Scientific and Statistical Committee Groundfish Subcommittee Report

Pacific Fishery Management Council
Online Meeting

October 1-2, 2024

The Groundfish Subcommittee (GFSC) of the Scientific and Statistical Committee (SSC) met via webinar October 1 and 2, 2024 to review the Suitability of Fish Age Estimates Developed Using Fourier Transformed Near-Infrared Spectroscopy (FT-NIRS) in Groundfish Stock Assessments. SSC participants are listed in Appendix A.

Review of FT-NIRS Approach to Ageing

Overview and basic principles of FT-NIRS, National Marine Fisheries Service (NMFS) Initiative

Tom Helser (NMFS Alaska Fisheries Science Center [AFSC]) provided an overview and basic principles of FT-NIRS to facilitate ageing of U.S. West Coast groundfish stocks. The AFSC is the lead center of a NOAA Fisheries strategic initiative to develop the methodology and application, and all seven NMFS science centers have been involved in developing methods to develop and evaluate the potential for FT-NIRS to estimate ages more efficiently than traditional methods. The overarching goal of the initiative is to operationalize the approach across NOAA ageing labs, but at the culmination of the 5-year effort in 2024 the AFSC and Northwest Fisheries Science Center (NWFSC) have sought additional technical review for near term implementation.

At AFSC, the FT-NIRS methodology has been rigorously evaluated for its potential to greatly increase ageing efficiencies for northern rockfish, walleye pollock, and Pacific cod, and their efforts have shown promise of substantial efficiency improvements. The method is best described as an interaction of electromagnetic energy with molecular bonds that vibrate at characteristic frequencies and absorb energy in relation to the fundamental dipole vibrations. Using a spectrometer, near infrared light is targeted at a sample in diffuse reflectance mode and absorbance related to the combinations and overtones of the fundamental dipole vibrations are measured providing a unique absorbance profile with each otolith in relation to age (Helser et al. 2019). Relating the absorbance profile to traditional age estimates (typically from break and burn methods) provides the basis for using FT-NIRS to estimate fish age. Relating predictions to the observed data is subject to uncertainty, which is quantified.

To provide the most robust age estimates from the spectral data, the AFSC has developed an approach that uses "deep learning" methods to improve age predictions. Specifically, this approach uses multimodal convolutional neural networks (MMCNN) that integrate FT-NIRS otolith spectra and biological or geospatial data to obtain model age predictions. The AFSC example of this approach used approximately 3,000 age estimates for northern rockfish, which were split into training and testing data and used to evaluate model performance (Benson et al. 2024). The relationship between observed and predicted ages in the test data resulted in an R^2 of 0.91 and a root mean square error (RMSE) of 3 yr. The inclusion of fish length, latitude, sex, and other variables as covariates improved model performance, although accounting for some of these covariates (e.g., length and sex) could represent "double use" of data in stock assessments, and complicate the means by which results are used within an assessment framework (discussed in more detail later in the meeting). Although the RMSE or coefficient of variation (CV) can be estimated by age group, observation error in the reference data makes fully accounting for uncertainty difficult without a neural network (NN) model. In the example for walleye pollock developed by the AFSC, which had a considerably larger sample size (N = 30,000), a MMCNN model with covariates produced an R^2 of 0.91-0.92. Uncertainty estimates were comparable to those associated with traditional ageing methods.

Monte-Carlo simulations with 10,000 iterations were used to estimate 95% confidence intervals with which future data could be evaluated. For the walleye pollock base model (2014-2018), less than 5% of model predictions were outside the confidence intervals for data from 2019 and 2021, indicating good model performance. Results were also presented from an AFSC study in which observation error from both FT-NIRS and traditional age estimates for walleye pollock were compared and quantified. Ages predicted from the MMCNN model (from otolith spectra) by 6 instrument scanners and several different instruments show greater agreement (98%) than traditional ages generated from 7 different readers (84%) indicating that very little error arises from the FT-NIRS predictive process itself.

Ageing error matrices are routinely incorporated into stock assessments for traditional age estimates, and additional study will be needed to quantify the total error arising from the FT-NIRS predictive process. The AFSC is currently conducting a simulation study to quantify the sum total of that error and developing an off-the-shelf method that can be used for other species. The framework for applying FT-NIRS in a production setting requires an assessment of the proportion of samples that need to be aged using the traditional method of ageing (TMA) (e.g., 20% for two readers and 5% for three readers) for CNN model retraining should it be necessary. ASFC is currently conducting simulation studies to evaluate the proportional sample requirements.

Overview of data and neural network modeling at NWFSC

John Wallace and Emily Wallingford (NWFSC) provided an overview of the FT-NIRS data and fully connected neural network (FCNN) modeling at the NWFSC. They reviewed procedures for scanning otoliths and developing the models necessary for analyzing spectral data. They also shared some procedural concerns related to the potential impact of excluding otoliths (e.g., those that were missing or showed greater than 30% damage) on model bias. It would be beneficial to quantify the proportion of otoliths that were excluded and its potential impact on bias in age compositions in the future.

Reference scans were collected for sablefish, Pacific hake (i.e., whiting), and rougheye/ blackspotted rockfish. They noted that two technicians can process an average of 250 scans per day with a single machine, greatly increasing efficiency compared to traditional methods, although this did not include the time necessary to develop the models needed to convert scanned data to age estimates. Moreover, this also did not include the time necessary to collect data on otolith weights and other metrics that have been shown to improve the FCNN models.

The proponents presented a suite of ways to quantify error in evaluating the performance of model-based FT-NIRS estimates relative to traditional age estimates, including the sum of the absolute deviations between FT-NIRS and traditional age estimates, mean relative absolute error (MRAE), mean absolute percent error (MAPE), and RMSE. Validation was conducted at multiple levels, using either 80% of the data for training and 20% for testing or ⅔ of the data for training and ⅓ for testing, depending upon the level of the FCNN model. In addition to several alternative formulations of the FCNN models, iterative partial least squares (iPLS) models were developed early on but deemed to be less promising. A comprehensive overview of the details related to alternative FCNN modeling algorithms and software were presented and discussed.

Overview of model results for sablefish age estimation

Jim Hastie (NWFSC) presented an analysis for U.S. West Coast sablefish otoliths and ageing using FT-NIRS. FT-NIRS ages were compared to traditional age estimates from sablefish collected on trawl surveys (2017 to 2019 and 2021 to 2022; there was no survey in 2020 and only two survey vessels in 2019, representing half of those used in a normal survey).

All available data were used to train FCNN models on the full time series (2017 to 2022) and individual years. Sensitivity runs for individual years were also carried out. A range of explanatory data was included, from just the spectral data to all sample data (including fish weight, otolith weight, fish length, and depth and latitude of capture). There was some discussion of whether including fish length in the analysis and then using those same length data directly in the assessment was a case of double counting and whether length data should not be included as model covariates. Inclusion or exclusion of length data had only small impacts on the results. Across the range of explanatory data and analyses, FT-NIRS predictions were biased high relative to traditional age estimates for ages 2 to 8 yr (particularly for ages 2 to 4 yr) and biased low for ages 11 and older. However, the ageing software designed to analyze both inter-reader (or inter-lab) bias and ageing error did not identify these biases. The double (between-reader) reads for sablefish showed less variation and inter-reader bias than the FT-NIRS to traditional age estimates comparison. Tom Helser noted that considerable tagging data has improved age estimates for sablefish off Alaska, which are difficult to age. However, tagging data from Alaska cannot be used to improve West Coast age estimates given the differences in ocean conditions and subsequent growth patterns.

Review of sablefish assessment sensitivity with FT-NIRS

Chantel Wetzel (NWFSC) presented some sensitivities of the 2023 sablefish assessment to using FT-NIRS ages in place of traditional age estimates. Due to various factors, it was not possible to get FT-NIRS ages for every fish with traditional age estimates from the survey. The number of fish without FT-NIRS ages represents a small portion of all ages from 2017 to 2022. Due to this issue, the first step was to run an assessment model using a subset of traditional age estimates (from the 2017 to 2022 surveys) with corresponding FT-NIRS ages. The second step was to replace those ages with FT-NIRS predictions. The third was to update the ageing error matrix for the assessment.

Each step resulted in marginal differences in the parameter estimates and assessment outputs. However, as the sensitivity only replaced the traditional age estimates with FT-NIRS age estimates in the last five years of survey data, it is not clear whether the differences in model parameter estimates and/or assessment outputs would be greater if all or the majority of ages were replaced with FT-NIRS ages. The GFSC discussed future work (e.g., MCMC) to evaluate the effect of replacing primary ages with FT-NIRS ages. Unfortunately, the current sablefish assessment is not amenable to that analysis. There was also discussion of the extra processing that is necessary to convert decimal FT-NIRS ages to integers. At present, whole number ages are necessary for use in stock assessment models.

Day 1 requests, responses, and discussion

The GFSC appreciates the efforts of the NWFSC to prepare responses to several minor requests from Day 1, which clarified the results and stimulated additional discussion. John Wallace clarified that MRE (mean relative error) in his Day 1 presentation was a typo and that he actually evaluated model accuracy using MARE (mean absolute relative error).

Chantel Wetzel provided a graphical approach for evaluating the bias of model predictions at each traditional age estimate and Jim Hastie provided a tabular approach for evaluating the same bias. The graphical approach provides histograms of the difference between the modeled and primary ages at each primary age, where a value of zero indicates no bias (perfect agreement). The tabular approach provided estimates of percent agreement and other metrics for each primary age. The GFSC agreed that both were useful approaches for quantifying bias of the model.

The GFSC and NWFSC discussed why the ageing error estimation program failed to detect bias when graphical and tabular evaluations clearly suggested a bias. There was agreement that developing a procedure for calculating ageing error when combining FT-NIRS and traditional age estimates is an important next step. One idea was to run the program on FT-NIRS and traditional age estimates independently and to characterize uncertainty as the standard deviation from independently estimated ages. This could potentially be accomplished as part of the sablefish stock assessment.

The GFSC and NWFSC also discussed potential reasons for why FT-NIRS tends to overpredict relative to traditional age estimates at younger ages (positive bias) and underpredict at older ages (negative bias). One potential explanation is that there is more scope for older predictions at younger ages whereas there is more scope for younger predictions at older ages. Another is that there have been fewer data from older fish available to include in the training models, thus there is more potential for bias or uncertainty for those ages. However, the GFSC noted that effects of bias on the older ages in stock assessment outputs may be minimal when aggregated into the “plus group” of older ages within the stock assessment model, provided the bias is primarily observed in fish older than the youngest “plus group” age.

The GFSC and NWFSC discussed how well the FT-NIRS approach is likely to perform well for species such as sablefish that have different growth patterns based on where they live. This is supported by the inclusion of latitude or statistical area as a covariate in the NN model, but it warrants further exploration.

Overview of model results for Pacific hake age estimation

John Wallace (Northwest Fishery Science Center) presented on the use of FT-NIRS for ageing Pacific hake. The model with NIRS spectra and metadata (i.e., otolith weight, fish length, fish weight) had marginally better skill than the model trained only on NIRS spectra. There was 65.4% agreement between ages generated by FT-NIRS and traditional age estimates. Agreement improved to 92.3% and 97.9% when widening agreement calculations to +/- 1 yr and +/- 2 yr, respectively. This is lower than agreement rates between human readers, who agreed on 80.2% of ages, which improved to 96.6% and 99.2% when widening agreement calculations to +/- 1 yr and +/- 2 yr, respectively. The model is negatively biased at older ages (≥ 10 yr), underpredicting age relative to traditional age estimates. It was also noted that a FCNN model trained on only the metadata (i.e., otolith weight, fish length, fish weight) performed well, suggesting that it could be used to flag and check suspicious traditional age estimate results.

The methodology proponents and the GFSC notes that the application of FT-NIRS for Pacific hake is advantageous because this species is assessed every year and has large sample sizes for otoliths. Even though Pacific hake is not assessed by the Council, the ageing of more Pacific hake with FT-NIRS would free up time and resources for ageing other Council-assessed stocks. The FT-NIRS approach will have to be reviewed and approved by other members of the Pacific Hake/Whiting Treaty to be used for the Pacific hake assessment.

Fish age is inherently continuous but uncertainty in birthdates, especially for species with protracted spawning seasons, complicates fractional estimates. The modeling frameworks chosen (e.g., various types of neural network models, PLS, etc.) all assume continuous data but ages are entered into the model as integers. The GFSC noted that model performance could also be evaluated using continuous FT-NIRS predictions and traditional age estimates. The age structure provided to the stock assessment could later be derived from a distribution of continuous ages. As groundfish stock assessments ultimately require integer ages, an evaluation of rounding versus truncation-based approaches for converting FT-NIRS predictions to integers should be included within the overall assessment of model performance.

The GFSC expressed concerns, shared by some of the analysts, about predicting ages using fish length and other metadata because these data are often (but not always) included elsewhere in the stock assessment (e.g., in length composition data and/or marginal age composition data). This would represent “double counting” of the data. If FT-NIRS predictions were only used to flag and review suspicious traditional age estimates, such concerns might be diminished. Finally, the GFSC discussed the potential value of systematically rather than randomly generating training datasets, specifically to ensure more even sample sizes across ages, as older ages are underrepresented in the current training dataset. However, it was also noted that sample balancing approaches can limit the unseen variability included in the test dataset if the construction of the training and testing datasets is not done carefully.

Overview of model results for rougheye/blackspotted rockfish age estimation

John Wallace (NWFSC) provided an overview of the model results for rougheye/blackspotted rockfish. Fewer FT-NIRS scans were available for this cryptic species complex, with approximately 1,000 samples that had paired ages and biological data (1,352 scans and paired traditional age estimates). Preliminary analysis indicated no qualitative differences in the spectra produced from a subset of samples of the two species for which genetic identification was available. The FCNN models performed similarly with and without covariates (i.e., length, otolith weight), but showed a lower percent agreement than the sablefish models, with less than 10% agreement between FT-NIRS and traditional age estimates. Double reads using traditional ageing methods have only about 20% agreement for this complex. Because rougheye/blackspotted rockfish is a longer-lived species and is more difficult to age, many more samples are required to cover the full age distribution.

The GFSC recommends further research and data collection for FT-NIRS to be useful for production ageing of rougheye/blackspotted rockfish. The low percent agreement between FT-NIRS and traditional age estimates partly reflects the fact that there are very few samples from fish younger than 15 years of age available for training the model. It was noted that a more complete characterization of variability of spectral information across the range of ages is particularly important to successful model development. The combination of fewer observations of younger individuals, the presence of two cryptic species, and the difficulty in obtaining traditional age estimates from older fish increases uncertainty of FT-NIRS estimates for this complex. Consulting with other Science Centers (e.g., AFSC) on ways to reduce uncertainty in the traditional age estimates using supplementary data such as morphometrics may be beneficial. If uncertainty in the training data can be improved in future years, FT-NIRS may prove useful for this cryptic species complex, but there was general agreement that the method was not ready to be applied for this species complex at this time.

Review of AFSC Pacific cod assessment sensitivity with FT-NIRS

Tom Helser (AFSC) presented a study that tested the sensitivity of a stock assessment for Pacific cod to FT-NIRS ageing methods. This study used the MMCNN method described in Day 1. The work was based on six years of sample data (over 8,000 otoliths) and involved subsampling 20% of the otoliths for training the models. Ageing error matrices were included in the assessment model that quantified both the traditional and FT-NIRS uncertainties. The Pacific cod assessment model output showed very little sensitivity when traditional ages were replaced with FT-NIRS age estimates, coupled with ageing error matrices. Error was quantified by comparing observation errors between FT-NIRS and traditional age estimates, which showed again the high repeatability with the machine-based ageing approach. Similar to the sablefish results shown by NWFSC analysts, there was very little sensitivity in the stock assessment model to replacing six years of ageing data with FT-NIRS predictions. Future work will explore the consequences of reduced subsampling of otoliths for traditional age estimates (e.g., 5 or 10%).

The GFSC discussed the timeline of FT-NIRS development for use in AFSC assessments and learned that a Center for Independent Experts (CIE) review will be conducted in early 2025. This review will focus on use of FT-NIRS in the Pacific cod and walleye pollock assessments. Increased sample sizes from a longer time series is desired. The AFSC is moving forward with sensitivity testing in an assessment when at least five years of data have been obtained. Moreover, incorporating more than one survey or source of fishery-dependent data will lead to more robust characterization of uncertainty and sensitivity testing.

The GFSC also discussed the development of workflows to incorporate FT-NIRS into full benchmark assessments. For example, the AFSC is exploring how often MMCNN models need to be re-trained. A toolbox to support neural network modeling decisions about when new validation is needed for existing models is under development by the NMFS. The GFSC encourages NWFSC analysts to incorporate these tools as they become available.

Panel Recommendations

The GFSC discussed the main messages from this methodology review and makes the following recommendations.

For 2025 stock assessments, the GFSC recommends:

- Incorporate a relatively limited number, as suggested by NWFSC staff, of FT-NIRS age estimates into the 2025 sablefish assessment base model, at the discretion of the stock assessment team (STAT). The benefits of being able to include more ages from the most recent years in the assessment are likely to outweigh the costs of additional ageing error. Sensitivities should be conducted similar to what was prepared for this review and the Pacific cod example presented by AFSC analysts.
- Do not incorporate FT-NIRS derived ages in the 2025 rougheye/blackspotted rockfish assessments due the issues described above.
- Provide an update at pre-assessment data workshops for FT-NIRS model diagnostics and results for FT-NIRS based ages that may be used in any of the upcoming assessments. This approach may be useful for chilipepper rockfish, but models for this species have not yet been developed. The GFSC defers to the STAT and responsible ageing lab as to whether there is sufficient time to develop the method for that particular species.

For assessments in 2027 and beyond, the GFSC recommends:

- Not using fish length or weight as covariates in the FCNN (or other) models used to develop FT-NIRS age estimates when an assessment relies on lengths or weights to estimate growth. This creates a circularity where, for example, length is informing age estimates, which in turn inform growth models that are fit to age and length data within the assessment.
- Not rounding age predictions from the FCNN model when quantifying error or generating age distributions. This rounding unnecessarily eliminates information and may inflate (or deflate) estimates of ageing error.
- Not introducing age estimates from the FT-NIRS method for the first time in an update assessment. A benchmark (or full) assessment provides more opportunity to explore new

data sources and ageing error assumptions and allows for more extensive review in a stock assessment review (STAR) panel.

References

Benson IM, Helser TE, and Barnett BK. 2024. Fourier transform near infrared spectroscopy of otoliths coupled with deep learning improves age prediction for long-lived northern rockfish. Fisheries Research. 278:107116.

Helser TE, Benson I, Erickson J, Healy J, Kastelle C, and Short JA. 2019. A transformative approach to ageing fish otoliths using Fourier transform near infrared spectroscopy: a case study of eastern Bering Sea walleye pollock (*Gadus chalcogrammus*). Canadian Journal of Fisheries and Aquatic Sciences. 76(5):780-789.

Appendix A.

SSC Member Participants

Dr. Cheryl Barnes (SSC Groundfish Subcommittee Vice-Chair), Oregon State University and Oregon Department of Fish and Wildlife, Newport, OR
Dr. John Budrick, California Department of Fish and Wildlife, San Carlos, CA
Dr. John Field (SSC Groundfish Subcommittee Chair), National Marine Fisheries Service (NMFS) Southwest Fisheries Science Center, Santa Cruz, CA
Dr. Chris Free, University of California Santa Barbara, Santa Barbara, CA
Dr. Owen Hamel, NMFS Northwest Fisheries Science Center, Seattle, WA
Dr. Kristin Marshall, NMFS Northwest Fisheries Science Center, Seattle, WA
Dr. Tommy Moore, Northwest Indian Fisheries Commission, Forks, WA
Dr. André Punt, University of Washington, Seattle, WA (Day 1 only)
Dr. Jason Schaffler (SSC Chair), Muckleshoot Indian Tribe, Auburn, WA
Dr. Andrew (Ole) Shelton, NMFS Northwest Fisheries Science Center, Seattle, WA
Dr. Tien-Shui Tsou, Washington Department of Fish and Wildlife, Olympia, WA