Comparison of Fish Age Data Using Traditional Counting of Annuli With Ages Derived Using Neural-Network Modeling and Fourier Transform Near-Infrared Spectroscopy for Potential Inclusion in Groundfish Stock Assessments

Dr. James Hastie, John Wallace, Chantel Wetzel: NW Fisheries Science Center
Patrick McDonald, Emily Wallingford, Denise Parker: Cooperative Ageing Project, Pacific States Marine Fisheries Commission

Background and Purpose

Modern age-structured stock assessments rely on data from biological sampling to characterize fishery and survey catch. Individual fish-length data represent the most commonly and easily collected biological information. However, the informational content of length data is limited, for many groundfish, as growth typically slows dramatically while fish are relatively young. Age data provide vital information for model estimates of recruitment strength and variability, which are critical for understanding population dynamics and accurately identifying abundance trends and forecasting near-term harvest guidance. Additionally, age data play a vital role in enabling research to discern environmental drivers of recruitment success and growth.

Despite their importance, the resources available to support age reading have not kept pace with the collection of ageing structures (primarily otoliths) for species that have been assessed using age data. Since the mid-2000s, the number of species with age-based assessments has increased from ~20 to 35, and the number of modeled areas in these assessments has increased from 25 to 46. In contrast, age-reading staff at the Pacific States Marine Fisheries Commission's Cooperative Ageing Project (CAP) in Newport, which is funded by the NW Fisheries Science Center, peaked during 2012-14. The mismatch between demand and traditional ageing capacity has created a backlog of roughly one million unread age structures from species with age-based assessments that were collected as part of fishery or NMFS survey sampling since 2001.

In response to the need to identify faster ways of ageing fish, staff from the NW Fisheries Science Center and have been actively engaged in NMFS's Strategic Initiative (SI) to evaluate the use of Fourier Transform Near-Infrared Spectroscopy (FT-NIRS) as a means of estimating fish ages more rapidly than can be done by manually reading the number of otolith rings (annuli). FT-NIRS has been used extensively in many industries. This exploration includes spectral scanning of otoliths and using neural-network (NN) models to estimate relationships between the spectral data and traditionally-determined fish ages. More recently, we have explored including other sample data, e.g. otolith weight and fish length, as additional explanatory variables in the NN models. Generally, the inclusion of some additional information has been observed to improve the correspondence of modeled and traditional age determinations.

The potential importance of this approach lies in the comparative speeds with which otoliths can be scanned rather than traditionally aged. For most of our groundfish species, traditional break-and-burn otolith reading can produce 25-50 ages per day, depending on the species' average age, difficulty in identifying annuli, and the reader's ability. During the current exploratory phase, CAP staff have been able to scan 175-250 otoliths per day. The NN model development and exploration represents an additional time component, but that work has been conducted by NWFSC staff (Wallace), and hence does not constrain the rate at which scanning can be completed. FT-NIRS will not replace all traditional age reading, but could eventually greatly reduce the reliance on that approach for assessment ageing, and reduce the backlog of unaged structures. In the near-term, given ageing workload considerations that led us to request removal of a redbanded rockfish assessment from the 2025 schedule, we are hopeful of being able to supplement traditional ages in 1-3 remaining groundfish assessments that will be reviewed in 2025.

Research Overview

Ideally, the NN models would be trained using samples from known-age fish. However, for west coast groundfish species, we lack a supply of otoliths from known-age fish that is of sufficient size (and age-coverage) to facilitate the training and testing of NN age models. Traditional counting of annuli from broken-and-burned (or thin-sectioned) otoliths routinely provides the "best scientific information available" regarding fish ages, for use in assessments and other scientific research. However, individual age determinations rely on the experience, judgment, and interpretation of age readers, in addition to cross-lab standards and methods agreed to by the Committee of Age Reading Experts (which includes participants from the US west coast. British Columbia, and Alaska). In addition to ages that will be used in assessments, CAP and agers in west coast state agencies routinely conduct "double-reads" of approximately 20% of those reads, to evaluate the consistency of reads among agers. Where multiple labs are involved in reading a particular species for an assessment, inter-lab double-reads are nearly always conducted. Disagreement between readers over the age of particular samples is an unavoidable reality, which is quantitatively characterized as 'ageing error' within most west coast groundfish assessments. Thus, it is important to remember that we are not training these NN models with, nor evaluating their performance against, 'true' values, but with age-determinations that are, themselves, approximations of the truth.

Funding to support the SI did not begin in earnest until FY2020. Consequently, scanning and NN model exploration for west coast groundfish were slowed initially by COVID restrictions, but then later by a problem with the initial spectrometer that necessitated its replacement, and resulted in little scanning throughout most of FY2023. Work over the past year has focused primarily on two species that had sizable inventories of unaged structures and were thought likely to be selected for assessment in 2025 at the June 2024 Council meeting: sablefish and rougheye/blackspotted rockfish (RE/BS).

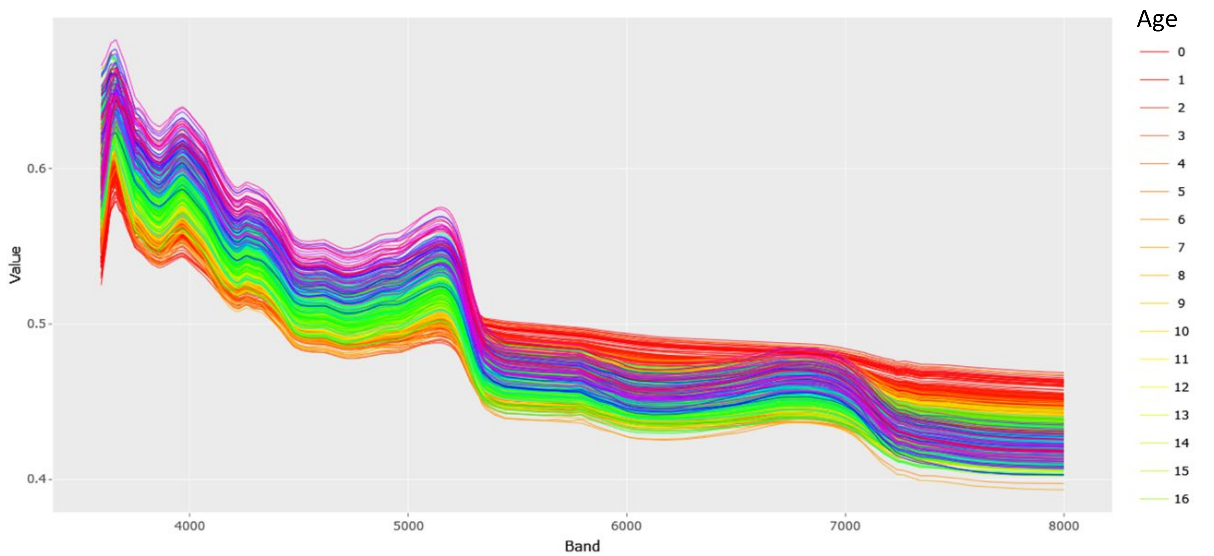
The extensive age reading of survey otoliths conducted for the 2023 sablefish update provided a base of nearly 7,000 traditional age reads from the 2017-2022 NWFSC trawl surveys. Where possible scanning was conducted before age reading, though in many cases, scanning was conducted on the second otolith, where available. The research has explored the use of single-year and multi-year training sets for the NN models, as well as the inclusion of varying amounts of supplemental sample data along with the FT-NIRS scan data. In addition to comparison of modeled and traditional ages on a sample-by-sample basis, we also include sensitivity results from replacing the age reads used in the 2023 sablefish update with modeled survey ages for the 2017-2022 period, using the NN model with the full set of selected supplemental sample data. Although there are now nearly 8,000 RE/BS scans, along with over 6,000 age reads, NN modeling of RE/BS was conducted earlier this year when fewer than 1,400 samples with ages and scans were available. Consequently, currently-available results for RE/BS provide a less robust basis for evaluating the correspondence of modeled and read ages. Limited results from earlier modeling of Pacific hake are also provided, even though we do not anticipate any use of modeled ages in the 2025 assessment.

One important question which is posed to the reviewers is whether, and under what circumstances, it may be appropriate to include fish length as supplemental information in the NN modeling of fish age. Possible scenarios for assessment inclusion of age and length data range from situations where ‘marginal’ ages and lengths are included as separate inputs to an assessment model, ages are added as conditional-ages-at-length (which is the case with survey ages and lengths in the current sablefish model), and situations in which length data are not included at all (which is currently the case for sablefish fisheries, and for all ‘fleets’ in the Pacific hake model). An additional question relates to whether correlations between variables included in the NN models are problematic, and whether a standard should be employed used to avoid correlations between explanatory variables above a threshold in these sorts of models.

Methods: Iterative Partial Least Squares and Neural Net Model Steps to Predict Ages from Spectral Data

Individual fish otoliths in their cleaned and dry native state were scanned with mid-infrared wavebands of light (limited to less than or equal to 8000 cm^-1) using a Bruker Tango Fourier Transform Near Infrared (FT-NIR) spectrometer and the absorbance recorded.

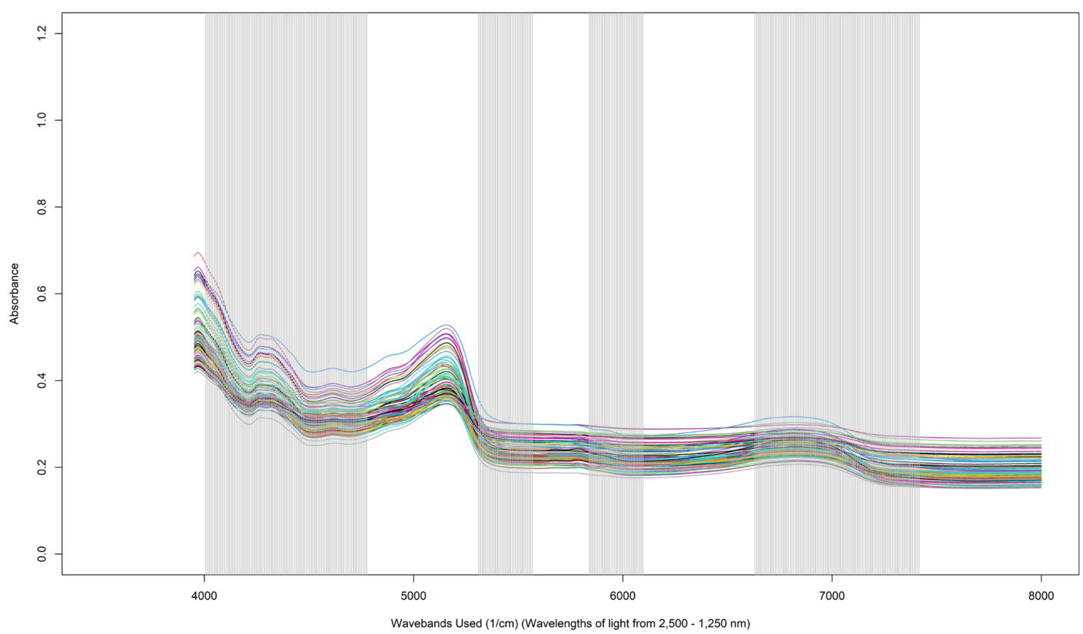
Absorbance output from scans of Sablefish otoliths using a Tango FT-NIR by traditional method of aging (TMA)



A predictive model that gives an estimate of annual age of the otolith was created in the following manner.

1) Savitzky-Golay smoothing (1st derivative of a 2^{nd}-order polynomial utilizing a 17 point window) was applied to the absorbance results from the Tango FT-NIR.

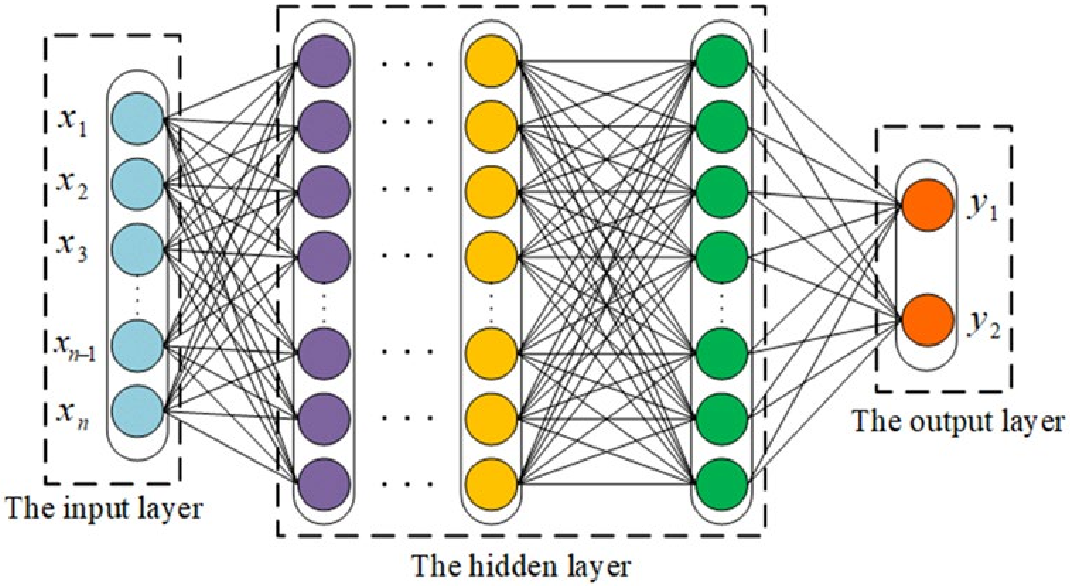
2) Iterative Partial Least Square (iPLS) was employed to identify which of those smoothed wavebands have the highest informational content. In the example figure below, the wavebands within the vertical bands were those selected for that model.



3) Fully connected neural net (FCNN) models with or without the smoothed absorbance output from the selected wavebands (spectra) and with or without selected normalized supplemental sample data were then fit.

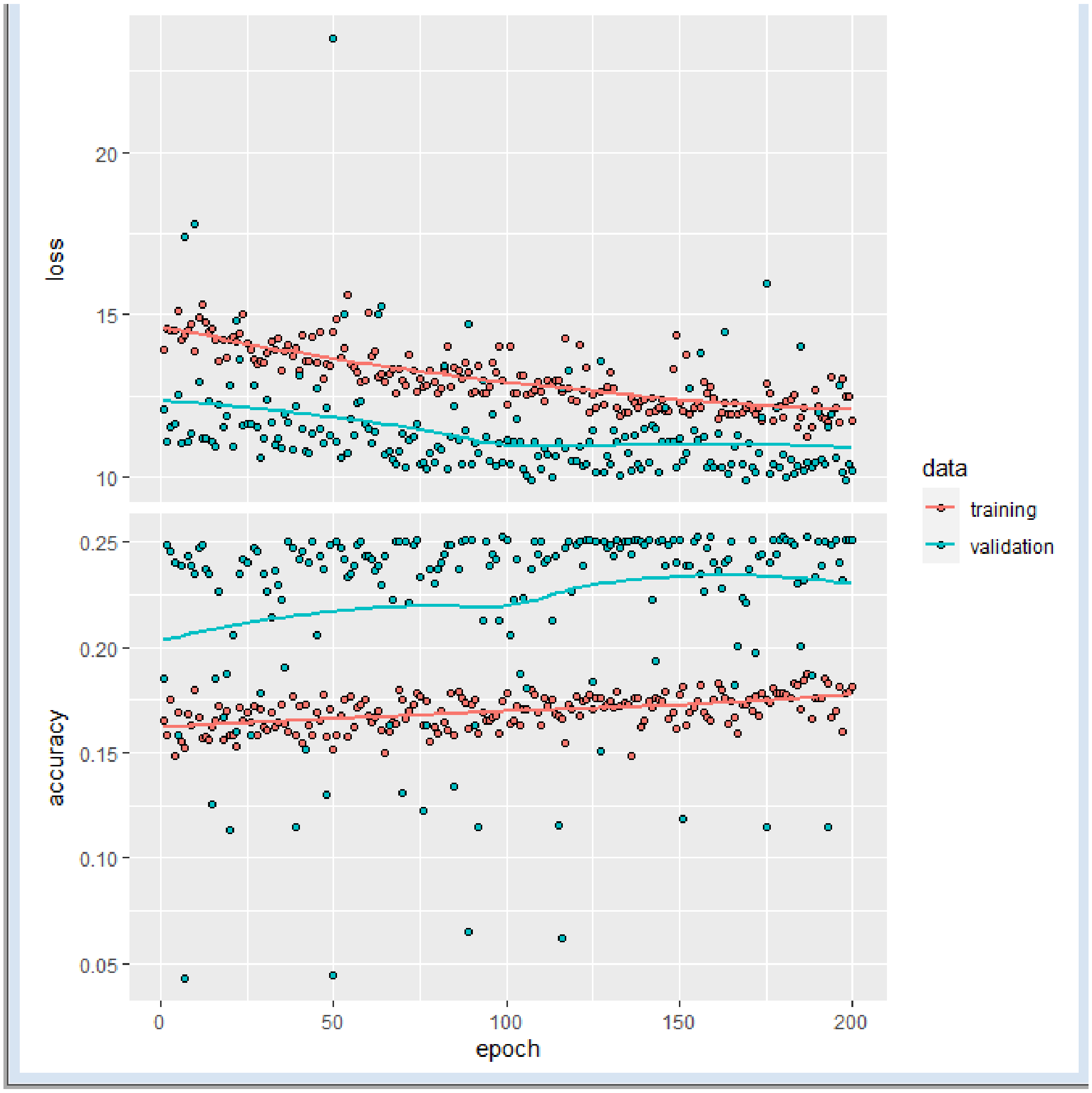
   a. The reference below was followed and testing corroborated that simple models worked best for NIR Spectroscopy on the NWFSC otolith data looked at. One dimensional (1D) spectroscopy is not 2D nor 3D (greyscale) image recognition for which more complex models work better.

      Use of Artificial Neural Networks and NIR Spectroscopy for Non-Destructive Grape Texture Prediction. Basile et al. Foods 2022, 11, 281.
      "We found that increasing the number of hidden layers resulted in a worsening of the prediction of our parameters."



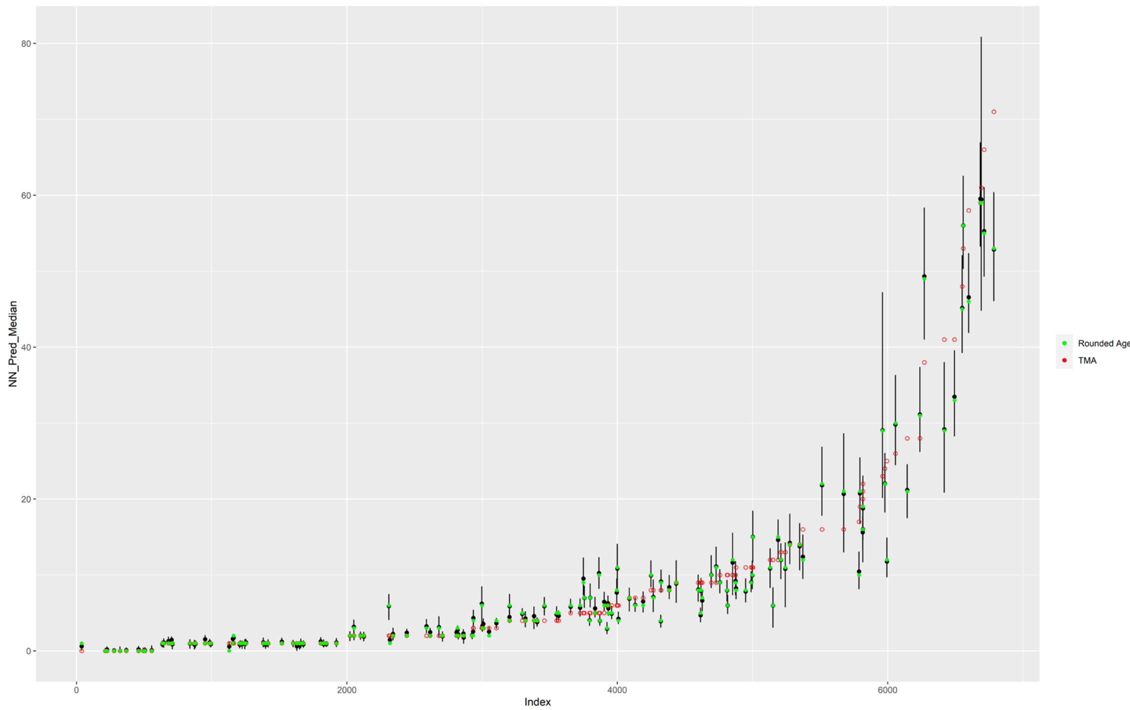
b. Neural net training is done in epochs.
   i. A neural net ‘epoch’ means training the neural network with all the training data for one cycle.
   ii. In an epoch, all of the data is used exactly once. A forward pass and a backward pass together are counted as one pass.

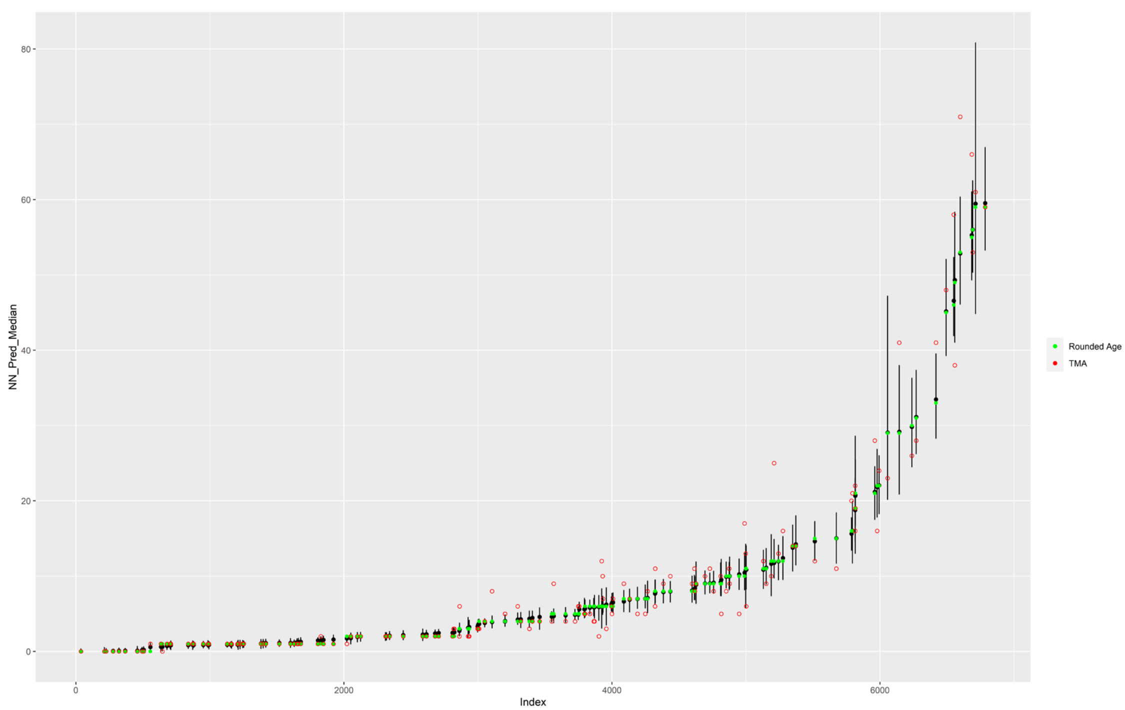
c. An epoch is made up of one or more batches. Batch size is the number of samples to work through before updating the internal model parameters.
   i. At the end of the batch, the predictions are compared to the expected output variables and an error is calculated. From this error, the update algorithm is used to improve the model, e.g. move down along the error gradient.
   ii. The training was structured into 8 iterations of 500 epochs each, with testing against the 1/3 test data done at the end on each iteration to view and record progress.
   iii. Batch sizes of 32 and a validation split of 0.2 (80% of the data was used to train and 20% to test the model) was used.
   iv. Loss and accuracy is evaluated while the NN model is running. Loss is defined as the difference between the predicted value by your model and the “true” value. Accuracy is defined as the number of correct predictions divided by the total number of predictions.



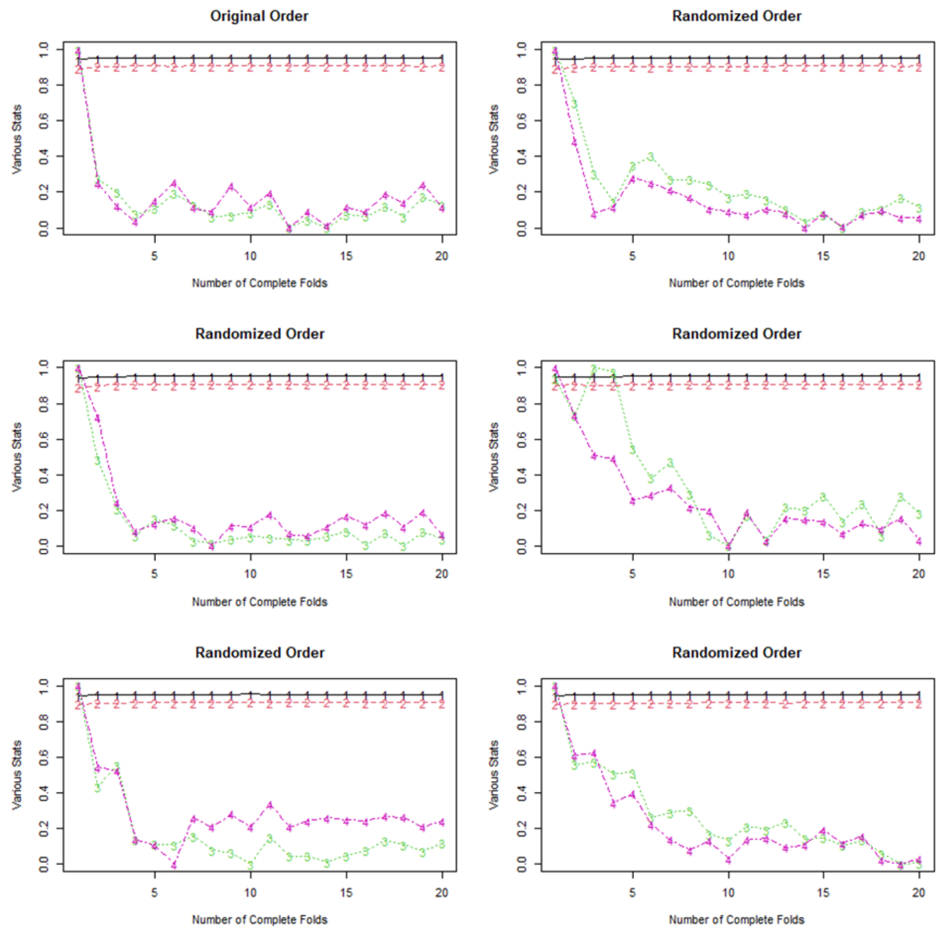
d. K- fold modeling was used on top of the single NN model.
   i. A 10-fold model format was used for single year and a 5-fold model for multi-year models (single species).
   ii. One tenth of the data was left untouched for testing a model from the remaining 90% of the data.

iii. Of that 90%, 2/3 was used for training and 1/3 for testing that particular model (as before).
iv. To train the sub-model, 500 neural net epochs were run on the training set and then tested against the 1/3 test set.
v. Eight such iterations, of 500 epochs each were performed.
vi. As before, a validation split of 20% and a batch size of 32 was used.
vii. However, eventually the model performs worse due to overfitting (almost always by the 8th iteration or 4,000 epochs for this FCNN model on the NWFSC species tested), and hence the best fitting iteration of the 8 is used for the current fold.
viii. After all 10% folds are set aside in turn, a complete fold set is finished, and each predicted point was never inside a model that predicted it.
ix. Twenty complete k-fold models were run, each with a different pseudo random number seed, controlled by a main seed for repeatability.
x. For those 20 replicates a median and 95% (0.025%, 0.975%) quantiles are calculated. A random subset of these statistics are presented below, first sorted by TMA and next sorted by predicted rounded age. The ‘Index’ on the x-axis is simply the numerical ordinal of the data after sorting.


NN_Pred_Median
Index
0
20
40
60
80
0
2000
4000
6000
Rounded Age
TMA


NN_Pred_Median
Index
0
20
40
60
80
0
2000
4000
6000
Rounded Age
TMA

e. Lastly, for a model diagnostic, medians over predicted ages where taken for each additional k-fold model added at each step from 1 to 20; first in the original order and then in randomized order.



1: Correlation: Black

2: R squared: Red

3: Standardized RMSE: Green

4: Standardized SAD: Purple

In the original run order, for this example, the 12th model addition had the best stats. An alternative approach (among others) would be to use those best stats and not the median over all 20 replicates, however that would mess up the quantile interval around the median and was quickly abandoned.

f. Rounding to the nearest integer age is done by first adding a Delta to the ℝ estimated age before rounding. All Deltas from 0 to -0.40 in steps of -0.05 are tried and the best fitting Delta is used. Often Delta is -0.05 or -0.10 for Sablefish and 0.0 for Hake.

g. The metadata examined included: otolith weight, fish length, fish weight, depth, latitude, sex, month of the year, and number of days into the year. Generalized linear model type indicator (dummy) variables were used for the categorical metadata. (referred to as 'one-hot encoding' in the neural net world.)

h. Where possible, metadata were not normalized with min/max. For example otolith weight was converted to decigrams (1/10 of a gram) to increase its value range by 10 fold. Average weight of the small otoliths of Sablefish is around 0.02 grams.

Results

Sablefish

Results for several NN models are compared to traditional age reads for several variations of the data included in the NN model and the timeframe from which samples were used to train and estimate ages. the former include runs in which 1) FT-NIRS data is the sole explanatory variable, 2) scan data are combined with all selected supplemental sample data, and 3) scan data are combined with all selected supplemental sample data except fish length data. For each of these specifications, three approaches for addressing times frames are included. In the first, 6,788 sample observations from the 2017-2022 surveys are using in training step, as described in the previous section, and rounded median age estimates are compared to all primary age reads in that set. In the second, the year-specific values form the first set of results for three years (2017, 2019, and 2022), and then compared to the primary age reads from each of those years. Finally, NN models were trained using only the data from those specific years.

Table 1 reports several metrics of comparison for those nine permutations, and well as for two sets of double-reads using traditional methods. Among the metrics reported are 1) the numbers of observations used in the comparison (sub-headers identify what samples were included in the training set); 2) the average primary age in the application set; 3) the Absolute Percent Error (APE), which is a commonly used metric of performance in the age-reading community; 4) R^2 from estimating the primary age with the modeled age; 5) percent (%) agreement between the two sets of ages; 6) the percentage of modeled ages within 1 year of the primary age, and 7) the percentage of modeled ages within 2 years of the primary age.

Within the first suite of modeled results (in which training and estimation used all 6,788 2017-22 survey otoliths/ages), lower (better) APE values and higher (better) R^2 and agreement percentages were associated with including more supplemental information (as restricted by the variable-selection process). However, even for the best of these models, that which included all of the selected sample data, APE values and the agreement percentages were not as good as observed in either of the two sets of double-read data, though they were closer to those from the larger double-read set, which included all available double-read data covering sablefish samples collected from 2004 to 2023. The R^2 values from the two NN models that included other sample data were slightly better than the large double-read set, but not as good as achieved by the two principal sablefish readers for samples collected during 2011-2023.

In the first set of columns, Table 2 reports the average of differences between the modeled and primary ages (modeled – primary) for each age in the primary set, across all years and for each individual year. The second set of columns expresses these differences as a percentage of the applicable primary age read. These results exhibit a tendency for the modeled ages to exceed the primary ages up through a primary age of 8 and to be lower thereafter. A number of the highest percentage differences, in absolute value, are clustered in ages 2-5, however percentages with similar absolute magnitude are scattered throughout the remaining ages.

The first set of columns in Table 3 reports the sample sizes associated with each cell in Table 1. Although 61 sablefish caught by the survey in this period have been assigned ages of at least 60 years, at least 78% of each year's total has been composed of fish with primary age assignments less than 11 years. The 2nd set of columns in Table 3 provides a comparison of how the differences between modeled a primary ages vary, for three specific years, depending on whether the full training set is used, or one that is restricted to samples from the same year.

Table 4 reports the percentage agreement, within 0, 1, or 2 years, for the same categories included in Tables 2 and 3, and compares those with percentage agreement statistics from the traditional double-reads on samples collected during the same years.

For entire 2017-2022 set of survey data, Figures 1a-3a plot the numbers of ages derived from NN modeling vs primary ages for the three NN model structures described above: FT-NIRS data only; FT-NIRS and other sample data except fish length, and all of the available selected data. The corresponding 'b' plots provide a zoomed-in focus on the counts for primary and modeled ages less than 30 years. Figure 4a and 4b provide a similar depiction of sablefish double-read data spanning samples collected from 2004 to 2023.

Figure 5a depicts the data that is presented in the first column of Table 2, showing average difference, for each primary age, between neural-network modeled ages and the primary ages, using the full 2017-2022 survey data set. Below that, Figures 5b and 5c show the same average of differences between double-reads, at each primary age. Figure 5b includes the full set of 2004-23 double-reads, while Figure 5c is restricted to double-reads between the two principal sablefish readers from 2011 through 2023.

Figure 6 provides an alternative comparative view of modeled vs traditional ages through plotting the numbers of fish ascribed to each age from 0 through 25 by the primary age reader vs the full NN model, as applied to survey samples collected in each of the five years. While there do appear to be some cohort effects among the younger ages, the propensity for the NN modeling to estimate older ages for fish with primary age assignment below 8 years produces a common pattern of higher modeled numbers at age for primary ages between 4 and 9 years.

In addition to the comparison of individual sample age assignments (and summaries, thereof) for one of the scenarios (2017-2022 training set and use of all selected sample data), Dr. Chantel Wetzel was able to provide a rigorous set of assessment runs, comparing results from the 2023 abbreviated update assessment with several alternatives. Due to the lack of availability of an intact otolith for all samples from which survey ages from 2017-2022 were included in the 2023 update, not all of the ages included in last year's modeling could be replaced with modeled ages.

Consequently, as a first step, she ran an adjusted version of the 2023 base model excluding those ages which could not be replaced with modeled ages. For the next step, she removed all traditional survey ages for 2017-2022 and replaced them with the available FT-NIRS ages, while applying the same default ageing error assumption. In the final step, she used the same data, but calculated and applied a NN-model-specific ageing error vector, based on the comparison of all modeled and primary ages from this period. A comprehensive reporting of parameters and estimates from these models is included in Excel workbook < Assessment sensitivity to replacing 2017-22 ages_SM_6788>.

A complete listing of parameters and estimated values is provided in the 'parameters' tab, along with columns, to the right, which show the percentage change in parameters from both the 2023 base model, and the version of that model without the ages that could not be replaced with modeled ages. Additional tabs provide full time series estimates for 1) absolute spawning output, 2) the fraction of unfished spawning output, 3) the number of recruits, and 4) recruitment deviations, along with similar columns showing percentage deviations from the 2023 base model and the version with restricted age data.

In the 'Figures' tab, Figures A1-A10 illustrate the time series across these four models for absolute and relate spawning output; the difference in those two metrics, relative to the 2023 base model; recruitment deviations and numbers of recruits; along with distributions of unfished biomass and Ln (R0) densities. Figure A11 shows the estimated standard deviation of age estimates between base model and the new version specific to the NN-modeled ages.
During the internal review stage, a request was made to document the impact on the base model of completely removing the survey ages (which are included as conditional ages at length) for the 2017-2022 period. Due to limited time, those time series results are not included in the tables, but an accompanying set of figures is included to the right of figures A1-A8. Results associated with the 'Remove 2017-22 CAAL' option in those 8 figures illustrate the dramatic impacts that removing all those survey ages have on various aspects of the 2023 base model.

<u>Rougheye/blackspotted rockfish</u>

A much more limited evaluation of performance is currently available for rougheye/blackspotted rockfish. When NN modeling was conducted earlier this year, only 1,364 scans and paired primary ages were available from several years of fishery data. Unavailability of fish length data for some samples meant that only 1,034 could be included in the model that use scans, otolith weight, and fish length. As reported in Table 5, although the APE values are lower than for sablefish, the percent agreement performance is much lower. (Summarization of double-read agreement was not available in time for this report, but will be shared during the review.)

As with sablefish, figures plotting modeled age vs primary age for each of these three modeling options are provided in Figures 7a to 9a, with zoomed versions of the area out to 50 years of age in the accompanying 'b' figures.

Pacific hake

Limited results are provided for NN-modeling of Pacific hake samples collected during 2011. The only results were generated using a NN model that included FT-NIRS scans, otolith weight, and fish length and weight. Training was conducted using 1,000 otoliths and applied to a total of 2,790. Table 6 reports agreement metrics, which are considerably better than for the other, longer-lived species, which are regarded as harder to age, traditionally. More than 90% of the modeled ages were within 1 year of the primary age reads. Table 7 reports the averages, by primary age, of individual modeled vs read age differences, and the differences relative to the primary age. All of the relative differences are less than 5% of the primary age up through age 8, and are consistently between 11% and 19% above age 9. Figure 10 provides a depiction of modeled vs traditional age reads for this set.

References

Arrington, M.B., Helser, T.E., Benson, I.M., Essington, T.E., Matta, M.E., Punt, A.E., et al. 2021. Rapid age estimation of longnose skate (*Raja rhina*) vertebrae using near-infrared spectroscopy. Mar. Freshw. Res.73: 71–80. doi:10.1071/MF21054. doi:10.1093/icesjms/fsaa131.

Basilone, et al, 2024. Fourier-transform near-infrared spectroscopy first application to age determination in European fish species: the case of the Atlantic horse mackerel from the central Mediterranean Sea. Frontiers Marine Sci., 14 August 2024, DOI 10.3389/fmars.2024.1429150.

Benson, I.M., Helser, T.E., Marchetti, G., and Barnett, B.K. 2023. The future of fish age estimation: deep machine learning coupled with Fourier transform near-infrared spectroscopy of otoliths. Can. J. Fish. Aquat. Sci. 00: 00–00. doi:10.1139/cjfas-2023-004.

Healy, J., Helser, T.E., Benson, I.M., and Tornabene, L. 2021. Aging Pacific cod (*Gadus macrocephalus*) from otoliths using Fourier-transformed near-infrared spectroscopy. Ecosphere. 12: e03697. doi:10.1002/ecs2.

Helser, T.E., Benson, I., Erickson, J., Healy, J., Kastelle, C., and Short, J.A. 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*). Can. J. Fish. Aquat. Sci. 76: 780–789. doi:10.1139/cjfas-2018-0112.

Matta and Helser ed, 2024. Proceedings of the Fourth Research Workshop on the Rapid Estimation of Fish Age. NOAA-AFSC Processed Report 2024-01.

Passerotti, M.S., Jones, C.M., Swanson, C.E. and Quattro, J.M. 2020. 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. doi:10.1016/j.fishres.2019.105439.

Passerotti, M.S., Helser, T.E., Benson, I.M., Barnett, B.K., Ballenger, J.C., Bubley, W.J., Reichert, M.J. and Quattro, J.M. 2020. Age estimation of red snapper (*Lutjanus campechanus*) using FT-NIR spectroscopy: feasibility of application to production ageing for management. ICES J. Mar. Sci. 77(6): 2144–2156.

Robins, J.B., et al. 2015. Revolutionising Fish Ageing: Using Near Infrared Spectroscopy to Age Fish. Department of Agriculture, Fisheries and Forestry, Brisbane, Aus, FRDC Project No 2012/011.

Sass, et al, 2024.Methods for collecting spectral data for three small coastal species. _NOAA Tech Memo, NOAA-TM-NMFS-SWFSC-694.