Age estimation of barramundi (Lates calcarifer) over multiple seasons from the southern Gulf of Carpentaria using FT-NIR spectroscopy

C. Wright A, B. B. Wedding B,D, S. Grauf B and O. J. Whybird C

A Rapid Assessment Unit, Horticulture and Forestry Science, Department of Agriculture and Fisheries, PO Box 1054, Mareeba, Qld 4880, Australia.
B Rapid Assessment Unit, Crop and Food Science, Department of Agriculture and Fisheries, PO Box 652, Cairns, Qld 4870, Australia.
C Fisheries Queensland, Department of Agriculture and Fisheries, PO Box 5396, Cairns, Qld 4870, Australia.
D Corresponding author. Email: brett.wedding@daf.qld.gov.au

Abstract. The age of whole otoliths from barramundi (Lates calcarifer) obtained from the southern Gulf of Carpentaria were estimated using Fourier transform near-infrared (FT-NIR) spectroscopy. Otoliths from 1716 barramundi collected in 2006, 2009 and 2012–2015 were used in this study. Partial least-squares regression models (PLS-R) and multiple linear regression models (MLR) were developed from the diffuse reflectance spectra and the age was obtained from traditional sectioned otoliths. Calibration models were built up over consecutive years (2012–2015) by using a subset of the samples and used to predict the age of the remaining samples and samples from the following year. Results suggest that when seasonal (temporal) variability is incorporated into the calibration model, FT-NIR has the ability to predict barramundi age (validation $R^2$ ranged from 0.73 to 0.78; RMSEP ranged from 6.92 to 7.64 months). The predicted age class was within 1 year of the reference age in over 96% of the samples. These models were also able to predict the age of otoliths from 2006 and 2009, which were retrieved from long-term storage (validation $R^2$ ranged from 0.77 to 0.84; RMSEP ranged from 8.66 to 10.88 months). The results from this study have shown the potential for barramundi from the southern Gulf of Carpentaria to be aged quickly and accurately by using FT-NIR.

Keywords: fish ageing, otolith, spectroscopy.

Received 8 October 2020, accepted 12 February 2021, published online 1 April 2021

Introduction

Data such as fish length, age and gender are collected annually as part of routine biological monitoring by many agencies. These data are often used to develop stock-specific parameters and for quantifying mortality and general population dynamics, which assist with determining the stock status of individual fisheries (Campana and Thorrold 2001; Ono et al. 2015; Streipert et al. 2019). Estimating the age structure of fish populations is an important component of sustainable fisheries management. This information is critical for age-structured stock assessments to identify whether a fishery is sustainable or if it is depleting. The most common method of ageing fish is to count the opaque bands in the ear bones, known as otoliths. The otoliths are primarily constructed of calcium carbonate and protein, which form concentric layers as the fish grows (Degens et al. 1969; Campana 1999; Hale and Swearer 2008). The periodicity of these concentric rings may be related to daily, seasonal or annual cycles and are thought to occur as a result of changes in many factors, including photoperiod, temperature, seasonal feeding and growth rates (Radtke and Shafer 1992; Secor et al. 1995; Chang and Geffen 2013). In many species, an otolith is age estimated by blocking it in resin, cutting a thin section (~250–500 μm), and then the section is mounted on a microscope slide for viewing under a microscope (Secor et al. 1992; Winkler et al. 2019). Although this approach is currently the most accepted method, it requires practice and experience to obtain accurate age estimates, is labour intensive, time consuming and expensive. Many otoliths are read several times, either by the same reader or independent readers, to verify the determined age. It is estimated that over 60 000 otoliths are collected and aged in Australia each year (Robins et al. 2015), New Zealand ages between 30 000–40 000 annually (Moore et al. 2019), whereas in the federally managed waters of Alaska, over 352 000 ages were estimated between 2009 and 2018 (Helset et al. 2019).

The traditional methods of ageing otoliths by counting the growth rings is known to contain an element of subjectivity (Campana 1999; Cardinale and Arthenius 2004). By developing a method that has increased efficiency and improved
repeatability, large savings with respect to time, resources and money could be made. One possible method is Fourier transform near-infrared spectroscopy (FT-NIR). Sir William Herschel is credited by many as discovering near-infrared (NIR) spectroscopy in 1800 (Herschel 1800), but it was not used for quantitative analyses until the 1960s (Norriss and Hart 1963). In more recent decades, FT-NIR has been used extensively in the food processing (Ozaki et al. 2003; Walsh et al. 2020), agricultural (Reeves et al. 1999), industrial and pharmaceutical sectors (Morisseau and Rhodes 1995; Burns and Ciurczak 2001). NIR spectroscopy is a non-destructive technique for determining chemical compositions through applying NIR light to the product which causes molecular vibrations. These vibrations occur at certain frequencies that coincide with the molecular grouping in the product being assessed (Murray and Williams 1987). As a secondary method of determination, and through the application of chemometric techniques, these responses to NIR light can be converted into simple calibration models relating to the property of interest (Blanco and Villarroya 2002). The developed calibration model can then be utilised to predict the property of interest of new samples. NIR spectroscopy has many advantages, including speed of sample throughput, a high level of repeatability and requires little sample preparation.

Recently FT-NIR has shown promise in ageing otoliths of saddletail snapper (Lutjanus malabaricus; Wedding et al. 2014), barramundi (Lates calcarifer) and snapper (Pagrus auratus; Robins et al. 2015), eastern Bering Sea walleye pollock (Gadus chalcogrammus; Helsel et al. 2019) and red snapper (Lutjanus campechanus; Passerotti et al. 2020a, b). FT-NIR has also been used to successfully age the vertebrae of hammerhead (Sphyra mokarran) and spot-tail sharks (Carcharhinus sorrah; Rigby et al. 2016), and to age the dorsal fin spines, vertebrae and fin clips of deepwater sharks (Squalus maggulops, Squalus montalbani; Rigby et al. 2014). Although these studies relating NIR spectra to the traditionally derived age estimate or increment count of the otolith or vertebrae have shown potential, there are limitations preventing the broad uptake of FT-NIR.

A possible limitation is the unknown explicit relationship between the FT-NIR spectra and the traditionally derived age. This relationship has been attributed to calcium carbonate (CaCO$_3$) and protein being deposited as the otolith is formed (Hale and Swearer 2008). Fish otoliths are composed of typically 90–96% CaCO$_3$, 0.01–10% organic matrix (protein complex) and ~1% non-organic trace elements (Campana 1999). However, the overall elemental composition of the otolith is influenced by many factors such as growth rate, seasonal cycles, environmental stress, salinity, temperature, food availability and reproductive stress (Radtke and Shafer 1992; Tabouret et al. 2011).

Many fish species have a complex life history, which may be expected to influence the ability of NIR spectroscopy to predict age with any precision. Barramundi are protandrous hermaphrodites, generally maturing as males at 2–5 years, and then becoming female at 5–7 years (Davis 1982; Saunders et al. 2018). Spawning occurs in saltwater, however barramundi can also survive in freshwater and estuarine habitats (Dunstan 1959), with the duration of time spent in the different aquatic habitats varying among individuals and among years (Millon et al. 2008; Crook et al. 2017). Freshwater flow is also known to influence the growth rates of barramundi, most likely because of fish exploiting improved access to temporary habitats and increased abundance of prey species (Robins et al. 2006; Milton et al. 2008). The irregular movement of barramundi among habitats and resulting changes in the water salinity are known to have an influence on the otolith microchemistry (Crook et al. 2017).

NIR spectroscopy may not be suitable to age otoliths for all fish species. As NIR is a secondary method of determination, any inaccuracies associated with the traditional method of ageing, will perpetuate through the NIR calibration model building process. Otoliths from some species cannot be accurately aged using increment counts (Beamish 1979). However, age validation studies on barramundi by Stuart and McKillup (2002) and Mc Dougall (2004) both suggested that sectioned otoliths can be used to predict the age of this species.

As part of the research by Robins et al. (2015), FT-NIR was applied to barramundi otoliths collected from the Archer River catchment in the southern Gulf of Carpentaria fishery in 2012. In that study, calibration models to predict the age of barramundi from otoliths were developed using samples collected in only one fishing season (February to October). The results presented in the present study are an extension of those from Robins et al. (2015) through the investigation of the temporal effects on barramundi ageing results from otoliths collected from 6 years over a 9-year period from the southern Gulf of Carpentaria. Age estimating calibration models were developed over four consecutive years (2012–2015) and were, subsequently, used to predict the age of 200 otoliths from the Archer River catchment that had been in long-term storage (2006 and 2009). The effect of predicting the age of samples for one fishing season using a calibration model developed using samples from a different season is discussed, along with the benefits of including temporal variability from multiple seasons. The overall advantages of using NIR spectroscopy in the ageing process is also discussed.

Materials and methods

Sample selection

Barramundi sagittal otoliths were collected from fish donated by commercial and recreational fishers and fish processors from the southern Gulf of Carpentaria genetic stock (Fisheries Queensland 2020; Department of Primary Industries and Fisheries 2005) in 2006, 2009, 2012, 2013, 2014 and 2015. The samples obtained in 2006, 2009 and 2012 are those used in the study by Robins et al. (2015). Where possible, both sagittal otoliths were collected, with one being used for traditional age estimation and the second being used to obtain the NIR spectra. As these fish were from wild stock and not of known age, strict protocols and quality control measures were employed by the laboratory to ensure a high quality of age predictions for all samples (Fisheries Queensland 2020). Reference age estimation was based on the traditional method of increment counts from sectioned otoliths (350 μm) examined under a microscope, but also taking into account the collection date, edge assessment and nominal birth date (Fisheries Queensland 2020). This allowed the age to be expressed in months or assigned to age class (number of birthdays a fish is assumed to have had) or age group (maximum age a fish would reach during a sampling season).
Data were collected from a total of 1716 otoliths in this study. Sample numbers and reference ages varied among the years (Table 1), with the median age ranging from 50 to 87 months. The distribution of the reference ages within each year was predominately skewed to the right, with a higher proportion of younger fish being sampled.

The management of the southern Gulf of Carpentaria fishery imposes a 120 cm maximum total length catch limit, which is usually reached by 11 years of age. As the samples used in this study were donated by commercial and recreational fishers and fish processors, there were few samples for older fish. Only 2% of otoliths across all years had a reference age greater than 120 months; therefore, the samples included in this research were restricted to otoliths with a reference age of 120 months or less.

The animal remains used in this study and that of Robins et al. (2015) were donated and the life and death of all animals were not altered as a result of the subsequent scientific use of the remains. Therefore, under Queensland Government policy, no animal ethics committee approval was required (https://www.business.qld.gov.au/industries/farms-fishing-forestry/agriculture/livestock/animal-welfare/animals-science/activities/dead-animals).

**NIR spectra acquisition**

Robins et al. (2015) found that calibration statistics for models based on barramundi otoliths collected from the Archer River catchment in the southern Gulf of Carpentaria stabilised after 6 months of storage. For consistency, all samples from 2012 to 2015 were scanned ~12 months after the otoliths were collected and are referred to as ‘fresh’ samples. The otoliths collected in 2006 and 2009 were not scanned until 2013. These otoliths are referred to as ‘historical’ samples, and after removal from the fish, they were dried and then stored at 25°C until the NIR spectra was collected.

The whole dried otolith from each barramundi that was not processed for traditional ageing, was scanned using a Bruker Multi-Purpose Analyser (MPA), FT-NIR spectrophotometer (Bruker Optics, Ettlingen, Germany; Bruker Optics operating software: OPUS v.6.5), with an integrating sphere in diffuse reflectance mode. The full wavelength range was from 12 500 to 3597 cm⁻¹ and in obtaining each sample spectrum, 16 scans at a resolution of 8 cm⁻¹ were collected and averaged. For consistency, all otoliths were scanned with a concave-up orientation (Fig. 1). A typical absorbance spectrum for a barramundi otolith from the southern Gulf of Carpentaria is shown in Fig. 2.

### Table 1. Summary of samples available for data analysis

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of otoliths</th>
<th>Age (months)</th>
<th>Mean age (months) (s.d.)</th>
<th>Median age (months) (IQR)</th>
<th>Total length (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>100</td>
<td>27–135</td>
<td>80.8 (21.8)</td>
<td>87 (12)</td>
<td>570–1070</td>
</tr>
<tr>
<td>2009</td>
<td>100</td>
<td>26–146</td>
<td>59.1 (24.0)</td>
<td>50 (24)</td>
<td>610–950</td>
</tr>
<tr>
<td>2012</td>
<td>100</td>
<td>28–148</td>
<td>63.0 (20.4)</td>
<td>64 (24)</td>
<td>610–960</td>
</tr>
<tr>
<td>2013</td>
<td>404</td>
<td>17–247</td>
<td>63.1 (24.1)</td>
<td>61 (14)</td>
<td>590–1210</td>
</tr>
<tr>
<td>2014</td>
<td>553</td>
<td>28–230</td>
<td>59.3 (25.6)</td>
<td>52 (27)</td>
<td>538–1216</td>
</tr>
<tr>
<td>2015</td>
<td>459</td>
<td>28–174</td>
<td>58.9 (16.5)</td>
<td>52 (14)</td>
<td>570–1182</td>
</tr>
</tbody>
</table>

**Fig. 1.** Concave-up orientation of an otolith on the integrating sphere window of the Bruker MPA.

**Fig. 2.** Typical absorbance spectra for a barramundi otolith sampled from the southern Gulf of Carpentaria.

**Data analysis**

Calibration models were developed to predict the reference age of otoliths on the basis of the FT-NIR spectra using partial least-squares regression (PLS-R) and multiple linear regression (MLR). The PLS-R models used several wide bands of wavelengths, whereas the MLR models were developed using a small number of individual wavelengths. Prior to model development, all spectra were pre-processed using a Savitzky-Golay (SG) smooth (25-point, second-order polynomial moving average) and a first derivative transformation (25-point SG smooth, second-order polynomial) to enhance the spectral features.
For both PLS-R and MLR, calibration models were developed from fresh samples (2012–2015) and built up over consecutive years. A base calibration model involving 2012 samples was initially developed and was used to predict the age of all fresh samples collected in 2013. Owing to the small number of fresh samples with spectra available in 2012, all samples less than 120 months were included in the development of the calibration models. The calibration model was then updated by including 100 fresh samples from 2013. The samples not used to develop the calibration model are referred to as the validation set. The samples from 2013, which were added to the calibration set, were selected on the basis of the total length of the fish. The range of total lengths was divided into classes of size 20 mm and, where possible, similar numbers of fish were selected randomly from each size class. This approach was adopted because (1) it is expected that the distribution of the total lengths, and subsequently the fish ages, would be closer to a box-car distribution than a non-uniform distribution and (2) in future, the age of the samples would not be known until they are aged using traditional techniques; hence, selection would need to be based on another attribute such as total length.

A calibration model was then developed using samples from 2012 and 2013 and this was used to predict the age of the 2013 validation set and all the 2014 fresh samples. This process was again repeated until the calibration model included samples from 2012 through to 2015. This combined fresh samples model was then used to predict the age of the 2013–2015 validation set. The combined fresh calibration model was also used to predict the age of historical samples collected in 2006 and 2009. No samples from these 2 years were included in the development of the calibration models. This systematic approach was taken to investigate the effect of predicting the age of samples for a different fishing season (year) and to highlight the benefits of including temporal (annual) variability.

Initially, all calibration models were developed using the same wavelength regions in the PLS-R and the same individual wavelengths in the MLR. The combined 2012–2015 calibration model was revised by selecting new wavelengths that reflect the variability introduced into the calibration set.

Segmented cross-validation, which involves deleting segments (or groups) of samples, was used to assess the PLS-R calibration models. Due to the smaller sample size for the 2012 model, full (leave-one-out) cross-validation was used. Leverage correction was used to assess the MLR models. Model performance was based on (1) the adjusted coefficient of determination ($R^2$) for the calibration and validation sets, (2) root mean-square error of cross-validation (RMSECV), (3) root mean-square error of prediction (RMSEP), (4) bias (average difference between predicted and reference values), (5) standard deviation ratio (SDR) calculated as the ratio of the standard deviation to RMSECV or RMSEP (Walsh et al. 2004), and (6) the ratio of performance to inter-quartile range (RPIQ) calculated as the ratio of the inter-quartile range to RMSECV or RMSEP (Bellon-Maurel et al. 2010). For skewed distributions, the standard deviation is not a suitable measure to describe the spread of the population and it is recommended to use the inter-quartile range (Bellon-Maurel et al. 2010). As the majority of validation sets follow a skewed distribution, the RPIQ is considered an appropriate measure of prediction accuracy. In general, better model performance is indicated by larger values of $R^2$, smaller RMSE values, and larger SDR and RPIQ values.

For models of stock assessment and population dynamics, the age class is the most important estimate of age. The age class for each sample in the PLS-R and MLR validation sets was calculated using the approach of Robins et al. (2015), by rounding down the predicted age from decimal years to a whole number. For example, a fish with a reference age of 7.0 years and predicted age of 7.9 years falls into the same age class of 7 years. Contrary to this, a fish with a reference age of 7.0 years and predicted as 6.9 years, will be placed in age class 6 years for the NIR predicted age, not age class 7 years.

The relative bias was investigated and calculated as the difference between the FT-NIR predicted age and the reference age (predicted age – reference age). The average percentage error (APE) between the two methods was calculated using the R package ‘FSA’ (Ogle et al. 2020). All spectroscopic data analysis was performed using the chemometrics software package ‘The Unscrambler’ v.10.5 (Camo, Oslo, Norway).

Results

Results for the PLS-R and MLR calibration models and subsequent predictions are shown in Tables 2 and 3 respectively. Two samples were excluded from the 2012 dataset as they had a reference age of 148 months, which was outside the predetermined range of the calibration model (≤120 months). The predicted ages for fresh samples in 2013 obtained from the PLS-R and the MLR calibration models based on the 98 otoliths in the 2012 fresh sample dataset performed extremely poorly. This is most likely to be due to the lack of spatial and temporal variability across the samples in the calibration model. Of the 393 samples in the corresponding validation set, the PLS-R resulted in 275 with a deviation greater than 30 months (2.5 years). The deviation is estimated as a function of the overall model error, sample leverage and the sample residual variance (De Vries and Ter Braak Cajo 1995). The predicted age associated with a large deviation is considered unreliable and it would be recommended that these samples be aged using traditional techniques to verify the prediction.

By adding samples from different years into the calibration model, the biological and temporal variability was increased. Tables 2 and 3 show that this led to an improvement in the predictive ability of the models. The final calibration model included approximately equal numbers of fresh samples from 2012 to 2015. Three latent variables were used in the PLS-R model, resulting in a validation $R^2$ of 0.76 with RMSEP of 7.29 months when predicting the 2013–2015 validation set. The MLR model produced similar validation statistics with an $R^2$ of 0.73 and RMSEP of 7.64 months.

The deviation associated with the 1083 predicted ages from the 2012–2015 PLS-R model had a median of 5.4 months. Only 14 predictions had a deviation greater than 30 months, with the traditional ages for these samples ranging from 27 to 99 months. In comparison, the 2012 calibration model that predicted the 2013 samples resulted in 70% of the predicted ages having a deviation greater than 30 months, suggesting unreliable age predictions. Similarly, the MLR predictions of the 2012–2015 validation set had predicted ages with a median deviation of 7.9
months and the same 14 samples had a deviation greater than 30 months. These samples included three from 2013, seven from 2014 and four from 2015.

The combined 2012–2015 models were also used to predict the ages of historical samples collected in 2006 and 2009. The PLS-R and MLR models performed similarly with RMSEP of 9.69 and 10.38 months and validation $R^2$ of 0.79 and 0.77 respectively.

Predictions from validation data showed considerable scatter around the 1:1 line for both the PLS-R (Fig. 3a) and MLR models (Fig. 3b). This level of scatter becomes less important when the ages are expressed as age classes. Figure 4 shows histograms of the reference ages and the predicted ages of the validation sets for the combined 2012–2015 PLS-R and MLR models. Predicted ages are shown for each year in the validation set (2013, 2014, 2015), as well as the predicted ages combined. Samples from only the validation set are included in these figures.

The difference between the predicted age class and the reference age class (predicted age class – reference age class) provides a measure of the relative bias. Of the 1083 samples in the combined 2012–2015 validation set, 691 (63.8%) were predicted by the PLS-R model to be in the same age class as the reference age and 366 (33.8%) were predicted to be in the adjacent age class, either one less than or one greater than the reference age (Fig. 5a). For the MLR model, 638 (58.9%) of the samples in the validation set were predicted in the correct age class, with a total of 1042 (96.2%) within one age class of the reference age class. The largest bias was a single otolith in the validation set that had a reference age of 8 years and was predicted by the MLR model to have a reference age of 4 years. The PLS-R model also underestimated the age of this sample by predicting the age as 6 years. The percentage agreement for the predicted age class was lower than that obtained by two readings by using traditional methods of a set of quality assurance samples. The routine biological monitoring program aged up to 800 fish annually from the southern Gulf of Carpentaria stock for the years included in this study. Of these, 200 otoliths each year were reread by the same reader for bias and precision testing of increment count and edge category

<table>
<thead>
<tr>
<th>Calibration data</th>
<th>Validation data</th>
<th>LV</th>
<th>$n$</th>
<th>s.d.</th>
<th>$R^2$</th>
<th>RMSECV</th>
<th>RMSEP</th>
<th>Bias</th>
<th>SDR</th>
<th>RPIQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td></td>
<td>3</td>
<td>98</td>
<td>16.5</td>
<td>0.89</td>
<td>5.48</td>
<td>-0.06</td>
<td>3.0</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td>3</td>
<td>393</td>
<td>15.9</td>
<td>0.72</td>
<td>8.36</td>
<td>2.08</td>
<td>1.9</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>2012–2013</td>
<td></td>
<td>3</td>
<td>198</td>
<td>17.9</td>
<td>0.83</td>
<td>6.02</td>
<td>0.01</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td>3</td>
<td>293</td>
<td>14.4</td>
<td>0.84</td>
<td>5.77</td>
<td>0.32</td>
<td>2.5</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td>4</td>
<td>538</td>
<td>17.7</td>
<td>0.65</td>
<td>10.43</td>
<td>-7.64</td>
<td>1.7</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>2012–2014</td>
<td></td>
<td>4</td>
<td>298</td>
<td>19.0</td>
<td>0.87</td>
<td>6.93</td>
<td>0.01</td>
<td>2.7</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>2013–2014</td>
<td></td>
<td>4</td>
<td>730$^a$</td>
<td>15.5</td>
<td>0.80</td>
<td>7.05</td>
<td>0.26</td>
<td>2.2</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td>4</td>
<td>455</td>
<td>14.3</td>
<td>0.69</td>
<td>8.07</td>
<td>-0.37</td>
<td>1.8</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>2012–2015</td>
<td></td>
<td>3</td>
<td>398</td>
<td>18.9</td>
<td>0.84</td>
<td>7.65</td>
<td>-0.01</td>
<td>2.5</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>2013–2015</td>
<td></td>
<td>3</td>
<td>1083</td>
<td>14.7</td>
<td>0.76</td>
<td>7.29</td>
<td>0.35</td>
<td>2.0</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>2006, 2009</td>
<td></td>
<td>3</td>
<td>188</td>
<td>21.5</td>
<td>0.79</td>
<td>9.69</td>
<td>6.28</td>
<td>2.2</td>
<td>3.8</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Two outliers were removed.

<table>
<thead>
<tr>
<th>Calibration data</th>
<th>Validation data</th>
<th>$n$</th>
<th>s.d.</th>
<th>$R^2$</th>
<th>RMSECV</th>
<th>RMSEP</th>
<th>Bias</th>
<th>SDR</th>
<th>RPIQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td></td>
<td>98</td>
<td>16.5</td>
<td>0.87</td>
<td>5.89</td>
<td>-0.03</td>
<td>2.8</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td>393</td>
<td>24.1</td>
<td>0.00</td>
<td>45.41</td>
<td>-44.41</td>
<td>0.4</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>2012–2013</td>
<td></td>
<td>197</td>
<td>17.9</td>
<td>0.86</td>
<td>6.62</td>
<td>-0.01</td>
<td>2.7</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td>293</td>
<td>25.6</td>
<td>0.75</td>
<td>7.22</td>
<td>-0.06</td>
<td>2.0</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td>538</td>
<td>25.6</td>
<td>0.69</td>
<td>9.88</td>
<td>-3.58</td>
<td>1.8</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>2012–2014</td>
<td></td>
<td>296</td>
<td>18.9</td>
<td>0.86</td>
<td>6.99</td>
<td>-0.003</td>
<td>2.7</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>2013–2014</td>
<td></td>
<td>731</td>
<td>26.1</td>
<td>0.77</td>
<td>7.47</td>
<td>-0.42</td>
<td>2.1</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td>357</td>
<td>17.5</td>
<td>0.71</td>
<td>7.71</td>
<td>-2.58</td>
<td>2.1</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>2012–2015</td>
<td></td>
<td>397$^a$</td>
<td>18.8</td>
<td>0.83</td>
<td>7.80</td>
<td>-0.01</td>
<td>2.4</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>2013–2015</td>
<td></td>
<td>1083</td>
<td>24.4</td>
<td>0.73</td>
<td>7.64</td>
<td>-0.38</td>
<td>1.9</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>2006, 2009</td>
<td></td>
<td>188</td>
<td>25.3</td>
<td>0.77</td>
<td>10.38</td>
<td>0.92</td>
<td>2.1</td>
<td>3.6</td>
<td></td>
</tr>
</tbody>
</table>

$^a$One outlier was removed.
Fig. 3. Predictions for the 2012–2015 validation data from the generic combined years (a) PLS-R and (b) MLR calibration models. Symbols represent the difference between the predicted age and the reference age: within ±6 months (●); within ±12 months (○); and more than 12 months (△).

Fig. 4. Histogram of the reference age class and predicted age class of the validation set for the generic combined 2012–2015 PLS-R and MLR models. (a) 2013 validation samples, (b) 2014 validation samples, (c) 2015 validation samples and (d) overall predicted age distribution.
For the years in this study, the percentage agreement between the age class resulting from the first read and the re-read was more than 92% for each year. The relative bias of the models was also investigated by considering age expressed in months. For the combined 2012–2015 validation set, the PLS-R and MLR models predicted 687 (63.4%) and 669 (61.8%) of the 1083 samples within ±6 months of the reference age, and 987 (91.1%) and 975 (90.0%) within ±12 months respectively (Fig. 5).

The APE between the predicted age class and the traditionally derived age class for the validation samples was 4.6% and 5.4% for the generic PLSR and MLR models respectively. These values are higher than the APE of age class for the re-reads of the quality assurance samples for the biological monitoring program testing, which was <1% for each year used in this study. The PLS-R wavelength regions and MLR individual wavelengths used in the calibration models in Tables 2 and 3 were revised for the combined 2012–2015 calibration models. It could be expected, with the addition of samples from different years, that an adjustment to wavelengths used in the calibration models would be required. Summary statistics for the revised combined years calibration model and resulting validation set predictions are shown in Table 4. A minor reduction of less than 1 month was observed in the RMSECV and RMSEP for both the PLS-R and the MLR calibration and prediction models. The RMSEP for the PLS-R and MLR models from predicting the historical samples improved by more than 1 month. Although this was only a minor improvement, it highlighted the importance of regular model maintenance.

The histogram of the overall reference ages and the validation set predicted ages from the revised combined 2012–2015 PLS-R and MLR models followed a trend similar to that of the original generic model (Fig. 6a), as does the histogram of bias (Fig. 6b). The 8-year-old sample that was incorrectly aged by 4 years using the generic MLR, was underestimated by 2 years in the revised MLR model. Within predictions from the revised PLS-R calibration, 681 (62.9%) of the 1083 samples in the validation set were classified in the same age class as the reference age class, with a total of 1056 (97.5%) within one age class. The revised MLR age class predictions were similar, with 679 (62.7%) of the 1083 samples in the validation set classified in the same age class as the reference age class.

(Fisheries Queensland 2020). For the years in this study, the percentage agreement between the age class resulting from the first read and the re-read was more than 92% for each year.

The relative bias of the models was also investigated by considering age expressed in months. For the combined 2012–2015 validation set, the PLS-R and MLR models predicted 687 (63.4%) and 669 (61.8%) of the 1083 samples within ±6 months of the reference age, and 987 (91.1%) and 975 (90.0%) within ±12 months respectively (Fig. 5).

The APE between the predicted age class and the traditionally derived age class for the validation samples was 4.6% and 5.4% for the generic PLSR and MLR models respectively. These values are higher than the APE of age class for the re-reads of the quality assurance samples for the biological monitoring program testing, which was <1% for each year used in this study.

The PLS-R wavelength regions and MLR individual wavelengths used in the calibration models in Tables 2 and 3 were revised for the combined 2012–2015 calibration models. It could be expected, with the addition of samples from different years, that an adjustment to wavelengths used in the calibration models would be required. Summary statistics for the revised combined years calibration model and resulting validation set predictions are shown in Table 4. A minor reduction of less than 1 month was observed in the RMSECV and RMSEP for both the PLS-R and the MLR calibration and prediction models. The RMSEP for the PLS-R and MLR models from predicting the historical samples improved by more than 1 month. Although this was only a minor improvement, it highlighted the importance of regular model maintenance.

The histogram of the overall reference ages and the validation set predicted ages from the revised combined 2012–2015 PLS-R and MLR models followed a trend similar to that of the original generic model (Fig. 6a), as does the histogram of bias (Fig. 6b). The 8-year-old sample that was incorrectly aged by 4 years using the generic MLR, was underestimated by 2 years in the revised MLR model.

Within predictions from the revised PLS-R calibration, 681 (62.9%) of the 1083 samples in the validation set were classified as being in the same age class as the reference age class, with a total of 1056 (97.5%) within one age class. The revised MLR age class predictions were similar, with 679 (62.7%) of the 1083 samples in the validation set classified in the same age class as the reference age class.

### Table 4. Model statistics for models developed from a revised set of wavelengths

<table>
<thead>
<tr>
<th>Model</th>
<th>Calibration data</th>
<th>Validation data</th>
<th>LV</th>
<th>n</th>
<th>s.d.</th>
<th>$R^2$</th>
<th>RMSECV</th>
<th>RMSEP</th>
<th>Bias</th>
<th>SDR</th>
<th>RPIQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS-R</td>
<td>2012–2015</td>
<td>4</td>
<td>397</td>
<td>18.9</td>
<td>0.86</td>
<td>7.11</td>
<td>-0.005</td>
<td>2.7</td>
<td>3.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2013–2015</td>
<td>4</td>
<td>1083</td>
<td>14.7</td>
<td>0.78</td>
<td>6.97</td>
<td>-0.21</td>
<td>2.1</td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2006, 2009</td>
<td>4</td>
<td>188</td>
<td>21.5</td>
<td>0.84</td>
<td>8.66</td>
<td>4.57</td>
<td>2.5</td>
<td>4.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLR</td>
<td>2012–2015</td>
<td>7</td>
<td>397</td>
<td>18.9</td>
<td>0.86</td>
<td>6.94</td>
<td>0.003</td>
<td>2.7</td>
<td>3.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2013–2015</td>
<td>7</td>
<td>1083</td>
<td>14.7</td>
<td>0.78</td>
<td>6.92</td>
<td>0.02</td>
<td>2.1</td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2006, 2009</td>
<td>7</td>
<td>188</td>
<td>21.5</td>
<td>0.83</td>
<td>8.73</td>
<td>3.70</td>
<td>3.1</td>
<td>4.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aOne outlier was removed.*
the reference age class and a further 379 (35.0%) in the age class one different from the reference age. The APE was 4.7% for the two revised models.

Both the revised and generic MLR calibration models included seven individual wavelengths. The seven wavelengths selected in the revised model in Table 4 were in the region from 7255 cm\(^{-1}\) to 4140 cm\(^{-1}\), which is a region similar to those selected for the generic models in Table 3, from 7460 cm\(^{-1}\) to 3930 cm\(^{-1}\). The main spectral regions selected in the combined 2012–2015 PLS-R calibration models were similar for the generic and revised models (Fig. 7), although the region from 7190 to 6640 cm\(^{-1}\) was not in the generic PLS-R model.

Discussion

The potential use of FT-NIR to predict the age of otoliths has been demonstrated in several recent publications (Wedding et al. 2014; Helser et al. 2019; Passerotti et al. 2020a, 2020b), including for barramundi (Robins et al. 2015). The results found in the present study, using barramundi otoliths from multiple seasons, are comparable to those of previous studies and to the traditional method of aging barramundi through reading sectioned otoliths. The generic PLS-R combined 2013–2015 calibration model predicted 97.6% of samples in the validation set to be within ±1 year, and the generic MLR model predicted 96.2% within ±1 year. These results are comparable with the barramundi age validation study of Mc Dougall (2004) who assigned 97% of samples to within 1 year of the known-age class and obtained an \(R^2\) of 0.89 between the estimated age from the sectioned otoliths and the known ages. Robins et al. (2015) applied FT-NIR to age ‘fresh’ barramundi caught in 2012 under 120 months of age from the Gulf of Carpentaria and central Queensland. The validation statistics for the Archer River catchment in the southern Gulf of Carpentaria were \(R^2 = 0.88\) and RMSEP = 5.9 months. The FT-NIR-predicted age classes were the same as the reference age class for 73% of samples and within ±1 year for 94% of samples. Helser et al. (2019) who applied FT-NIRS to eastern Bering Sea walleye pollock (Gadus chalcogrammus) reported a model combining more than 1500 samples from 2 years having \(R^2 = 0.89\), with RMSEP = 0.96 years. Helser et al. (2019) also reported that 75% of the FT-NIR-predicted ages were the same as the results from traditional ageing techniques, with this increasing to 94% for ages within ±1 year. Similar results were obtained by Wedding et al. (2014) who developed models for predicting increment counts of saddletail snapper (Lutjanus malabaricus) collected over a post-wet and post-dry seasons. They achieved validation statistics of \(R^2 = 0.94\) with RMSEP = 1.54 increments. The percentage agreement for age class in the study of red snapper (Lutjanus campechanus) by Passerotti et al. (2020a) was −45% for the combined region models, increasing to nearly 90% of samples within ±1 year of the reference age.

As the fish used in this study were not of known age, the predictive ability of the calibration models can be only as good...
as the accuracy of the ages derived from the traditional methods. An age validation study by Stuart and McKillup (2002) on barramundi with known ages from 1 to 5 years found that estimated age from sectioned otoliths produced an APE of 5.8%. Mc Dougall (2004) obtained an APE of 3.2% for barramundi with known ages up to 8.6 years. The ageing laboratory used in the present study, had a between read APE of <1% with percentage agreement of >92% of age class for a set of quality assurance samples within each year. In the present study, the APE between the FT-NIR-predicted age class and the traditionally derived age class was 4.6% and 5.4% for the generic PLSR and MLR validation sets respectively, and 4.7% for the validation sets in the two revised models. An APE below 5% is considered acceptable in reading accuracy and obtainable by experienced readers (Robertson and Morison 1999). The acceptability of the observed difference between the FT-NIR-predicted age and the traditionally derived age will depend on the purpose for which the age is required.

Some differences in the predicted age and the reference age may be attributed to the method of converting the predicted age in months to an age class. In this study, the predicted age in decimal years was truncated (rounded down) such that the integer of the predicted age corresponded with the age class. Alternative options are possible, such as conventional rounding, as used by Passerotti et al. (2020a), or taking into account the capture and spawning dates (birth date). This is an area that requires further research and may lead to improved FT-NIR age class predictions.

Studies by Wedding et al. (2014), Robins et al. (2015), Helser et al. (2019) and Passerotti et al. (2020a) all showed indications that, for some fish species, the predicted age of older otoliths was under-estimated. In all of these studies, the sample size for older fish was limited. Wedding et al. (2014) found that saddletail snappers with more than 18 increment counts were under-estimated, and models developed by Helser et al. (2019) showed a positive bias for eastern Bering Sea walleye pollocks that were greater than 10 years of age. Similarly, models developed by Passerotti et al. (2020a) under-estimated the ages of the oldest red snapper otoliths by as much as 22%. There was also a general tendency for the older barramundi otoliths to be under-estimated by Robins et al. (2015); however, in the same study, this positive bias was not observed for Gulf of St Vincent snapper. The under-estimation of older barramundi was also observed in the present study. From a modelling perspective this may suggest that separate models are required for young and old fish. The biological process that may be contributing to the under-estimation of older-age fish by FT-NIR is unknown. The deposition of chemical elements in the otolith is influenced by both intrinsic and extrinsic factors (Chang and Geffen 2013). Passerotti et al. (2020a) suggested that light penetration may be a factor in the under-estimation of the older samples. Further research is required to understand the changes in otolith microchemistry that occur as the fish ages and how this may interact with NIR spectra.

Given the complex life history of many species, such as barramundi, and the effect of the environmental conditions on the otolith composition, it is not unexpected that many of the previous studies using NIR to age otoliths have shown the presence of spatial and temporal variability (Wedding et al. 2014; Robins et al. 2015; Helser et al. 2019; Passerotti et al. 2020a). The presence of spatial and temporal variability leads to the need for ongoing model maintenance. Periodic model maintenance is an important component in the use of NIR models because they are not perpetual models (Mercader and Puigdomènech 2014). The process of model maintenance should not be viewed as a disadvantage of the method. Model maintenance is a form of continually improving knowledge on the factors influencing spectral variability and optimising the modelling process. The spectral and reference data used to develop the NIR calibration models may vary over time due to temporal effects, genetic changes, sample preparation changes, refinements in the reference method and spectrometer effects such as instrument drift and changes in the signal to noise ratio. These changes can have an impact on the validity of the calibration models to predict samples from different seasons or locations.

Two forms of model maintenance were performed in this study. First, the calibration dataset was updated by adding 100 new samples from the subsequent season, across the range of total lengths. This is a simple approach that ensures that the samples being added to the calibration set have a uniform distribution (Fearn 1992), and does not require expertise with sophisticated mathematical techniques. Research has been conducted regarding alternate methods to select new samples to add to the calibration model, such as applying a weighting scheme to the samples (Stork and Kowalski 1999; Capron et al. 2005), Kennard-Stone algorithms (Kennard and Stone 1969), ridge regression (Kailivas et al. 2009; Zhang et al. 2019) and Mahalanobis distance (Shenk et al. 2001). These methods can play an influential role in the predictive ability of the calibration model when only a few new samples are available (Stork and Kowalski 1999). This may be the situation when samples are expensive to obtain or from a limited resource. In this study, the traditional ages of several hundreds of fish were available. The addition of 100 new samples each year did not negatively affect the predictive ability of the model, nor substantially increase the number of latent variables, suggesting that an adequate number of samples was included in the model maintenance process. Further investigations using alternative sample selection methods may determine that fewer samples are required, or that by including more samples in the model maintenance process, the predictive ability improves. For some species, model maintenance may not be required on an annual basis, but at a different frequency.

The second form of model maintenance undertaken was to revise the wavelength selection used in the generic models. The generic and revised combined 2012–2015 PLS-R calibration models had similar wavelength regions, although the region from 7190 to 6640 cm⁻¹ was an additional region in the revised model. Helser et al. (2019) identified wavelengths in the 6821–5269 and 5022–4171 cm⁻¹ region to be most informative for ageing eastern Bering Sea walleye pollock otoliths. These spectral regions correspond to −CH₂, −OH and −NH functional groups (Helser et al. 2019). The regions selected in the generic PLS-R calibration models, as shown in Fig. 7, fall within these wider spectral regions.

A major benefit of NIR spectroscopy is the speed of throughput. Using traditional ageing techniques, a sample of 500 barramundi otoliths takes experienced fishery monitoring...
technicians from the laboratory used in the present study ~40 h to read. This includes the technician training and testing against a reference collection so as to qualify to read and re-read the quality assurance samples. Robins et al. (2015) reported that more than 80 otoliths can be scanned per hour using a Bruker Multi-Purpose Analyser with an integrating sphere and up to 180 samples per hour using a 30-sample carousel. In 6 h of actively reading ages, Helset et al. (2019) suggested that 360 walleye pollock otoliths could be assessed by FT-NIR, compared with just 35 by using traditional techniques. Similarly, Passerotti et al. (2020a) reported that a few hundred red snapper otoliths could be assessed using the traditional sectioning method per week, but the 1357 otoliths used in the study could be aged by FT-NIR in 34 h.

Further benefits are that little preparation of the otolith is required before obtaining the spectra. The traditional method is time-consuming, with considerable preparation of the otolith required before being assessed by a trained reader, and in many situations, read more than once or by multiple readers. To obtain spectra from the whole otolith, it needs only to be cleaned and dried, thus saving time and resources. The otolith also remains whole, allowing it to be used for other research purposes.

It is not intended to promote NIR spectroscopy as a complete replacement method for ageing otoliths. It is recommended that the traditional ageing of samples continues to ensure that the NIR calibration model remains robust when additional temporal and spatial spectral variability is identified. However, a major benefit of using NIR spectroscopy to age otoliths is that only a subset of the otoliths collected each year needs to be aged using traditional methods. The remaining samples can be aged quickly using the calibration model and, for many agencies, this could result in a large saving in resources, time and money.

The results obtained in this study have shown the use of FT-NIR to age barramundi from the southern Gulf of Carpentaria genetic stock. Barramundi have a complex life history, which may have been expected to negatively influence the ability of FT-NIR to predict age on the basis of otolith micro-chemistry. By incorporating temporal variability into the calibration models, the influence of these external factors can be reduced. This is very important as it is known that the prediction accuracy becomes less sensitive to changes when more variability, biological and temporal, is built into the model (Bobelyn et al. 2010).

The results from this study have shown that barramundi from the southern Gulf of Carpentaria could be assigned an age class to within ±1 year of the reference age class more than 96% of the time. This suggests that FT-NIR spectroscopy, used in collaboration with traditional ageing methods, may be a suitable method to predict the age of barramundi with vastly reduced resources. However, further research is required to improve the overall precision and understand why there is a tendency in multiple species for the otoliths from older fish to be underestimated. A study using known-age barramundi has the potential to improve the calibrations and would be beneficial to uncovering the true accuracy of FT-NIR spectroscopy for ageing otoliths.

**Declaration of funding**

This research did not receive any specific funding.

**Acknowledgements**

We thank staff from Fisheries Queensland’s Fishery Monitoring for providing the reference ages and access to the otoliths for the purpose of scanning. We are also very grateful for the constructive comments from the anonymous reviewers, which has resulted in an improved manuscript.

**References**


**Conflicts of interest**

The authors declare that they have no conflicts of interest.


Handling Editor: Lee Baumgartner