A transformative approach to ageing fish otoliths using Fourier transform-near infrared spectroscopy (NIRS): a case study of eastern Bering Sea walleye pollock (*Theragra chalcogramma*)

Thomas E. Helser[1], Irina Benson[1], Jason Erickson[2], Jordan Healy[3], Craig Kastelle[1], Jonathan A. Short[1]

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

[2] Bruker Optics, Inc. 5465 East Cheryl Parkway Madison WI 53711, USA

[3] University of Washington, School of Aquatic and Fisheries Science, 1122 NE Boat Street, Seattle, WA 98195

Abstract

We investigated the use of Fourier transform near Infrared spectroscopy (FT-NIRS), which is a method of measuring light absorbance signatures in organic substances, to derive ages from eastern Bering Sea (EBS) walleye pollock (*Theragra chalcogramma*) otoliths. The primary advantage of FT-NIRS ageing over traditional methods is the speed and repeatability in which age estimates are generated. The EBS walleye pollock supports the most valuable fishery in Alaska, with an annual catch of over 2.0 M tons worth $1.0 billion. In this study, the application of FT-NIRS to age EBS walleye pollock yielded $r^2 = 0.90$ between the traditional ages (best age produced between two age readers) and ages based on FT-NIRS spectral information in external validation samples. This approach can be expected to predict fish age within ± 1.0 year of age 67% of the time (RMSEP = 0.96). When comparing approaches, the FT-NIRS had as good or slightly better precision than the traditional ageing method; 75% of the NIR ages agreed with the traditional ages while 66% agreement is typical between two (traditional) age readers. Also, both methods showed little of no bias at age before 10 years of age where sample numbers were robust enough for inference. Once an initial set of otoliths are scanned for spectral data, traditionally aged, and a predictive model calibrated, FT-NIRS has the capacity to generate 120 age estimates per hour compared to 12 ages per hour from traditional methods. The use of FT-NIR spectroscopy for the rapid assessment of fish ages from otoliths, either in the laboratory or during ship-board research surveys, could be transformative for providing a key data source for fish stock assessments and ultimately fisheries management advice. Given the large number of spectra collected and favorable results reported from this study, we suggest the next step is the operationalization the FT-NIRS age production process for EBS walleye pollock stock assessments.

Introduction

Fish ages are one of the most fundamental data elements of integrated stock assessments because they provide information on recruitment, growth, maturity and production (Maunder and Punt 2013; Ono et al. 2015). Estimation of age composition, size at age, maturity at age and longevity of a fish population is critical for assessing the overfishing or overfished status of a stock (Ricker 1975; Hilborn and Walters 1992; Campana and Thorrold 2001; Chen et al. 2003; Conn et al. 2010; Coggins Jr et al; 2013, Maunder and Punt 2013; Ono et al. 2015). Fish age has historically been determined by microscopically counting pairs of annual opaque and translucent growth zones in a number of different hard structures including scales, vertebrae, opercula, spines, and most commonly otoliths (Bagenal and Tesch 1978; Chilton and Beamish 1982).

In federally managed waters of the U.S., over 1.1 M hard structures (minimum estimate) have been examined for age estimates during the period from 2008 to 2015 (Helser et al. 2017) and most management regions have shown a steady increase in the number otoliths aged per year. In federally managed waters of Alaska, 633,000 otoliths were collected from research surveys and commercial fisheries during the last 9 years, from which over 352,000 fish ages were estimated and used in assessment models to support of scientific advice to the North Pacific Fishery Management Council (Helser et al. 2017). While most population assessment models employ the use of age data, depending on species and complexity of the model, the most dramatic case is eastern Bering Sea (EBS) and Gulf of Alaska (GOA) walleye pollock (*Theragra chalcogramma*) with over 12,000 ages annually, followed by Pacific cod (*Gadus macrocephalus*) with over 3,300 annually. EBS walleye pollock supports the most valuable fishery in Alaska with an annual catch of over 2.0 M tons worth $1.0 billion (Ianelli et al. 2017). Other species in federally managed waters of Alaska account for the remaining total of otoliths aged which has average about 20,000 per year (Helser et al. 2017).

A large investment is made in the age determination and production process of age data to support stock assessments. While the traditional production approach varies in the procedures used to prepare and count the number of growth zones in the otoliths depending on species, the hard structures require some degree of handling time in the process of embedding, sectioning, cutting, burning to enhance growth zones, and microscopic examination. Otoliths are processed and aged using a variety of methods including surface reading, breaking the otolith transversely and either baking it or burning the edge, and thin sectioning. The breaking and thin-sectioning methods expose annuli that form on the proximal side of the otolith which are not be visible by surface reading (Beamish and McFarlane 1995). Each method has its advantages and pit-falls in terms of associated costs and quantity and quality of outputs (Begg et al. 2005). For instance, an analysis of 10 years of age production data from the Alaska Fisheries Science Center (AFSC) showed that the rate of output depended on the maximum age of the species, method of processing, and the age reading difficulty (Helser et. al. 2017). Comparatively, there was more than a 50% reduction in efficiency for a long-lived rockfish that required sectioning, such as Pacific Ocean Perch, versus a short-lived groundfish, such as walleye pollock, where age can be estimated by a combination of surface and break-and-burn techniques. Hence, when considering the time from when the otolith is collected at sea to the time a trained expert age reader produces an age estimate, quite an investment has gone into the generation of a single age. Age data have been referred to as one of the most expensive sources of data collected for stock assessments (Campana 1999). Furthermore, most production ageing laboratories employ quality control procedures (to estimate precision and bias) where some percentage (usually between 10 – 20%) of the total specimens are handled and read a second time by another independent expert age reader to estimate age reading precision (Kimura and Anderl 2005). This is a crucial stage in the process because not including ageing error can lead to bias in estimates of growth, weights or maturity-at-age, and estimates of natural mortality by affecting the estimate of productivity (i.e. less older fish means higher natural mortality and higher productivity). Hence, ageing error can in turn affect management indicators or reference points (Reeves 2003; De Pontual et al. 2006; Bertignac and De Pontual 2007).

In the US, and elsewhere in the world, where collecting fish age structures is extensive and expensive, there is a growing need to prioritize data collection needs. This was highlighted at the National level in the results of the 2013 NOAA Fisheries science program review of data used in fishery stock assessments (http://www.st.nmfs.noaa.gov/science-program-review/) which suggested the agency undertake an evaluation of the number of otoliths needed for stock assessments. There are likely multiple reasons for the increased pressure to produce more age data, some of which are efforts to improve stock assessments for data-poor species for which little age data exits, for existing integrated age-structured models to account for sex-specific growth, for spatially-explicit management units, and incorporation of historic age compositions to extend model time-series. For most, if not all, production fish ageing laboratories in the U.S., the demand for age data has outstripped the capacity to produce them, which is why there is a need to rationalize the amount of age data for stock assessments and/or explore new methods and technologies with significantly higher efficiency.

It is widely accepted that traditional age reading is a subjective process where two analysts examining the same structure microscopically interpret and produce different age estimates. There are two aspects to ageing error, precision and accuracy (Kimura and Lyons 1991; Beamish and McFarlane 1995). Precision reflects the repeatability of age estimates on one specimen from one age reader to another. It can be expressed as percentage agreement (PA) between independent readers, average percent error (APE), or CV. The two latter being functionally equivalent although CV's efficiency has been favored in some studies (Beamish and Fournier 1981; Chang 1982; Kimura and Lyons 1991; Campana 2001). Precession is an aspect of the ageing error that will always occur as it is inherent to the nature of the structure and species being assessed (Beamish and McFarlane 1995; Campana 2001). Accuracy is the relationship between estimated and true age (Beamish and McFarlane 1995), i.e., it is an estimation of bias. It is often estimated by age validation methods (e.g., bomb radiocarbon, mark-recapture, marginal increment analysis, etc.) (see review in Campana 2001; Kimura et al 2006).

Efforts to streamline, increase the efficiency, and improve repeatability of age determination has been the focus on a number machine-based technologies during the last several decades. Historically, machine-based or computerized fish age determination has been attempted with various levels of success and fall into several general categories: from otolith morphometrics and image analysis or both. Since otolith growth continues beyond the age at which somatic growth stops (Campana 1999), otolith morphometrics have been used to predict fish age. For instance, otolith weight in many species is significantly related to age, however, the general consensus is that reliable predictions fail for individual ages, particularly of longer-lived fish, due to the level of overlap in otolith weight between age classes (Pilling et al. 2003, Cope et al. 2010). Otolith shape analysis and otolith morphometrics have been used with some success at ageing (Fablet et al. 2009). This has functioned better on young fish (Ye et al. 2015) where otolith weight and a shape produced fish ages with 93.5% agreement to conventional growth zone based ageing, up to an age of 2 years. The otolith diameter proved to be a good predictor of age in sardines up to an age of 280 day. Otolith weigh alone was a good predictor of fish age for oblique-banded grouper, (*Epinephelus radiatu*) up to an age of 13 years and in young of the year bluefin tuna (Megalofonou et al. 2006). Otolith gray scale, opacity, or light intensity has been used in many applications. In scales this worked well to identify daily circuli. (Szedlmayer et al. 1991), and Fablet and Le Josse (2005) used shape statistical classifiers and light intensity of otoliths to determine the age of fish using a learning strategy of a neural network. Advancements in machine based interoperations of images have been made. Nasreddine et al. (2013) used reference images, in a form of pattern matching, to interpret shape sequences and account for inter-individual variability. See Troadec and Benzinou (2002) for a good review on otolith growth zone shape analysis. When a large number of fish need to be aged, such as for stock assessments, the goal may be to develop fish ages more efficiently, while not hindering age accuracy or precision (Fablet and Le Josse 2005). While many of the methods listed above provided assistance to experts who were previously trained in otolith interoperation, they were not developed as a large-scale production age determination tool.

Recently, new applications in the established technology of Fourier transform near infrared spectroscopy (FT-NIRS) have been applied to fish otoliths (Wedding et al. 2014; Robins et al., 2015) and shark vertebra (Rigby et al. 2015) for age determination. FT-NIRS is used in a wide variety of industries including pharmaceutical, chemical, petrochemical, agricultural and food, feed and dairy manufacturing. FT-NIRS is capable of providing both qualitative and quantitative information about the material or product being analyzed. It has applications ranging from raw material identification, to quality control monitoring, and to determination of product composition. FT-NIRS itself is a vibrational spectroscopy technique based on the interaction of electromagnetic energy of a specific frequency range with the covalent bonds in organic molecules. The bonds associated with different functional groups (C-H. N-H, O-H) will absorb energy at unique and characteristic frequencies, and the relative amount of each functional group present is proportional to the amount of energy absorbed. This relationship is what allows this technique to be utilized for quantitative analytical work.

Why does NIRS work on fish otoliths? Fish otoliths are composed of > 90% calcium carbonate ($CaCO_3$), 3-4% organic matrix (protein called otolin), and about 1% inorganic trace elements (Campana 1999). While morphological properties (size, weight, growth) and chemical composition vary widely among fish species (Zorica et al. 2010), the quantity or proportion of protein in the calcium carbonate matrix changes as fish grow older. NIR spectroscopy quantitatively measures the absorption of near infrared wavelengths from excited chemical bonds in protein molecules. Specifically, when protein molecules are excited or irradiated by an energy source they begin to vibrate at different energy levels related to the different chemical bonds. The instrument scans the object and measures the "spectra" of the material as proportions over a specific frequency domain (i.e. near infrared). The spectra most meaningful for otoliths are the groups of molecular constituents in proteins such as carbon-hydrogen (C-H), oxygen-hydrogen (O-H), and nitrogen-hydrogen (N-H) groups. As such, the spectral information is a function of the chemical properties of the otoliths and the quantity of the absorbed NIR energy within those specific spectral regions is a proxy for fish age. Analyses of Saddletail snapper (*Lutjanus malacaricus*) in Australia found that NIRS predictive models could predict the age of fish from otoliths with a high degree of accuracy leading to reduced costs per sample aged and reduced subjectivity (Wedding et al. 2014).

The objective of this study was to evaluate the use of FT-NIRS to age EBS walleye pollock. Specifically our objectives were to; 1) Scan walleye pollock otoliths to gather spectral data and develop a calibration model that predicts new NIR ages from that spectral data from traditionally aged calibration specimens, and to validate model skill; 2) Explore spatial and temporal stability in the calibration-validation models; and 3) compare the relative accuracy and precision between ages generated between the FT-NIR spectroscopy and "traditional" production procedures.

Materials and Methods

For this study walleye pollock otoliths along with biological and station location data were taken from the AFSC's 2016-2017 EBS shelf bottom trawl survey (BTS). The EBS BTS is conducted annually using a 20 nm fixed-grid design, stratified by depths corresponding to inner (0 – 50 m), middle (50 – 100 m) and outer (> 100 m) domains and further subdivided into a southeast to northwest strata. The standard survey area encompasses 492,898 km^2 and annually averages approximately 350 hauls, each made at the center of a grid cell. Details of the EBS BTS design and catch sampling can be found in Conner and Lauth (2017) and Stauffer (2004). Sampling for walleye pollock otoliths was random within each haul. The survey area was divided into low- and high-density strata based on historical density and an isobath of approximately 70 m. Otoliths were collected from all hauls in which the total number of walleye pollock was greater than 19. Five pairs of otoliths were collected in high-density strata and three in low-density strata. Approximately 1,700 walleye pollock otolith pairs were collected in each year and transported to the AFSC's Age & Growth Laboratory (AGL) for analysis.

To facilitate exploration of spatial variability we conducted a statistical spatial analysis to stratify the EBS survey area into distinct regions based on fish growth and morphometrics. Since geographic regions with differing water temperatures, depths and growth are thought to influence biomineraliztion (Chang and Geffen 2013) in otoliths we sought to stratify using an objective measure readily available in the data. In particular, walleye pollock condition was calculated from individual length and weight data and analyzed using Gettis-Ord (Getis et al. 1992) statistic which measures the concentration or lack of concentration (referred to as "hotspots") in a spatially ordered variable.

Otolith samples and age determination

Prediction of fish age from FT-NIRS spectral data requires development of a "model calibration" or "training" set that encompasses the full range of ages, spectral variation (space and time) and instrument measurement conditions. Here the goal is to ensure the calibration model is robust and minimizes extrapolation outside the boundaries of the reference data. A second "external validation" set is also chosen to test the predictive model. Details of walleye pollock age determination can be found in Matta and Kimura (2012), however a brief explanation is provided here related to processing otoliths both for age and spectral data. For age determination, walleye pollock sagittal otoliths are first cut in the transverse plane using a low speed saw, and then roasted either in an oven or alcohol flame to enhance growth patterns. Age is estimated by counting the consecutive pairs of opaque and translucent zones corresponding to the summer and winter growths zones, respectively. A 20% random subsample (test set) of the total AFSC's annual collection is aged a second time by an independent expert analyst for calculating precision and bias (relative) statistics associated with the age determination process. For the 2016 and 2017 test data we computed percent agreement (PA), average percent error (APE; Beamish and Fournier, 1981) and the coefficient of variation (CV; Chang, 1982) to evaluate inter-reader precision for a given sample. High PA values and low CV and APE (< 5.0) values indicate good precision between age estimates. All precision statistics are relayed to data end users at the AFSC. For collection of spectral data (both calibration and validation sets), we used the unroasted sagittal otoliths from among the AFSC test set so as to provide the best possible reference ages (i.e. evaluated by two independent analysts). The average age was used when age estimates from independent analysts disagreed.

We employed two different sampling procedures for the 2016 and 2017 survey otolith collections to partition the calibration and validation data sets. For the 2016 calibration set, otoliths were selected on the basis of a uniform sampling distribution on age, giving about 10 samples per each age group (ages 1 – 12+). This ensured sparse samples on the extremes of the age distribution are represented in the calibration. The external validation collection, which consisted of about 66% more otoliths, was taken as a random selection of the entire 2016 collection. For 2017, the test set of 20% of otoliths from the entire AFSC's 2017 collection was used for spectral data acquisition, and the calibration and external validation sets were equal in number.

Spectral Data Acquisition and Pretreatment

Spectral data from all otoliths were collected on a Bruker Tango-R spectrometer using an integrating sphere in diffuse reflectance mode. Otoliths were placed on the stage distal side up and scanned at three separate orientations relating to 0^o, 45^o and 90^o positions. Data was acquired at $16 cm^{-1}$ resolution with 64 co-added scans for each otolith position, resulting in three separate spectra for each specimen. Here we wanted to analyze the variation in the spectral information associated with the orientation of the otolith. Statistical tests were performed by analysis of covariance of FT-NIR age estimates with the model calibration samples with position as a class variable.

Once all spectral data was acquired, we ran pretreatment procedures to the raw FT-NIR spectral data, which is considered the first step to model development and optimization. First, variation in all spectral data was examined by principal component analysis (PCA) to identify unusual spectra as outliers. Then the spectral data was normalized to remove light scattering effects and small path length differences due to sample presentation anomalies. Furthermore, specific spectral regions with the highest correlation to changes in reference values can be identified and isolated for the model and to enhance subtle spectral variation. To maximize the amount of information across the entire bandwidth of the NIR spectrum we used the 25-point Savitsky-Golay smoothing (2^{nd} order polynomial) and a 1^{st} derivative transformation to remove noise and baseline drift. All of this data pre-processing and model generation was carried out in the chemometric software package OPUS.

FT-NIRS Calibration and Validation

We used Partial Least Squares (PLS) regression, which employs a multivariate calibration method, to develop a quantitative NIR model to predict age. In PLS, the information contained in the spectral data (the X data matrix) is compared to reference values for the component of interest (the Y data matrix) and changes that occur in both matrices are correlated to each other. The spectral data is broken down into principal components or factors (also called loadings) and scores, which allows the most relevant spectral information to be retained while the "noise" or non-relevant information in the data can be disregarded. Evaluation of the loadings and regression coefficients is used to determine the optimal number of factors (or ranks) to include in the model without "overfitting" the data.

Practically speaking, a set of representative samples (the model calibration samples) was measured with an NIR spectrometer to obtain the NIR spectra and the corresponding reference value (in this case age) was determined (in this case by the reader). The spectral data and corresponding model calibration ages were input into the chemometric software package and the calibration model was generated. This was done using the process of *cross validation* where each calibration sample is temporarily removed from the data set, a PLS model is created from the remaining samples and the age of the sample that was temporarily removed is predicted as an unknown. This leave-one-out procedure provided and iterative method was used to refine the predictive calibration model and develop a cross validation model set of FT-NIR ages. Once a cross validated calibration model is obtained, we further assessed the performance (robustness) of this model by using the model to predict a completely new set of representative unknown samples. This process is called *external validation*. In this case, ages for the entire external validation set was predicted by the model and the residual (FT-NIR prediction age vs. traditional age) for each sample is calculated. The mean error of prediction for the external validation set is calculated as the Root Mean Square Error of Prediction or RMSEP. The mean error of prediction for the cross validation is calculated as the Root Mean Square Error of Cross Validation or RMSECV and is an indication of the model accuracy. For a robust model the RMSEP and RMSECV should be similar, however the accuracy of the NIR calibration model is always dependent upon the accuracy of the reference data that was used to create and validate it. Another statistic used to evaluate the performance of calibration model is the Residual Prediction Deviation (or RPD). The RPD is equal to the standard deviation of the reference values divided by the prediction error of the calibration model, either the RMSECV or RMSEP. There are no universal guidelines for determining an acceptable RPD, but in general a minimum value of 3 is desired.

The assessment of model performance and robustness (ability to predict independent samples) from an NIRS perspective is based on the following partial least squares statistics:

1) Coefficient of determination (r^2) of cross calibration ($r_c{}^2$) and external validation/prediction ($r_v{}^2$);
2) Root mean square error of cross validation (RMSECV) and root mean square error of prediction (RMSEP);

$$RMSEP = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(Y_i^{Meas} - Y_i^{Pred}\right)^2}$$

   Where Y is the age of the i th fish, and M is the number of samples
3) Model Bias (B_M - average difference between predicted and observed values); and
4) Slope of the calibration/validation model.

We also assessed the accuracy between ages generated between the FT-NIR and traditional production age estimation procedures by calculating the frequency of the differences as a measure of relative bias (Bias = B); $B^{NIR} = (Age^{Prod} – Age^{NIR})$ and $B^{Prod} = (Age^{read} – Age^{test})$.

Results

Walleye pollock condition was spatially related to latitudinal gradient proceeding from high conditions in the southeastern Bering Sea to northwestern Bering Sea (Fig. 1). Concentrations of generally higher fish condition (> 1.0, a "hot spot") were clustered between the Aleutian Islands in the south up to Pribilof Islands along the middle and outer depth domains. Above approximately 57^{o}N latitude fish conditions were generally close to or below 1.0 and non-significant, and further to the northwest above 58^{o}N latitude significant concentrations were associated with low condition ("cold spots"). Hence, analysis of 2016 and 2017 FT-NIRS validation test data were stratified north and south of 57^{o}N latitude.

Walleye pollock population age compositions in the 2016 EBS BTS were dominated by relatively large 2012 and 2008 year classes (Fig. 2a). In 2017, the 2012 year class was again large as age-5 along with another age-4 2013 strong year class (Fig 2b). Although walleye pollock can grow older than 20 years of age relatively few fish are captured greater than 12 years of age. Sampling schemes for NIR analysis, while different for 2016 and 2017, were reasonably representative of the population age composition overall (Fig. 2c & 2d). The primary difference is the calibration sample selection, where in 2016 otoliths were based on a uniform sampling distribution with target total sample size and in 2017 both calibration and validation samples were the 20% test selected at random. Age reading precision from traditional methods were typical of EBS walleye Pollock with the PA = 67% (± 0 years), APE = 3.14 and CV = 4.44 both 2016 and 2017 (Fig. 3). The scatter shows a good correspondence to a 1:1 line indicating little relative bias.

The PLS loadings from the calibration models for 2016, 2017, and 2016-2017 combined were similar for the majority of the wavelength regions, suggesting that similar molecular compounds correlated with traditional ages. The relevant spectral information for all calibration-validation models was isolated primarily in the 6821 to 5269 cm^{-1} and 5022 to 4171 cm^{-1} spectral regions. Molecular bonds related to these vibrational group frequencies correspond to –CH, –OH, and –NH combination tones. The absorbance spectra for the most informative regions from a selection of 12 walleye pollock otoliths are shown in Figure 4. Given that the exact nature of molecular species of organic compounds in walleye pollock otoliths are unknown, the characteristic of molecules related to absorbers in the near infrared regions can only be inferred from texts such as that of Workman and Weyer (2008). We did not find statistical differences ($p > 0.10$) in the PLS models with regard to position orientation on the FT-NIR spectrometer stage so future studies could reliably include only a single spectral analysis (Fig. 3). The repeatability of FT-NIR ages from trial to trial was also very good with the RSD, which is equivalent to the APE, was generally below 3.0 % which is shown in Figure 3 as tight spread along the 1:1 line.

In total, over 1,300 walleye pollock otoliths were scanned for spectral data. Three calibration models were developed ($n = 655$) and their performance evaluated using 9 separate external validation data sets ($n = 655$). Calibration models were developed for 2016 and 2017 separately, and then 2016-2017 combined (Table 1). Model performance for the 2016 and 2017 calibration models were evaluated for the north and south spectral data. In addition, model performance was evaluated for 2016-2017 combined calibration model with external validation data form 2016 and 2017 individually then combined. The 2016, 2017 and 2016-2017 combined calibration models yielded good predictability, producing an $r_c^2 = 0.91$ to 0.95 (Table 1). Calibration models were qualitatively similar in appearance and had roughly equivalent calibration statistics, so only the 2016-2017 combined PLS calibration model is shown in Figure 5, although detailed statistics are given in Table 1. Graphically the calibration shows a tight correspondence with a 1:1 line indicating exact equality between the predicted FT-NIRS age in the cross validation and the traditional age estimates. Age estimates included in the calibration model ranged from age-1 to age-15 encompassing all age classes observed in the 2016-2017 surveys. Other measures of calibration model fit were favorable as well: RMSECV = 0.78 to 0.97, RPD = 2.99 to 3.85, and very low overall bias = -0.003 to 0.002. In simple terms this means that in the cross validation 90% (or better) of the variability in FT-NIRS ages generated from the spectral data in walleye pollock otoliths was explained by ages estimated by analysts using traditional methods. Also, in the cross validation the predicted FT-NIRS ages from the calibration model were accurate within ± 1 year based on the RMSECV.

In the external validation the predictive calibration model performance for 2016 and 2017 was better when combining the north and south external validation data (Table 1). The predictive performance measures were r_v^2 = 0.93, RMSEP = 0.88 years, and RPD = 3.21 for 2016 and r_v^2 = 0.91, RMSEP = 1.04 years, and RPD = 3.09 for 2017 (Table 1). In general, calibration model performance degraded slightly when externally validated with south data where the r_c^2 dropped to r_v^2 = 0.82 and 0.85 for 2016 and 2017, respectively. In both cases, the RPD declined below a value of 3.0. When combining 2016 and 2017 spectral data into a single calibration model (Fig. 5a), and validated against each year, performance was also overall good; r_v^2 = 0.91 for 2016 and r_v^2 = 0.88 for 2017 (Table 1; Fig. 5c & 5d). In each case where data were combined over space and time the RPD was 3.0 or greater. Model bias was lower than nearly all previous model evaluations in this case (-0.046 and -0.036) probably due to the large increase in spectral data for calibration. RMSEP was 0.817 and 1.12 years for the 2016 and 2017 validation tests, respectively, and indicates there is only a .2 of a year difference in prediction accuracy between years. Hence the north and south data for both 2016 and 2017 were combined to develop a global calibration model for EBS walleye pollock representing that spectral variability.

The accuracy (or relative bias) between ages generated between the FT-NIR and traditional production age estimation procedures were evaluated for the 2016 and 2017 separately and then for 2016-2017 combined. Qualitatively there was little difference in relative bias comparisons between the FT-NIR and traditional ageing approaches, so only the combined 2016-2017 analysis is shown in Figure 6. Frequency histograms of relative bias between the two age determination procedures were nominal. For the 2016-2017 combined data, 75% of the FT-NIR age estimates were the same as traditional production ages while 68% of the reader and tester ages from the traditional age reading procedure were identical (Fig. 6a). Similarly, 94% and 92% of the FT-NIR and standard ageing procedure, respectively, were the same by ± one year of age. Relative bias by age were largely indistinguishable between the ageing procedures up to 10 years of age, after which FT-NIR ages were consistently less bias than the standard production procedure in 2016 (Fig. 6b). Samples numbers diminish rapidly after 10 years of age so estimates are less reliable. In general, these results suggest that age estimation using FT-NIR spectroscopy of walleye pollock otoliths replicated the traditional age estimation procedure with equal or slightly better precision.

Discussion

While FT-NIR spectroscopy has its roots in the manufacturing/industrial sciences, there are numerous examples of the application of NIR spectrometry in ecological and wildlife research in recent years for which Vance et al. (2016) provide citations in both invertebrate and vertebrate studies. The most common application of NIR spectroscopy in vertebrate wildlife ecology, particularly mammals, is the analysis of fecal samples for questions of nutrition, diet composition and parasitology (Vance et. al. 2016). For instance, investigators examined the use of NIR spectroscopy to analyze marine mammal diets via fecal analysis (Kaneko and Lawler 2006). In fisheries ecology research, we found only a few published studies for which NIR spectroscopy has been used to age fish from hard structures. Wedding et al. (2014) aged otoliths from a demersal snapper (*Lutjanus malabaricus*) while Rigby et al. (2016) aged two shark species (*Sphyrna mokarran* and *Carcharhinus sorrah*) from vertebrae. Two other species of fish, Barramundi (*Lates calcarifer*) and Snapper (*Pagrus auratus*) have also been analyzed (Robins et al. 2015).

Compared to other studies for fishes, our results for ageing EBS walleye pollock from FT-NIR spectroscopy of otoliths was equally promising. For instance, for saddletail snapper (Wedding et al. 2014) the calibration model yielded r^2 = 0.93 and RMSECV of 1.35 with equally good validation statistics (r^2 of 0.94 and RMSECV of 1.54). For the two species of sharks (Rigby et al. 2016) FT-NIR analyses were slightly poorer where calibration models r^2 yielded 0.84 to 0.89. Of the three calibration models developed for walleye pollock in this study from over 1,500 spectra analyzed r^2 in the cross validation ranged between 0.91 and 0.95 for calibration models (which combined entire survey area) and nearly equally good external validation statistics (r^2 = 0.93 and RMSECV of 0.88 for 2016; 0.91 and RMSECV of 1.0). This indicates that walleye pollock age could be predicted within ± 1.0 increment count or year of age about 67% of the time.

The accuracy of predicting ages from future samples depends on the validity of the calibration models which must be assessed for every species analyzed. The validity of a calibration model will need to be assessed as to how robust model results are to variability in otolith microchemistry which is thought to be an interaction between fish physiology and environment

(Chang and Geffen 2013). Ontogenetic changes of fish related to growth rate, reproduction, feeding, diet changes, and stress are known to influence trace element assimilation in otoliths (Radtke and Shafer 1992). Furthermore, the chemical composition of the water varies geographically and over time so environmental factors such as depth, salinity, and temperature will also play a role (Chang and Geffen 2013). Wedding et al. (2014) found conditions in otolith microchemistry were so variable between post-wet and post-dry seasons of saddletail snapper that calibration models were not robust across seasons, but recommended combining the data to account for this variability. In another study of geographic variability, Robins et al. (2015) found calibration models for a specific location and season are not as successful at predicting age across at differing locations or seasons for the river species of barramundi (*Lates calcarifer*) or coastal snapper. This highlights the fact that there is a lack of deep understanding relating to the biological and physical factors contributing to spectral variability in fish otoliths. Insight from other disciplines using FT-NIR would suggest that prediction accuracy becomes less sensitive to unknown factors when more variability is accounted for (Bobelyn et al. 2010). The flip side is that too much biological variability may lead to reduced prediction accuracy.

In most research applications, knowledge of water chemistry and precise biological factors affecting the molecular constituents activated by NIR will not be known, so efforts to evaluate model robustness might consider broad scale geographic and temporal factors. We spatially stratified the 2016 and 2017 EBS otoliths samples on the basis of walleye pollock body condition to evaluate model robustness, but found only nominal difference between the north and south in 2016 (RMSEP was 33% higher in the south compared to the north). Where the south otoliths were incorporated into the 2016 and 2017 validation the RPD fell 35% and 13%, respectively. The extent to which the otoliths from the south affected the model performance each year is unclear, however, the presence of the very large 2012 year class, most of which occurred in the south may be influencing the results (i.e., over prediction of FT-NIR ages for 2016 south validation). Subsequent analysis of 2016 samples where associated otolith weights were considered, revealed that larger model residuals were related ($r = 0.5$, $p < 0.001$) to heavier otolith weights at age-4. This may suggest that some otoliths estimated to be age 4 by the traditional methods had an otolith weight more consistent with age-5 walleye pollock. Misassigning the age of fish to a particularly large cohort when present in the population age composition has been identified as a source of ageing error (Kimura et al. 1992). Nevertheless, model robustness was still very promising when combining 2016 and 2017, as well and two survey year's data, over the entire survey area which seemed to account for the unequal distribution of the dominant age 4 fish in the EBS.

Prediction accuracy of FT-NIR ages will only be as good as the accuracy in the calibration samples that are provided and whenever possible an age validation (estimation of true ageing bias) should be conducted to facilitate the use of FT-NIR age estimation. EBS walleye pollock ages are not without error nor has the age determination method used been unequivocally validated. From over hundreds of thousands of historic walleye pollock double readings the ageing precision at AFSC has been well characterized and the FT-NIR ageing process has replicated this quite well. However, the impact of that observation error on the FT-NIR model calibration results needs further investigation. In terms of ageing (true) bias, the general age range and longevity of walleye pollock has been determined from Pb/Ra studies (Kastelle and Kimura 2006), they found an average of a 0.8 year bias between Pb/Ra ages and traditional ages. Also, recent work using $\Delta^{14}C$ indicates (Kastelle unpublished) ageing bias using traditional methods are within the accuracy of at least +/- 1-3 years, which is the limit of resolution associated with that technique (Campana 1999). Other measures of ageing bias, such as known age fish, are not available for walleye pollock. Therefore, it is quite possible that age misassignment by one year less than the true age (i.e. misassigning age 5 fish to age 4) may have resulted in observation error rather than prediction error in some cases, particularly in the south data which had an effect on calibration. Efforts to derive true walleye pollock ages will only improve the calibration model and studies should be undertaken to support this.

Precision and efficiency is one of the main benefits of FT-NIR age estimation procedures over traditional approaches. Age determination precision using the FT-NIR spectroscopy was as good, or better, than the traditional ageing methods. Generally we found a slightly greater percentage of the sample in agreement between the FT-NIR age and the reference age than between two independent age reading analysts. Also relative bias by age was as good or slightly better using FT-NIR up to, in most cases, twelve years of age which represents the vast majority of population numbers (Ianelli et al. 2017). One reason for possibly better precision with FT-NIR is that traditional age determination methods of growth zone counts are inherently a subjective process of pattern recognition between two people (Campana 1999). Many factors are at play that can result in a discrepancy in the age estimates between analysts (even assuming use of a standardized age determination criteria) including variation in sample processing, sample presentation, microscopic errors as well as personal well-being and stress. Most production ageing laboratories apply rigorous quality control measures to maintain consistency and guard against relative bias. By its nature, FT-NIR spectroscopy is a quantitative technique (once the parameters for scanning and wavenumber ranges selected) that measures the vibrational frequencies of covalent bonds in the otolith molecules. Since the instrument is frequently internally calibrated the spectral data is acquired with high repeatability (Roggo et al. 2017). With a robust model, efficiency is one of the major benefits of FT-NIR over traditional production ageing procedures. Samples can be analyze whole without labor intensive sample processing such as embedding in resin, sectioning, cutting, roasting, etc., and final microscopic examination. We estimate that, on average, it takes an age reading analyst approximately 3-5 minutes to estimate a single fish age using traditional procedures compared to approximately 30 seconds to estimate an age from a walleye pollock otolith using FT-NIR spectroscopy. Thus, once an initial set of 120 otoliths were scanned for spectral data, age estimated using traditional methods and model calibrated, FT-NIR had the capacity to generate 120 ages per hour compared to 12 ages per hour. The use of FT-NIR spectroscopy for the rapid assessment of fish ages from otoliths, either in the laboratory or ship-board research surveys, could be transformative for providing a key data source for stock assessments and ultimately management advice. The large number of spectra collected and favorable results reported from this study indicate that operationalizing the age production process for EBS walleye pollock stock assessments is feasible and protocols for implementation are the next step.

References

Bagenal, T., and F. Tesch. 1978. Age and Growth in Method of Assessment of Fish Production in Fresh water. IBP Handbook. Blackwell Scientific Press, Oxford.

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. Recent developments in fish otolith research. University of South Carolina Press, Columbia 545-565.

Beamish, R., and D. Fournier. 1981. A method for comparing the precision of a set of age determinations. Canadian Journal of Fisheries and Aquatic Sciences **38**: 982-983.

Begg, G. A., S. E. Campana, A. J. Fowler, and I. M. Suthers. 2005. Otolith research and application: current directions in innovation and implementation. Marine and Freshwater Research **56**: 477-483.

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 Journal of Marine Science: Journal du Conseil **64**: 981-988.

Bobelyn, E., Serban, A.-S., Nicu, M., Lammertyn, J., Nicolai, B.M., Saeys, W. 2010. Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effects of biological variability on specta and model performance. Postharvest biology and Technology 55: 133-143.

Campana, S.E., 1999. Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Marine Ecology Progress Series* **188**: 263-297.

Campana, S. E., and S. R. Thorrold. 2001. Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations? Canadian Journal of Fisheries and Aquatic Sciences **58**: 30-38.

Chang, W. Y. 1982. A statistical method for evaluating the reproducibility of age determination. Canadian Journal of Fisheries and Aquatic Sciences **39**: 1208-1210.

Chang M-Y, Geffen, A. J., 2013. Taxonomic and geographic influences on fish otolith microchemistry. *Fish and Fisheries* **14**, 458-492.

Chilton, D. E., and R. J. Beamish. 1982. Age determination methods for fishes studied by the groundfish program at the Pacific Biological Station. Department of Fisheries and Oceans.

Coggins Jr, L. G., D. C. Gwinn, and M. S. Allen. 2013. Evaluation of age–length key sample sizes required to estimate fish total mortality and growth. Transactions of the American Fisheries Society **142**: 832-840.

Conn, P. B., E. H. Williams, and K. W. Shertzer. 2010. When can we reliably estimate the productivity of fish stocks? Canadian Journal of Fisheries and Aquatic Sciences **67**: 511-523.

Conner, J., and R. R. Lauth. 2017. Results of the 2016 eastern Bering sea continental shelf bottom trawl survey of groundfish and invertebrate resources. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC352,159 p.

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 Journal of Marine Science: Journal du Conseil **63**: 1674-1681.

Fablet, R., Chessel, A., Carbini, S., Benzinou, A., and de Pontual, H. 2009. Reconstructing individual shape histories of fish otoliths: A new image-based tool for otolith growth analysis and modeling. Fisheries Research **96**(2): 148-159.

Fablet, R., and Le Josse, N. 2005. Automated fish age estimation from otolith images using statistical learning. Fisheries Research **72**(2-3): 279-290.

Getis, A., Ord, J.K. 1992. Analysis of spatial association by use of distance stastistics. Geographical Analysis , Vol. 24, No. 3. Ohio State University Press.

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Reviews in fish biology and fisheries **2**: 177-178.

Hussy, K., Groger, J., Heidemann, F., Hinrichsen, H.H., and Marohn, L. 2015. Slave to the rhythm: seasonal signals in otolith microchemistry reveal age of eastern Baltic cod (*Gadus morhua*). ICES J. Mar. Sci

Kaneko, H., Lawler, I.R. 2006. "Can near infrared spectroscopy be used to improve assessment of marine mammal diets via fecal analysis?:. Marine Mammal Science 22: 261.

Kastelle, C.R., and Kimura, D.K. 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.

Kimura, D.K., and Anderl, D.M. 2005. Quality control of age data at the Alaska Fisheries Science Center. Mar Freshw Res **56**(5): 783-789.

Kimura, D., J. Lyons, S. MacLellan, and B. Goetz. 1992. Effects of year-class strength on age determination. Marine and Freshwater Research **43**: 1221-1228.

Mapp, J.J.I., Fisher, M.H., Atwood, R.C., Bell, G.D., Greco, M.K., Songer, S., and Hunter, E. 2016. Three-dimensional rendering of otolith growth using phase contrast synchrotron tomography. J. Fish Biol. **88**(5): 2075-2080.

Matta, M.E., Kimura, D.K (editors). 2012. Age determination manual of the Alaska Fisheries Science Center Age and Growth Program. NOAA Professional Paper NMFS 13, 97 p.

Maunder, M. N., and A. E. Punt. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research **142**: 61-74.

Megalofonou, P. 2006. Comparison of otolith growth and morphology with somatic growth and age in young-of-the-year bluefin tuna. J. Fish Biol. **68**(6): 1867-1878.

Nasreddine, K., Benzinou, A., and Fablet, R. 2013. Geodesics-Based Image Registration: Applications To Biological And Medical Images Depicting Concentric Ring Patterns. Ieee T Image Process **22**(11): 4436-4446.

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, and C. S. Szuwalski. 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES Journal of Marine Science: Journal du Conseil **72**: 31-43.

Reeves, S. A. 2003. A simulation study of the implications of age-reading errors for stock assessment and management advice. ICES Journal of Marine Science: Journal du Conseil **60**: 314-328.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Dept. of Fisheries and Oceans.

Rigby, C.L., Wedding, B.B., Grauf, S., Simpfendorfer, C.A., 2014. The utility of near infrared spectroscopy for age estimation of deepwater sharks. Deep-Sea Res Pt I 94, 184-194.

Rigby, C.L., Wedding, B.B., Grauf, S., and Simpfendorfer, C.A. 2016. Novel method for shark age estimation using near infrared spectroscopy. Mar Freshw Res **67**(5): 537-545.

Robins J.B., Wedding, B.B., Wright C., Grauf S., Sellin M., Fowler A., Saunders T. and Newman S. 2015. Department of Agriculture, Fisheries and Forestry, Revolutionising Fish Ageing: Using Near Infrared Spectroscopy to Age Fish. Brisbane, April, 2015. CC BY 3.0. **http://frdc.com.au/research/Final_reports/2012-011-DLD.pdf**

Roggo, Y., Chalus, P. Maurer, L., Lema-Martinez, C., Edmond, A., Jent, N. 2007. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. Journal of Pharmaceutical and Biomedical Analysis 44: 683-700.

Szedlmayer, S.T., Szedlmayer, M.M., and Sieracki, M.E. 1991. Automated Enumeration by Computer Digitization of Age-O Weakfish Cynoscion-Regalis Scale Circuli. Fish. Bull. **89**(2): 337-340.

Troadec, H., and Benzinou, A. 2002. computer assisted age estimation. *In* Manual of Fish Sclerochronology. *Edited by* J. Panfili, H. de Pontual, H. Troadec and P.J. Wright. Ifremer-IRD coedition, Brest, France. pp. 199-241.

Vance, C.K., Tolleson, D.R., Kinoshita, K., Rodriguez, J., Foley, W.J. 2016. Revew: Near infraread spectroscopy in wildlife and biodiversity. Journal of Near Infrared Spectroscopy 24:1-25.

Wedding BB, Forrest, A.J., Wright, C., Grauf, S., Exley P. 2014. A novel method for the age estimation of Saddletail snapper (*Lutjanus malabaricus*) using Fourier Transform-near infrared (FT-NIR) spectroscopy. *Marine and Freshwater Research* **65**, 894-900.

Workman, J., Weyer, L. 2008. “Practical Guided to Interpretive Near Infrared Spectroscopy”. CRC Press Taylor & Francis Group: Boca Raton, FL.

Ye, Z.J., Zhang, C., Panhwar, S.K., Li, Z.G., and Wan, R. 2015. Ageing Belanger's croaker, Johnius belangerii (Cuvier, 1830), based on otolith shape analysis. J. Appl. Ichthyol. **31**(1): 27-31.

Zorica B., Sinovcic, G., Kec, V.C., 2010. Preliminary data on the study of otolith morphology of five pelagic fish species from the Adriatic Sea (Croatia). *Acta Adriatica* **51**, 89-96.

Table 1. Results of PLS model to EBS walleye pollock showing 3 calibration and 9 validation model statistics.

	Calibration Model Stats								External Validation Stats							
Calibration	r^2	n	**RMSECV**	**RPD**	**Bias**	**Slope**	**Offset**	**Validation Set**	r^2	n	**RMSEP**	**Bias**	**SEP**	**RPD**	**Offset**	**Slope**
2016 North	94.6	202	0.78	3.85	0.002	0.99	-0.17	2016 North	0.87	217	0.667	0.187	0.641	4.21	0.44	1.00
and South								2016 South	0.82	232	1.040	-0.551	0.886	2.13	1.65	0.82
								2016 North and South	0.93	449	0.882	-0.195	0.860	3.21	0.44	0.96
2017 North	92.4	340	0.966	2.99	-0.003	0.89	0.69	2017 North	0.89	173	1.090	0.231	1.070	3.21	0.79	0.88
and South								2017 South	0.85	171	0.975	-0.240	0.945	2.31	0.60	0.94
								2017 North and South	0.91	344	1.040	-0.004	1.040	3.09	0.69	0.89
2016 and 2017	91.3	654	0.826	3.26	-0.001	0.91	0.56	2016	0.91	328	0.817	-0.046	0.817	3.28	0.51	0.92
								2017	0.88	338	1.120	-0.036	1.120	2.65	1.07	0.84
								2016 and 2017	0.89	665	0.957	-0.050	0.956	3.14	0.87	0.86

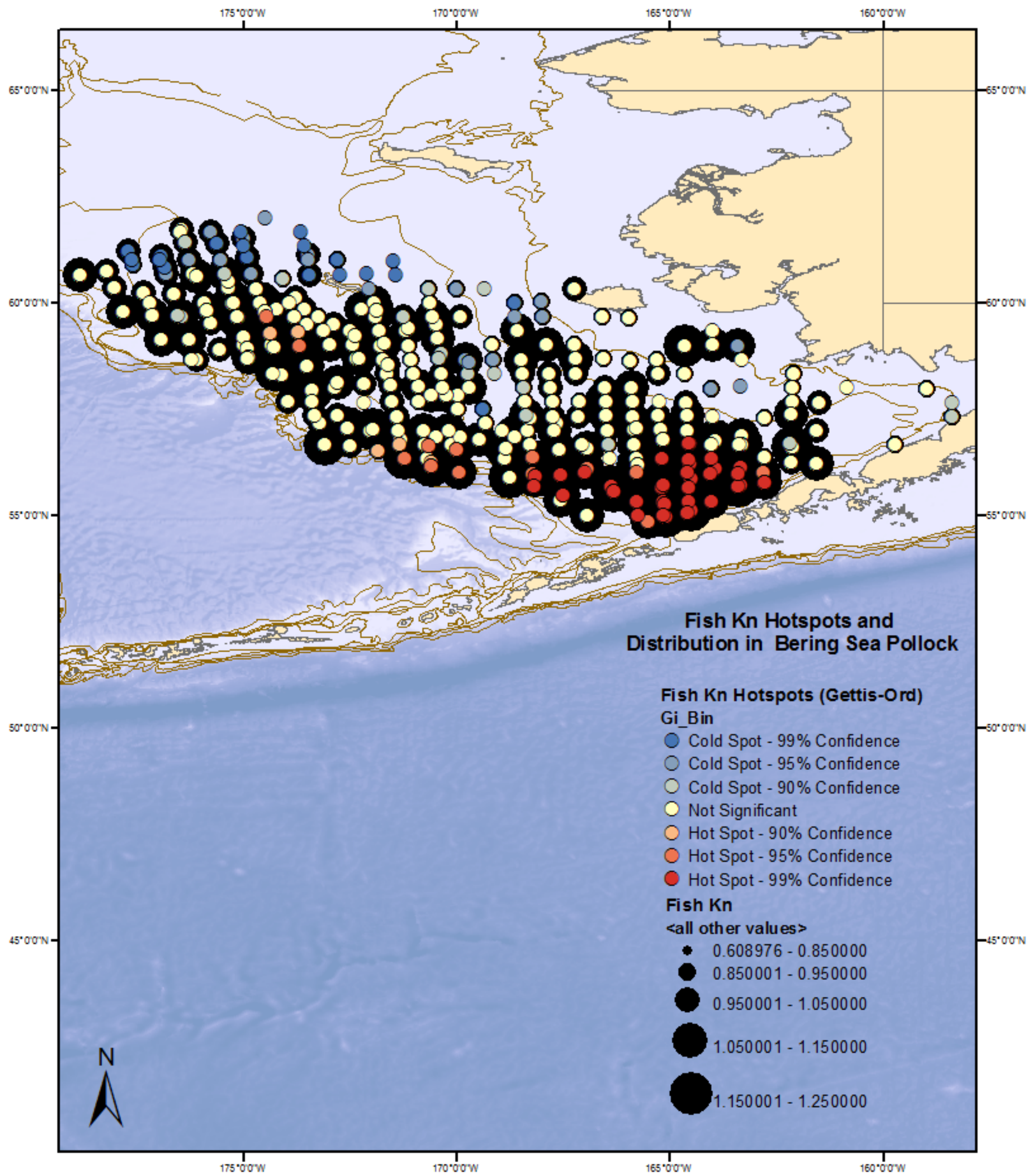

Figure 1. Spatial statistical analysis using Gettis-Ord (Getis et al. 1992) statistic which measures the concentration or lack of concentration (referred to as "hotspots") in a spatially ordered variable. Here walleye pollock fish condition is used to evaluate stratification in the EBS shelf survey for PLS analysis.

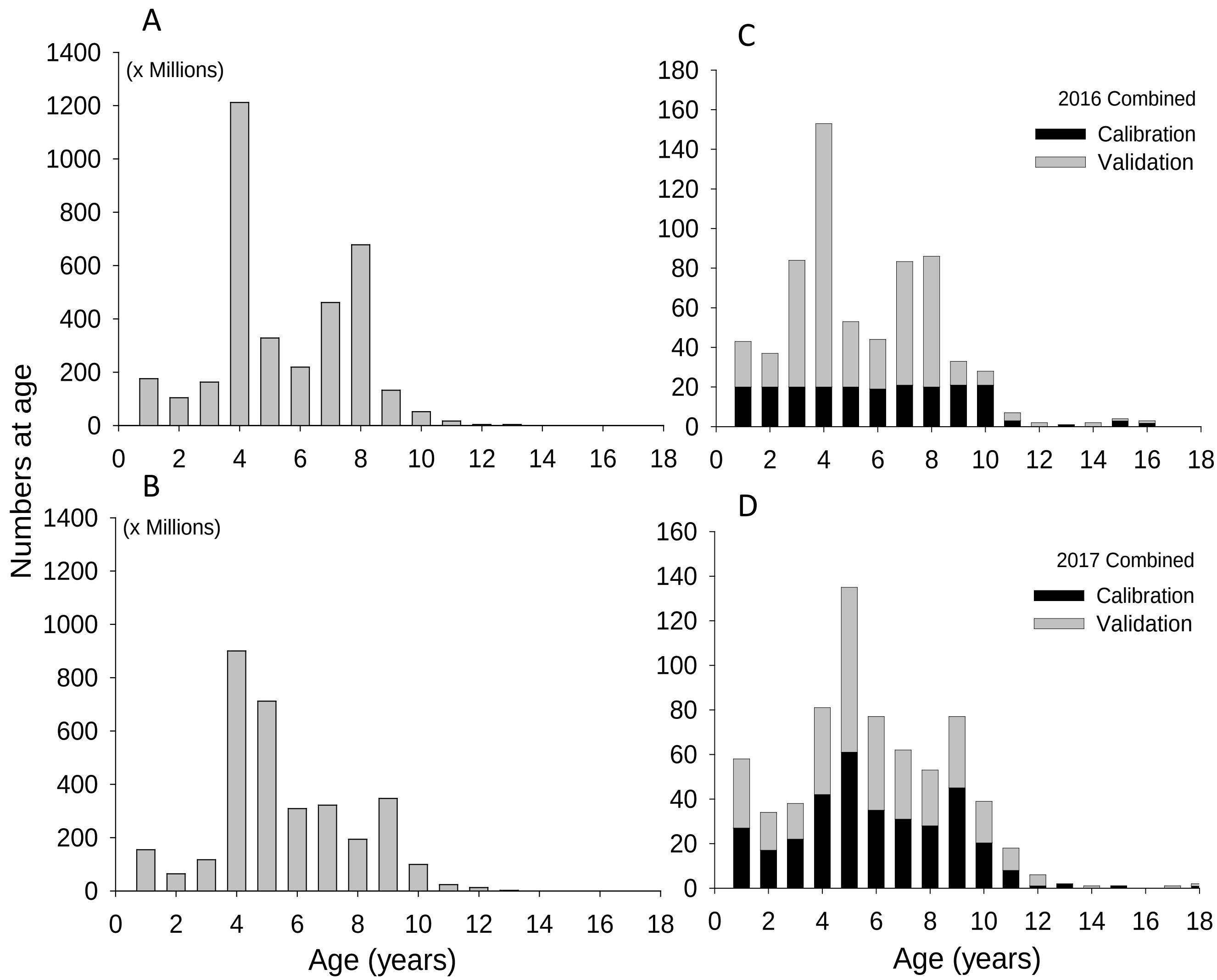

Figure 2. Eastern Bering Sea walleye pollock abundance by age in 2016 (A) and 2017 (B). A large 2012 year class as age-4 fish is evident. Numbers of 2016 (C) and 2017 (D) walleye pollock otoliths selected for FT-NIR analysis in the calibration and validation data sets.

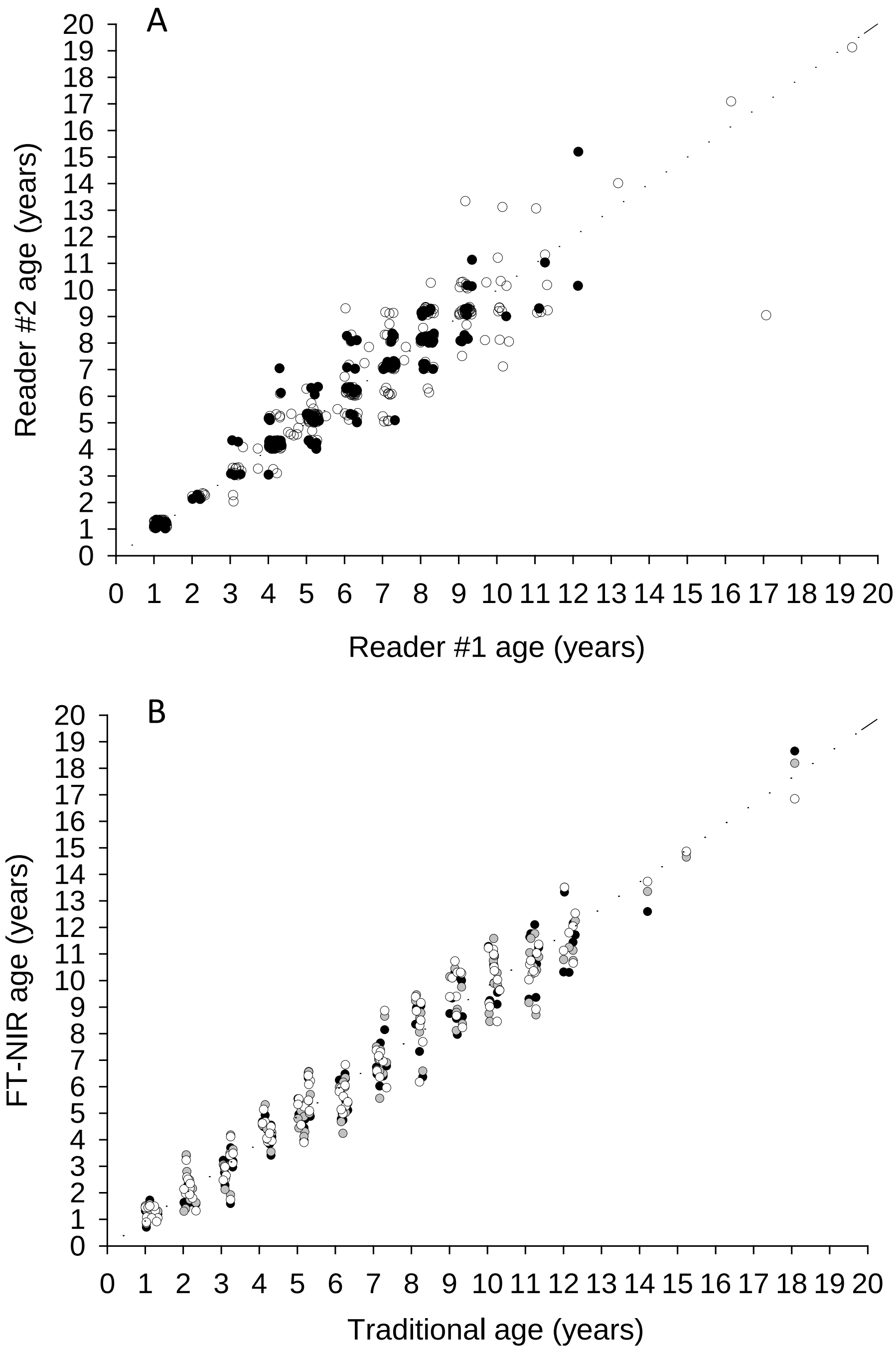

Figure 3. Walleye pollock ageing precision shown as age-to-age plots from traditional ageing methods (A) for 2016 (open circles) and 2017 (black circles); and triplicate FT-NIR spectroscopy ages associated with a single traditionally estimated age.

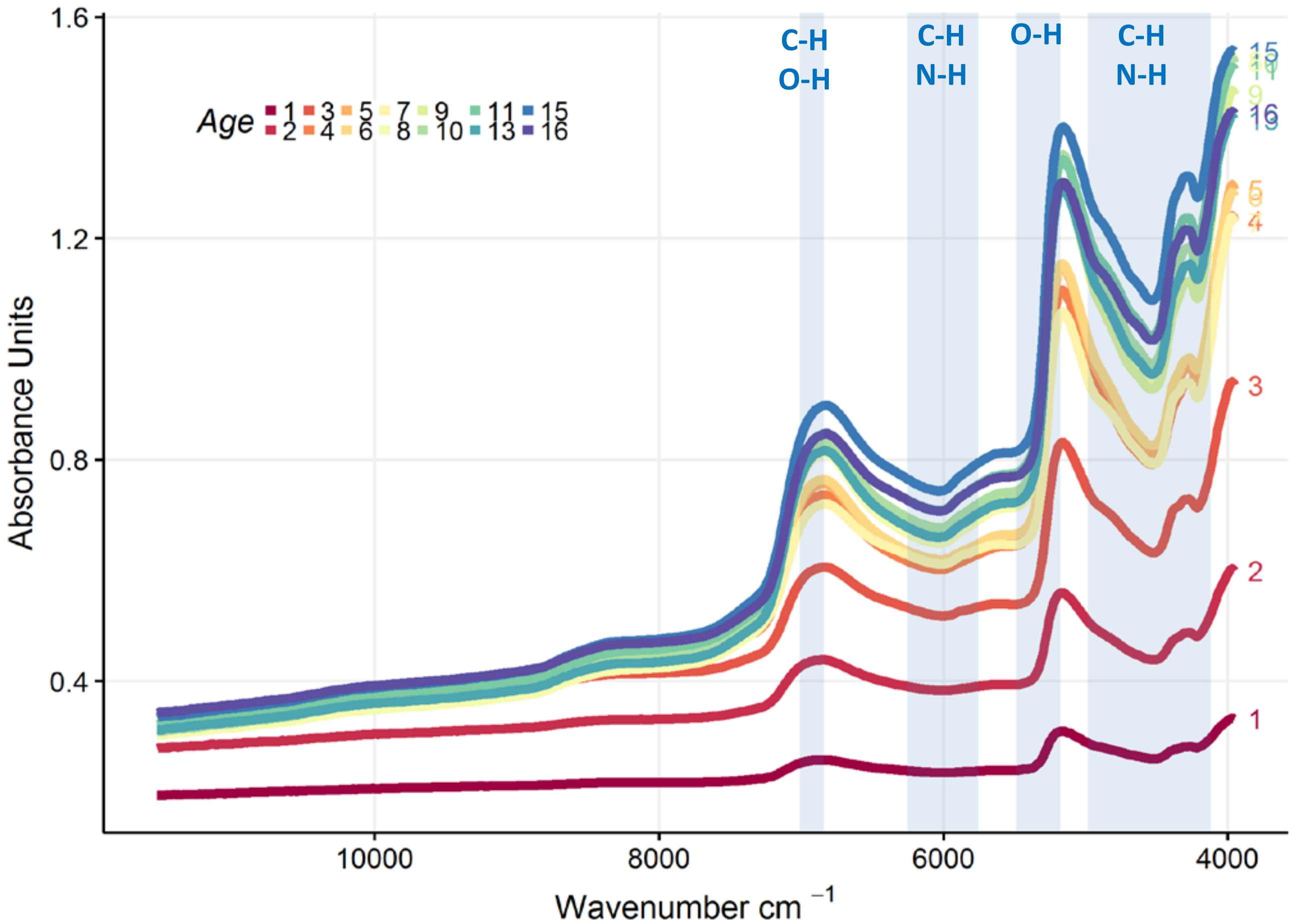

Figure 4. Selected spectral data for age 1 through age 12 EBS walleye pollock otoliths. Shaded regions of absorbance signatures by wavenumber show molecular vibrational group frequencies most relevant to age estimation from FT-NIR.

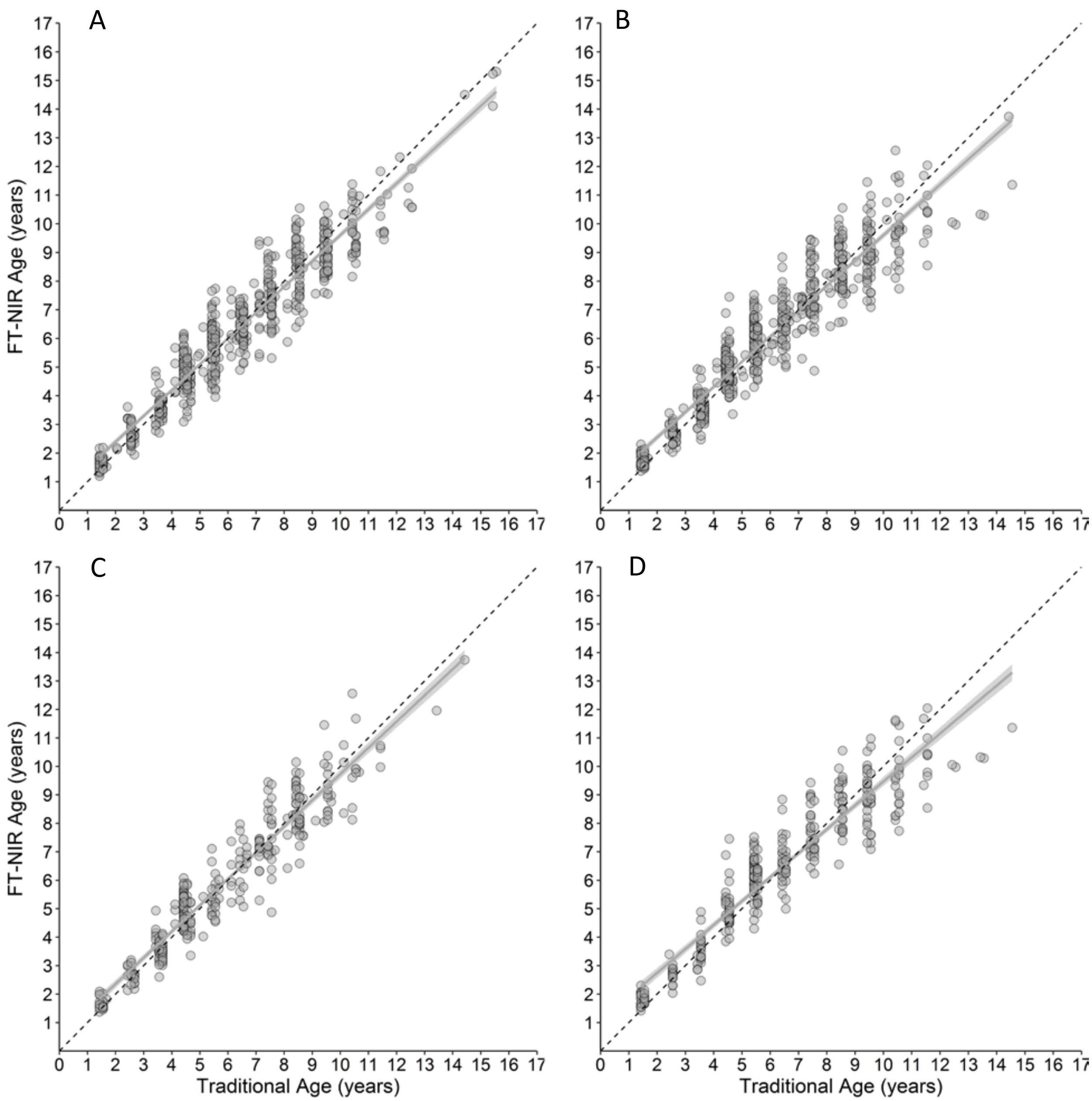

Figure 5. Calibration model (A) from 2016-2017 combined EBS walleye pollock otoliths showing predicted of FT-NIR spectroscopy ages from traditional age estimation. External validation of calibration model performance in shown for 2016-2017 data combined (B), and from 2016 (C) and 2017 (D) data separately.

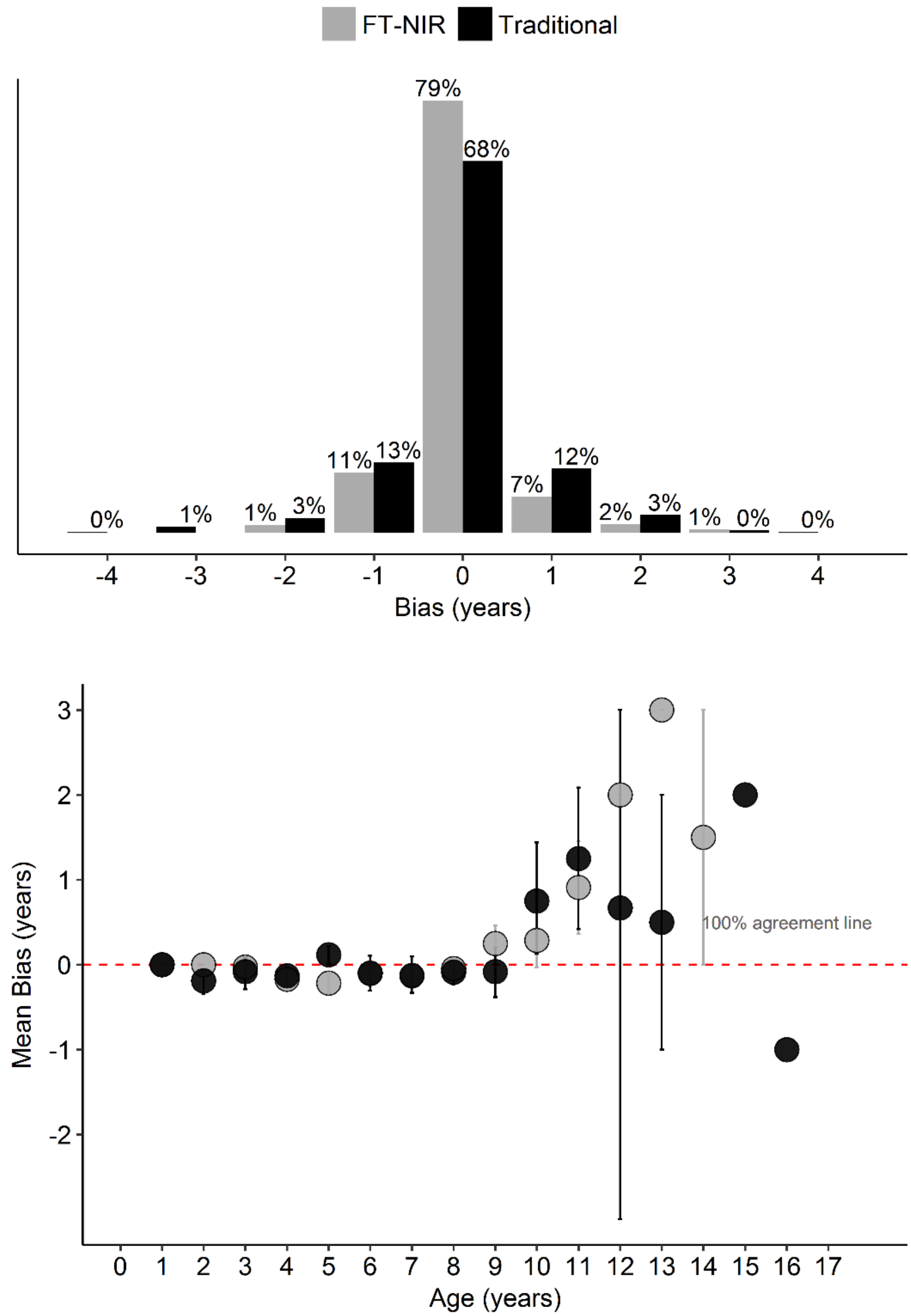

Figure 6. Comparison of accuracy between ages generated between the FT-NIR and traditional age estimation procedures by calculating the frequency of the differences (Bias = B); $B^{FT\text{-}NIR} = (Age^{Trad} – Age^{FT\text{-}NIR})$ and $B^{Trad} = (Age^{read1} – Age^{read2})$.