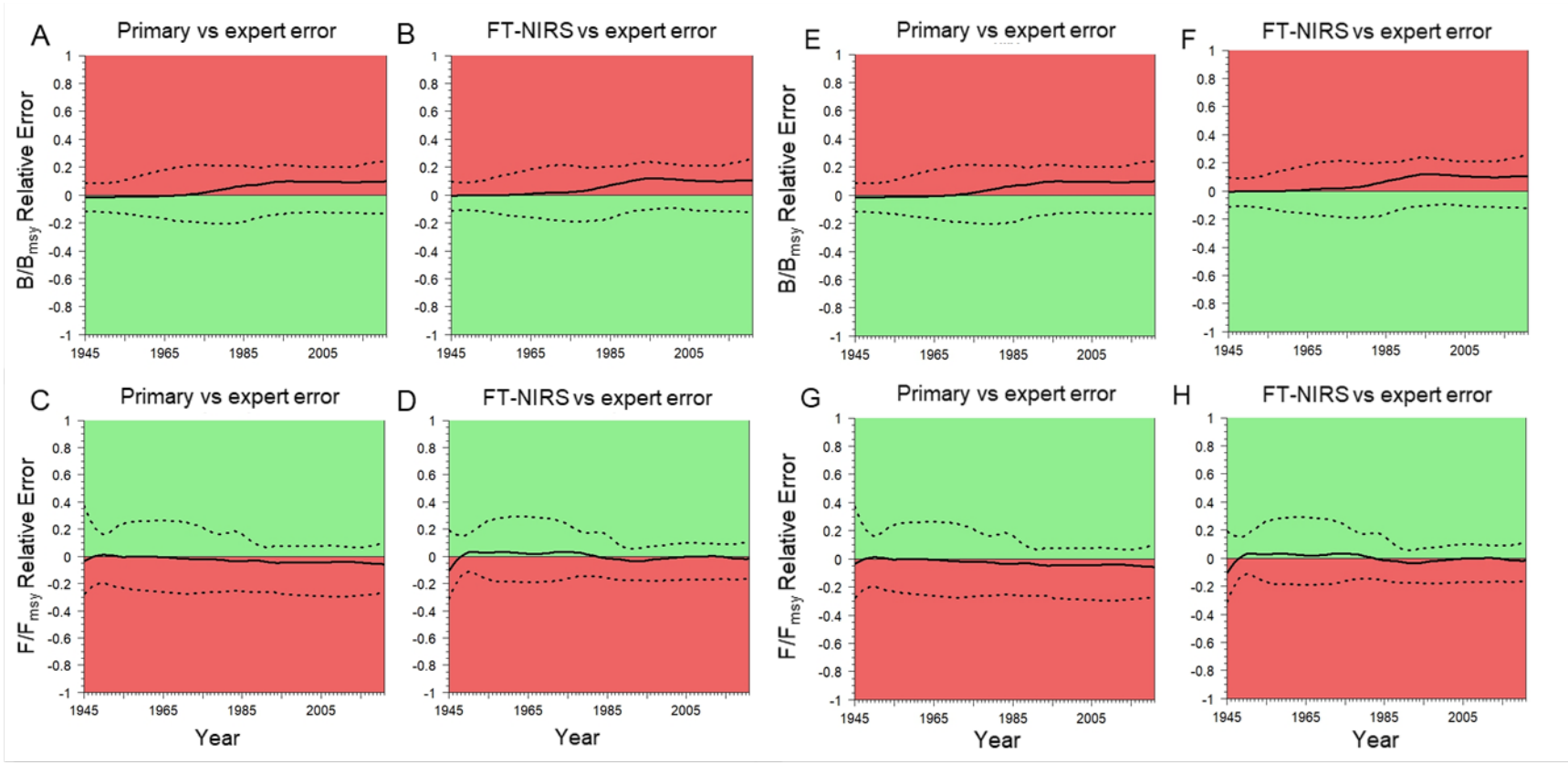


Figure 25.3. -- Relative error in reference points (B/B_{MSY} or F/F_{MSY}) from simulation set 1 (observed vs. Fourier transform near infrared spectroscopy (FT-NIRS)-predicted ages; Panels A-D) or simulation set 2 (resampled observed vs resampled FT-NIRS-predicted ages; Panels E-H). Solid black lines show the median values while the dashed lines indicate the 95% quantiles for each set of simulation runs (n = 10,000 iterations).

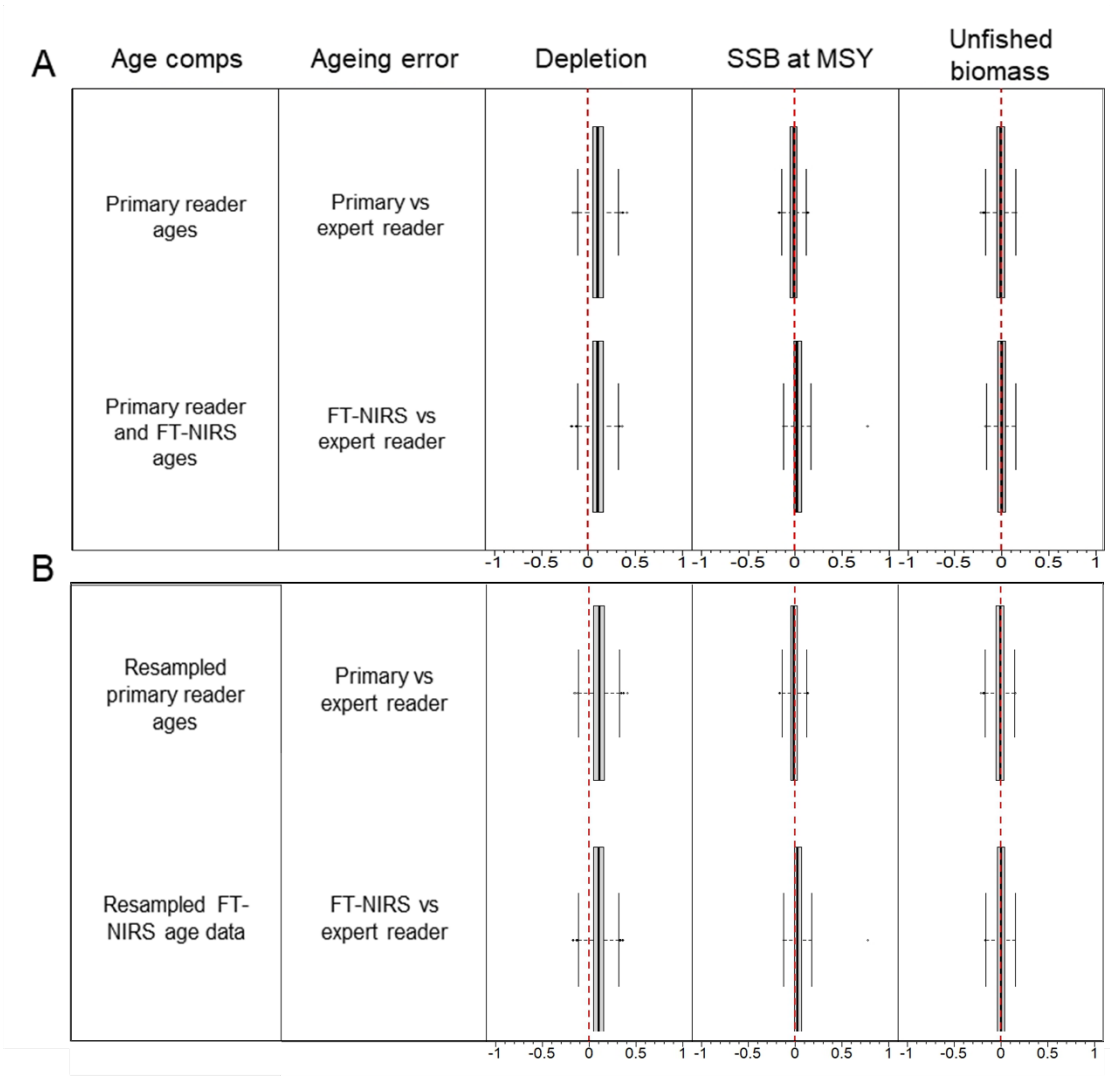


Figure 25.4. -- Boxplots of relative error in median values for model outputs (depletion, spawning stock biomass SSB at maximum sustainable yield MSY, or unfished biomass) from gray snapper (*Lutjanus griseus*) stock assessment models in A) simulation set 1 (primary reader ages vs. substituted Fourier transform near infrared spectroscopy (FT-NIRS)-predicted ages) and B) simulation set 2 (resampled primary reader vs fully substituted, resampled FT-NIRS-predicted ages). Box ends indicate the 25th and 75th quantiles while whiskers indicate the 5th and 95th quantiles. The solid black line on each box indicates the median value while red dashed lines indicate the 0 value of relative error.

26. Quality Control and Assurance of Reference Age Data

at the Alaska Fisheries Science Center

Mary Elizabeth Matta, John D. Brogan, Jonathan A. Short, and Thomas E. Helser

Alaska Fisheries Science Center
NOAA, National Marine Fisheries Service
7600 Sand Point Way NE
Seattle, WA USA

INTRODUCTION

Fish age data are a critical component of fisheries stock assessment models. Every year, the Age and Growth Program at the Alaska Fisheries Science Center (AGP-AFSC) is tasked with providing upwards of 30,000 age estimates to support stock assessments of commercially important groundfish species in the federal waters of Alaska. Historically, the AGP-AFSC has met these requirements through microscopic examination and visual interpretation of otolith growth patterns, but we are in the process of implementing the application of Fourier transform near infrared (FT-NIR) spectroscopy to rapidly predict age from otolith spectral scans for some of our key stocks. Models that predict age from FT-NIR spectra of fish otoliths are reliant on high-quality reference age data. The AGP-AFSC has long had in place a framework to maintain quality control and assurance of fish age data (Kimura and Anderl 2005), a process that will continue following adoption of the FT-NIR spectroscopy method to check model performance and preserve consistency in reference information over time. Here, using bottom-trawl survey collections of walleye pollock (*Gadus chalcogrammus*) otoliths from the eastern Bering Sea as an example, we describe how traditional (microscopic) age estimates are generated and vetted by the AGP-AFSC prior to being included in predictive models of age from FT-NIR spectra (Helser et al. this volume-a).

METHODS

Otoliths were collected from walleye pollock during the 2019 and 2021 AFSC Resource Assessment and Conservation Engineering (RACE) summer bottom trawl surveys in the eastern

Bering Sea and preserved in glycerin-thymol solution until age determination in the laboratory. Otoliths were aged from the otolith surface or using the break-and-burn technique, wherein otoliths are sectioned transversely through the core using a Buehler Isomet low-speed saw and heated using an alcohol burner or toaster oven to enhance contrast between growth zones (Matta and Kimura 2012). Otolith annuli were enumerated and recorded as integer ages in the AGP-AFSC database, *AGE3*.

Otoliths were assigned an initial age estimate (termed the "Read Age") by age readers that had prior experience and had demonstrated proficiency in ageing walleye pollock. A random test subsample of 20% was aged independently by a second experienced age reader ("Test Age") to evaluate precision and bias between readers (Kimura and Anderl 2005). Because survey otolith collections are rather large, they were split into sub-collections among several reader-tester pairs in each year. The precision statistics percent agreement (PA), average percent error (APE; Beamish and Fournier 1981), and Chang's coefficient of variation (CV; Chang 1982) were calculated. Bias was evaluated statistically using Bowker's test of symmetry (Bowker 1948, Kimura and Anderl 2005) and visually using contingency tables and age bias plots (Campana 1995). Discrepancies between readers and testers were discussed and resolved. In cases where systematic differences were detected, specimens were re-aged by the reader and reevaluated by the tester as a data check. Any ages that were reconciled to an age that differed from the original Read Age were indicated as a separate field ("Update Age") in *AGE3*. Upon completion of age reading, all ages were finalized. In cases where there was an Update Age, it became the "Final Age"; otherwise, the Read Age was accepted as the Final Age.

Prior to releasing age data to end users, length-at-age observations were graphically examined for potential outliers. Otoliths identified as outliers were reexamined by the reader to

determine if there were any obvious age reading errors and compared to otoliths from fish of similar body length to ascertain instances of collection recording error at sea; corrections were made as necessary, and fast- or slow-growing fish were noted in *AGE3*. One last inspection was made on each age collection by the production team lead to ensure that Final Age estimates were satisfactory. The team lead examined test results to ensure that the level of precision was acceptable and that any biases were addressed, and also confirmed that dominant year classes were consistent with previously aged collections. Following the team lead's assessment, Final Age data were provided to the RACE survey team and stock assessment scientists to perform age-length expansions and incorporate age information into the statistical age-structured model for the Bering Sea walleye pollock stock (Ianelli et al. 2022).

RESULTS

The 2019 and 2021 Bering Sea walleye pollock RACE bottom trawl otolith collections were split into smaller sub-collections of ten and seven parts, respectively. Sub-collections consisted of 78-280 otolith pairs and comprised a variety of reader-tester combinations. Readers' experience level in walleye pollock age determination varied from <1 year to over 30 years, and all testers had at least 2 years of experience.

The overall precision between paired readings was higher in 2019 (PA = 65%, APE = 3.19%, CV = 4.51%) than in 2021 (PA = 56%, APE = 4.34%, CV = 6.14) (Table 26.1). This was likely attributable to several factors: the average age of the 2021 survey collection was older (6.44 years for the full collection, compared to 5.87 years in 2019), the growth patterns in this year were judged to be more difficult to interpret due to checks and unusual amounts of growth

on the otolith edge given the time of capture, and one of the age readers in 2021 was relatively new at walleye pollock age determination. During the age determination process, a bias was detected between the new reader and her testers on the two age sub-collections she was tasked with reading. Other experienced age readers were consulted to provide advisory ages, and it was agreed that the tester ages were likely more reliable. The bias appeared to be limited to specimens approximately 6 years and older; consequently, the reader re-aged all untested specimens that were originally aged as ≥6 years old, and the testers aged an additional 20-50% of these re-aged specimens as a quality control check. Ages were reconciled to Final Age estimates based on these results. The effects of the testing and reconciliation process are evident looking at the pooled test sample results for 2019 and 2021 (Fig. 26.1). Figure 26.1a shows paired Read and Test Age estimates, with a clear bias occurring after age-8 for the 2021 collection that was also confirmed with Bowker's test of symmetry (Table 26.1). Figure 26.1b shows paired Test and Final Age estimates, where there is no evidence of systematic bias (Bowker's $\chi^2 = 22.25$, df = 17, $p = 0.18$).

Despite the difficulty associated with visual age interpretation for these samples, overall precision between reader and tester in both years was considered satisfactory based on comparison to precision values from previous historical collections (Fig. 26.2). Furthermore, age readers successfully identified the dominant 2013 year class (Ianelli et al. 2022) in both collection years (Fig. 26.3), providing corroborative support for the age estimates (Kimura et al. 2006). Estimated length-at-age was similar in both 2019 and 2021 (Fig. 26.4). In total, 20 observations were identified by age readers as outliers and were deemed to represent fast- or slow-growing fish. Only two observations were identified as probable collector errors at sea based on the size of the otolith relative to the recorded fish length.

DISCUSSION

This example illustrates the power of the random test subsample as an early detector of problems with age interpretation prior to providing data to end users. Further, splitting otolith collections into sub-collections among multiple readers and testers can reduce overall bias while providing data to end users in a timelier manner. As the AGP-AFSC workforce ages and employee onboarding and retention is becoming increasingly hampered by budgetary constraints, these tools are perhaps our best assets in terms of maintaining high quality age data. Continuing the test subsample protocol will also ensure consistency in reference age data and can quantify prediction error following implementation of the FT-NIR spectroscopy method (Helser et al. this volume-a).

The AGP-AFSC is currently investigating making slight improvements to how we evaluate quality and check traditional age data for errors. We will continue to calculate PA, APE, and CV and to visually assess bias, but we may discontinue Bowker's test of symmetry in favor of McNemar's and Evans and Hoenig's tests of symmetry, which have recently been shown to outperform Bowker's test under certain scenarios (McBride 2015, Nesslage et al. 2022). We are also investigating ways to more objectively detect age-length outliers.

In addition to the quality control procedures outlined in this case study, our program also employs several forms of quality assurance for assessed Alaska stocks. Prior to producing age estimates, new age readers are trained extensively using a combination of physical reference collections, annotated image libraries, and written and visually recorded species-specific age determination manuals. All new age readers are initially tested at a rate of 100% until they become proficient at ageing a given species. The AGP-AFSC is also an active participant in the

Committee of Age Reading Experts (CARE), a coalition of international, state, and federal agencies from the U.S. West Coast and Canada. The CARE group conducts biennial meetings and performs regular blind structure exchanges to standardize and maintain consistency in age determination methodology across laboratories. Finally, the AGP-AFSC has completed a number of age verification and validation studies to assess and quantify accuracy of our age estimates for assessed stocks, including for walleye pollock (Kastelle and Kimura 2006, Kimura et al. 2006), the subject of the largest commercial fishery in Alaska (Ianelli et al. 2022).

Though implementation of the FT-NIR spectroscopy method will result in smaller numbers of traditional microscopic age estimates, we anticipate that the general workflow for maintaining quality control of the reference age estimates will remain much the same. The AGP-AFSC will continue to traditionally age and test subsample a certain percentage of otolith collections to evaluate precision and bias in reference age data and to check FT-NIR spectroscopy model prediction performance. In this way, we can ensure consistency of age estimates as we transition to this new technology.

Table 26.1. -- Paired otolith age reading results for walleye pollock *Gadus chalcogrammus* from the 2019 and 2021 Bering Sea bottom trawl survey. *Year* is year of collection, *n* is the number of paired readings, $\overline{T}$ is the average tester age, *PA* is percent agreement, *APE* is the mean average percent error, *CV* is the mean Chang's coefficient of variation, and χ^2, *df*, and *p* are the test statistic, degrees of freedom, and *p* value for Bowker's test of symmetry.

Year	*n*	$\overline{T}$	PA (%)	APE (%)	CV (%)	$\chi^2(df)$	*p*
2019	307	5.86	64.82	3.19	4.51	23.69 (17)	0.128
2021	359	6.57	56.27	4.34	6.14	50.09 (28)	0.006

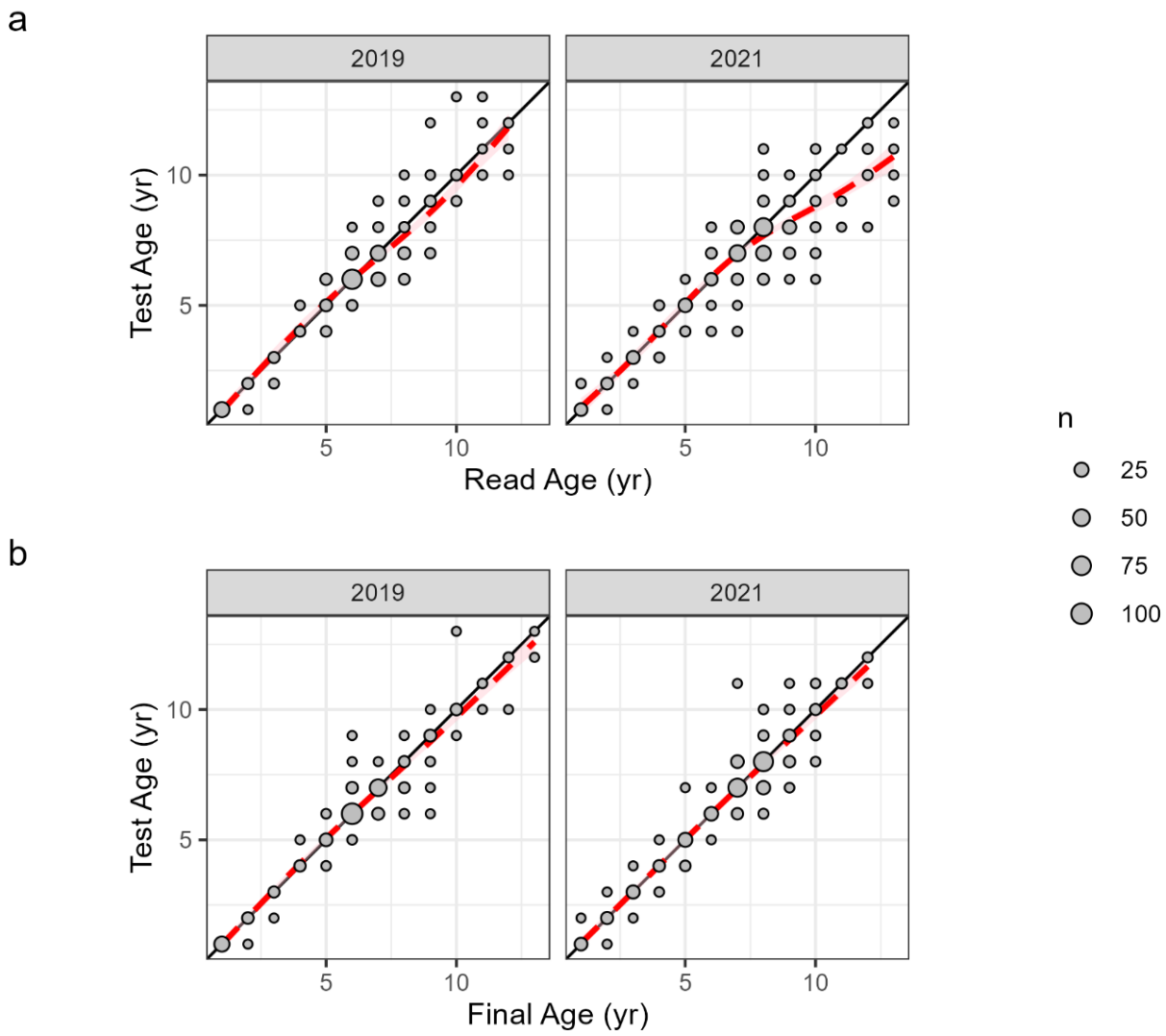


Figure 26.1. -- Comparisons of Read and Test Age (a) and Final and Test Age estimates (b) for walleye pollock *Gadus chalcogrammus* otoliths collected from the eastern Bering Sea bottom trawl survey in 2019 and 2021. The black solid diagonal lines represent a 1:1 relationship. The red dashed lines represent fitted generalized additive models. The size of the points represents the number of observations.

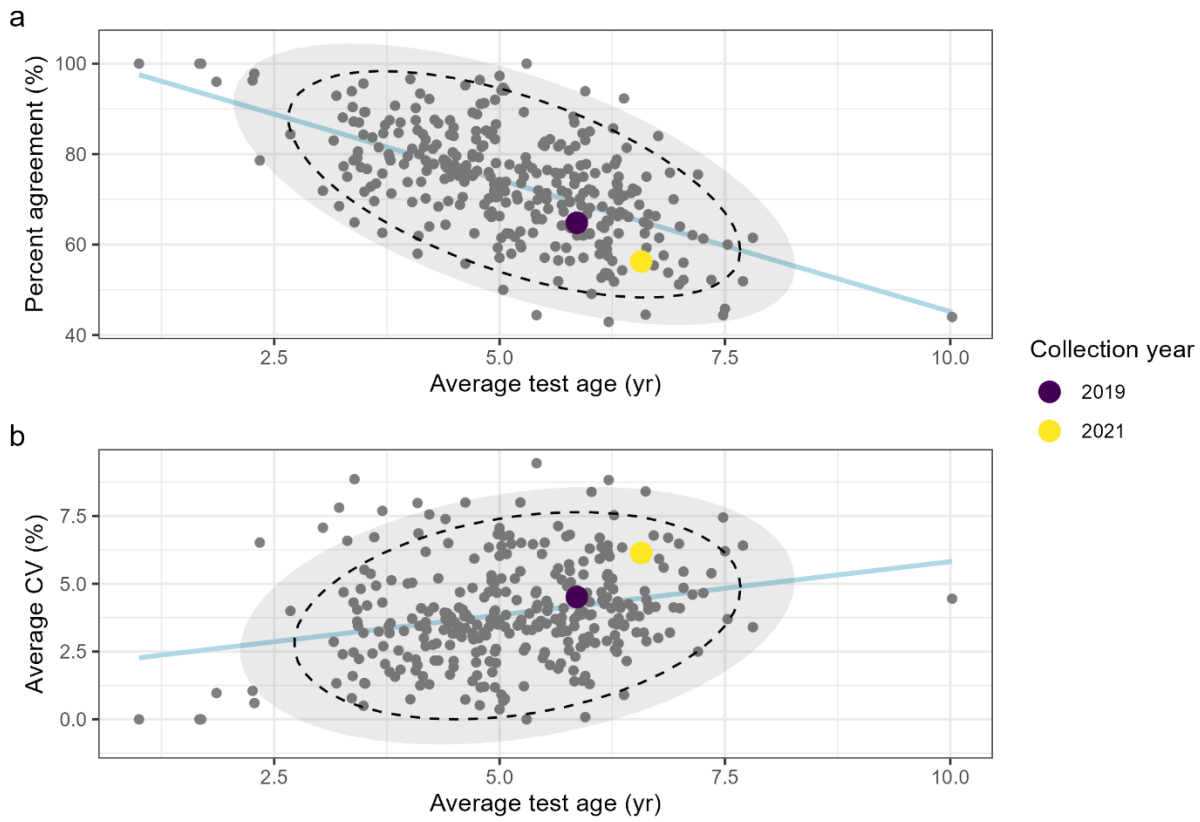


Figure 26.2. -- Comparison of reader agreement metrics between the 2019 (purple), 2021 (yellow), and historical collections of walleye pollock *Gadus chalcogrammus* (gray points). Average Test Age is shown versus a) percent agreement and b) mean values of Chang's coefficient of variation (CV). The dashed lines and gray areas indicate 95% and 99% confidence ellipses, respectively. Blue lines indicate fitted generalized additive models (GAMs). Points outside the ellipse but above the GAM in the case of percent agreement or below the GAM in the case of CV would still be desirable outcomes because they indicate better precision than expected for this species.

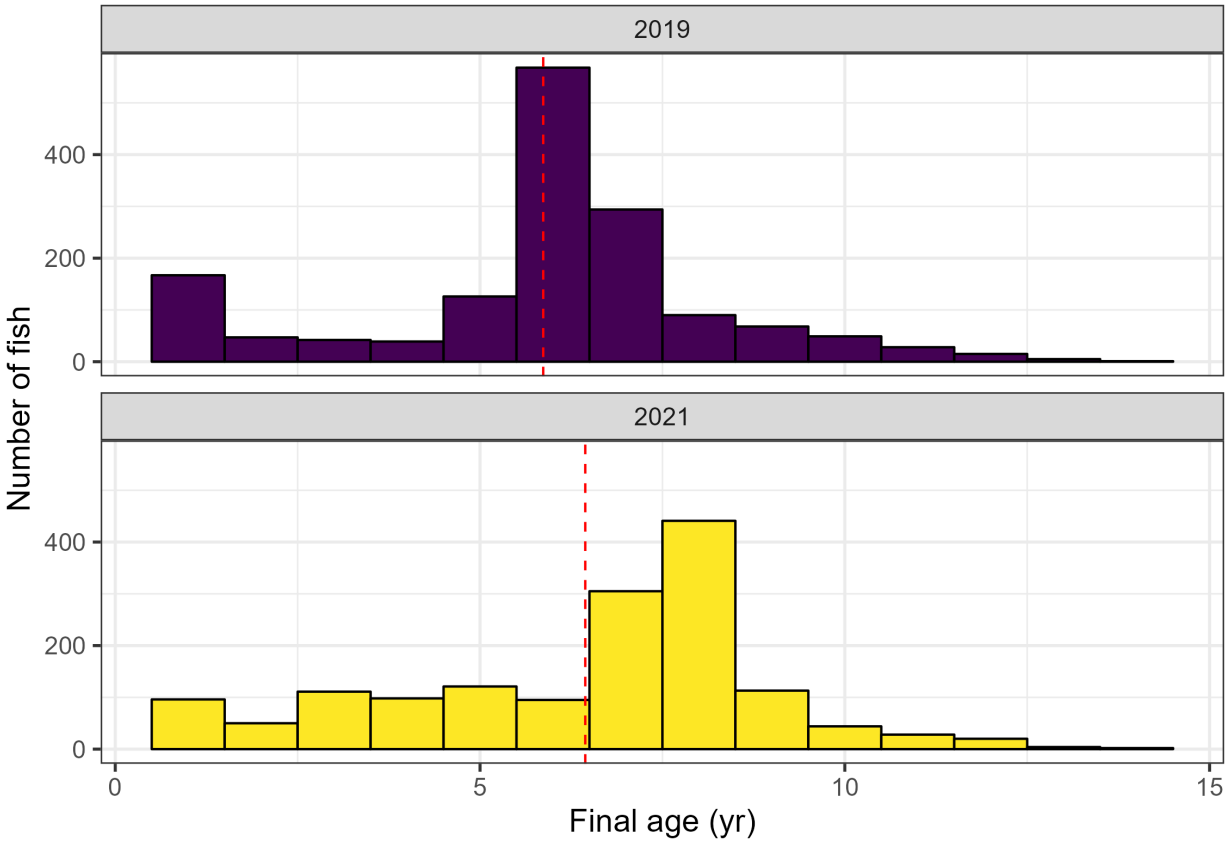


Figure 26.3. -- Finalized age distributions of walleye pollock *Gadus chalcogrammus* otoliths collected from the 2019 and 2021 bottom trawl surveys in the eastern Bering Sea. Dashed red lines represent the mean age values in each year.

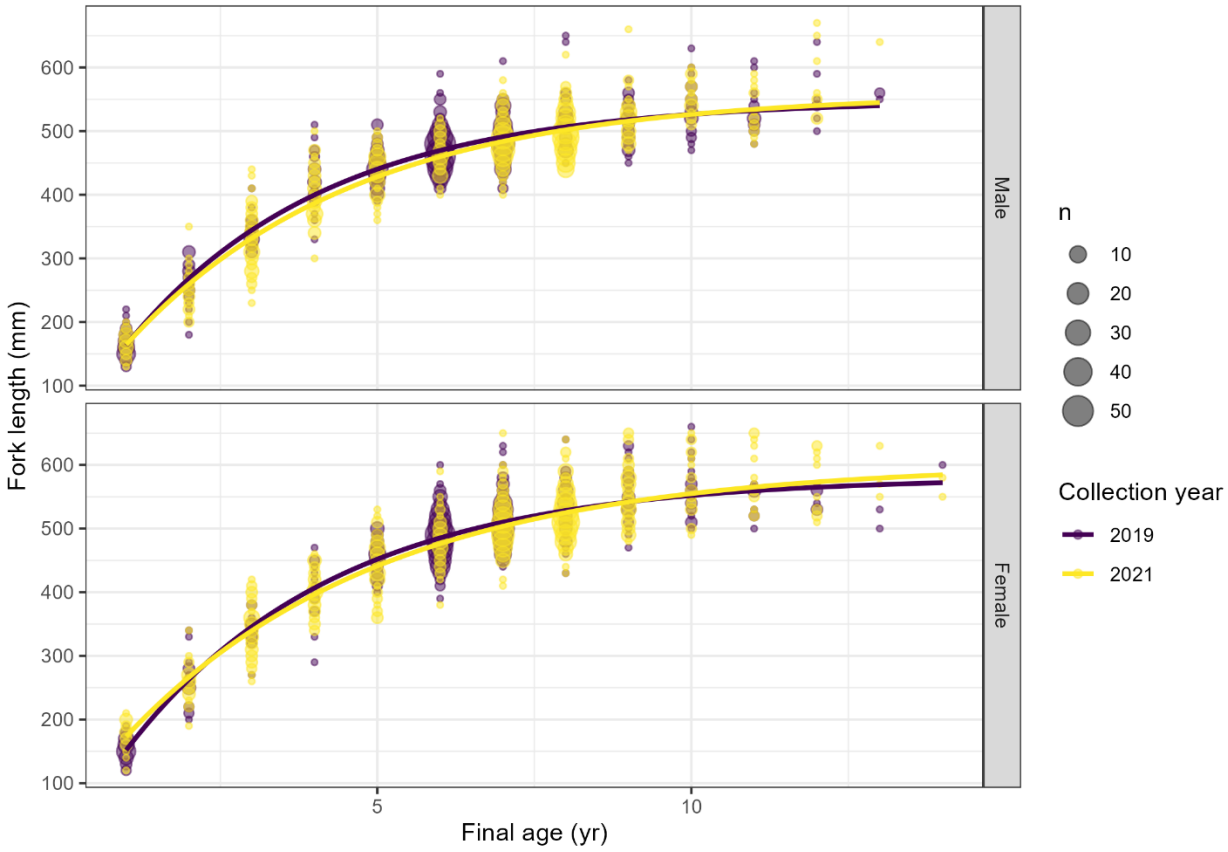


Figure 26.4. -- Observations of length at age fitted with von Bertalanffy growth functions for male (top panel) and female (bottom panel) walleye pollock *Gadus chalcogrammus* from the 2019 (purple) and 2021 (yellow) eastern Bering Sea bottom trawl survey.

27. Quality Assurance and Quality Control of Fourier Transform Near Infrared Spectroscopy Data for Age Prediction

Esther D. Goldstein[1], Brenna C. Hsieh[1], Morgan B. Arrington[1,2], and Thomas E. Helser[1]

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

[2] Cooperative Institute for Climate, Ocean, and Ecosystem Studies (CICOES)
University of Washington
3737 Brooklyn Ave NE
Seattle, WA USA

INTRODUCTION

Implementation of Fourier transform near infrared (FT-NIR) spectroscopy to obtain fish ages for stock assessment requires rigorous quality assurance and quality control (QA/QC) practices to ensure that new data sources maintain the caliber of traditional age data. Unlike traditional ageing approaches with established QA/QC protocols (Bland and Altman 1986, Kimura and Lyons 1991, Matta and Kimura 2012, McBride 2015, Ogle et al. 2021), spectral data for fish age prediction lack well-defined best practices. Additionally, fish ages derived from FT-NIR spectroscopy require spectral acquisition from otoliths and subsequent calibration model development (Helser et al. 2019b), eliciting the need for QA/QC at multiple junctures and increasing the potential to propagate error related to data quality.

Quality assurance and quality control of spectral data must address several potential issues:

(1) Spectral quality related to the specific sample such as crystallization, broken otoliths, or tissue remnants obscuring the otolith.
(2) Spectral quality related to instrument issues such as instrument drift, maintenance requirements, or malfunction.
(3) User errors such as sample presentation on the scan window or data entry errors.
(4) Differentiation of unusual spectra that expand the data domain and may be important for inclusion in the calibration model for age prediction (Wise and Roginski 2015).

Issues 1 and 2 can primarily be recognized through visualization and analysis of spectra alone to identify unusual patterns and outliers (Pomerantsev and Rodionova 2014). Issues 3 and

4 may require exploration of spectra in relation to other data such as age or body length to recognize mismatches between spectra and specimen information (Rodionova and Pomerantsev 2020), an approach that is often applied to traditional age data (see Matta et al. this volume-a).

With the goal of developing QA/QC tools to address the four issues outlined above, we present a first step toward standardized QA/QC practices and reporting for spectral data acquisition prior to calibration model development. We also offer suggestions for next steps to obtain data from spectral outliers and considerations for unusual (henceforth, "extreme") spectra that do not qualify as outliers. The proposed best practices build from user experience throughout the National Marine Fisheries Service-funded strategic initiative, particularly at the Alaska Fisheries Science Center (AFSC), and borrow insights from the spectroscopy discipline as well as QA/QC of traditional age data.

METHODS

We present a case study using walleye pollock (*Gadus chalcogrammus*) collected during fishery-independent surveys conducted by the AFSC in the Bering Sea from 2014-2021 to demonstrate FT-NIR spectra QA/QC. Fish were measured (fork length) at sea and traditional age data were collected following standard protocols (Matta and Kimura 2012). Spectra were collected from otoliths using a Bruker MPA II FT-NIR spectrometer (Bruker Optics, Ettlingen, Germany) with a wavenumber range between 11,500 to 4,000 cm^{-1} and a 22-mm diameter sample window. An *a priori* decision was made to focus only on fisheries independent data and spectral data collected from a single spectrometer. While datasets obtained from multiple spectrometers and data sources can be combined for analysis, fishery-independent and dependent

data may encompass different data domains (e.g., fish size or age), and spectra collected from different instruments may require additional data processing (Oliveri et al. 2013). Since the goal of QA/QC is to identify poor quality data with scrutiny, we opted to exclude these extraneous and potentially confounding factors from the work flow during this initial development. Future work could evaluate the inclusion of spectra from different instruments and from fishery-dependent surveys. Data were collected with 64 scans per sample, a resolution of 16 cm^{-1}, and a scanner velocity of 7.5 kHz. Detailed scan methods and sample presentation are presented in Benson et al. (2020). Wavenumbers >7,500 cm^{-1} are primarily noise and were excluded from the analyses. For analyses with pre-processed data, a Savitsky-Golay smooth (first-derivative, 17-point, 2^{nd} order polynomial; Benson et al. 2020) was used.

QA/QC practices (see Fig. 27.1 for schematic) were developed using multivariate analyses that are frequently applied to spectral data due to data multicollinearity: principal component analysis (PCA) and partial least squares regression (PLS) (Carrascal et al. 2009, Haenlein and Kaplan 2004, Kvalheim 2010). PCA is an unconstrained ordination technique that reduces dimensionality in the spectral dataset, whereas PLS decomposes both X (spectra) and Y (response variable) matrices while constraining the components to maximize covariance between X and Y. PCA and PLS provide complementary information about spectral data for QA/QC. PCA potentially addresses issues 1 and 2 (outlined above) by providing a view of spectral characteristics. PLS potentially addresses issue 3 and 4 (outlined above) by offering additional information about whether new spectra relate to a meaningful characteristic (response variable) in an expected way. For example, traditional age data may not be available for newly acquired spectral data, but relationships between spectra and fish length, which is presumed to increase

with age, should follow a consistent trend among all data sets and may provide a QA/QC check in the absence of age data.

Both PCA and PLS QA/QC are context dependent such that outlier spectra may only be identifiable when they are compared with other data. This is particularly true for potential instrument issues that may affect a large portion of a dataset or the majority of recently scanned data. Therefore, we developed QA/QC practices that compare "new" data with previously collected spectra. To emulate a typical data collection process, we selected 2019 and 2021 as focal years that represent "new" data. We then built PCA and PLS models from data excluding each focal year in turn, using the larger dataset from 2016-2021. For PCAs, meaningful components were selected based on the Broken Stick Model (Legendre and Legendre 2012), or if only the first principal component met the criteria, then the first two were retained in the PCA.

When FT-NIR spectroscopy is operationalized, "new" spectral data lack paired age data from traditional methods. Therefore, we built our base QA/QC PLS models (on all but "focal" years) to relate spectra with fish length. Optimal components for PLS models were selected by comparing root mean square error of cross validation among models. If the optimal number of components exceeded 10, then components were further reduced by selecting only those that explained greater than 0.5% of the variability in fish length. This further reduction in components reduced model overfitting (modeling noise in the spectral data), but alternative approaches for selecting components should be explored in future studies. We retained all wavenumber variables >7500 cm^{-1} in the final models rather than incorporating variable selection (Farrés et al. 2015) because the focus of this analysis was not to optimize length prediction, but rather to provide a benchmark for assessing "new", incoming spectral data quality. Outliers were also removed from the base PLS model dataset following the iterative

procedures using the "robust" approach outlined in (Rodionova and Pomerantsev 2020). This was to ensure that PLS models for comparison of "new" data were representative of length-spectra relationships. These same outliers were excluded from the PCA dataset so that the both baseline PCA and PLS models contained equivalent data. We then utilized the respective PCA and PLS models that excluded our focal years to predict (or project in the case of PCA) our new spectral data and determine outliers (Fig. 1).

PCA and PLS were paired with multivariate distance metrics (Hotelling's T^2 and Q residuals) and comparisons of observed versus predicted values (henceforth, Y-distance) to quantitatively determine extreme and outlier data points in "new" scan data (Legendre and Legendre 2012, Wise and Roginski 2015). Hotelling's T^2 is a multivariate extension of a Student's *t*-test and provides information about how far a data point is from the center scores of a PCA or PLS model. Q residuals are the magnitude of variation (distance) between the actual data and the model predictions (PCA or PLS) in projected space. Y-distance applies to only PLS since PCA is not a predictive modeling approach. Extreme and outlier spectra were defined using outlier analyses, following Pomerantsev and Rodionova (2014) for PCA and Rodionova and Pomerantsev (2020) for PLS as implemented in R package mdatools (Kucheryavskiy 2020b) using statistical moments (mean and standard deviation) and a 5% significance threshold for extreme spectra and 1% threshold for outliers based on a chi-square distribution. Analyses were conducted using R statistical software (R Core Team 2023) using packages "prospectr" (Stevens and Ramirez-Lopez 2022), "mdatools" (Kucheryavskiy 2020a), and "opusreadr2" (Baumann et al. 2023). Standardized QA/QC reporting for record keeping and reproducibility were developed using Quarto implemented via Posit (https://docs.posit.co/).

RESULTS AND DISCUSSION

Outlier detection case studies for Bering Sea walleye pollock from 2019 and 2021 highlight the utility of employing multiple tools to discern spectral data quality. We present a detailed description and example of interpretation of the QA/QC process from the 2021 dataset. We do not discuss the 2019 dataset in detail, but an example QA/QC draft report using the 2019 dataset is included in Appendix A of these proceedings.

Dataset Metadata

QA/QC reporting begins with a summary of metadata for the "new" scans. This includes the total number of scans, the number of scan sessions, and the number of scans per session (see Appendix A Tables A1 and A2 for examples). At the AFSC, scan sessions typically refer to a group of samples that were scanned by a single individual. For implementation of the QA/QC procedure, each scan session could be checked upon completion and/or QA/QC procedures can be implemented for a full collection year once all sessions are complete. This decision may depend on scanning timelines or collection sizes for different species. For the 2021 dataset, there were a total of 1,500 scans and 7 scan sessions that range in sample size from 133-277 otoliths, along with a small proportion of otoliths that were unscannable or had other identified problems such as damage, crystallization, or dried tissue on the otolith.

QA/QC Using PCA – 2021 Collection Year

PCA plots of unprocessed spectra provide an overview of whether "new" scan data ordinate differently than previous scans. Separation by length quantile for PCAs minimizes the number of obscured data points and allows for a visual representation of whether length classes differ among "new" and "old" scans. For the 2021 data, PCA data clouds from unprocessed data differed slightly from other years, but pre-processing addressed such shifts in the spectra (Fig. 27.2). Further, PC scores visualized by scan date suggest that there were no issues with the spectrometer or user errors associated with a particular time period (only showing pre-processed plots for demonstration; Fig. 27.3). While pre-processing spectra is typically done prior to modeling and data analysis (Rinnan et al. 2009), PCAs with unprocessed and pre-processed data maximize the ability to identify scan data quality issues. For example, stray light correction miscalculations can be more easily identified using unprocessed spectra (M. Arrington, pers. comm.).

Quantitative designations of spectral outliers were performed only using pre-processed data, under the assumption that these data exclude any correctable instrument-based issues (Rinnan et al. 2009). Outlier designation from PCA used Hotelling's T^2 and Q residuals to provide a definitive cut-off for inclusion in analyses (Fig. 27.4). While the 5% threshold for extreme values and 1% threshold for outliers present a decision step for the analyst, this approach provides a repeatable tool and threshold to apply across all datasets. Of 1,500 total specimens scanned from the 2021 dataset, 30 were identified as outliers from PCA and a greater number were identified as extremes (Fig. 27.4). Many of these spectra differed visually from others (Fig. 27.4b), but plots of PC scores indicate that thresholds based on PC1 and PC2 scores

would not be sufficient to identify outliers to the same degree as Hotelling's T^2 and Q residuals (Fig. 27.5).

QA/QC Using PLS – 2021 Collection Year

Fish lengths from "new" scan data were predicted from PLS models that excluded each respective focal year. Outlier identification using Y-distance and full X-distance, that combined Hotelling's T^2 and Q residuals into a single metric (Rodionova and Pomerantsev 2020), identified 52 outliers. Many outlier spectra diverged from the 1:1 observed and predicted line, following expectations for outliers detected by Y-distance (Fig. 27.6). For the 2021 case study, these outliers were often larger fish compared to most previous observations, indicating that these larger fish might be important for expanding the data domain for future outlier detection, and potentially for age prediction (Fig. 27.6). However, other outliers did not diverge from the 1:1 line, suggesting that X-distance metrics are also necessary to identify extreme and outlier spectra. While these extreme and outlier values may not impact length prediction, the impacts on final age models require additional exploration. Outliers detected from PLS exceeded those from PCA alone, but also excluded some outliers that were identified by PCA based on comparisons of individual specimen identifiers (see Appendix A Table A5 for example). This highlights the benefit of a combined PCA and PLS approach to QA/QC as both provide different information about outliers based on spectra (PCA) and through variation in relationships with a meaningful biological characteristic (PLS).

Final QA/QC Step with Combined PCA and PLS – 2021 Collection year

Outliers identified from PCA and PLS were used to generate a series of diagnostic plots and generate a list of outlier spectra for further inspection by a data manager. Plots of pre-processed spectra showing "new" scan data served as a visual check for a data analyst and indicated that many of the spectra identified as outliers using the two methods in conjunction had some visually discernible spectral irregularities (i.e., shifted peak, additional peaks, or unusual spectral shape). The outliers detected by PLS also encompassed spectra with both larger and smaller first derivative pre-processed peak heights (differences in slope in raw data) compared to most spectra (Fig. 27.7).

Other standardized QA/QC output includes file names for outlier spectra as well as details about potential issues with the otolith (broken, crystallized, or other problems) to aid in determination of next steps for a particular specimen (see Appendix A Table A5 for example). For the 2021 case study, 24 of the 61 total outliers were identified as being broken, crystallized, or other. This number far exceeds the proportion of the 1,500 total scans with these issues or that were unscannable. This report tool would then signal the QA/QC analyst that these otoliths should be aged using traditional methods. In addition to outliers, a large proportion of spectra were identified as "extreme" from both PCA and PLS approaches. These data points will require additional exploration to determine their impact on age prediction models. However, one potential approach is to consider "extreme" spectra as values that may increase the spectral data domain across additional ages or lengths, thus representing data points to include in calibration model updates (see Helser et al. this volume-a). The next course of action for the remaining outliers may depend on the workflow and focal species. A suggested workflow outlined in Figure 27.1 is: (1) Check the metadata for each sample to determine if comments were included

in data collection that may explain the unusual spectra. For example, a fish may be collected from an unusual geographic region, the fish may be smaller or larger in size than historic collections, or there may be a data entry error. (2) Rescan the otolith if there is no indication of other potential issues. (3) Use traditional ageing methods to obtain an age.

Considerations

The QA/QC process will be ongoing with new data collection, and documentation is essential for record keeping purposes and reproducibility. One option to ensure good record keeping and documentation is to utilize markdown documentation and code in R statistical software or Python to produce standardized reports and to maintain records of datasets used for analysis. In practice, analysts may not be well versed in a particular programming language. Therefore, interactive output such as html documents, user-friendly interfaces, or specialized functions or packages may facilitate the use of QA/QC procedures and documentation (examples shown in Appendix A are exported as pdf documents derived from html with interactive plots). PCA and PLS models built with large datasets may also be constrained by computing power. Our recommendation is to build these models and store them (e.g., R object or workspace) to load for QA/QC of "new" data. An additional consideration is to update these baseline models annually with "new" scan data.

As FT-NIR spectroscopy approaches for fish age prediction continue to develop and implementation proceeds across agencies, these QA/QC practices will likely be refined. Ultimately, these tools can develop into standardized best practices across agencies with a potential goal to implement programmatic packages for universal use. In light of this goal, we

document a first step toward standardization at the AFSC and highlight considerations for implementation.

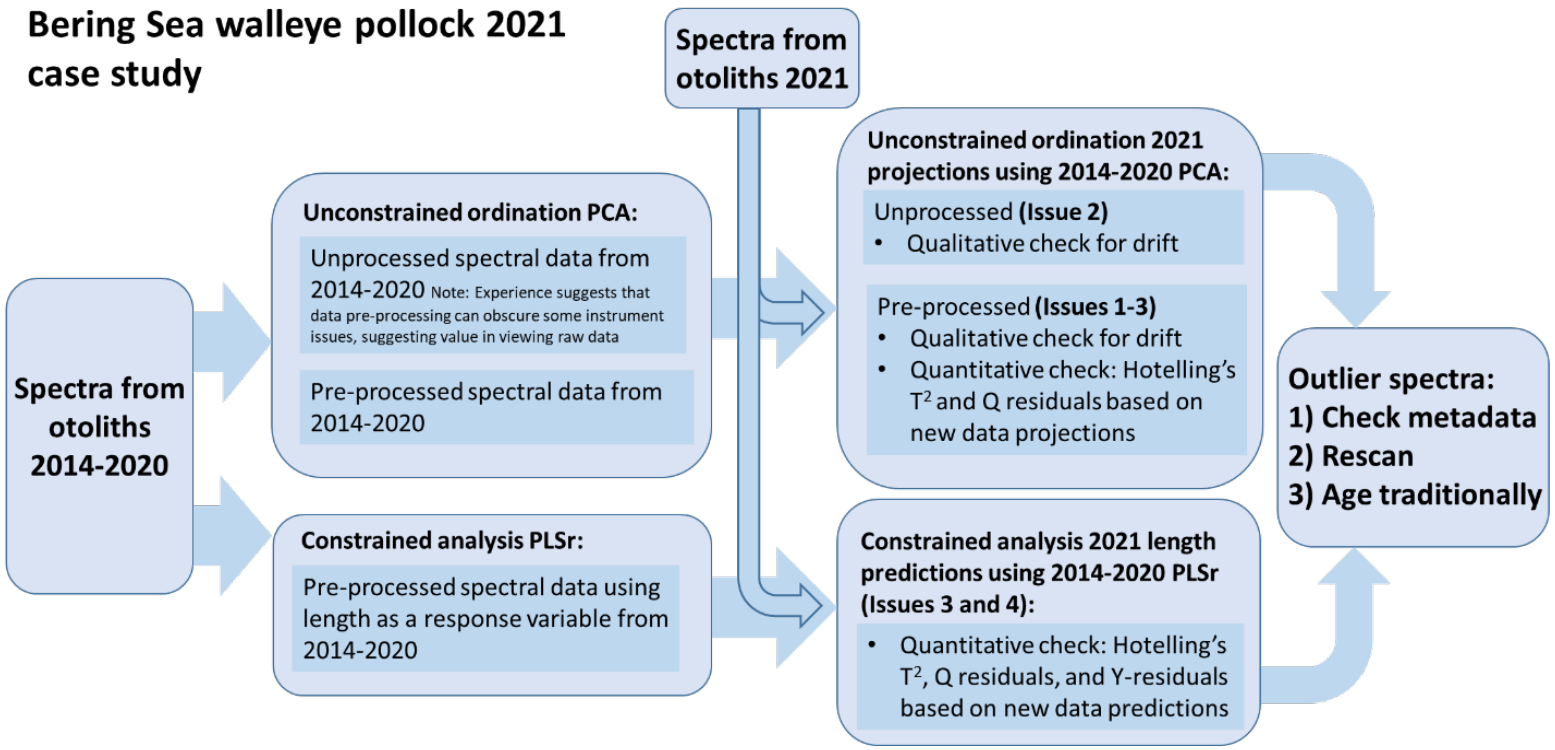


Figure 27.1. -- Schematic of the quality assurance/quality control (QA/QC) process conducted by a data manager and final actions to address outlier spectra in order of operation. The schematic uses 2021 walleye pollock *Gadus chalcogrammus* otolith data as an example. For a description of Issues 1-4, see the Introduction.

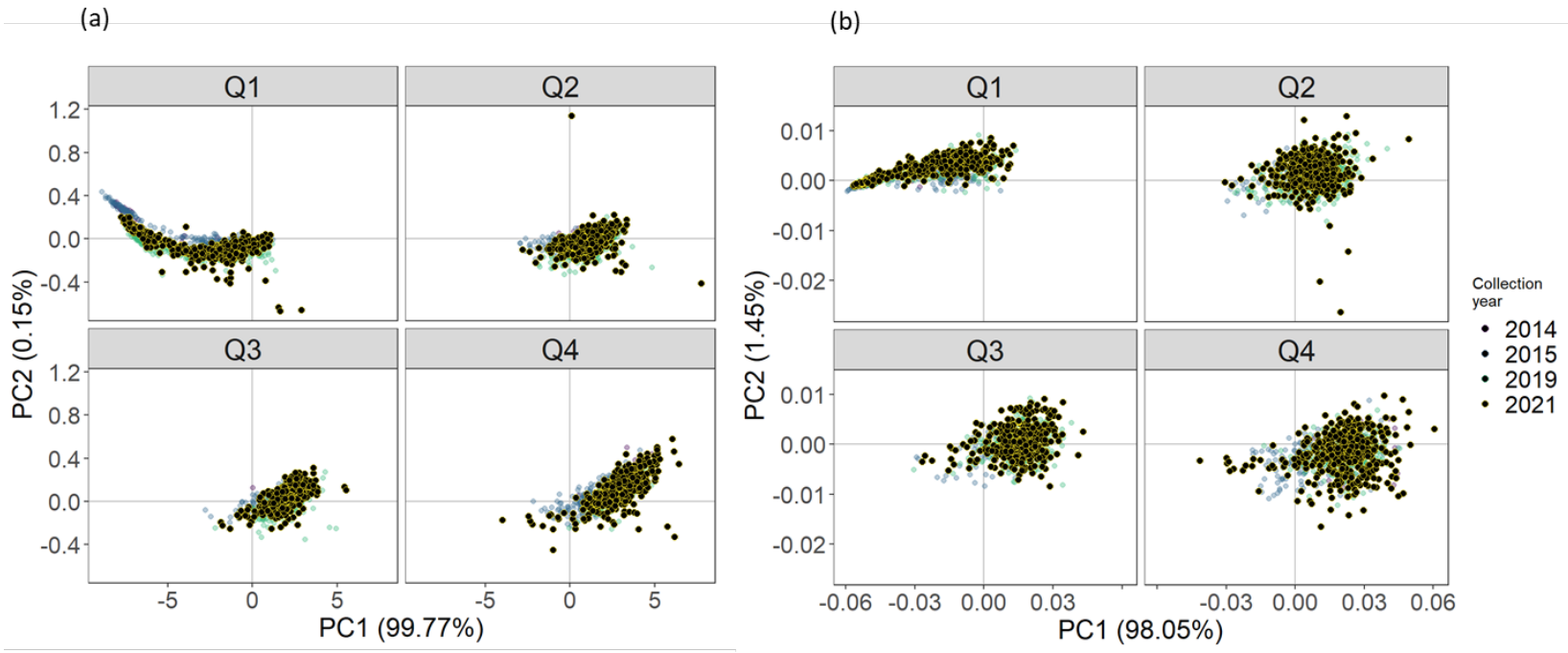


Figure 27.2. -- Principal component analysis (PCA) of a) unprocessed and b) pre-processed spectra by length quartile for the 2021 case study where data from 2021 were projected using the PCA built with all other data.

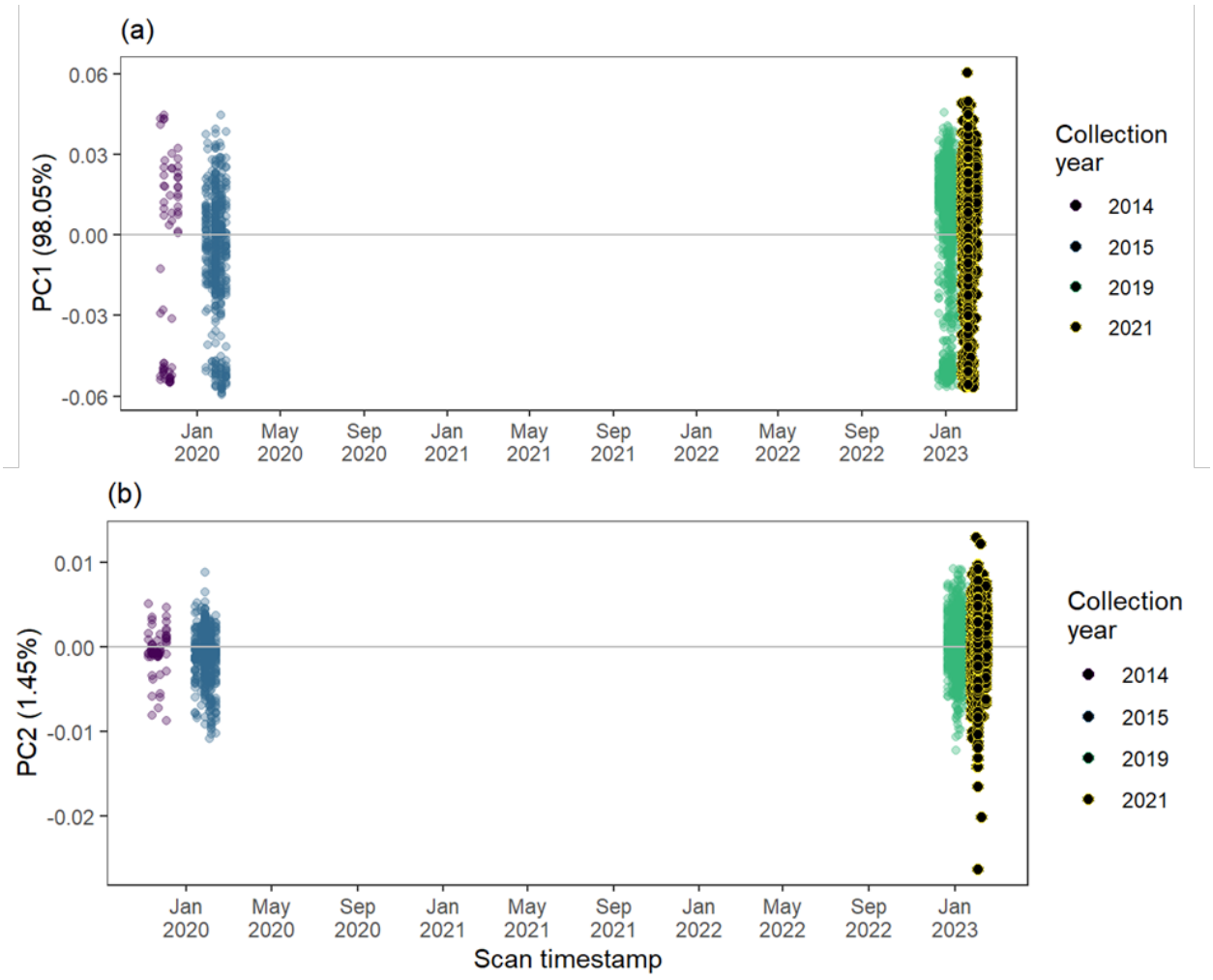


Figure 27.3. -- a) Principal component axis-1 (PC1) and b) PC2 of pre-processed spectra by scan date for the 2021 case study where data from 2021 were projected using the PCA built with all other data.

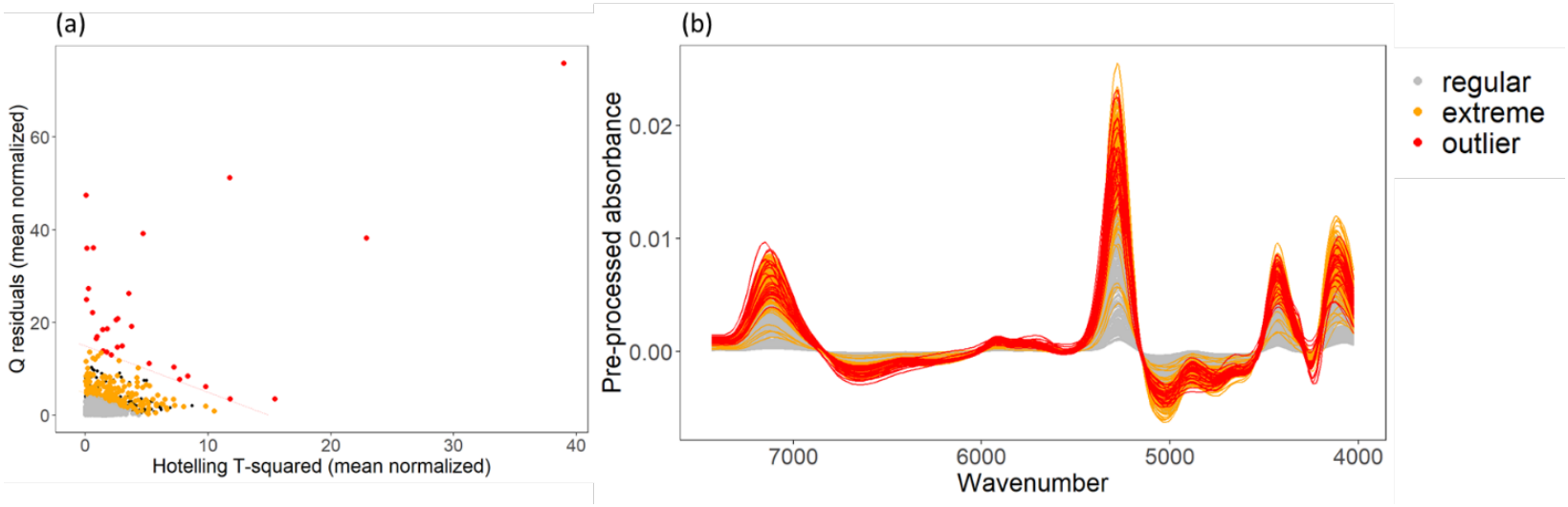


Figure 27.4. -- Principal component analysis (PCA) extreme values and outliers for the 2021 case study using pre-processed data showing thresholds based on a) Hotelling T^2 (T-squared) and Q residuals and b) plots of new spectra based on regular, extreme, and outlier categories. Extreme = 5% significance threshold, outliers = 1% threshold, and regular did not exceed either threshold. Black dots in (a) are observations from the previous scan data used to build the PCA.

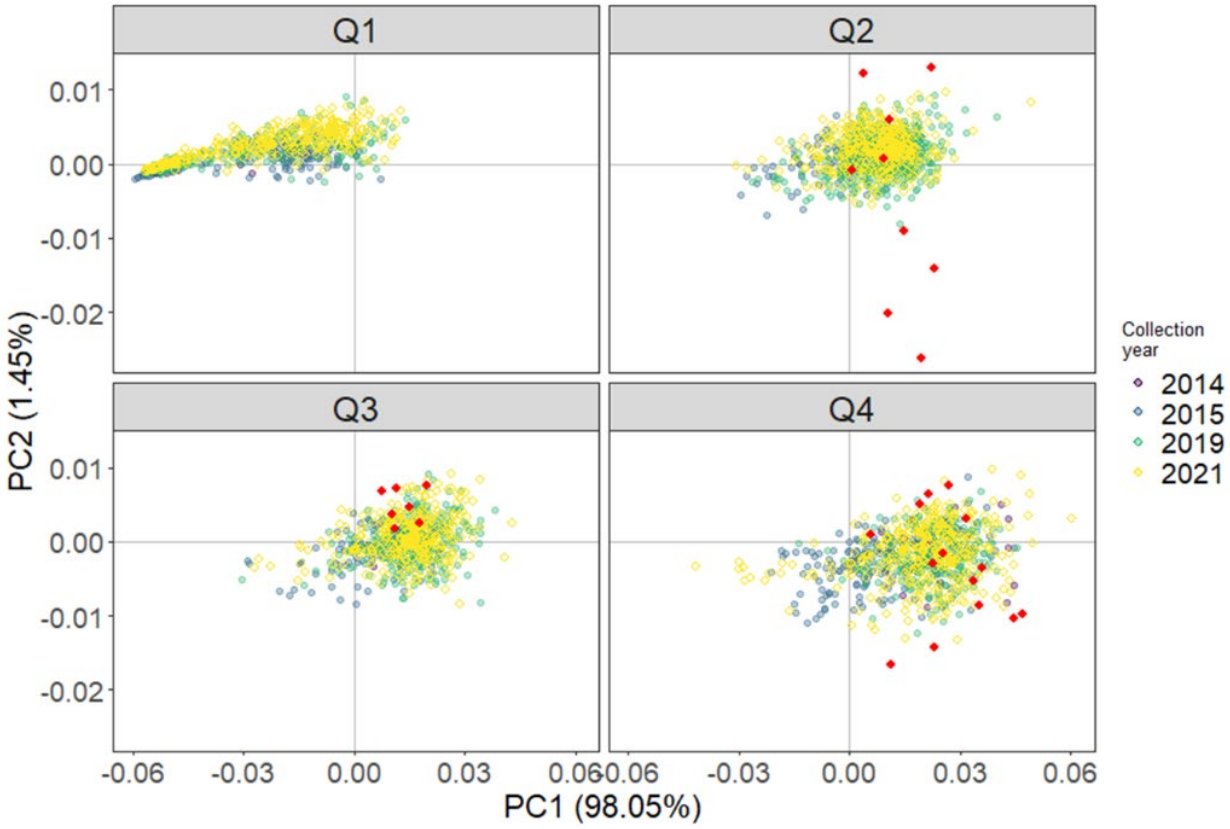


Figure 27.5. -- Outlier spectra (red points) principal component analysis (PCA) scores in the context of all other scan data for the 2021 case study where data from 2021 were projected using the PCA built with all other data.

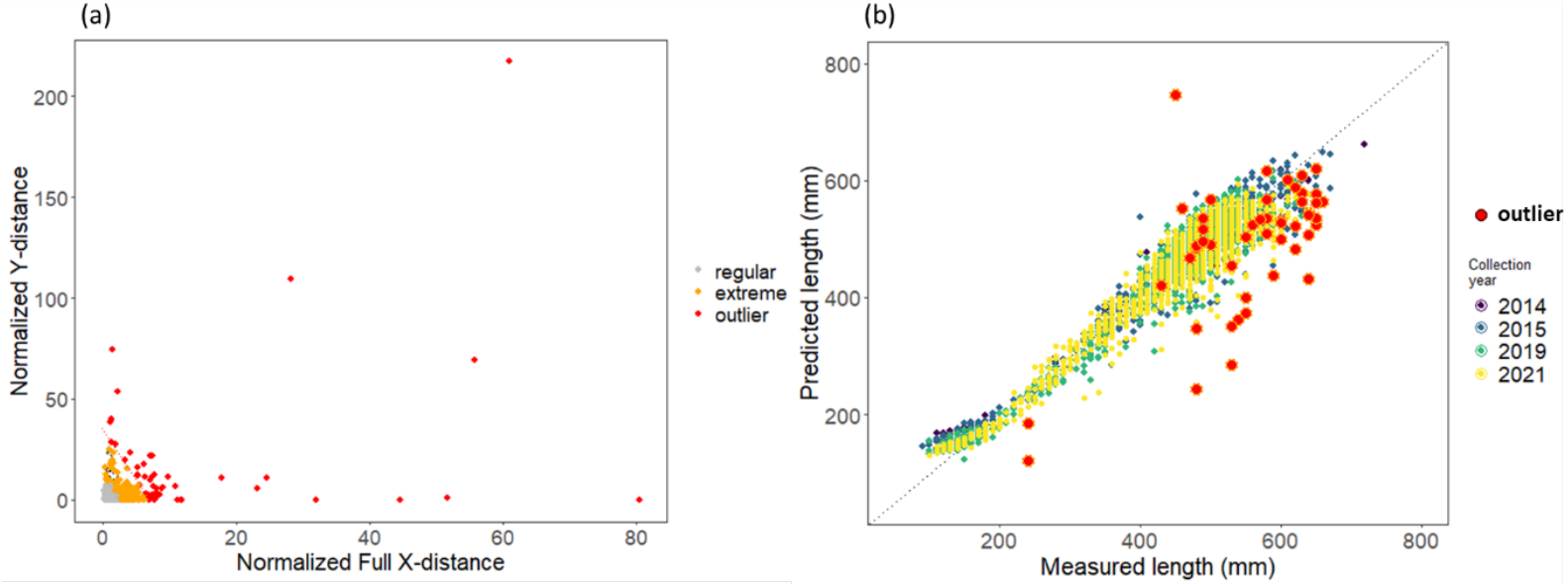


Figure 27.6. -- Extreme and outlier spectra identified by partial least squares regression (PLS) models showing a) outlier thresholds and b) length predictions. “New” data from 2021 were predicted using a PLS model built with all other years. In (a) “new” spectral data are shown in the color scale and previous data are black points.

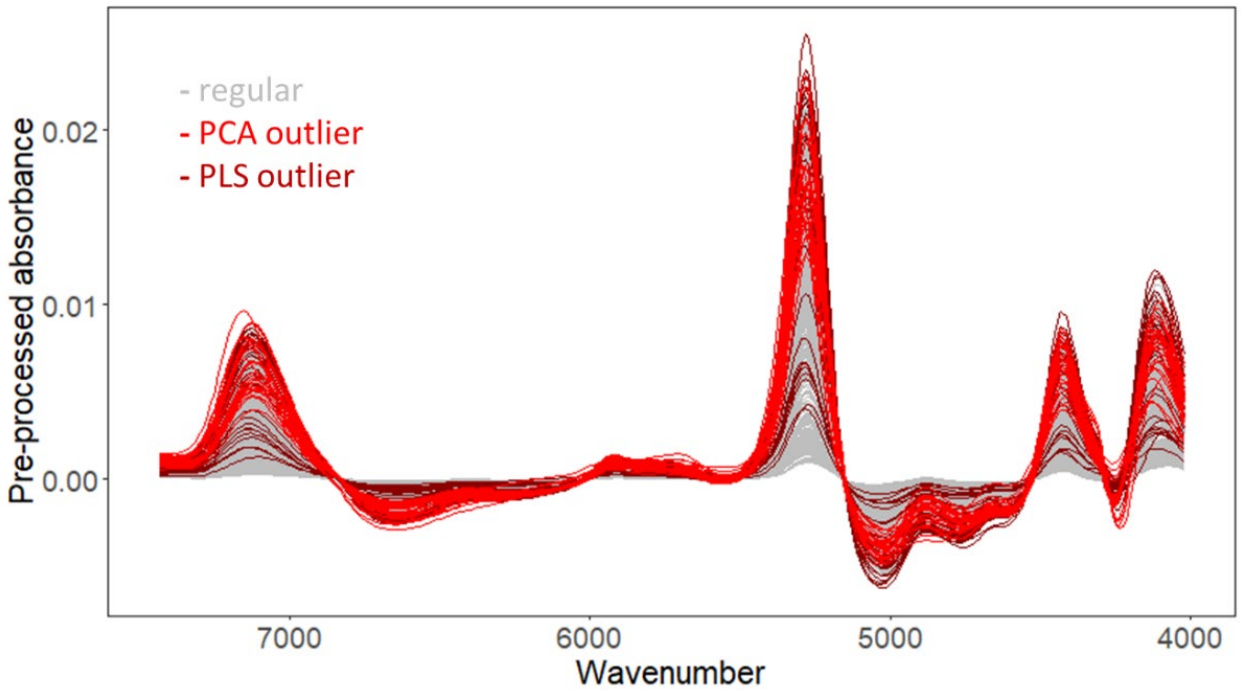


Figure 27.7. -- Pre-processed spectra from "new" scan data (2021) showing outliers identified by principal component analysis (PCA) and partial least squares regression (PLS) models.

DISCUSSION TOPICS AND RECOMMENDATIONS

The Fourier Transform Near Infrared (FT-NIR) Spectroscopy Workshop held in Seattle April 3-7, 2023 at the Alaska Fisheries Science Center (AFSC) included research presentations organized into three main sections: *Application and Development*, *Spectroscopy Implementation*, and *Stock Assessment Integration*, with open discussions during the last two days of the workshop pertaining to these and other relevant FT-NIR spectroscopy topics. During the presentation Q&A sessions, as well as the open discussion sessions, several salient themes were identified, some of which were also recurring topics of discussion during the first in-person FT-NIR Spectroscopy Fish Age Determination Workshop held in Seattle in 2019 (Helser et al. 2019a). The major themes outlined in the 2019 workshop were: 1) Instrument Settings and Optimization, 2) Sample Presentation – Preparation and Storage, 3) Calibration Models – Statistical Approach to Quantitative Analysis, 4) Communication to Assessment Reviewers, the Fishing Industry, and Policy Makers, and 5) Questions or Concerns about Technology. In the 2023 workshop, these themes of discussion remained largely the same, but with a greater emphasis on establishing best practices in scanning methodology, model development, and prediction evaluation, identifying barriers and building pathways towards implementation and stock assessment integration, and communicating with stakeholders. Below we describe the key points raised during the 2023 workshop discussions, with recommendations for further development.

ISSUE #1: SCANNING METHODOLOGY AND SAMPLE PRESENTATION

Since the 2019 workshop, most of the laboratories on the FT-NIR Spectroscopy Strategic Initiative Development Team (SIDT) have experimented with various sample presentation methods, tissue standardization methods, and scanning accessories to reduce excess light scattering and optimize spectral absorbance profiles of otoliths and other tissues. This has been particularly important for small otoliths, which generally have low signal-to-noise ratios when scanned directly on the instrument window alone. Using various transflectance stamps in conjunction with steel rings and Teflon discs to reduce the aperture size and focus the light on the sample has yielded promising results (Passerotti et al. 2020a, TenBrink et al. this volume, Matta et al. this volume-b). The choice of accessories is likely to be dependent on the application and may be species- or life stage-specific, requiring thorough investigation.

The discussion shifted focus towards exploring methods to optimize scanning procedures and improve the efficiency of the scanning process. Variations in scanning practices among different SIDT laboratories were noted. Thomas Helser noted the extensive effort by the AFSC laboratory to refine sample presentation methods and establish best practices for scanning, including wearing gloves when handling otoliths, using plastic forceps, and standardizing the orientation and position of specimens on the sample window. While there are certain center-specific steps that may affect efficiency (e.g., weighing, imaging, wiping away preservation media), there are commonalities that can be shared among centers through a best-practices scanning manual that Brenna Hsieh (AFSC) is developing, allowing standardization of

procedures. Development of standard operating procedures could also be expanded to other tissues (e.g., gonads, liver, muscle) and applications (e.g., hormone analysis, disease detection), where appropriate.

ISSUE #2: CALIBRATION MODEL DEVELOPMENT

Discussions pertaining to calibration model development revolved around two main topics: ensuring representative sampling of traditional microscopic ages (TMA) and employing the most appropriate model type to achieve unbiased predictions. Because FT-NIR spectroscopy is a secondary method in which age predictions from otolith spectra are "calibrated" against TMA reference age estimates, choice of those reference data is an important issue. It was stressed during the workshop that calibration models used to predict future data need to be representative of any new observations. Research may be needed to evaluate otolith spectral variability, both temporally and or spatially, before a base model is accepted for future predictions. The sample size must be large enough and the age distribution must be adequately covered in each calibration data set. The simulation work of Arrington et al. (this volume-a) suggests diminishing returns for calibration data sets when sample sizes exceed 900 for eastern Bering Sea walleye pollock (*Gadus chalcogrammus*) collections between 2014 and 2021. They recommend evaluating optimal sample sizes based on a species-specific basis due to varying levels of ageing error among species. To date, the AFSC has not employed a pre-specified method or sampling distribution for selecting samples for model calibration. This is because different sample selection approaches will depend on a number of factors such as data quantity (TMA ages and associated otolith spectra), data variability, and the model employed. It can also

be said that the model employed for training or calibration will be dependent upon data quantity. In cases of data limitations or exploratory calibration model development, a boxcar distribution may be a reasonable approach, as it represents ages and associated otolith spectra equally over the age range. Arrington et al. (this volume-a) suggest algorithms that maximize variability in otolith spectra in principal components space, particularly in cases where spectral variability increases with age, are preferable for calibration model sample selection. Their results indicate that the Kennard-Stone algorithm performs the best. However, in instances where this algorithm is not suitable, random selection is a viable alternative for sample selection. Morgan Arrington (University of Washington) noted during the discussions that she is happy to share her code with the rest of the SIDT.

Standard chemometric approaches using partial least squares (PLS) regressions are the models most typically used by all the laboratories to predict age and other metrics. These models are easy to implement directly in the Bruker OPUS software, increasing their appeal. However, several SIDT members are exploring modeling alternatives that may generate more accurate predictions. High-performance computing alternatives, such as cloud computing, graphics processing units, and virtual machines, were identified as potential needs when employing more sophisticated deep learning tools and larger quantities of data. Wallace (this volume) at NWFSC and Benson et al. (2023, this volume) at AFSC have employed deep learning methods as alternatives to PLS regressions for developing calibration models. Wallace (this volume) has found the *keras* R package (https://keras.io) that sits on top of Google's TensorFlow software (https://www.tensorflow.org) to be useful for neural network modeling for age predictions of sablefish (*Anoplopoma fimbria*) and Pacific hake (*Merccluccius productus*). When asked about the number of otoliths required to construct a deep learning model, Irina Benson (AFSC)

referenced a subject matter expert who recommended a minimum of 1,000 samples, although it is generally recognized that having more data is beneficial for improved model training. The ongoing deep machine learning research being conducted by the University of Washington's Information Processing Laboratory, under the guidance of Professor Jenq-Neng Hwang, integrates FT-NIR spectra, images, and metadata. This work suggests potential advantages of using separate models for older and younger fish (Zheng et al. this volume). The rapid expansion of deep learning and cloud computing in spectral data analysis across multiple industries highlights the SIDT's future focus on leveraging these tools as a central point of the strategic initiative. The SIDT suggested that any new proprietary or open source software for chemometric analysis should be vetted for compatibility and reliability and incorporated into a "tool box" for FT-NIR spectroscopy fish age modeling in NOAA Fisheries. Also notable is that the SI could greatly benefit from personnel with expertise in data science and deep machine learning to enhance and improve the development of tools.

ISSUE #3: AGE PREDICTION AND MODEL EVALUATION

Considerable time was dedicated to discussing the assessment of FT-NIR spectroscopy's reliability in predicting fish age, as well as determining the optimal timing for model updates. To reach this objective, maintaining some level of traditional age determination to verify model predictions will be essential, although the exact amount is still uncertain. Since FT-NIR spectroscopy is a secondary method reliant on TMA for model training and updating, TMA will continue to be necessary. This includes species that have transitioned into the FT-NIR process and those that are still completely dependent on TMA for stock assessments. Papers on TMA

ageing error have demonstrated that characteristics such as the coefficient of variation (CV), the number of samples, and age class distribution can have significant effects on bias detection (McBride 2015, Nesslage et al. 2022). Arrington et al.'s (this volume-b) simulation work using Pacific cod (*Gadus macrocephalus*) as a case study demonstrated that bias may have more of an impact on model prediction than precision, though these impacts also depend on the level of spectral variation. Extending Arrington et al.'s work to other species with different error structures would be very informative towards understanding the interaction between ageing error and spectral variation. Here again, the SIDT noted that it would be helpful to bring in a data scientist who understands predictive analytics to evaluate different model-updating scenarios. A team member highlighted the longstanding acceptance of TMA among stakeholders, recognizing the challenges of communicating the complicated methods and earning acceptance of the new approach. The SIDT might benefit from further discussion on collaborating with a dedicated communications expert to effectively convey scientific concepts to the public.

The next topic of discussion focused on how to best compare FT-NIR spectroscopy decimal age predictions with TMA ages, which are typically in integer form. The predictions are continuous random data, but TMA data are multinomial. There are several different options for how to deal with this problem: the decimal ages can be rounded (in different ways) to integer ages, or the reference TMA ages can be converted to decimal ages if marginal increment information is available (Barnett et al. this volume). It was noted that the AFSC tested these methods on walleye pollock, but it did not have much effect on the overall outcome, most likely because predictions were within 0.6 years of TMA ages. Treating the response, TMA age, as a categorical outcome would produce an integer age. Linear methods and deep learning methods can be both used to produce integer ages with probabilities, though this may be problematic for

long-lived stocks that have many age classes Such a method using otolith images and NIR spectra using convolutional neural networks was presented at the workshop (Zheng et al. this volume).

An SIDT member involved in stock assessment raised concerns regarding the alignment between TMA and FT-NIR spectroscopy ages. Our current analysis suggests a potential tendency of underfitting at older age groups. For some species, we have also observed overfitting at the youngest ages, although both underfitting and overfitting across various age classes are possible. This pattern may be attributable to inherent observation uncertainty stemming from TMA errors or natural variation in otolith spectra. Tom Helser (AFSC) noted that models derived from otolith spectra collected among different analysts are more precise than TMA double-reads and that variability in agreement between TMA readers can reduce agreement between TMA and FT-NIR spectroscopy. Additionally, spectra collected from species with a broad range of otolith sizes might challenge the proportionality assumption of the Beer-Lambert law concerning analyte concentration and absorbance, potentially causing non-linearity. To address these challenges, exploring different modeling approaches such as leveraging deep learning models, which are capable of handling non-linearity using specialized activation functions, may offer promising solutions. Benson et al. (this volume) demonstrated the potential of such methodologies. Furthermore, it is important to consider the extent of disagreement between TMA and FT-NIR spectroscopy ages. Predictions within 1 year may have little impact on stock assessment outcomes, especially in cases where ageing imprecision is high (CV > 5%) or the deviation is near or beyond the plus group. Our focus remains on striving for the most accurate predictive model using the best goodness-of-fit tools available. However, it is essential to evaluate impacts on stock assessments, as outlined in the next section.

Additionally, a mixed method approach could be established whereby FT-NIR spectroscopy is used to predict age for the majority of specimens, while ages of the remaining group of fish, often comprising the oldest and least numerous in the sample, are estimated by TMA. Such an approach could still result in reasonably good efficiency of age estimate production for the vast majority of otoliths. It is not uncommon to see mixed methods employed within the TMA determination process itself. Typically, ages are derived from otolith surfaces of the youngest fish or those with the clearest growth patterns, while otoliths from older fish are sectioned and sometimes burned or stained. Issues related to stock assessment models are worthy of further consideration. If mixed methods are employed to generate ages for age compositions, it may be necessary to construct separate ageing error vectors to account for the uncertainty arising from the different methods. A simulation using the R package ss3sim (https://cran.r-project.org/web/packages/ss3sim/) would be a good framework to investigate these effects.

ISSUE #4: STOCK ASSESSMENT INTEGRATION

Overall, FT-NIR spectroscopy age prediction shows promise, as was demonstrated by a number of case studies in these proceedings. Yet, a better understanding of the sensitivity of population dynamics models to changes in the input data type to evaluate stock assessment outcomes is needed. It will be necessary to evaluate model sensitivity on a case-by-case basis because assessments can vary greatly in model structure, parameter specification, data quantity, and data use. Nevertheless, some general criteria for acceptance of substituting FT-NIR spectroscopy age data could be defined and universally applied. For instance, Ianelli (this volume) used changes in log likelihoods to evaluate goodness-of-fit from substitution of age

compositions based on the different ageing methods. However, depending on the gradient of the likelihood space, small differences in likelihoods may not be that informative. Garner et al. (this volume) employed the R package ss3sim to simulate age data using an operating model and an estimation model under ageing error assumptions based on the two ageing methods. Their choice for using that approach was based on limited data for which there were both TMA and FT-NIR spectroscopy ages for direct substitution of age compositions. Finally, Helser et al. (this volume-b) used MCMC simulation in Stock Synthesis (V3.30.16.02) to evaluate the probability in the difference of the marginal posterior density of model parameters and management quantities when substituting a 6-year time series of age compositions from the two ageing methods. While there may not be a universal solution, criteria and benchmarks to evaluate assessment model outputs could be defined and used to provide the evidence needed to accept FT-NIR spectroscopy age data as a credible substitute for TMA.

Another question is how to best blend different data types (i.e., TMA and FT-NIR spectroscopy age data) in a stock assessment. Historical age estimates are largely generated through TMA, so combining them with FT-NIR spectroscopy age predictions for more recent years may require a transition matrix necessary to estimate the probabilistic difference between them. Alternatively, most modern statistical age-structured assessment models can incorporate an evolution of data type properties, such as age composition and ageing error, within a time series through blocking structure mitigating a need to make modifications to the past data. In such a case, a clean transition to age composition data generated from FT-NIR spectroscopy ageing can be made if such data result in no discernable assessment model outcomes as discussed above.

Different rounding scenarios for the continuous FT-NIR spectroscopy age predictions could also be explored in more detail. Several workshop participants presented simulations

integrating both data types and different ageing error scenarios into assessments. These case studies demonstrated varying effects on recruitment, spawning stock biomass, and natural mortality, all of which impact biological reference points (Garner et al. this volume, Ianelli this volume, Helser et al. this volume-b). Notably, these simulations all focused on relatively short-lived stocks (walleye pollock, Pacific cod, and gray snapper). Considering species with diverse life histories could prove beneficial, as longevity is expected to impact these estimates.

A full evaluation of stock assessment model sensitivity may be challenging, particularly in cases when intact otoliths are not available for FT-NIR scanning. For example, typically, only one of each otolith pair is collected in fishery samples of gray snapper. Given the frequently destructive nature of TMA methods, a second otolith is not usually available for scanning historical samples. Because the SEFSC laboratory has not evaluated whether there are differences between left and right otolith spectral data, predictive models were based on only left otoliths, further reducing sample size. It was noted that the AFSC laboratory has not detected significant differences between scans of right and left otoliths for a number of species including flatfish, which have asymmetrical otoliths, suggesting that the otolith side may be inconsequential. To address situations where sample sizes are relatively small, an SIDT member suggested performing an analysis outside the stock assessment whereby bootstrapped age compositions are calculated from each dataset and the uncertainty is summarized as a weighting factor.

In general, the initial work evaluating sensitivity of stock assessment outcomes from different ageing methods shows promise, but there is still more work to do to improve predictions, particularly given the concerns about lack of fit. Deep learning models have largely been able to rectify non-linearity in TMA reference data and FT-NIR spectroscopy age

predictions. However, how many ages and how many years of age composition data are needed for a sensitivity analysis in stock assessment models remain pertinent questions. Again, these issues may need to be evaluated on a case-by-case basis because models vary considerably in the quantity of age data used, age-related model specifications (e.g., time-varying growth), and the competing influence of other data types. Clearly, SIDT members need to work closely with stock assessment scientists to define the most critical assessment model outcomes, the criteria used to test the different ageing methods, and analytical platforms for the sensitivity analysis.

ISSUE #5: PATHWAYS AND BARRIERS TOWARDS IMPLEMENTATION

The development of Standard Operating Procedures (SOPs) for all the steps associated with generating FT-NIR spectroscopy age predictions is critical for creation of a data pipeline and providing a framework for data quality assurance and quality control (QA/QC). These SOPs would include standardized QA/QC procedures to 1) maintain consistency during the scanning process itself, 2) evaluate spectral data, 3) determine the percentage of otoliths that need to be aged using TMA methods, 4) evaluate model predictions, and 5) determine when model updating is necessary. Staff at the AFSC are actively working towards creating protocols to standardize both the scanning process (Hsieh et al. in prep.) and data/model evaluation procedures (Goldstein et al. this volume-a). These SOPs will be disseminated to other laboratories participating in the SI as either NOAA publications (e.g., NOAA Tech Memos or Processed Reports) or in peer-reviewed journals.

Having an establish database infrastructure will also facilitate more rapid adoption and implementation of FT-NIR spectroscopy to age fish. Jon Short (AFSC) has built a relational

database (*AGE3*) using SQL Server and a Python front end for all of AFSC's age and related data (Short this volume). Not every other laboratory in the SI has a database like *AGE3*, but there was great interest in patterning new database structures based on *AGE3*. It was noted that replicating *AGE3* exactly is not feasible due to its reliance on various data sources, making a direct "plug and play" into the workflows of other laboratories unlikely. However, Jon is willing to share *AGE3*'s general database table structure and suggested that other laboratories start with Microsoft *ACCESS* to build new tables copying *AGE3*'s structure. Code sharing and furthering open communication among database managers from different laboratories would greatly facilitate the process. Discussion also centered on funding for expertise in developing an enterprise-level database within NOAA Fisheries that can harmonize the different data types, standardize data storage and archiving, and develop data discovery and analytic tools for quality control and modeling applications for age prediction.

Achieving buy-in from stock assessment scientists, the fisheries management councils, and stakeholders could pose a potential challenge in the adoption and transition of the FT-NIR spectroscopy ageing method. Active collaboration between SIDT members and assessment scientists is imperative to test effects of integrating FT-NIR spectroscopy age data into population models and on biological reference points. The SIDT will continue to work directly with assessment scientists to determine criteria and benchmarks that must be met before the new method is adopted. Additional workshops and meetings with the council bodies, such as the Scientific and Statistical Committee, are envisioned as is critical review of the methodology by the Center for Independent Experts.

Resource and staffing limitations were a major concern brought up by member of the SIDT during the discussion. Many SIDT laboratories have limited personnel and resources to

dedicate to FT-NIR spectroscopy SI research and development while maintaining their core responsibility of meeting annual production ageing demands. Potential avenues of communication to alleviate these concerns are discussed in detail in the next section.

ISSUE #6: COMMUNICATION WITH DATA END USERS, STAKEHOLDERS, AND NOAA LEADERSHIP

Modern stock assessments of commercial stocks often rely heavily on fish age data. While the FT-NIR spectroscopy method shows great promise for producing age data more efficiently, gaining full support from data end users and stakeholders will require continued serious investments in research and development of FT-NIR spectroscopy age estimation methods. Improving communication among the SIDT, the NOAA Science Board, stock assessment community, and stakeholders will be key. At a few laboratories, efforts are already underway where SIDT members are working side by side with assessment teams to communicate and evaluate assessment models with regard to the new technology. However, more outreach is needed to communicate differences between ageing methods, develop a transition plan which defines the data pipeline with quality control processes and data controls, and finalize data products that will enable users to better understand, and therefore accept, FT-NIR spectroscopy age data. Direct conversations with the Science Board are also necessary to understand their expectations regarding the accomplishments at the conclusion of the SI, expected at the end of FY2025. Several SIDT laboratories, particularly the smaller laboratories that have fewer dedicated personnel, are resource-limited and unable to fully perform the research and development needed to implement the technology. These laboratories voiced the

concern that the timeline to operational readiness may not be feasible. Additional support from individual science centers would be helpful to offset the disruption caused by this technological innovation. Sharing resources among the laboratories and bringing in outside experts to assist with development and integration were discussed as possible solutions towards improving efficiency and reducing redundancy. Dedicated funds for staff training could also assist in FT-NIR spectroscopy integration but would require cutbacks related to other outputs and tasks. Open science tools (e.g., Google groups, GitHub, NOAA Virtual Lab, Confluence pages) were mentioned as ways to improve communication within the SIDT. The monthly SIDT meetings have provided valuable opportunities for SIDT members to discuss progress, setbacks, and troubleshooting. Since the April 2023 workshop, the SIDT meetings have refocused on topical themes each month, resulting in rich, in-depth discussions that have helped advance and share the SIDT's collective knowledge. The SIDT has concluded that sharing documentation of best practices as they develop among members would be helpful in proceeding towards implementation. We have used NOAA Virtual Lab as a platform for knowledge transfer among SIDT members and will continue to do so until a potentially better platform is found.

In short, great progress has been made since the Strategic Initiative began, and the SIDT's continued efforts have forged a path towards implementing and integrating FT-NIR spectroscopy data collections into fisheries stock assessments and communicating research products.

ACKNOWLEDGMENTS

We thank the staff of the AFSC Age and Growth Program for their help in coordinating the fourth workshop of the NOAA Strategic Initiative on Rapid Estimation of Fish Age Using Fourier Transform Near Infrared Spectroscopy, particularly Andrew Chin and Irina Benson. We also thank members of the Strategic Initiative Development Team, many of whom were also workshop attendees, for their participation in monthly progress meetings to discuss advancing the technology towards implementation. We are grateful to workshop participants for sharing their research and lessons learned, helping us towards our goal of operational readiness. Many thanks to Morgan Arrington, Esther Goldstein, and Todd TenBrink, who took notes during the workshop sessions that facilitated writing the Discussion Topics and Recommendations section (additional thanks to them and to Irina Benson and Brenna Hsieh for providing helpful suggestions that greatly improved that section of this document). Thanks also to Rebecca White of the AFSC Communications Team, who designed the beautiful graphics for the workshop agenda. Finally, we thank the leadership of the Alaska Fisheries Science Center for their continued support and the NOAA Fisheries Science Board for funding this Strategic Initiative and making this workshop possible.

CITATIONS

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng. 2016. TensorFlow: a system for large-scale machine learning. *In* Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation (OSDI'16). USENIX Association, USA, 265-283. Available at https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

Alix, M., O. S. Kjesbu, and K. C. Anderson. 2020. From gametogenesis to spawning: how climate-driven warming affects teleost reproductive biology. J. Fish Biol. 97:607-632. https://doi.org/10.1111/jfb.14439

Alzubaidi, L., J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamarfa, M. A. Fadhel, M. Al-Amidie, and L. Farhan. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8:53 https://doi.org/10.1186/s40537-021-00444-8

Anastasiadi, D., and F. Piferrer. 2020. A clockwork fish: Age prediction using DNA methylation-based biomarkers in the European seabass. Mol. Ecol. Resour. 20(2):387-397. https://doi.org/10.1111/1755-0998.13111

Andersen, C. M., and R. Bro. 2010. Variable selection in regression-a tutorial. J. Chemom. 24(11-12):728-737. https://doi.org/10.1002/cem.1360

Anderson, S. C., C. C. Monnahan, K. F. Johnson, K. Ono, and J. L. Valero. 2014. ss3sim: An R package for fisheries stock assessment simulation with Stock Synthesis. PLoS ONE 9:e92725. https://doi.org/10.1371/journal.pone.0092725

Anderson, W. D. 1981. A new species of Indo-West Pacific *Etelis* (Pisces: Lutjanidae), with comments on other species of the genus. Copeia 4:820-825. https://doi.org/10.2307/1444183

Andrews, K. R., A. J. Williams, I. Fernandez-Silva, S. J. Newman, J. M. Copus, C. Wakefield, J. E. Randall, and B. W. Bowen. 2016. Phylogeny of deepwater snappers (Genus *Etelis*) reveals a cryptic species pair in the Indo-Pacific and Pleistocene invasion of the Atlantic. Mol. Phylogenetics Evol. 100:361-371. https://doi.org/10.1016/j.ympev.2016.04.004

Andrews, K. R., I. Fernandez-Silva, J. E. Randall, and H. C. Ho. 2021. *Etelis boweni* sp. nov., a new cryptic deepwater eteline snapper from the Indo-Pacific (Perciformes: Lutjanidae). J. Fish Biol. 99(2):335-344. https://doi.org/10.1111/jfb.14720

Arrington, M. B., T. E. Helser, I. M. Benson, T. E. Essington, M. E. Matta, and A. E. Punt. 2022. Rapid age estimation of longnose skate (*Raja rhina*) vertebrae using near-infrared spectroscopy. Mar. Freshwat. Res. 73(1):71-80. https://doi.org/10.1071/MF21054

Arrington, M. B., J. T. Healy, T. E. Helser, E. D. Goldstein, I. M. Benson, B. C. Hsieh, and A. E. Punt. This volume-a. A simulation-based approach to evaluate best practices for estimating fish age using FT-NIR spectroscopy.

Arrington, M. B., T. E. Helser, I. M. Benson, M. E. Matta, E. D. Goldstein, and A. E. Punt. This volume-b. A simulation framework to examine the effect of ageing error on FT-NIR model-based age predictions.

Barbeaux, S. J., K. Holsman, and S. Zador. 2020. Marine heatwave stress test of ecosystem-based fisheries management in the Gulf of Alaska Pacific cod fishery. Front. Mar. Sci. 7. https://doi.org/10.3389/fmars.2020.00703

Barnett, B. K., I. M. Benson, T. E. Helser, S. K. Lowerre-Barbieri, and H. S. Menendez. This volume. Investigating the use of FT-NIR spectroscopy to age gag grouper (*Mycteroperca microlepis*), a protogynous hermaphroditic species.

Baumann, P., T. Knecht, and P. Roudier. 2023. Opusreader2: read spectroscopic data from Bruker OPUS binary files. R package version 0.5.0.9000. https://github.com/spectral-cockpit/opusreader2

Beamish, R. J., and D. E. Chilton. 1977. Age determinations of lingcod (*Ophiodon elongatus*) using dorsal fin-rays and scales. J. Fish. Res. Board Can. 34:1305-1313. https://doi.org/10.1139/f77-192

Beamish, R. J., and D. A. Fournier. 1981. A method for comparing the precision of a set of age determinations. Can. J. Fish. Aquat. Sci. 38(8):982-983. https://doi.org/10.1139/f81-132

Beamish, R. J., and G. A. McFarlane. 1983. The forgotten requirement for age validation in fisheries biology. Trans. Amer. Fish. Soc. 112(6):735-743.

Beamish, R., and G. A. McFarlane. 1995. A discussion of the importance of aging errors, and an application to walleye pollock: the world's largest fishery. *In* Recent developments in fish otolith research (D. H. Secor, S. E. Campana, and J. M. Dean, eds.), p. 545-565. Univ. South Carolina Press, Columbia, SC.

Beamish, R. J., G. A. McFarlane, and A. Benson. 2006. Longevity overfishing. Prog. Oceanogr. 68:289-302. https://doi.org/10.1016/j.pocean.2006.02.005

Beattie, J. R., and F. W. L. Esmonde-White. 2021. Exploration of principal component analysis: deriving principal component analysis visually using spectra. Appl. Spectrosc. 75(4):361-375. https://doi.org/10.1177/0003702820987847

Benson, I. M., B. K. Barnett, and T. E. Helser. 2020. Classification of fish species from different ecosystems using the near infrared diffuse reflectance spectra of otoliths. J. Near Infrared Spectrosc. 28:224-235. https://doi.org/10.1139/cjfas-2023-0045

Benson, I. M., T. E. Helser, M. Marchetti, and B. K. Barnett. 2023. The future and fish age estimation: deep machine learning coupled with Fourier transform near infrared spectroscopy. Can. J. Fish. Aquat. Sci. https://doi.org/10.1139/cjfas-2023-0045

Benson, I. M., B. K. Barnett, and T. E. Helser. This volume. Deep learning coupled with Fourier transform near infrared spectroscopy of otoliths improves age prediction for long-lived fish.

Bertignac, M., and H. De Pontual. 2007. Consequences of bias in age estimation on assessment of the northern stock of European hake (*Merluccius merluccius*) and on management advice. ICES J. Mar. Sci. Journal du Conseil 64:981-988. https://doi.org/10.1093/icesjms/fsm039

Bird, A. P. 1993. Functions for DNA methylation in vertebrates. Cold Spring Harb. Symp. Quant. Biol. 58:281-285. https://doi.org/10.1101/sqb.1993.058.01.033

Bland, J. M. and D. G. Altman. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307-310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bland, J. M., and D. G. Altman. 1999. Measuring agreement in method comparison studies. Stat Methods Med. Res. 8:135-160. https://doi.org/10.1177/096228029900800204

Boulesteix, A.-L. 2004. PLS dimension reduction for classification with microarray data. Stat. Appl. Genet. Mol. 3(1). https://doi.org/10.2202/1544-6115.1075.

Bowker, A. H. 1948. A test for symmetry in contingency tables. J. Amer. Stat. Assoc. 43(244):572-574. https://doi.org/10.2307/2280710

Bullock, L. B., and G. B. Smith. 1991. Seabasses (Pisces: Serranidae). Memoirs of the Hourglass cruises 8(2). Florida Marine Research Institute, St. Petersburg, FL, USA.

Buonaccorsi, J. P. 1996. Measurement error in the response in the general linear model. J. Am. Stat. Assoc. 91(434):633-642. https://doi.org/10.2307/2291659

Butler, H. J., L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone, and F. L. Martin. 2016. Using Raman spectroscopy to characterize biological materials. Nat. Protoc. 11(4):664-687. https://doi.org/10.1038/nprot.2016.036

Campana, S. E. 1999. Chemistry and composition of fish otoliths: pathways, mechanisms and applications. Mar. Ecol. Prog. Ser. 188:263-297. https://doi.org/10.3354/meps188263

Campana, S. E. 2001. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. J. Fish. Biol. 59:197-242. https://doi.org/10.1111/j.1095-8649.2001.tb00127.x

Campana, S. E., M. C. Annand, J. I. McMillan. 1995. Graphical and statistical methods for determining the consistency of age determinations. Trans. Am. Fish. Soc. 124(1):131-138.

Carrascal, L. M., I. Galván, and O. Gordo. 2009. Partial least squares regression as an alternative to current regression methods used in ecology. Oikos 118:681-690. https://doi.org/10.1111/j.1600-0706.2008.16881.x

Carruthers, T. R., J. F. Walter, M. K. McAllister, and M. D. Bryan. 2015. Modelling age-dependent movement: an application to red and gag groupers in the Gulf of Mexico. Can. J. Fish. Aquat. Sci. 72:1159-1176. https://doi.org/10.1139/cjfas-2014-0471

Cass, A. J., and R. J. Beamish. 1983. First evidence of validity of the fin-ray method of age determination for marine fishes. N. Am. J. Fish. Manag. 3(2):182-188.

Chang, M.-Y., and A. G. Geffen. 2013. Taxonomic and geographic influences on fish otolith microchemistry. Fish Fish. 14(4):458-492. https://doi.org/10.1111/j.1467-2979.2012.00482.x

Chang, W. Y. B. 1982. A statistical method for evaluating the reproducibility of age determination. Can. J. Fish. Aquat. Sci. 39(8):1208-1210. https://doi.org/10.1139/f82-158

Chen, D. G. and D. M. Ware. 1999. A neural network model for forecasting fish stock recruitment. Can. J. Fish. Aquat. Sci. 56:2385-2396. https://doi.org/10.1139/f99-178

Chen, J., and X. Z. Wang. 2001. A new approach to near-infrared spectral data analysis using independent component analysis. J. Chem. Inf. Comput. Sci. 41:992-1001. https://doi.org/10.1021/ci0004053

Chen, L.-D., M. Santos-Rivera, I. J. Burger, A. Kouba, D. Barber, and C. K. Vance. 2021. Near-infrared spectroscopy (NIRS) as a method for biological sex discrimination in the endangered Houston toad (*Anaxyrus houstonensis*). Methods Protoc. 5(1):4-16. https://doi.org/10.3390/mps5010004

Chen, L. D., M. A. Caprio, D. M. Chen, A. J. Kouba, and C. K. Vance. 2023. Optimizing predictive performance for wildlife spectroscopy applications through a multi-model framework. PLoS Comput. Biol. In revision.

Chen, Y.-Y., and Z.-B. Wang. 2019. End-to-end quantitative analysis modeling of near-infrared spectroscopy based on convolutional neural network. J. Chemom. 33 (5):e3122. https://doi.org/10.1002/cem.3122

Chollet, F. 2015. Keras [Online]. Available at: https://github.com/fchollet/keras

Choo-Smith, L.-P., H. G. M. Edwards, H. P. Endtz, J. M. Kros, F. Heule, H. Barr, J. S. Robinson Jr., H. A. Bruining, and G. J. Puppels. 2002. Medical applications of Raman spectroscopy: From proof of principle to clinical implementation. Biopolymers 67:1-9. https://doi.org/10.1002/bip.10064

Claiborne, A., S. Rosenfield, J. Topping, D. Downs, and T. Tsou. 2016. Comparison of ages determined from vertebrae, dorsal fin rays, and otoliths in lingcod (*Ophiodon elongatus*). Poster presented at 19th Western Groundfish Conference; February 8-12, 2016, Newport, OR, USA.

Clausen, D. M. and J. Heifetz. 2002. The northern rockfish, *Sebastes polyspinis*, in Alaska: commercial fishery, distribution, and biology. Mar. Fish. Rev. 64(4):1-28.

Committee of Age-Reading Experts (CARE). 2006. Manual on Generalized Age Determination Procedures for Groundfish. Technical Subcommittee of the Canada/U.S. Groundfish Committee. Available at https://care.psmfc.org/

Conzen, J. 2014. Multivariate calibration, 3rd English ed. Bruker Optik GmbH, Drachten, The Netherlands.

Cook, T. R. 2020. Neural networks. *In* Advanced studies in theoretical and applied econometrics 161-189. https://doi.org/10.1007/978-3-030-31150-6_6

Cope, J. M., and A. E. Punt. 2007. Admitting ageing error when fitting growth curves: an example using the von Bertalanffy growth function with random effects. Can. J. Fish. Aquat. Sci. 64(2):205-218. https://doi.org/10.1139/f06-179

Cui, C., and T. Fearn. 2018. Modern practical convolutional neural networks for multivariate regression: applications to NIR calibration. Chemometr. Intell. Lab. Syst. 182:9-20. https://doi.org/10.1016/j.chemolab.2018.07.008

Cuvier, M. L. B., and M. Valenciennes. 1828. Chapter VI. *Etelis*. In: Histoire naturelle des poissons. Tome II, Museum Histoire Naturelle, Paris, pp. 127-131.

de Carvalho, C. C. C. R., and M. J. Caramujo. 2017. Carotenoids in aquatic ecosystems and aquaculture: a colorful business with implications for human health. Front. Mar. Sci. 4:93. https://doi.org/10.3389/fmars.2017.00093

Delwiche, S., and J. Reeves. 2010. A graphical method to evaluate spectral preprocessing in multivariate regression calibrations: example with Savitzky-Golay filters and partial least squares regression. Appl. Spectrosc. 64:73-82. https://doi.org/10.1366/000370210790572007

Dempsey, D. P., P. Pepin, M. Koen-Alonso, and W. C. Gentleman. 2020. Application of neural networks to model changes in fish community biomass in relation to pressure indicators and comparison with a linear approach. Can. J. Fish. Aquat. Sci.77:963-977. https://doi.org/10.1139/cjfas-2018-0411

De Paoli-Iseppi, R., B. E. Deagle, C. R. McMahon, M. A. Hindell, J. L. Dickinson, and S. N. Jarman. 2017. Measuring animal age with DNA methylation: From humans to wild animals. Front. Genet. 8:106. https://doi.org/10.3389/fgene.2017.00106

De Pontual, H., A. L. Groison, C. Piñeiro, and M. Bertignac. 2006. Evidence of underestimation of European hake growth in the Bay of Biscay, and its relationship with bias in the agreed method of age estimation. ICES J. Mar. Sci. 63:1674-1681. https://doi.org/10.1016/j.icesjms.2006.07.007

Deriso, R. B., T. J. Quinn II, and P. R. Neal. 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42(4):815-824. https://doi.org/10.1139/f85-104

Devlaming, V. G. Grossman, and F. Chapman. 1982. On the use of gonosomatic index. Comp. Biochem. Physiol. A 73:31-39. https://doi.org/10.1016/0300-9629(82)90088-3

Dorval, E. D. Porzio, B. D. Schwarzkopf, K. C. James, L. Vasquez, and B. Erisman. 2022. Sampling methodology for estimating life history parameters of coastal pelagic species along the U.S. Pacific Coast. U.S. Dep. Commer., NOAA Tech. Memo NMFS-SWFSC-660, 46 p.

Dudoit, S., J. Fridlyand, and T. P. Speed. 2002. Comparison of discrimination methods for the classification of tumors using gene expression data. J. Am. Stat. Assoc. 97(457):77-87. https://doi.org/10.1198/016214502753479248

Eisner, L. B., Y. I. Zuenko, E. O. Basyuk, L. L. Britt, J. T. Duffy-Anderson, S. Kotwicki, C. Ladd, and W. Cheng. 2020. Environmental impacts on walleye pollock (*Gadus chalcogrammus*) distribution across the Bering Sea shelf. Deep Sea Res. Pt. II: Top. Stud. Oceanogr. 181-182. https://doi.org/10.1016/j.dsr2.2020.104881

Eschmeyer, W. N., E. S. Herald, and H. Hammann. 1983. A field guide to Pacific coast fishes of North America from the Gulf of Alaska to Baja California. Houghton Mifflin Company, Boston.

Evans, G. T., and J. M. Hoenig. 1998. Testing and viewing symmetry in contingency tables, with application to readers of fish ages. Biometrics 54:620-629. https://doi.org/10.2307/3109768

Fablet, R., and N. Le Josse. 2005. Automated fish age estimation from otolith images using statistical learning. Fish. Res. 72(2-3):279-290. https://doi.org/10.1016/j.fishres.2004.10.008

Fablet, R., A. Chessel, S. Carbini, A. Benzinou, and H. de Pontual. 2009. Reconstructing individual shape histories of fish otoliths: A new image-based tool for otolith growth analysis and modeling. Fish. Res. 96(2):148-159. https://doi.org/10.1016/j.fishres.2008.10.011

Farrés, M., S. Platikanov, S. Tsakovski, and R. Tauler. 2015. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. J. Chemom. 29: 528-536. https://doi.org/10.1002/cem.2736

Fearn, T. 2008. The interaction between Standard Normal Variate and derivatives. NIR News 19:16-17. https://doi.org/10.1255/nirn.1098

Fennie, H. W., K. Grorud-Colvert, and S. Sponaugle. 2023. Larval rockfish growth and survival in response to anomalous ocean conditions. Sci. Rep. 13:4089. https://doi.org/10.1038/s41598-023-30726-5

Ferreira, D. S., J. A. L. Pallone, and R. J. Poppi. 2013. Fourier transform near-infrared spectroscopy (FT-NIRS) application to estimate Brazilian soybean [*Glycine max* (L.) Merril] composition. Food Res. Int. 51:53-58. https://doi.org/10.1016/j.foodres.2012.09.015

Ferreira, R. A., G. Teixeira, and L. A. Peternelli. 2021. Kennard-Stone method outperforms the random sampling in the selection of calibration samples in SNPs and NIR data. Ciência Rural 52:e20201072. https://doi.org/10.1590/0103-8478cr20201072

Field, J. C., E. J. Dick, M. Key, M. Lowry, Y. Lucero, A. MacCall, D. Pearson, S. Ralston, W. Sydeman, and J. Thayer. 2007. Population dynamics of an unexploited rockfish (*Sebastes jordani*) in the California Current. *In* Biology, Assessment, and Management of North Pacific Rockfishes (J. Heifetz, J. DiCosimo, A. J. Gharrett, M. S. Love, V. M. O'Connell, and R. D. Stanley, eds.), pp. 451-472. Alaska Sea Grant College Program, Fairbanks. https://doi.org/10.4027/bamnpr.2007.25

Fournier, D., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39:1195-1203. https://doi.org/10.1139/f82-157

Fournier, D. A., H. J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. N. Maunder, A. Nielson, and J. Sibert. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27:233-249. https://doi.org/10.1080/10556788.2011.597854

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33:1-22. https://doi.org/10.18637/jss.v033.i01

Funamoto, T., and Y. Mugiya. 1998. Binding of strontium vs. calcium to 17β-estradiol-induced proteins in the plasma of the goldfish *Carassius auratus*. Fish. Sci. 64(2):325-328. https://doi.org/10.2331/fishsci.64.325

Gal, Y., and Z. Ghahramani. 2016. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. Proceedings of the 33rd International Conference on Machine Learning, in Proceedings of Machine Learning Research 48:1050-1059. Available from https://proceedings.mlr.press/v48/gal16.html

Gal, Y., P. Koumoutsakos, F. Lanusse, G. Louppe, and C. Papadimitriou. 2022. Bayesian uncertainty quantification for machine-learning models in physics. Nat. Rev. Phys. 4:573-577. https://doi.org/10.1038/s42254-022-00498-4

Gallagher, N. B., and D. O'Sullivan. 2020. Selection of representative learning and test sets using the onion method. White paper, Eigenvector Research Incorporated, Manson, WA. Available at https://eigenvector.com/wp-content/uploads/2022/10/Onion_SampleSelection.pdf

Garner, S. B., B. K. Barnett, D. W. Chamberlin, F. C. Forrestal, T. E. Helser, I. M. Benson, W. Kline, and W. F. Patterson III. This volume. Sensitivity of gray snapper (*Lutjanus griseus*) stock assessment models to age inputs estimated with near infrared spectroscopy.

Goetz, B. J., C. E. Piston, and C. M. Gburski. 2012. Rockfish (*Sebastes*) species, p. 49-64. *In* M.E. Matta and D. K. Kimura, (editors), Age determination manual of the Alaska Fisheries Science Center Age and Growth Program. U.S. Dep. Commer., NOAA Professional Paper NMFS 13.

Goldstein, E.D., T. E. Helser, J. J. Vollenweider, A. Sreenivasan, and F. F. Sewall. 2021. Rapid and reliable assessment of fish physiological condition for fisheries research and management using Fourier transform near-infrared spectroscopy. Front. Mar. Sci. 8:690934. https://doi.org/10.3389/fmars.2021.690934

Goldstein, E. D., B. C. Hsieh, M. B. Arrington, and T. E. Helser. This volume-a. Quality assurance and quality control of Fourier transform near infrared spectroscopy data for age prediction.

Goldstein, E. D., C. D. Waters, H. K. Fulton-Bennett, M. E. Matta, T. E. Helser, J. J. Vollenweider, C. M. Hinds, K. W. McNeel, C. R. Kastelle, S. K. Neidetcher, D. S. Oxman, and F. J. Mueter. This volume-b. Developing spectroscopy approaches to measure life history characteristics of fish throughout ontogeny.

Gorbatenko, K.M., and A. E. Lazhentsev. 2016. The biochemical composition and calorie density of the walleye pollock *Theragra chalcogramma* in the Sea of Okhotsk. Russ. J. Mar. Biol. 42(7):591-601. https://doi.org/10.1134/S1063074016070038

Gulf of Mexico Fishery Management Council (GMFMC). 2013. Gulf of Mexico gag management history. SEDAR33-RD06. SEDAR, North Charleston, SC. 16 pp. Available at https://sedarweb.org/documents/s33rd06-gulf-of-mexico-gag-management-history/

Haenlein, M., and A. M. Kaplan. 2004. A beginner's guide to partial least squares analysis. Underst. Stat. 3:283-297. https://doi.org/10.1207/s15328031us0304_4

Healy, J., T. E. Helser, I. M. Benson, and L. Tornabene. 2021. Aging Pacific cod (*Gadus macrocephalus*) from otoliths using Fourier-transformed near-infrared spectroscopy. Ecosphere 12(8):e03697. https://doi.org/10.1002/ecs2.3697

Heather, J. M., and B. Chain. 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107:1-8. https://doi.org/10.1016%2Fj.ygeno.2015.11.003

Heino, M., B. Díaz Pauli, and U. Dieckmann. 2015. Fisheries-induced evolution. Annu. Rev. Ecol. Evol. Syst. 46:461-480. https://doi.org/10.1146/annurev-ecolsys-112414-054339

Helser, T. E., H. Lai, and B. A. Black. 2012. Bayesian hierarchical modeling of Pacific geoduck growth increment data and climate indices. Ecol. Model. 247:210-220. https://doi.org/10.1016/j.ecolmodel.2012.08.024

Helser, T. E., I. M. Benson, and B. K. Barnett. 2019a. Proceedings of the Research Workshop on the Rapid Estimation of Fish Age Using Fourier Transform Near-Infrared Spectroscopy (FT-NIRS). AFSC Processed Rep. 2019-06, 195 pp. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115. Available at http://www.afsc.noaa.gov/Publications/ProcRpt/PR2019-06.pdf

Helser, T. E., I. Benson, J. Erickson, J. Healy, C. Kastelle, and J. A. Short. 2019b. 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. 76(5):780-789. https://doi.org/10.1139/cjfas-2018-0112

Helser, T. E., I. M. Benson, M. E. Matta, E. D. Goldstein, B. C. Hsieh, M. B. Arrington, and J. A. Short. This volume-a. Envisioning the future of production fish ageing: End-to-end integration of the FT-NIR spectroscopy age estimation enterprise at the Alaska Fisheries Science Center.

Helser, T. E., M. R. Siskey, I. M. Benson, and S. J. Barbeaux. This volume-b. Sensitivity of Pacific cod stock assessment to alternative age composition data derived from Fourier transform near infrared spectroscopy of otoliths.

Herlau, T., M. N. Schmidt, and M. Morup. 2022. Bayesian dropout. The 3rd International Workshop on Statistical Methods and Artificial Intelligence. Elsevier B.V.

Herrero, A. M. 2008. Raman spectroscopy a promising technique for quality assessment of meat and fish: a review. Food Chem. 107:1642-1651. https://doi.org/10.1016/j.foodchem.2007.10.014

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Springer, New York, NY.

Hollowed, A. B., M. Barange, R. Beamish, K. Brander, K. Cochrane, K. Drinkwater, M. Foreman, J. Hare, J. Holt, S.-I. Ito, S. Kim, J. King, H. Loeng, B. MacKenzie, F. Mueter, T. Okey, M. A. Peck, V. Radchenko, J. Rice, M. Schirripa, A. Yatsu, and Y. Yamanaka. 2013. Projected impacts of climate change on marine fish and fisheries. ICES J. Mar. Sci. 70:1023-1037. https://doi.org/10.1093/icesjms/fst081

Holsman, K. K., E. L. Hazen, A. Haynie, S. Gourguet, A. Hollowed, S. J. Bograd, J. F. Samhouri, and K. Aydin. 2019. Towards climate resiliency in fisheries management. ICES J. Mar. Sci. 76:1368-1378. https://doi.org/10.1093/icesjms/fsz031

Horvath, S. 2013. DNA methylation age of human tissues and cell types. Genome Biol. 14:R115. https://doi.org/10.1186/gb-2013-14-10-r115

Ianelli, J., S. Stienessen, T. Honkalehto, E. Siddon, and C. Allen-Akselrud. 2022. Assessment of the walleye pollock stock in the eastern Bering Sea. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions. North Pacific Fishery Management Council, Anchorage, AK.

Ianelli, J. N. This volume. Fourier transform near infrared spectroscopy data and application within the eastern Bering Sea pollock assessment model.

Ingle, P. D., R. Christian, P. Purohit, V. Zarraga, and E. Handley. 2016. Determination of protein content by NIR spectroscopy in protein powder mix products. J. of AOAC Intl. 99(2):360-363. https://doi.org/10.5740/jaoacint.15-0115

Ishigaki, M., S. Kawasaki, and D. Ishikawa. 2016. Near-infrared spectroscopy and imaging studies of fertilized fish eggs: in vivo monitoring of egg growth at the molecular level. Sci. Rep. 6:20066. https://doi.org/10.1038/srep20066

Jackson, J. R. 2007. Earliest references to age determination of fishes and their early application to the study of fisheries. Fish. 32 (7):321-328.

Javor, B., N. Lo, and R. Vetter. 2011. Otolith morphometrics and population structure of Pacific sardine (*Sardinops sagax*) along the west coast of North America. Fish. Bull. 109: 402-415.

Jones, C. M. 2013. Growth and mortality of pre and post-settlement age-0 red snapper, *Lutjanus campechanus* (Poey 1860), in the Gulf of Mexico. PhD dissertation, University of South Alabama, 128 pp.

Jung, M., and G. P. Pfeifer. 2015. Aging and DNA methylation. BMC Biol. 13(1):1-8. https://doi.org/10.1186/s12915-015-0118-4

Kastelle, C. R., and D. K. Kimura. 2006. Age validation of walleye pollock (*Theragra chalcogramma*) from the Gulf of Alaska using the disequilibrium of Pb-210 and Ra-226. ICES J. Mar. Sci. 63(8):1520-1529. https://doi.org/10.1016/j.icesjms.2006.06.002

Kastelle, C. R., T. E. Helser, S. G. Wischniowski, T. Loher, B. J. Goetz, and L. A. Kautzi. 2016. Incorporation of bomb-produced ^{14}C into fish otoliths: a novel approach for evaluating age validation and bias with an application to yellowfin sole and northern rockfish. Ecol. Modell. 320:79-91. https://doi.org/10.1016/j.ecolmodel.2015.09.013

Kennard, R. W., and L. A. Stone. 1969. Computer aided design of experiments. Technom. 11(1): 137-148. https://doi.org/10.2307/1266770

Kimura, D. K., and D. M. Anderl. 2005. Quality control of age data at the Alaska Fisheries Science Center. Mar. Freshw. Res. 56(5):783-789. https://doi.org/10.1071/MF04141

Kimura, D. K., and J. J. Lyons. 1990. Choosing a structure for the production ageing of Pacific cod (*Gadus macrocephalus*). Int. N. Pac. Fish. Comm. Bull. 50:9-23.

Kimura, D. K., and J. J. Lyons. 1991. Between-reader bias and variability in the age-determination process. Fish. Bull. 89:53-60.

Kimura, D. K., C. R. Kastelle, B. J. Goetz, C. M. Gburski, and A. V. Buslov. 2006. Corroborating the ages of walleye pollock (*Theragra chalcogramma*). Mar. Freshw. Res. 57(3):323-332. https://doi.org/10.1071/MF05132

Kucheryavskiy, S. 2020. *mdatools – R package for chemometrics*. Chemom. Intell. Lab. Syst. 198:103937. https://doi.org/10.1016/j.chemolab.2020.103937

Kucheryavskiy, S. 2020. Projection based methods for preprocessing, exploring and analysis of multivariate data used in chemometrics. https://doi.org/10.1016/j.chemolab.2020.103937

Kuriyama, P. T., J. P. Zwolinski, K. T. Hill, and P. R. Crone. 2020. Assessment of the Pacific sardine resource in 2020 for U.S. management in 2020-2021. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-SWFSC-628.

Kvalheim, O. M. 2010. Interpretation of partial least squares regression models by means of target projection and selectivity ratio plots. J. Chemom. 24:496-504. https://doi.org/10.1002/cem.1289

Lai, H.-L. 1987. Optimum allocation for estimating age composition using age-length key. Fish. Bull. 85:179-186.

Lai, H. L., and D. R. Gunderson. 1987. Effects of ageing errors on estimates of growth, mortality and yield per recruit for walleye pollock (*Theragra chalcogramma*). Fish. Res.5:287-302. https://doi.org/10.1016/01657836(87)90048-8

Laidig, T. E., K. R. Silberberg, and P. B. Adams. Age validation of the first, second, and third annulus from the dorsal fin rays of lingcod (*Ophiodon elongatus*). NOAA Tech. Memo. NMFS-SWFSC-306.

Lam, L. S., B. L. Basnett, M. A. Haltuch, J. Cope, K. Andrews, K. M. Nichols, G. C. Longo, J. F. Samhouri, and S. L. Hamilton. 2021. Geographic variability in lingcod *Ophiodon elongatus* life history and demography along the US West Coast: oceanographic drivers and management implications. Mar. Ecol. Prog. Ser. 670:203-222. https://doi.org/10.3354/meps13750

Lang, C., F. R. C. Costa, J. L. C. Camargo, F. M. Durgante, and A. Vicentini. 2015. Near infrared spectroscopy facilitates rapid identification of both young and mature Amazonian tree species. PLoS One 10: e0134521. https://doi.org/10.1371/journal.pone.0134521

Langseth B., J. Syslo, A. Yau, and F. Carvalho. 2019. Stock assessments of the bottomfish management unit species of Guam, the Commonwealth of the Northern Mariana Islands, and American Samoa, 2019. NOAA Tech Memo. NMFS-PIFSC-86, 177 pp. (+ supplement, 165 pp.). https://doi.org/10.25923/bz8b-ng72

Lauth, R. R., E. J. Dawson, and J. Conner. 2019. Results of the 2017 eastern and northern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate fauna. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-396, 260 p. Available at https://repository.library.noaa.gov/view/noaa/20734

Lavin, A., C.M. Gilligan-Lee, A. Visnjic, S. Ganju, D. Newman, S. Ganguly, D. Lange, A.G. Baydin, A. Sharma, A. Gibson, S. Zheng, E.P. Xing, C. Mattmann, J. Parr, and Y. Gal. 2022. Technology readiness levels for machine learning systems. Nat. Comm. 13(1):6039. https://doi.org/10.1038/s41467-022-33128-9

Legendre, P., and L. Legendre. 2012. Numerical ecology. Elsevier, Amsterdam, The Netherlands.

Li, L., K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar. 2017. Hyperband: a novel bandit-based approach to hyperparameter optimization. J. Mach. Learn. Res. 18:6765-6816.

Lieber, C. A., and A. Mahadevan-Jansen. 2003. Automated method for subtraction of fluorescence from biological Raman spectra. Appl. Spectrosc. 57(11):1363-1367. https://doi.org/10.1366/000370203322554518

Liu, A., G. Li, Z. Fu, Y. Guan, and L. Lin. 2018. Non-linearity correction in NIR absorption spectra by grouping modeling according to the content of analyte. Sci. Rep. 8:8564. https://doi.org/10.1038/s41598-018-26802-w

Liu, D., X.-A. Zeng, and D.-W. Sun. 2013. NIR spectroscopy and imaging techniques for evaluation of fish quality – a review. Appl. Spectrosc. Rev. 48(8):609-628. https://doi.org/10.1080/05704928.2013.775579

Loeun, K. L., S. Goldstien, D. Gleeson, S. J. Nicol, and C. J. Bradshaw. 2014. Limited genetic structure among broad-scale regions for two commercially harvested, tropical deep-water snappers in New Caledonia. Fish. Sci. 80(1):13-19. https://doi.org/10.1007/s12562-013-0673-y

Love, M. S., M. Yoklavich, and L. Thorsteinson. 2002. The rockfishes of the Northeast Pacific. University of California Press, Berkeley and Los Angeles, CA.

Lowerre-Barbieri, S., H. Menendez, J. Bickford, T. S. Switzer, L. Barbieri, and C. Koenig. 2020. Testing assumptions about sex change and spatial management in the protogynous gag grouper, *Mycteroperca microlepis*. Mar. Ecol. Prog. Ser. 639:199-214. https://doi.org/10.3354/meps13273

Lu, A., Z. Fei, A. Haghani, T. R. Robeck, J. A. Zoller, C. Z. Li, J. Zhang, J. Ablaeva, D. M. Adams, J. Almunia, et al. 2023. Universal DNA methylation age across mammalian tissues. Nat. Aging 3:1144-1166. https://doi.org/10.1038/s43587-023-00462-6

Lu, J., A. Johnston, P. Berichon, K. Ru, D. Korbie, and M. Trau. 2017. PrimerSuite: a high-throughput web-based primer design program for multiplex bisulfite PCR. Sci. Rep. 7:1-12. https://doi.org/10.1038/srep41328

Lundberg, S. M., and S.-I. Lee. 2017. A unified approach to interpreting model predictions. *In* 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA. Curran Associates, Inc., Red Hook, NY. https://doi.org/10.48550/arXiv.1705.07874

Luthria, D. L., S. Mukhopadhyay, L. Z. Lin, and J. M. Harnly. 2011. A comparison of analytical and data preprocessing methods for spectral fingerprinting. Appl. Spectrosc. 65(3):250-259. http://doi.org/10.1366/10-06109

Mahe, K., K. Rabhi, E. Bellamy, R. Elleboode, Y. Aumond, J. Huet, P. Cresson, P., and D. Roos. 2016. Growth of the oblique-banded grouper (*Epinephelus radiatus*) on the coasts of Reunion Island (SW Indian Ocean). Cybium 40(1):61-65.

Martens, H., and E. Stark. 1991. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. J. Pharm. Biomed. Anal. 9(8):625-635. https://doi.org/10.1016/0731-7085(91)80188-F

Martin, B. T., R. Heintz, E. M. Danner, and R. M. Nisbet. 2017. Integrating lipid storage into general representations of fish energetics. J. Anim. Ecol. 86:812-825. https://doi.org/10.1111/1365-2656.12667

Matta, M. E., and D. K. Kimura (editors). 2012. Age determination manual of the Alaska Fisheries Science Center Age and Growth Program. NOAA Professional Paper NMFS 13, 97 p.

Matta, M. E., K. M. Rand, M. B. Arrington, and B. A. Black. 2020. Competition-driven growth of Atka mackerel in the Aleutian Islands ecosystem revealed by an otolith biochronology. Estuar. Coast. Shelf Sci. 240:106775. https://doi.org/10.1016/j.ecss.2020.106775

Matta, M. E., J. D. Brogan, J. A. Short, and T. E. Helser. This volume-a. Quality control and assurance of reference age data at the Alaska Fisheries Science Center.

Matta, M. E., E. D. Goldstein, I. M. Benson, H. K. Fulton-Bennett, C. D. Waters, B. C. Hsieh, and T. E. Helser. This volume-b. Rapid daily age estimation of juvenile walleye pollock in the Gulf of Alaska using FT-NIR spectroscopy.

Maunder, M. N., and A. E. Punt. 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142:61-74. https://doi.org/10.1016/j.fishres. 2012.07.025

Mayne, B., D. Korbie, L. Kenchington, B. Ezzy, O. Berry, and S. Jarman. 2020. A DNA methylation age predictor for zebrafish. Aging (Albany NY) 12:24817-24835. https://doi.org/10.18632/aging.202400

Mayne, B., T. Espinoza, D. Roberts, G. L. Butler, S. Brooks, D. Korbie, and S. Jarman. 2021. Nonlethal age estimation of three threatened fish species using DNA methylation: Australian lungfish, Murray cod and Mary River cod. Mol. Ecol. Resour. 21:2324-2332. https://doi.org/10.1111/1755-0998.13440

McBride, R. S. 2015. Diagnosis of paired age agreement: a simulation of accuracy and precision effects. ICES J. Mar. Sci. 72(7):2149-2167. https://doi.org/10.1093/icesjms/fsv047

McCarthy, A., S. Stienessen, M. Levine, and D. Jones. 2018. Results of the acoustic-trawl surveys of walleye pollock (*Gadus chalcogrammus*) in the Gulf of Alaska, February-March 2017 (DY2017-01, DY2017-02, and DY2017-03). AFSC Processed Rep. 2018-04, 126 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

McFarlane, G. A., and J. R. King. 2001. The validity of the fin-ray method of age determination for lingcod (*Ophiodon elongatus*). Fish. Bull. 99(3):459-459.

Melnychuk, M. C., H. Kurota, P. M. Mace, et al. 2021. Identifying management actions that promote sustainable fisheries. Nature Sustainability 4:440-449. https://doi.org/10.1038/s41893-020-00668-1

Methot, R. D., and C. R. Wetzel. 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142:86-99. https://doi.org/10.1016/j.fishres.2012.10.012

Miller, C. E. 1993. Sources of non-linearity in near infrared methods. NIR News 4:3-5. https://doi.org/10.1255/nirn.216

Moen, E., N. O. Handegard, V. Allken, O. T. Albert, A. Harbit, and K. Malde. 2018. Automatic interpretation of otoliths using deep learning. PLoS One 13(12):e0204713. https://doi.org/10.1371/journal.pone.0204713

Mujica, L., J. Rodellar, A. Fernandez, and A. Guemes. 2011. Q-statistic and T^2-statistic PCA-based measures for damage assessment in structures. Struct. Health Monit. 10:539-553. https://doi.org/10.1177/1475921710388972

Murie, D. J., and D. C. Parkyn. 2005. Age and growth of white grunt (*Haemulon plumieri*): a comparison of two populations along the west coast of Florida. Bull. Mar. Sci. 76:73-93.

Murray, I., and P. Williams. 1987. Chemical principles of near-infrared technology. *In* Near infrared technology in the agricultural and food industries (P. Williams and K. H. Norris, eds.). American Association of Cereal Chemists and University of Wisconsin, Madison, WI.

Nasreddine, K., A. Benzinou, and R. Fablet. 2013. Geodesics-based image registration: Applications to biological and medical images depicting concentric ring patterns. IEEE Trans. Image Process. 22(11):4436-4446. https://doi.org/10.1109/TIP.2013.2273670

Neidetcher, S. K., M. B. Arrington, T. E. Helser, E. D. Goldstein, and I. M. Benson. This volume. A novel approach for determining the reproductive status of walleye pollock (*Gadus chalcogrammus*) using Raman spectroscopy.

Neil, J. C., and K. W. McNeel. 2018. Age Determination Unit age-structure pattern interpretation: lingcod *Ophiodon elongatus*. Unpublished Manuscript.

Nesslage, G., A. M. Schueller, A. R. Rezek, and R. M. Mroch. 2022. Influence of sample size and number of age classes on characterization of ageing error in paired-age comparisons. Fish. Res. 249:106255. https://doi.org/10.1016/j.fishres.2022.106255

Norgaard, L., A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, and S. B. Engelsen. 2000. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. Appl. Spec. 54:413-419. https://doi.org/10.1366/0003702001949500

Northeast Fisheries Science Center (NEFSC). In review. Labcut 250B Saw Processing Protocol. https://www.fisheries.noaa.gov/new-england-mid-atlantic/science-data/age-and-growth-studies-northeast

O'Donnell, T. P., M. J. Reichert, and T. L. Darden. 2019. Genetic population structure of white grunt in the southeastern United States. N. Am. J. Fish. Manag. 39(4):725-737. https://doi.org/10.1002/nafm.10306

Ogle, D. H. 2017a. Age bias plot changes in FSA. Available from https://derekogle.com/fishR/2017-04-26-AgeBias_inFSA [accessed 10 October 2022].

Ogle, D. H. 2017b. fishR Vignette - precision and accuracy in ages. Available from http://derekogle.com/fishR/examples/oldFishRVignettes/AgeComparisons.pdf [accessed 16 August 2023].

Ogle, D. H. 2018. Introductory fisheries analyses with R. Chapman and Hall/CRC, New York. 338 p. https://doi.org/10.1201/9781315371986

Ogle, D. H., T. O. Brenden, and J. L. McCormick. 2017. Growth estimation: growth models and statistical inference. *In* Age and growth of fishes: principles and techniques (M. C. Quist and D. A. Isermann, eds.), p. 265-359. American Fisheries Society, Bethesda, Maryland.

Ogle, D., J. Doll, P. Wheeler, and A. Dinno. 2021. FSA: Fisheries Stock Analysis. R package version 0.9.1. https://github.com/droglenc/FSA

Ohlberger, J., Ø. Langangen, and L. Chr. Stige. 2022. Age structure affects population productivity in an exploited fish species. Ecol. Appl. 32(5):e2614. https://doi.org/10.1002/eap.2614

Oliveri, P., M. C. Casolino, M. Casale, L. Medini, F. Mare, and S. Lanteri. 2013. A spectral transfer procedure for application of a single class-model to spectra recorded by different near-infrared spectrometers for authentication of olives in brine. Anal. Chim. Acta 761:46-52. https://doi.org/10.1016/j.aca.2012.11.020

Ono, K., R. Licandeo, M.L. Muradian, C.J. Cunningham, S.C. Anderson, F. Hurtado-Ferro, K.F. Johnson, C.R. McGilliard, C.C. Monnahan, C.S. Szuwalski, J.L. Valero, K.A. Vert-Pre, A.R. Whitten, and A.E. Punt. 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES J. Mar. Sci. 72(1):31-43. https://doi.org/10.1093/icesjms/fsu007

Ordoñez, A., L. Eikvil, A.-B. Salberg, A. Harbitz, S. M. Murray, and M. C. Kampffmeyer. 2020. Explaining decisions of deep neural networks used for fish age prediction. PLoS ONE 15(6):e0235013. https://doi.org/10.1371/journal.pone.0235013

Padalkar, M. V., and N. Pleshko. 2015. Wavelength-dependent penetration depth of near infrared radiation into cartilage. Analyst 140(7):2093-2100. https://doi.org/10.1039/C4AN01987C

Passerotti, M. S., C. M. Jones, C. E. Swanson, and J. M. Quattro. 2020a. Fourier-transform near infrared spectroscopy (FT-NIRS) rapidly and non-destructively predicts daily age and growth in otoliths of juvenile red snapper *Lutjanus campechanus* (Poey, 1860). Fish. Res. 223:105439. https://doi.org/10.1016/j.fishres.2019.105439

Passerotti, M. S., T. E. Helser, I. M. Benson, B. K. Barnett, J. C. Ballenger, W. J. Bubley, M. J. M. Reichert, and J. M. Quattro. 2020b. Age estimation of red snapper (*Lutjanus campechanus*) using FT-NIR spectroscopy: feasibility of application to production ageing for management. ICES J. Mar. Sci. 77:2144-2156. https://doi.org/10.1093/icesjms/fsaa131

Passerotti, M. S., M. J. M. Reichert, B. A. Robertory, Z. Marsh, M. Stefik, and J. M. Quattro. 2022. Physicochemical mechanisms of FT-NIRS age prediction in fish otoliths. Mar. Freshwat. Res. 73(6):846-865. https://doi.org/10.1071/MF21341

Pearson, D. E., J. E. Hightower, and J. T. H. Chan. 1991. Age, growth, and potential yield for shortbelly rockfish *Sebastes jordani*. Fish. Bull. 89(3):403-409.

Pilling, G. M., R. S. Millner, M. W. Easey, D. L. Maxwell, and A. N. Tidd. 2007. Phenology and North Sea cod *Gadus morhua* L.: has climate change affected otolith annulus formation and growth? J. Fish Biol. 70(2):584-599. https://doi.org/10.1111/j.1095-8649.2007.01331.x

Politikos, D. V., G. Petasis, A. Chatzispyrou, C. Mytilineou, and A. Anastasopoulou. 2021. Automating fish age estimation combining otolith images and deep learning: the role of multitask learning. Fish. Res. 242:106033. https://doi.org/10.1016/j.fishres.2021.106033

Pomerantsev, A. L., and O. Y. Rodionova. 2014. Concept and role of extreme objects in PCA/SIMCA. J. Chemom. 28(5):429-438. https://doi.org/10.1002/cem.2506

Potts, J. C. 2000. Population assessment of two stocks of white grunt, *Haemulon plumieri*, from the southeastern coast of the United States. NOAA Tech. Memo. NMFS-SEFSC-442, 67 p.

Potts, J. C., and C. S. Manooch. 2001. Differences in the age and growth of white grunt (*Haemulon plumieri*) from North Carolina and South Carolina compared with southeast Florida. Bull. Mar. Sci. 68:1-12.

Punt, A. E., D. C. Smith, K. KrusicGolub, and S. Robertson. 2008. Quantifying age-reading error for use in fisheries stock assessments, with application to species in Australia's southern and eastern scalefish and shark fishery. Can. J. Fish. Aquat. Sci. 65(9):1991-2005. https://doi.org/10.1139/F08-111

Quinn, T. J., and R. B. Deriso. 1999. Growth and fecundity. *In* Quantitative fish dynamics, p. 128-207. Oxford University Press, New York.

R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Raml, C., X. He, M. Han, D. R. Alexander, and Y. Lu. 2011. Raman spectroscopy based on a single-crystal sapphire fiber. Opt. Lett. 36(7):1287-1289. https://doi.org/10.1364/OL.36.001287.

Rasco, B. A., C. E. Miller, and T. L. King. 1991. Utilization of NIR spectroscopy to estimate the proximate composition of trout muscle with minimal sample pre-treatment. J. Agri. Food Chem. 39:67-72. https://doi.org/10.1021/jf00001a012

Reeves, S. A. 2003. A simulation study of the implications of age-reading errors for stock assessment and management advice. ICES J. Mar. Sci. 60:314-328. https://doi.org/10.1016/S1054-3139(03)00011-0

Richards, L. J., J. T. Schnute, A. Kronlund, and R. J. Beamish. 1992. Statistical models for the analysis of ageing error. Can. J. Fish. Aquat. Sci. 49:1801-1815. https://doi.org/10.1139/f92-200

Rideout, R. M., T. N. Youssef, A. T. Adamack, R. John, A. M. Cohen, T. D. Fridgen, and J. H. Banoub. 2023. Qualitative shotgun proteomics strategy for protein expression profiling of fish otoliths. BioChem 3:102-117. https://doi.org/10.3390/biochem3030008

Rigby, C. L., B. B. Wedding, S. Grauf, and C. A. Simpfendorfer. 2014. The utility of near infrared spectroscopy for age estimation of deepwater sharks. Deep-Sea Res. Pt. I 94:184-194. https://doi.org/10.1016/j.dsr.2014.09.004

Rigby, C. L., B. B. Wedding, S. Grauf, and C. A. Simpfendorfer. 2016. Novel method for shark age estimation using near infrared spectroscopy. Mar. Freshw. Res. 67(5):537-545. https://doi.org/10.1071/MF15104

Rinnan, Å., F. van den Berg, and S. B. Engelsen. 2009. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal. Chem. 28:1201-1222. https://doi.org/10.1016/j.trac.2009.07.007

Roberson, N. E., D. K. Kimura, D. R. Gunderson, and A. M. Shimada. 2005. Indirect validation of the age-reading method for Pacific cod (*Gadus macrocephalus*) using otoliths from marked and recaptured fish. Fish. Bull. 103:153-160.

Robins, J. B., B. B. Wedding, C. Wright, S. Grauf, M. Sellin, A. Fowler, T. Saunders, and S. Newman S. 2015. Revolutionising fish ageing: using near infrared spectroscopy to age fish. Department of Agriculture, Fisheries and Forestry. Brisbane, April 2015. http://frdc.com.au/research/Final_reports/2012-011-DLD.pdf

Rodionova, O. Y., and A. L. Pomerantsev. 2020. Detection of outliers in projection-based modeling. Anal. Chem. 92:2656-2664. https://doi.org/10.1021/acs.analchem.9b04611

Rogers, L. A., and A. B. Dougherty. 2019. Effects of climate and demography on reproductive phenology of a harvested marine fish population. Glob. Change Biol. 25(2):708-720. https://doi.org/10.1111/gcb.14483

Rogers, L. A., M. T. Wilson, J. T. Duffy-Anderson, D. G. Kimmel, and J. F. Lamb. 2021. Pollock and "the Blob": impacts of a marine heatwave on walleye pollock early life stages. Fish. Oceanogr. 30:142-158. https://doi.org/10.1111/fog.12508

Rothschild, B. J., and M. J. Fogarty. 1989. Spawning-stock biomass: A source of error in recruitment/stock relationships and management advice. ICES J. Mar. Sci. 45(2):131-135. https://doi.org/10.1093/icesjms/45.2.131

Rubin, A. M., M. S. Passerotti, and E. M. Robillard. This volume. Fourier transform near infrared spectroscopy ageing of finfish and shark species in the northwest Atlantic.

Savitzky, A., and M. J. E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36:1627-1639. https://doi.org/10.1021/AC60214A047

Shapley, L. S. 1953. A value for n-person games. *In* Contributions to the theory of games (H. Kuhn and A. Tucker, eds.), p.:307-317. Princeton University Press, Princeton, NJ. https://doi.org/10.1515/9781400881970-018

Short, J. A. This volume. Database design and considerations with FT-NIR spectra data collection and management.

Shotwell, S. K., J. L. Pirtle, J. T. Watson, and others. 2022. Synthesizing integrated ecosystem research to create informed stock-specific indicators for next generation stock assessments. Deep Sea Res. Part II: Top. Stud. Oceanogr. 198:105070. https://doi.org/10.1016/j.dsr2.2022.105070

Siesler, H. W., S. Kawata, H. M. Heise, and Y. Ozaki. 2008. Near-infrared spectroscopy: principles, instruments, applications. John Wiley & Sons, Hoboken, NJ, 361 p.

Smith, M. K. 1992. Regional differences in otolith morphology of the deep slope red snapper *Etelis carbunculus*. Can. J. Fish. Aquat. Sci. 49(4):795-804. https://doi.org/10.1139/f92-090

Smith, M. K., and E. Kostlan. 1991. Estimates of age and growth of ehu *Etelis carbunculus* in four regions of the Pacific from density of daily increments in otoliths. Fish. Bull. 89(3):461-472.

Smith, N., C. Donato-Hunt, V. Allain, S. McKechnie, B. Moore, and I. Bertram. 2017. Developing a Pacific community marine specimen bank. 10th SPC Heads of Fisheries Meeting. 5 pp.

Solo. 2021. Eigenvector Research, Inc., Manson, WA, version 8.9.2. Software available at http://www.eigenvector.com

Sorvaniemi, J., A. Kinnunen, A. Tsados, and Y. Malkki. 1993. Using partial least-squares regression and multiplicative scatter correction for FT-NIR data evaluation of wheat flours. J. Food Sci. Tech. 26(3):251-258. https://doi.org/10.1006/fstl.1993.1053

Southeast Data Assessment and Review (SEDAR). 2018. SEDAR 52 Stock Assessment Report Gulf of Mexico red snapper. SEDAR, North Charleston, SC 434 p. https://sedarweb.org/assessments/sedar-52/

Southeast Data Assessment and Review (SEDAR). 2021. SEDAR 72 Gulf of Mexico gag grouper final stock assessment report. SEDAR, North Charleston, SC. 319 pp. http://sedarweb.org/sedar-72

Southeast Data Assessment and Review (SEDAR). 2022. SEDAR 74 Gulf of Mexico red snapper Section III: Data Workshop Report. SEDAR, North Charleston, SC 225 p. https://sedarweb.org/assessments/sedar-74/

Spencer, P.D., and M. W. Dorn. 2013. Incorporation of weight-specific relative fecundity and maternal effects in larval survival into stock assessments. Fish. Res. 138:159-167. https://doi.org/10.1016/j.fishres.2012.05.003

Spies, I., L. Barnett, R. Haehn, J. Ianelli, E. Markowitz, Z. Oyafuso, E. Siddon, and C. Yeung. 2022. Assessment of the yellowfin sole stock in the Bering Sea and Aleutian Islands. *In* 2022 North Pacific groundfish stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions. North Pacific Fishery Management Council, Anchorage, AK, 117 p.

Stevens, A., and L. Ramirez-Lopez. 2022. An introduction to the prospectr package. R package vignette package version 0.2.6. https://cran.r-project.org/web/packages/prospectr/vignettes/prospectr.html

Switzer, T. S., T. C. MacDonald, R. H. McMichael Jr., and S. F. Keenan. 2012. Recruitment of juvenile gags in the eastern Gulf of Mexico and factors contributing to observed spatial and temporal patterns of estuarine occupancy. Trans. Amer. Fish. Soc. 141:707-719. https://doi.org/10.1080/00028487.2012.675913

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. 2016. Rethinking the inception architecture for computer vision, p. 2818-2826. *In* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV. https://doi.org/10.1109/CVPR.2016.308

Talari, A. C. S., Z. Movasaghi, S. Rehman, and I. U. Rehman. 2015. Raman spectroscopy of biological tissues. Appl. Spectrosc. Rev. 50(1):46-111. https://doi.org/10.1080/05704928.2014.923902

TenBrink, T., S. Neidetcher, M. Arrington, I. Benson, C. Conrath, and T. Helser. 2022. Fourier transform near infrared spectroscopy as a tool to predict spawning status in Alaskan fishes with variable reproductive strategies. J. Near Infrared Spectrosc. 30(4):179-188. https://doi.org/10.1177/09670335221097005

TenBrink, T. T., I. M. Benson, B. C. Hsieh, M. E. Matta, and T. E. Helser. This volume. Preliminary analysis of Fourier transform near Infrared (FT-NIR) spectroscopy and convolutional neural network (CNN) models to predict ages of yellowfin sole.

Thomas, O. R. B., S. E. Swearer, E. A. Kapp, P. Peng, G. Q. Tonkin-Hill, A. Papenfuss, A. Roberts, P. Bernard, and B. R. Roberts. 2019. The inner ear proteome of fish. FEBS J. 286(1):66-81. https://doi.org/10.1111/febs.14715

Thompson, G. G. 2018. Assessment of the Pacific cod stock in the eastern Bering Sea. *In* 2018 Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea and Aleutian Islands regions. North Pacific Fishery Management Council, 605 West 4th Ave., Suite 306, Anchorage, AK. Available at https://www.fisheries.noaa.gov/resource/data/2018-assessment-pacific-cod-stock-eastern-bering-sea

Thompson, G. G., J. Conner, S. K. Shotwell, B. Fissel, T. Hurst, B. Laurel, L. Rogers, and E. Siddon. 2020. Assessment of the Pacific cod stock in the eastern Bering Sea. *In* 2020 Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea and Aleutian Islands regions. North Pacific Fishery Management Council, 605 West 4th Ave., Suite 306, Anchorage, AK. Available at https://www.fisheries.noaa.gov/resource/data/2020-assessment-pacific-cod-stock-eastern-bering-sea

Townsend H., C. J. Harvey, Y. deReynier, D. Davis, S. G. Zador, S. Gaichas, M. Weijerman, E. L. Hazen, and I. C. Kaplan. 2019. Progress on implementing ecosystem-based fisheries management in the United States through the use of ecosystem models and analysis. Front. Mar. Sci. 6:641. https://www.frontiersin.org/articles/10.3389/fmars.2019.00641

Troadec, H., and A. Benzinou. 2002. Computer assisted age estimation. *In* Manual of fish sclerochronology (J. Panfili, H. de Pontual, H. Troadec, and P. J. Wright, eds.), p. 199-241. Ifremer-IRD coedition, Brest, France.

Vance, C. K., A. J. Kouba, and S. T. Willard. 2014. Near infrared spectroscopy applications in amphibian ecology and conservation: gender and species identification. NIR News 25:5-10. https://doi.org/10.1255/nirn.1444

Vance, C. K., A. J. Kouba, H. X. Zhang, H. Zhou, Q. Wang, and S. T. Willard. 2016a. Near infrared reflectance spectroscopy (NIRS) studies of Chinese giant salamanders in aquaculture production. NIR News 26(2):4-7. http://doi/10.1255/nirn.1510

Vance, C. K., D. R. Tolleson, K. Kinoshita, J. Rodriguez, and W. J. Foley. 2016b. Near infrared spectroscopy in wildlife and biodiversity. J. Near Infrared Spectrosc. 24(1):1-25. https://doi.org/10.1255/jnirs.1199

Van Rossum, G., and F. L. Drake. 2009. Python 3 reference manual. CreateSpace, Scotts Valley, CA.

Wakefield, C. B., A. J. Williams, S. J. Newman, M. Bunel, C. E. Dowling, C. A. Armstrong, and T. J. Langlois. 2014. Rapid and reliable multivariate discrimination for two cryptic Eteline snappers using otolith morphometry. Fish. Res. 151:100-106. https://doi.org/10.1016/j.fishres.2013.10.011

Wakefield, C. B., A. J. Williams, E. A. Fisher, N. G. Hall, S. A. Hesp, T. Halafihi, and others. 2020. Variations in life history characteristics of the deep-water giant ruby snapper (*Etelis* sp.) between the Indian and Pacific Oceans and application of a data-poor assessment. Fish. Res. 230:105651. https://doi.org/10.1016/j.fishres.2020.105651

Wallace, J. R. This volume. Exploring the use of FT-NIR spectroscopy for ageing sablefish (*Anoplopoma fimbria*) and Pacific hake (*Merluccius productus*) off the U.S. West Coast.

Wedding, B. B., A. J. Forrest, C. Wright, S. Grauf, and P. Exley. 2014. A novel method for the age estimation of saddletail snapper (*Lutjanus malabaricus*) using Fourier transform–near infrared (FT-NIR) spectroscopy. Mar. Freshw. Res. 65(10):894-900. https://doi.org/10.1071/MF13244

West, G. 1990. Methods of assessing ovarian development in fishes: a review. Aust. J. Mar. Freshw. Res. 41:199-222. https://doi.org/10.1071/MF9900199

Williams, A. J., K. Loeun, S. J. Nicol, P. Chavance, M. Ducrocq, S. J. Harley, G. M. Pilling, V. Allain, C. Mellin, and C. J. A. Bradshaw. 2013. Population biology and vulnerability to fishing of deep-water Eteline snappers. J. Appl. Ichthyol. 29(2):395-403. https://doi.org/10.1111/jai.12123

Williams, A. J., S. J. Newman, C. B. Wakefield, M. Bunel, T. Halafihi, J. Kaltavara, and S. J. Nicol. 2015. Evaluating the performance of otolith morphometrics in deriving age compositions and mortality rates for assessment of data-poor tropical fisheries. ICES J. Mar. Sci. 72(7):2098-2109. https://doi.org/10.1093/icesjms/fsv042

Williams, A. J., C. B. Wakefield, S. J. Newman, E. Vourey, F. J. Abascal, T. Halafihi, J. Kaltavara, and S. J. Nicol. 2017. Oceanic, latitudinal, and sex-specific variation in demography of a tropical deepwater snapper across the Indo-Pacific region. Front. Mar. Sci. 4:382. https://doi.org/10.3389/fmars.2017.00382

Williams, K. 2007. Evaluation of the macroscopic staging method for determining maturity of female walleye pollock *Theragra chalcogramma* in Shelikof Strait, Alaska. Alaska Fish. Res. Bull. 12(2):252-263.

Williams, T., and B. C. Bedford. 1974. The use of otoliths for age determination. *In* The ageing of fish. Proceedings of an International Symposium (T. B. Bagenal, ed.), p. 114-123. Unwin Brothers Limited, Surrey, England.

Wise, B. M., and R. T. Roginski. 2015. A calibration model maintenance roadmap. IFAC-PapersOnLine 48(8):260-265. https://doi.org/10.1016/j.ifacol.2015.08.191

Wold, S., A. Ruhe, H. Wold, and W. J. Dunn III. 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM J. Sci. Stat. Comput. 5(3):735-743. https://doi.org/10.1137/0905052

Wright, C., B. B. Wedding, S. Grauf, and O. J. Whybird. 2021. Age estimation of barramundi (*Lates calcarifer*) over multiple seasons from the southern Gulf of Carpentaria using FT-NIRS spectroscopy. Mar. Freshwat. Res. 2:1268-1279. https://doi.org/10.1071/MF20300

Wuenschel, M. J., W. D. McElroy, K. Oliveira, and R. S. McBride. 2018. Measuring fish condition: an evaluation of new and old metrics for three species with contrasting life histories. Can. J. Fish. Aquat. Sci. 76:886-903. https://doi.org/10.1139/cjfas-2018-0076

Wyllie Echeverria, T. 1987. Thirty-four species of California rockfishes: maturity and seasonality of reproduction. Fish. Bull. 85(2):229-250.

Yaremko, M. L. 1996. Age determination in Pacific sardine, *Sardinops sagax*. U.S. Dep. Commer., NOAA Tech. Memo. NOAA-TM-NMFS-SWFSC-223.

Yoklavich, M.M., and K.M. Bailey. 1990. Hatching period, growth and survival of young walleye pollock *Theragra chalcogramma* as determined from otolith analysis. Mar. Ecol. Prog. Ser. 64(1-2):13-23. https://doi.org/10.3354/meps064013

Zheng, A., Y. Li, A. Vardanyan, S. Arsov, J.-N. Hwang, I. M. Benson, T. E. Helser, J. A. Short, K. A. Bayer, C. R. Kastelle, B. K. Barnett, A. R. Rezek, F. R. Wallace, R. L. Hill, W. Kline, L. A. Thornton, N. Evou, R. J. Allman, and N. Willett. This volume. Automatic fish age prediction using deep machine learning – combining otolith images, NIR spectra, and metadata features.

Zwolinksi, J. P., and D. A. Demer. 2023. An updated model of potential habitat for northern stock Pacific sardine (*Sardinops sagax*) and its use for attributing survey observation and fishery landings. Draft working document for PFMC SSC CPS Sub-committee Meeting, March 20-21, 2023. Available at: https://www.pcouncil.org/documents/2023/03/draft-update-model-of-potential-habitat-for-northern-stock-pacific-sardine.pdf/

APPENDIX A. Supplemental Tables and Figures

Supplemental information to "Quality Assurance and Quality Control of Fourier Transform Near Infrared Spectroscopy Data for Age Prediction"

Esther D. Goldstein (Esther.Goldstein@noaa.gov)[1], Brenna C. Hsieh[1], Morgan B. Arrington[1,2], and Thomas E. Helser[1]

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

[2] Cooperative Institute for Climate, Ocean, and Ecosystem Studies (CICOES)
University of Washington
3737 Brooklyn Ave NE
Seattle, WA USA

Quality assurance and quality control (QA/QC) procedures utilize both principal component analysis (PCA) and partial least-squares regression (PLS) to visually identify unusual patterns in spectra and quantitatively determine outliers from PCA and PLS. This approach relies on a comparison of new scan data with previous scans to ensure that data remain comparable across surveys and years.

1. Scan session information

Table A1. -- New scan data information.

Year	Total scans all sessions
2019	1,478

Table A2. -- New scan data summary information by scan session.

Scan session	Year	Scans per session
NIR_162201901201A	2019	138
NIR_162201901201B	2019	261
NIR_162201901201C	2019	140
NIR_162201901201D	2019	140
NIR_162201901201E	2019	85
NIR_94201901201A	2019	124
NIR_94201901201B	2019	236
NIR_94201901201C	2019	138
NIR_94201901201D	2019	138
NIR_94201901201E	2019	78

Table A3. -- New scan data summary of potential otolith issues.

	broken	crystallized	other problem	unscannable
	0	0	0	0
	0	0	0	7
	0	0	49	0
	0	0	29	29
	0	31	0	0
	0	14	0	14
	0	4	4	0
	0	3	3	3
	11	0	0	0
	13	0	0	13
	2	0	2	0
	1	0	1	1
	1	1	0	1
	28	53	88	68
Total	56	106	176	136

2. PCA

2.1 Unprocessed spectra visual inspection

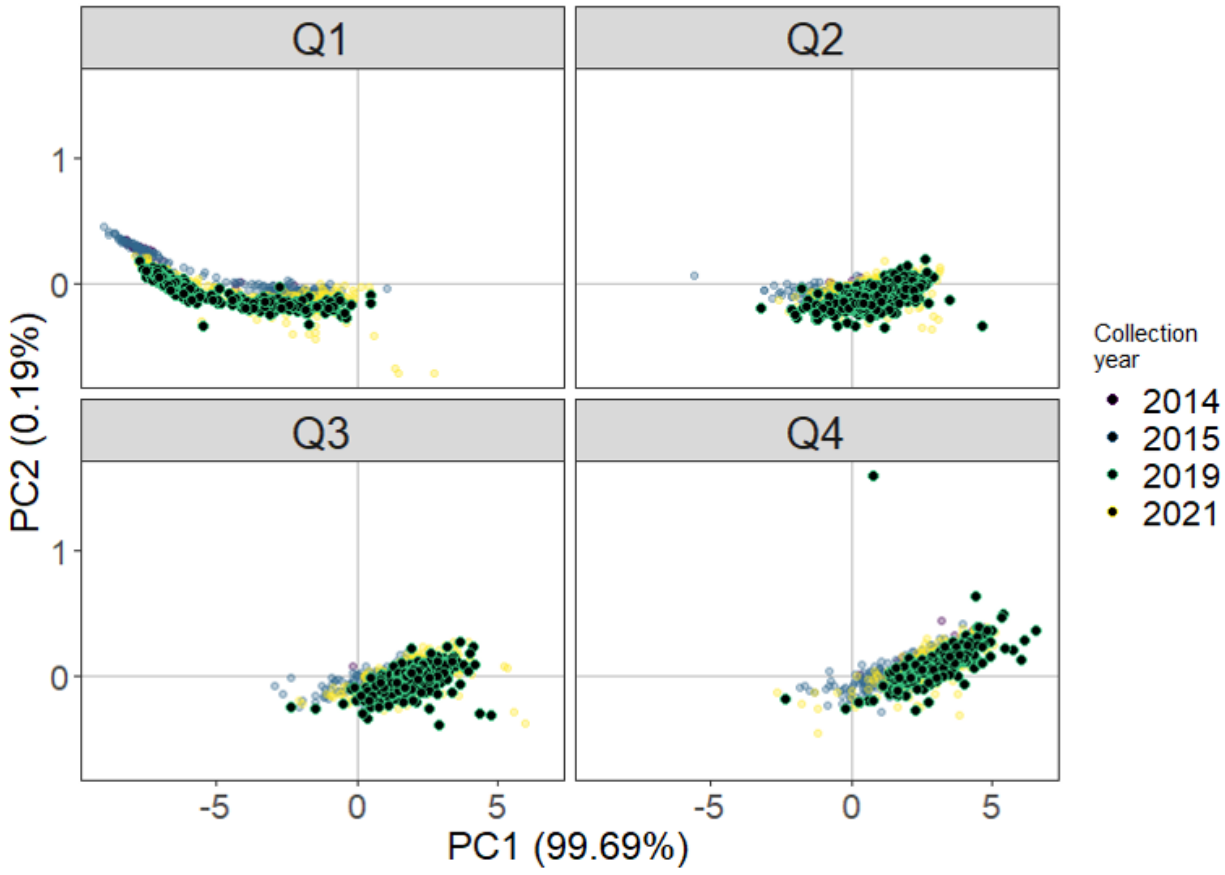


Figure A1. -- PCA of unprocessed spectra with new spectra (2019) projected using PCA from the other collection years. Data are divided into fish fork length quartiles (Q1-Q4) to better discern patterns in data "clouds".

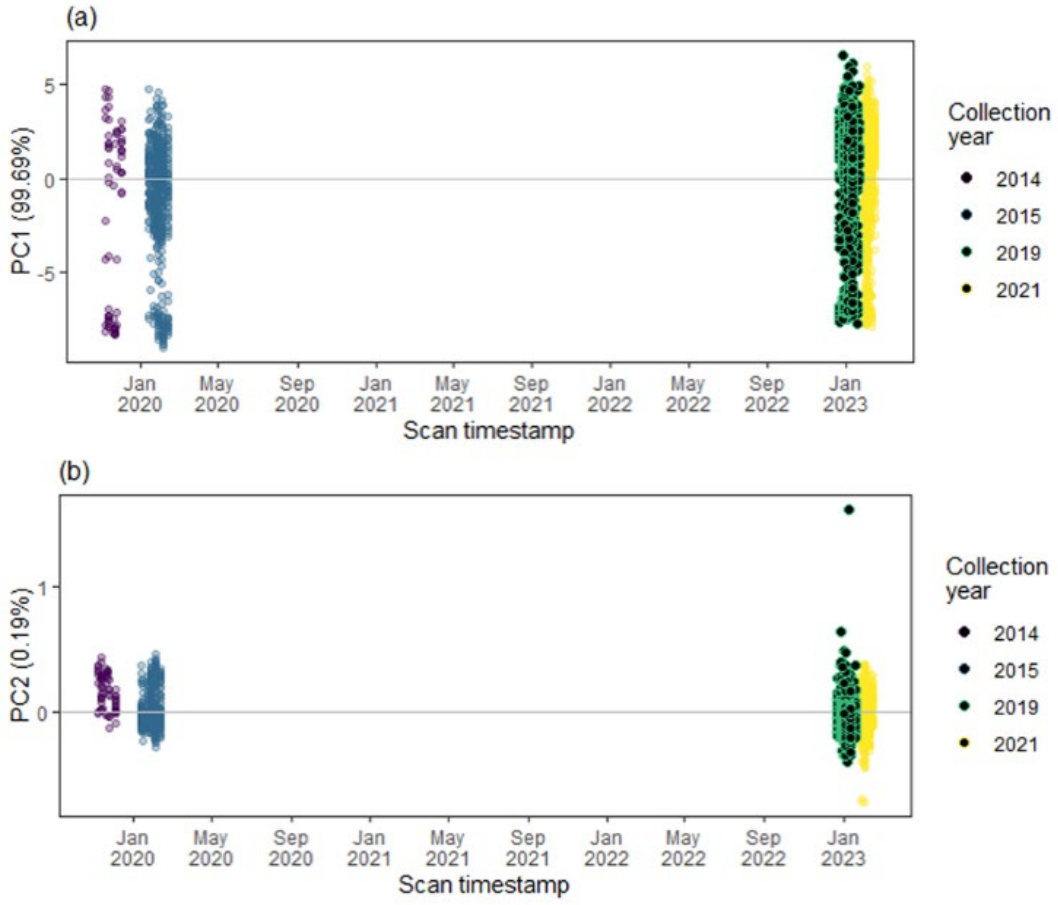


Figure A2. -- PC1 (a) and PC2 (b) of unprocessed spectra by scan date to visually assess drift over time related to instrumentation or scan procedures. New spectra (2019) were projected using PCA from the other collection years.

2.2 Pre-processed spectra visual inspection

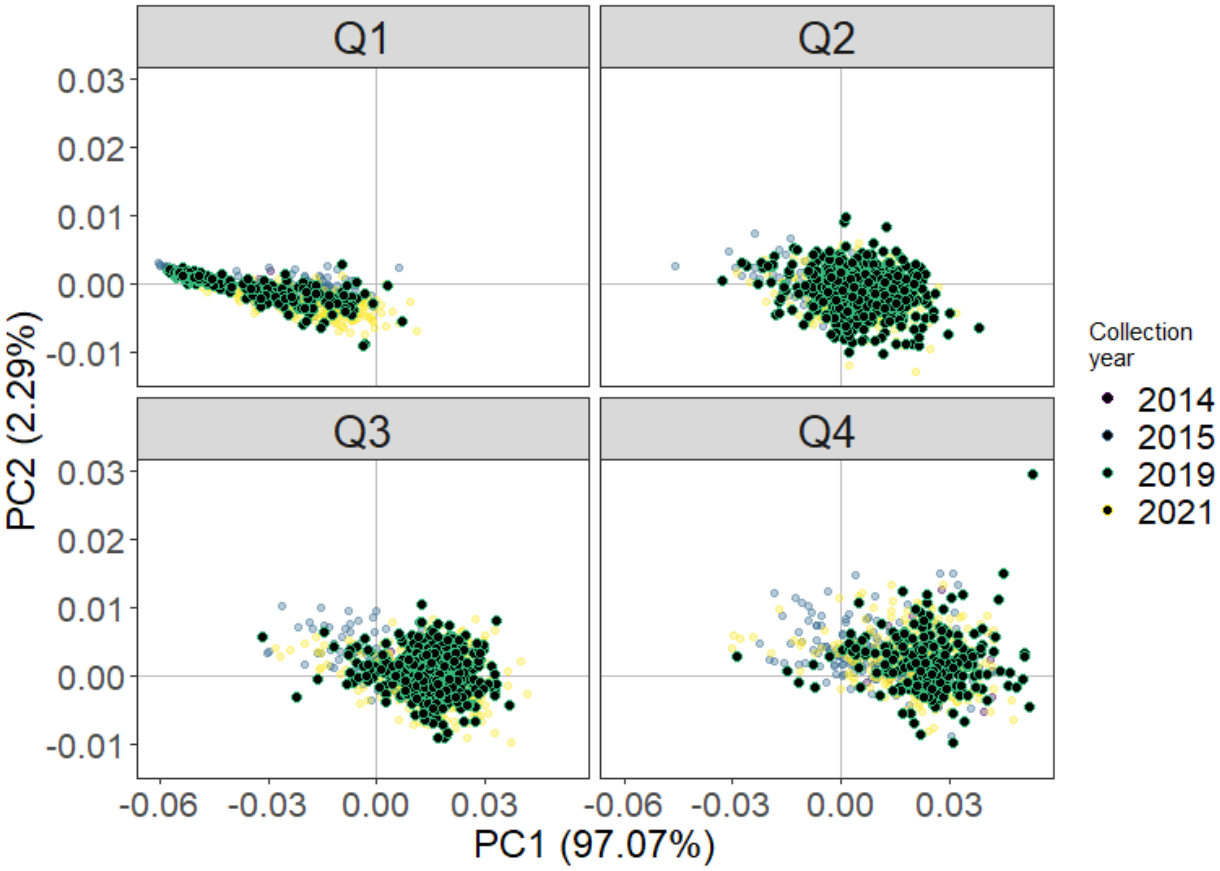


Figure A3. -- PCA of pre-processed spectra with new spectra (2019) projected using PCA from other scan years. Data are divided into fish fork length quartiles (Q1-Q4) to better discern patterns in data "clouds".

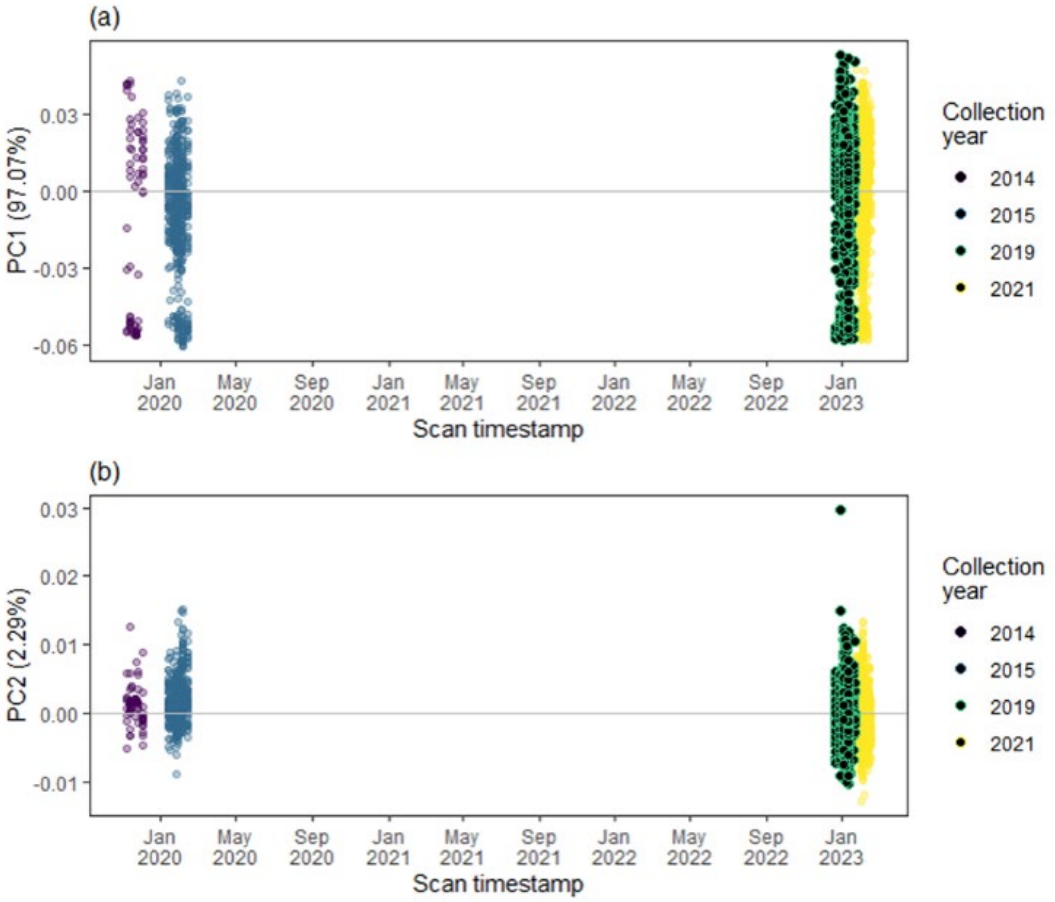


Figure A4. -- PC1 (a) and PC2 (b) of pre-processed spectra by scan date to visually assess drift over time related to instrumentation or scan procedures. New spectra (2019) were projected using PCA from the other collection years.

2.3 PCA outliers

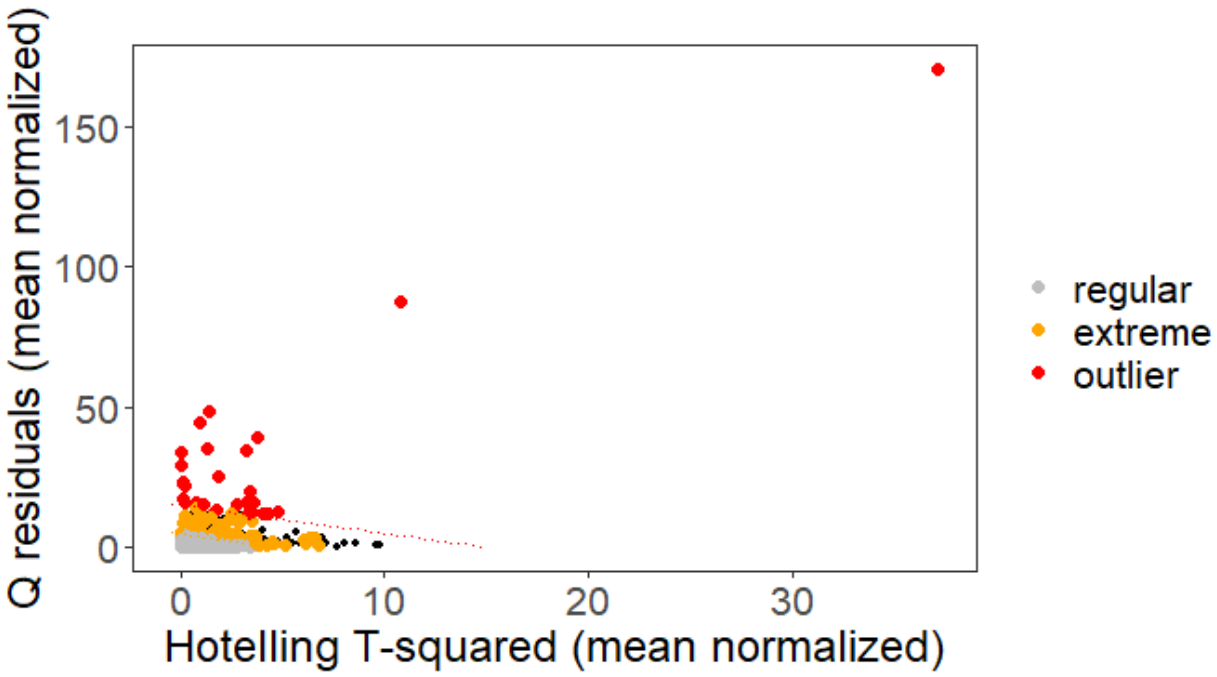


Figure A5. -- PCA outliers using pre-processed data. PCA from other collection years data was used to set the thresholds for new data (2019) using the approach outlined in Pomerantsev and Rodionova (2014) using statistical moments. Extreme = 5% significance threshold and outliers= 1%.

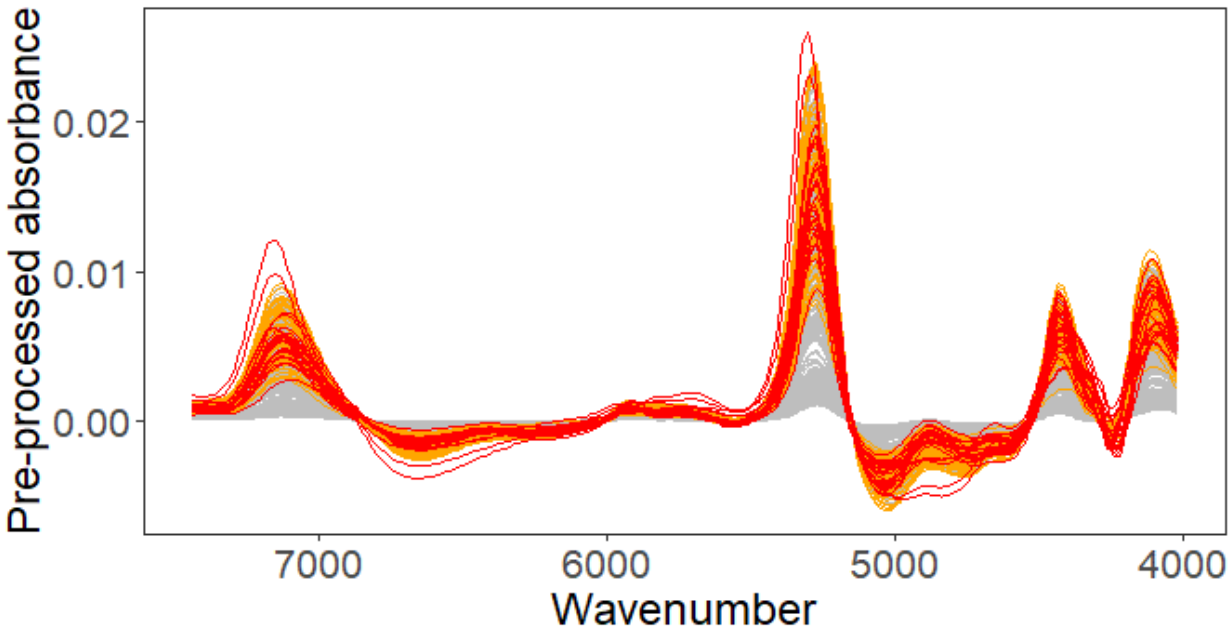


Figure A6. -- Pre-processed new data (2019) showing normal (gray), extreme (orange), and outlier (red) spectra from PCA-based outlier designations shown in Fig. A5.

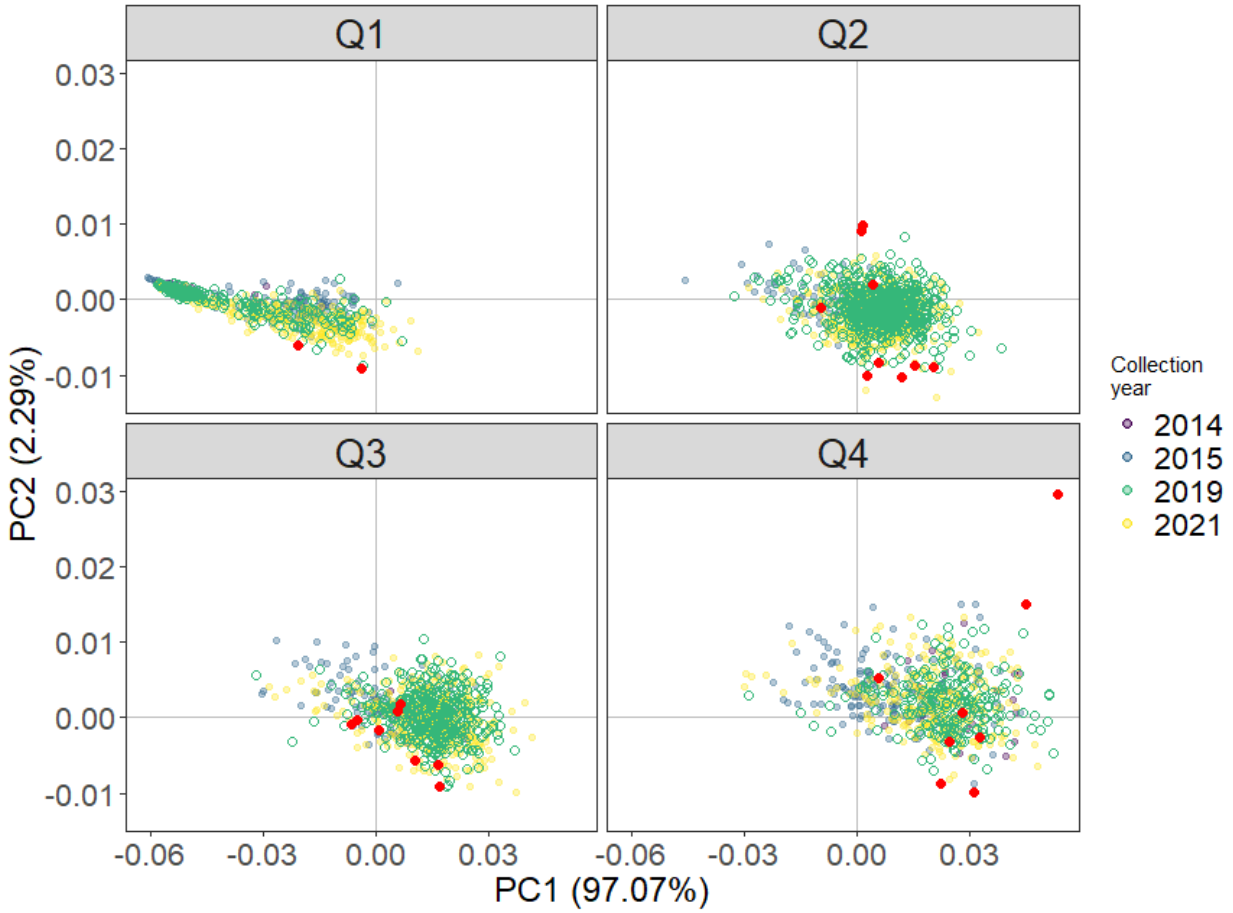


Figure A7. -- Outliers (red) identified from PCA of pre-processed spectra plotted in PCA biplot space. Data are divided into fish length quartiles (Q1-Q4) to better discern patterns in data "clouds".

3. PLS

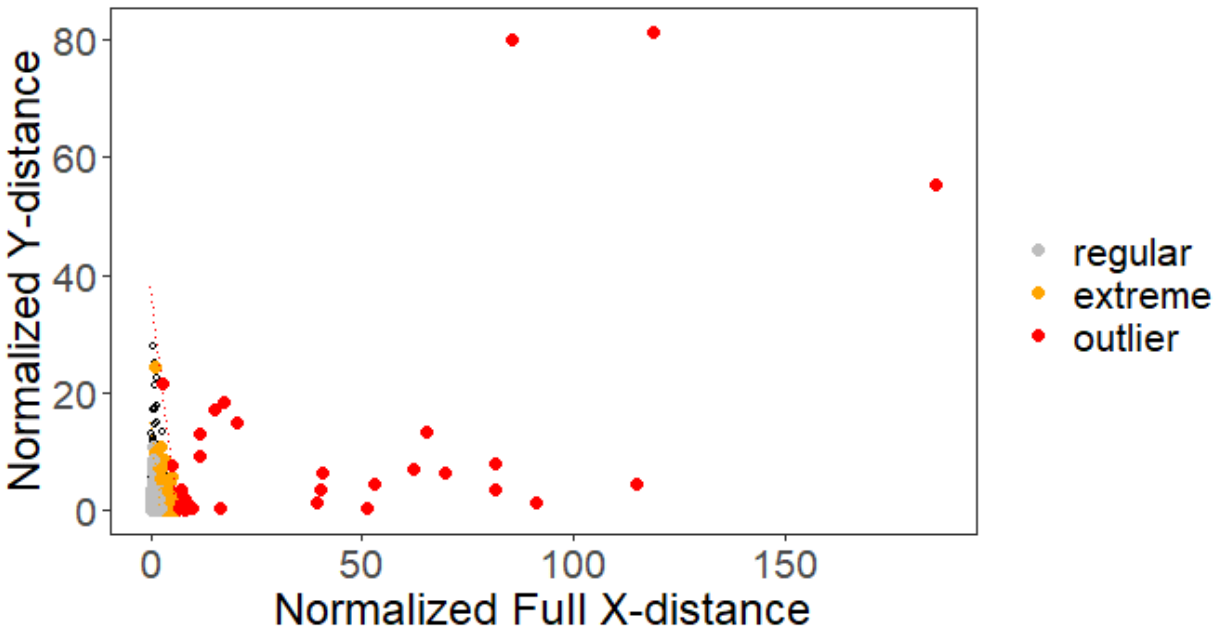


Figure A8. -- Outlier, extreme, and all other (regular) data identified from PLS. Pre-processed spectra were used to build a PLS model from 2014, 2015, and 2021 collection years (black dots) and fish lengths were predicted from spectra from the 2019 dataset to determine outlier and extreme data points. Normalized full X-distance is a combination of Hotelling T-squared and Q-residuals, and normalized Y-distance is a comparison of observed and predicted values (Rodionova and Pomerantsev 2020). Critical limit extreme = 0.05 and critical limit outlier = 0.01.

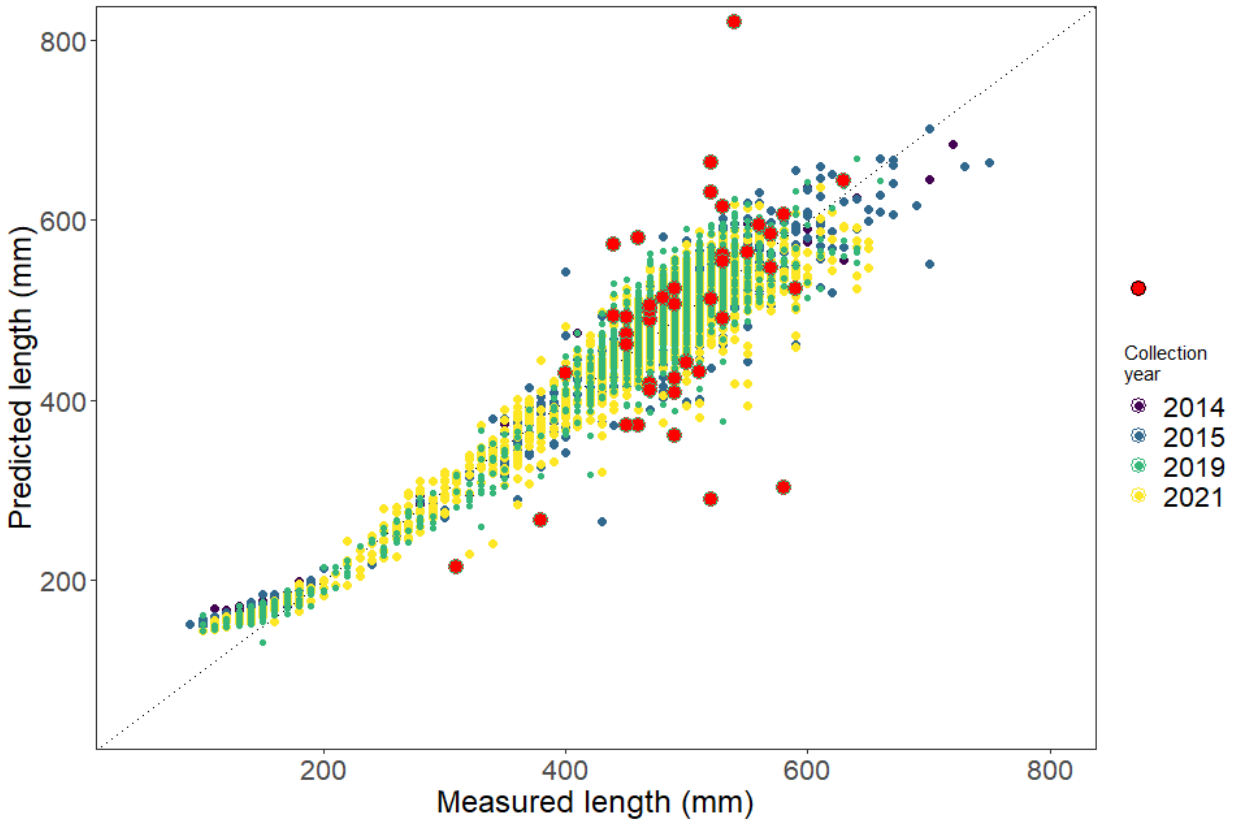


Figure A9. -- Observed and predicted fish lengths from the PLS. Red points show outliers from new data (2019) based on Y-distance and full X-distance.

4. PCA and PLS outliers combined

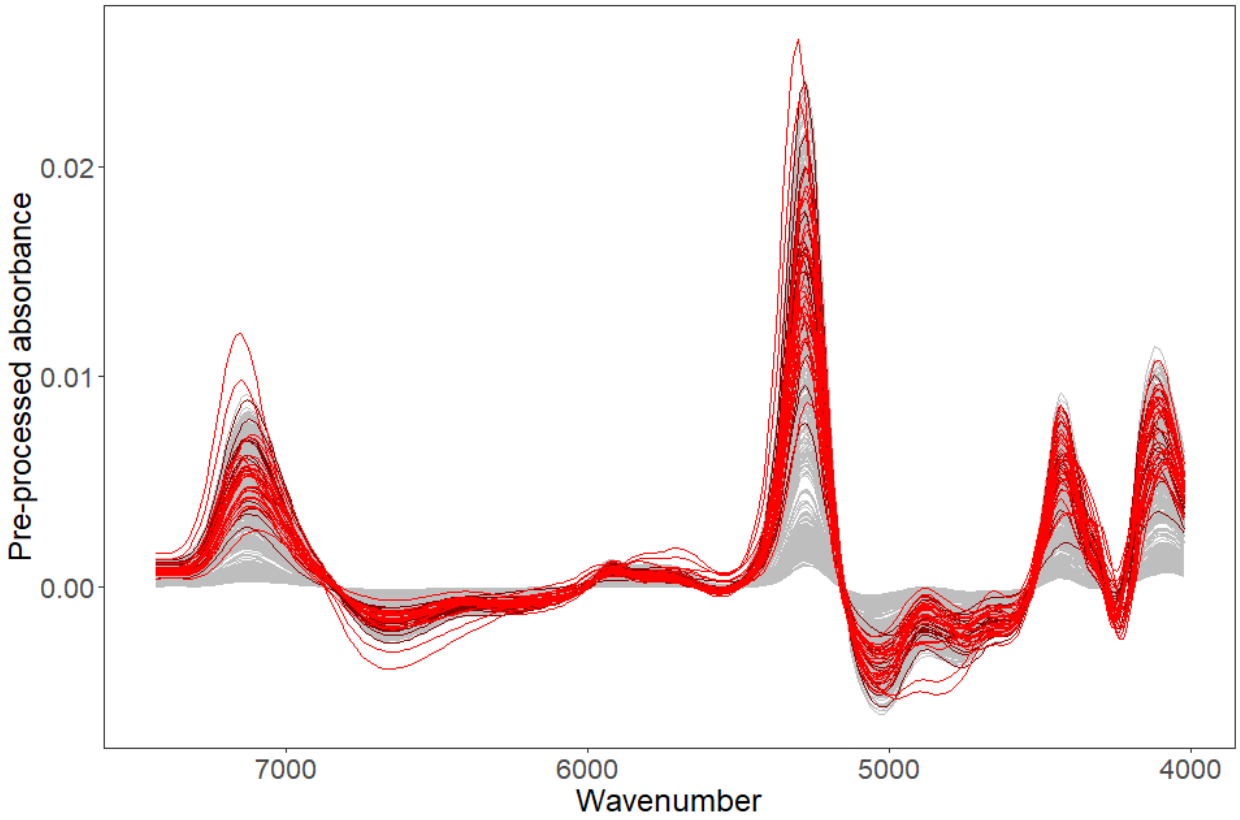


Figure A10. -- Pre-processed spectra from 2019 showing PCA-based outliers in red and PLS-based outliers in dark red. Spectra that were not identified as outliers are in gray.

Table A4. -- Summary of outlier spectra from PCA and PLS using pre-processed data.

Total outliers	PCA outliers	PLS outliers
42	27	42

Table A5. -- List of file names for spectra identified as outliers from PCA and PLS, and information about potential issues with otoliths that may have impacted scan quality.

File name	PCA	PLS	broken	crystallized	other problem
WALLEYE_POLLOCK_94201901201_186_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_94201901201_212_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_94201901201_218_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_94201901201_248_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_94201901201_280_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_94201901201_435_OA1.0	outlier	outlier			X
WALLEYE_POLLOCK_94201901201_442_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_94201901201_566_OA1.0	outlier	outlier		X	
WALLEYE_POLLOCK_94201901201_589_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_94201901201_605_OA1.0	outlier	outlier			X
WALLEYE_POLLOCK_94201901201_609_OA1.0	outlier	outlier		X	
WALLEYE_POLLOCK_94201901201_620_OA1.0	outlier	outlier		X	
WALLEYE_POLLOCK_94201901201_632_OA1.0	outlier	outlier		X	
WALLEYE_POLLOCK_94201901201_728_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_94201901201_776_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_162201901201_8_OA1.0	outlier	outlier		X	
WALLEYE_POLLOCK_162201901201_55_OA1.0	outlier	outlier		X	
WALLEYE_POLLOCK_162201901201_135_OA1.0	outlier	outlier			X
WALLEYE_POLLOCK_162201901201_136_OA1.0	outlier	outlier		X	X
WALLEYE_POLLOCK_162201901201_196_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_162201901201_324_OA1.0	outlier	outlier			X
WALLEYE_POLLOCK_162201901201_334_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_162201901201_339_OA1.0	outlier	outlier			X
WALLEYE_POLLOCK_162201901201_464_OA1.0	outlier	outlier		X	
WALLEYE_POLLOCK_162201901201_504_OA1.0	outlier	outlier		X	
WALLEYE_POLLOCK_162201901201_596_OA1.0	outlier	outlier		X	
WALLEYE_POLLOCK_162201901201_764_OA1.0	outlier	outlier			
WALLEYE_POLLOCK_94201901201_246_OA1.0		outlier			

File name	PCA	PLS	broken	crystallized	other problem
WALLEYE_POLLOCK_94201901201_338_OA1.0		outlier			
WALLEYE_POLLOCK_94201901201_347_OA1.0		outlier			
WALLEYE_POLLOCK_94201901201_409_OA1.0		outlier			
WALLEYE_POLLOCK_94201901201_410_OA1.0		outlier			
WALLEYE_POLLOCK_94201901201_421_OA1.0		outlier		X	
WALLEYE_POLLOCK_94201901201_497_OA1.0		outlier		X	
WALLEYE_POLLOCK_94201901201_705_OA1.0		outlier			
WALLEYE_POLLOCK_94201901201_755_OA1.0		outlier			
WALLEYE_POLLOCK_162201901201_163_OA1.0		outlier			X
WALLEYE_POLLOCK_162201901201_459_OA1.0		outlier		X	
WALLEYE_POLLOCK_162201901201_485_OA1.0		outlier	X		X
WALLEYE_POLLOCK_162201901201_561_OA1.0		outlier		X	
WALLEYE_POLLOCK_162201901201_640_OA1.0		outlier		X	
WALLEYE_POLLOCK_162201901201_737_OA1.0		outlier		X	

CITATIONS

Rodionova, O. Y., and A. L. Pomerantsev. 2020. Detection of outliers in projection-based modeling. Anal. Chem. 92:2656-2664. https://doi.org/10.1021/acs.analchem.9b04611

Pomerantsev, A. L., and O. Y. Rodionova. 2014. Concept and role of extreme objects in PCA/SIMCA. J. Chemom. 28(5):429-438. https://doi.org/10.1002/cem.2506

APPENDIX B. Workshop Agenda

WORKSHOP: Rapid Estimation of Fish Age Using Fourier-transform Near Infrared Spectroscopy (FT-NIRS)

Virtual attendance via Webex will also be available

April 3-7, 2023 NOAA Western Regional Campus, Alaska Fisheries Science Center
7600 Sand Point Way, NE, Building 9 Auditorium, Seattle, WA 98115

Monday, 3 April

1:00 PM – 1:30 PM	**WELCOME AND INTRODUCTIONS**
1:30 PM – 2:30 PM	NOAA Fisheries FT-NIRS Fish Ageing Strategic Initiative: Past, present & future (Thomas Helser, AFSC — Seattle Lab)
2:30 PM – 4:30 PM	Session 1: AFSC Spectroscopy Lab Tour & demonstration (Bruker Applications Scientist - Jason Erikson). AFSC, Building 4 Room 1111
5:00 PM	Informal social: **Burke Gilman Brewing Company**, 3626 NE 45th Street, Suite 102

Tuesday, 4 April

Presentations	**FT-NIRS APPLICATION DEVELOPMENT**
9:00 AM – 9:30 AM	FT-NIRS age prediction of eastern Bering Sea yellowfin sole: a case study of a moderately long-lived flatfish (Todd TenBrink, AFSC – Seattle Lab)
9:30 AM – 10:00 AM	A preliminary analysis and review of FT-NIR spectroscopy and DNA CpG site methylation for fish age prediction (Laurel Lam, NWFSC – Seattle Lab)
10:00 AM – 10:30 AM	Applying FT-NIRS predictive ageing to different genetic stocks of white grunt within the U.S. south Atlantic (Jamie Clark, SEFSC – Beaufort Lab)
Break	
11:00 AM – 11:30 AM	A novel approach for determining the spawning phenology of walleye pollock using Raman spectroscopy (Sandi Neidetcher, AFSC – Seattle Lab)
11:30 AM – 12:00 PM	FT-NIRS ageing of finfish and shark species in the northwest Atlantic (Alex Rubin, NEFSC – Narragansett Lab)
12:00 PM – 12:30 PM	Exploration of FT-NIRS for shortbelly rockfish (*Sebastes jordani*): an ecologically important forage fish off the coast of California (Jessica Choi, SWFSC – Santa Cruz Lab)
Lunch	
2:00 PM – 2:30 PM	Investigating the use of FT-NIR spectroscopy to age gag (*Mycteroperca microlepis*), a protogynous hermaphroditic species (Beverly Barnett, SEFSC – Panama City Lab)
2:30 PM – 3:00 PM	Exploring the use of FT-NIRS for ageing sablefish (*Anoplomona fimbria*) and Pacific hake (*Merluccius Productus*) of the U.S. west coast (John Wallace, NWFSC – Seattle)

3:00 PM – 3:30 PM	Developing spectroscopy approaches to measure life history characteristics of fish throughout ontogeny (Esther Goldstein, AFSC – Seattle, Lab)
Break	
4:00 PM – 4:30 PM	Rapid daily age estimation of juvenile walleye Pollock in the Gulf of Alaska using FT-NIR spectroscopy (Beth Matta, AFSC – Seattle Lab)
4:30 PM – 5:00 PM	Fish otolith proteomics and its relationship and impact on the use of FT-NIR spectroscopy for fish age prediction (Oliver Thomas, University of Melbourne)
Close	

Wednesday, 5 April

Presentations cont.	
9:00 AM – 9:30 AM	Developing NIR sampling methodology for modeling species discrimination in live catfish for aquaculture (Carrie Vance – Mississippi State University)
9:30 AM – 10:00 AM	Fourier transform near infrared spectroscopy discriminates archived otoliths of newly detected cryptic species, *Etelis carbunculus* and *Etelis boweni* (Kristen Dahl, PIFSC)
	FT-NIRS IMPLEMENTATION
10:00 AM – 10:30 AM	Trials and tribulations of using FT-NIRS on Pacific Sardine: Method development for scanning Pacific Sardine otoliths (Emma Saas, SWFSC – La Jolla Lab)
10:30 AM – 11:00 AM	Benefits and challenges of using FT-NIRS for production age estimation at the Northeast Fisheries Science Center (Eric Robillard, NEFSC – Woods Hole Lab)
Break	
11:30 AM – 12:00 PM	FT-NIR spectroscopy of otoliths coupled with deep machine learning to improve age prediction (Irina Benson, AFSC – Seattle Lab)
12:00 PM – 12:30 PM	Automatic fish age prediction using deep machine learning: combining otolith image, NIR spectra and metadata features (Aotian Zheng – University of Washington)
12:30 PM – 1:00 PM	Calibration and variation of FT-NIRS otolith spectra among NIR spectrometers and species (Andy Ostrowski, SEFSC – Beaufort Lab)
Lunch	
2:00 PM – 2:30 PM	A simulation study exploring best practices for model development and updating (Morgan Arrington, CICOES-AFSC)
2:30 PM – 3:00 PM	Database architecture and management envisioned for the FT-NIRS paradigm at Alaska Fisheries Science Center (Jon Short, AFSC – Seattle Lab)
3:00 PM – 3:30 PM	Group discussion: Likelihood of operational success, timeline for operational readiness and communication to stake holders (follow up on Friday, April 7 at 9:00 AM).

3:30 PM – 5:30 PM	Session 2: AFSC Spectroscopy Lab Tour & demonstration (Bruker Applications Scientist – Jason Erikson). AFSC, Building 4 Room 1111
6:00 PM – 10:00 PM	Organized social mixer and casual dinner – **Elliot Bay Public House & Brewery** 12537 Lake City Way NE., Seattle WA 98125

Thursday, 6 April

Presentations cont.	**FT-NIRS STOCK ASSESSMENT INTEGRATION**
9:00 AM – 10:00 AM	Envisioning the future of production fish ageing: end-to-end integration of the FT-NIRS age estimation enterprise (Thomas Helser, AFSC – Seattle Lab)
10:00 AM – 10:30 AM	A simulation framework to examine the effect of ageing error on model-based age predictions (Morgan Arrington, CICOES-AFSC)
Break	
11:00 AM – 11:30 PM	Integration of FT-NIRS age data products into the eastern Bering Sea walleye pollock and Pacific cod stock assessments (Thomas Helser, AFSC – Seattle Lab)
Lunch	
1:00 PM –1:30 PM	Efficacy of FT-NIRS predicted ages for use in the Gulf of Mexico gray snapper stock assessment (Steve Garner, SEFSC – Panama City Lab)
1:30 PM – 5:00 PM	Open discussion with national stock assessment forum

Friday, 7 April

	FT-NIRS STRATEGIC INITIATIVE BUSINESS
9:00 AM – 10:30 AM	Wrap up - open discussion
Break	
10:30 AM – 12:00 PM	FT-NIR strategic initiative business and budget meeting
Meeting Adjourn	

APPENDIX C. Workshop Participants

Name	Email	Organization	Home office location
Alex Rubin	Alexander.Rubin@noaa.gov	NOAA, NEFSC, Apex Predators Program	Narragansett, RI
Amanda Rezek	amanda.rezek@noaa.gov	NOAA, SEFSC, Biology & Life History	Beaufort, NC
Anderson, Alena	aqa5@msstate.edu	Mississippi State University	Starkville, MS
Andrew Chin	andrew.chin@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Andrew Claiborne	Andrew.Claiborne@dfw.wa.gov	WDFW	Olympia, WA
Andy Ostrowski	andy.ostrowski@noaa.gov	NOAA, SEFSC, FATES, Life History	Beaufort, NC
Audrey Ty	Audrey.Ty@dfo-mpo.gc.ca	DFO/PBS	Nanaimo, BC, CA
Austin Anderson	Austin.Anderson@dfw.wa.gov	WDFW	Olympia, WA
Beth Matta	beth.matta@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Beverly Barnett	beverly.barnett@noaa.gov	NOAA, SEFSC, FATES, Biology & Life History	Panama City, FL
Bill Kline	bill.kline@noaa.gov	CIMAS, NOAA, SEFSC, Biology & Life History	Panama City, FL
Brenna Groom	brenna.groom@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Brittany Schwartzkopf	brittany.schwartzkopf@noaa.gov	NOAA, SWFSC, Life History	La Jolla, CA
Carrie Kim Vance	ckv7@msstate.edu	Mississippi State University	Starkville, MS
Cathy Mattson	cathy.mattson@alaska.gov	ADF&G-Juneau	Juneau, AK
Charlie Piston	charlie.piston@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Chelsea Rothkop	Chelsea.Rothkop@dfo-mpo.gc.ca	DFO/PBS	Nanaimo, BC, CA
Cheryl Barnes	cheryl.barnes@oregonstate.edu	ODFW-MRP	Newport, OR
Chris Hinds	chris.hinds@alaska.gov	ADF&G-Juneau	Juneau, AK
Chris Johnston	chris.johnston@iphc.int	IPHC	Seattle, WA
Christina Jump	Christina.Jump@dfw.wa.gov	WDFW	Olympia, WA
Christopher Gburski	christopher.gburski@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Claire Stuart	claire.stuart@humboldt.edu	NOAA, SWFSC, Habitat and Groundfish Ecology	Santa Cruz, CA
Cooke, Chelsea	Chelsea.Cooke@dfo-mpo.gc.ca	DFO/PBS	Nanaimo, BC, CA
Craig Kastelle	craig.kastelle@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Denise Parker	dparker@psmfc.org	NWFSC/PSMFC; Cooperative Ageing Lab	Newport, OR
Emma Saas	emma.saas@noaa.gov	Saltwater Inc. and NOAA SWFSC	La Jolla, CA
Emmanis Dorval	Emmanis.dorval@noaa.gov	NOAA, SWFSC, Life History Program	La Jolla, CA
Eric Robillard	eric.robillard@noaa.gov	NOAA, NEFSC, Age and Growth Program	Woods Hole, MA
Esther Goldstein	Esther.Goldstein@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Han Ju Kim	hj0528@korea.kr	Fisheries Resources Research Center	Tongyoung, KR
Heather Moncrief-Cox	heather.moncrief-cox@noaa.gov	CIMAS, NOAA, SEFSC, Biology & Life History	Panama City, FL
Hwan Sung Ji	ninise9@nfrdi.go.kr	Fisheries Resources Research Center	Tongyoung, KR
Irina Benson	irina.benson@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
James Hale	jhale@psmfc.org	NWFSC/PSMFC; Cooperative Ageing Lab	Newport, OR
Jamie Clark	jamie.clark@noaa.gov	NOAA, SEFSC, FATES, Biology & Life History	Beaufort, NC
Jason Erikson	Jason.Erickson@bruker.com	Bruker rep	Madison, WI
Jennifer Potts	jennifer.potts@noaa.gov	NOAA, SEFSC, FATES, Biology & Life History	Beaufort, NC
Jenny Topping	Jennifer.Topping@dfw.wa.gov	WDFW	Olympia, WA
Jessica Choi	jessica.choi@noaa.gov	FCP/UCSC, NOAA-SWFSC, Groundfish	Santa Cruz, CA
Jessica Horn	jessica.horn@alaska.gov	ADF&G-Kodiak	Kodiak, AK
Jim Ianelli	jim.ianelli@noaa.gov	AFSC/SSMA	Seattle, WA
Joan Forsberg	joan.forsberg@iphc.int	IPHC	Seattle, WA
John Brogan	john.brogan@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
John Wallace	john.wallace@noaa.gov	NWFSC	Seattle, WA
Jon Short	jon.short@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA

Name	Email	Organization	Home office location
Josh Dore	josh.dore@alaska.gov	ADF&G-Juneau	Juneau, AK
Julie Pearce	julie.pearce@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Kali Stone	kali.stone@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Kathrin Bayer	kathrin.bayer@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Kathryn Berry	Kathryn.Berry@dfo-mpo.gc.ca	DFO/PBS	Nanaimo, BC
Kelsey Magrane	kelsey.magrane@iphc.int	IPHC	Seattle, WA
Kevin McNeel	kevin.mcneel@alaska.gov	ADF&G-Juneau	Juneau, AK
Kouba, Andy	a.kouba@msstate.edu	Mississippi State University	Starkville MS
Kristen Dahl	kristen.dahl@noaa.gov	NOAA, PIFSC, Life History Program	Honolulu, HI
Leif Rasmuson	leif.k.rasmuson@odfw.oregon.gov	ODFW-MRP	Newport, OR
Li-Dunn Chen	lc1817@msstate.edu	Mississippi State University	Starkville MS
Liz Ortiz	lortiz@psmfc.org	NWFSC/PSMFC; Cooperative Ageing Lab	Newport, OR
Marian Ford	marian.ford@alaska.gov	ADF&G-Homer	Homer, AK
Mark Plumb	mark.plumb@alaska.gov	ADF&G-Juneau	Juneau, AK
Mark Terwilliger	mark.r.terwilliger@odfw.oregon.gov	ODFW-MRP	Newport, OR
McAree, Danielle	dmm787@msstate.edu	Mississippi State University	Starkville, MS
Melissa Monk	melissa.monk@noaa.gov	NOAA, SWFSC, Habitat and Groundfish Ecology	Santa Cruz, CA
Meredith Emery Boeck	memeryboeck@psmfc.org	NWFSC/PSMFC; Cooperative Ageing Lab	Newport, OR
Merrie Schultz	Merrie.Schultz@dfw.wa.gov	WDFW	Olympia, WA
Michelle Passerotti	michelle.passerotti@noaa.gov	NOAA, NEFSC, Apex Predators Program	Narragansett, RI
Morgan Arrington	morgan.arrington@noaa.gov	UW CICOES-AFSC, Age and Growth Program	Seattle, WA
Naeem Willett	naeem.willett@noaa.gov	CIMAS, NOAA, SEFSC, Biology & Life History	Panama City, FL
Nikki Paige	npaige@psmfc.org	NWFSC/PSMFC; Cooperative Ageing Lab	Newport, OR
Oliver Thomas	oliverrbthomas@gmail.com	The University of Melbourne	Melbourne, VIC, AU
Patrick McDonald	pmcdonald@psmfc.org	NWFSC/PSMFC; Cooperative Ageing Lab	Newport, OR
Poudel, Ashmita	ap2751@msstate.edu	Mississippi State University	Starkville MS
Robert Tobin	robert.tobin@iphc.int	IPHC	Seattle, WA
Sandi Neidetcher	sandi.neidetcher@noaa.gov	NOAA,AFSC, Age and Growth Program	Seattle, WA
Seung Hwan Lee	hwan2915@korea.kr	West sea Fisheries Resources Research Division	Incheon, KR
Sonya Elmejjati	sonya.elmejjati@alaska.gov	ADFG	Kodiak, AK
Steve Barbeaux	steve.barbeaux@noaa.gov	AFSC/SSMA	Seattle, WA
Steve Garner	steven.garner@noaa.gov	CIMAS, NOAA, FATES, Life History	Panama City, FL
Tera Winters	tera.winters@noaa.gov	CIMAS, NOAA, SEFSC, Biology & Life History	Panama City, FL
Thomas Helser	thomas.helser@noaa.gov	NOAA, AFSC, AGP	Seattle, WA
Tien-Shui Tsou	Tien-Shui.Tsou@dfw.wa.gov	WDFW	Olympia, WA
Todd TenBrink	todd.tenbrink@noaa.gov	NOAA, AFSC, Age and Growth Program	Seattle, WA
Tracey Loewen	Tracey.Loewen@dfo-mpo.gc.ca	DFO/Freshwater Institute	Winnipeg, MB, CA
Tracy McCulloch	Tracy.McCulloch@noaa.gov	NOAA, SEFSC, FATES, Biology & Life History	Beaufort, NC
Tyler Johnson	tjohnson@psmfc.org	NWFSC/PSMFC; Cooperative Ageing Lab	Newport, OR
Walt Rogers	Walt.rogers@noaa.gov	NOAA, SEFSC, FATES, Biology & Life History	Beaufort, NC

U.S. Secretary of Commerce
Gina M. Raimondo

Under Secretary of Commerce for
Oceans and Atmosphere
Dr. Richard W. Spinrad

Assistant Administrator,
National Marine Fisheries Service.
Janet Coit

January 2024

www.fisheries.noaa.gov

OFFICIAL BUSINESS

National Marine
Fisheries Service
Alaska Fisheries Science Center
7600 Sand Point Way N.E.
Seattle, WA 98115-6349