Accounting for uncertainty due to data processing in virtual population analysis using Bayesian multiple imputation
Accounting for uncertainty due to data processing in virtual population analysis using Bayesian multiple imputation

Thomas Carruthers\textsuperscript{1}, Laurence Kell\textsuperscript{2}, and Carlos Palma\textsuperscript{2}.

1. Fisheries Center, AERL, 2202 Main Mall, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z4

2. International Commission for the Conservation of Atlantic Tunas. Corazón de María, 8. 28002 Madrid, Spain. Laurie.kell@iccat.int; Carlos.palma@iccat.int

Corresponding Author

Thomas R. Carruthers

Tel: (+1) 604 822-6903

Fax: (+1) 604 822-8934

Email: t.carruthers@oceans.ubc.ca.
Abstract

Virtual Population Analysis (VPA) is used in many stock assessment settings and requires a total catch-at-age dataset where an age is assigned to each fish that has been caught. These datasets are typically constructed using ad-hoc methods that rely on numerous assumptions. Although approaches are available to account for observation error in these data, no statistically rigorous methods have been developed to account for uncertainty from data processing. To address this a Bayesian multiple imputation approach to filling missing size data was investigated. Using Atlantic yellowfin tuna and bigeye tuna as case-studies the hypothesis was evaluated that data processing is as important in determining management reference points in stock assessments as conventional sources of uncertainty. Size imputation models accounting for location, season and year provided good predictive capacity. Uncertainty from data processing could be large however the circumstances for this were unpredictable and varied depending on the stock. These results indicate that VPA assessments should attempt to account for uncertainty in data-processing to avoid potentially large compression of uncertainty in assessment results.

Keywords

Data imputation; stock assessment, size data, length data, tuna, Atlantic, virtual population analysis, Bayesian.
**Introduction**

Fishery catch-at-size (CAS) and catch-at-age (CAA) data are commonly required by stock assessment approaches that explicitly model the age-structure of fish populations. These assessment models may be desirable in instances where there have been strong changes in cohort strength or changes in the vulnerability of different age classes to fishing (vulnerability-at-age). These CAS and CAA data are often patchy, particularly for highly migratory species exploited by numerous and varied fishing fleet types. Statistical catch-at-age and catch-at-length assessments (e.g. Methot and Wetzel 2012; Fournier et al. 1990) typically include models that approximate vulnerability dynamics and therefore do not require complete records of CAS and CAA data. However vulnerability may be irregular and poorly approximated by parametric models and difficult to distinguish from the confounding influence of other time-varying processes on the observed CAS or CAA data, such as recruitment strength, natural mortality, growth, availability (spatial overlap), encounter rate (vertical overlap) and changes in management measures (e.g. Atlantic bigeye tuna and yellowfin tuna, ICCAT 2015; 2016 respectively).

An older and simpler approach is Virtual Population Analysis (VPA) which uses the total CAA dataset (all historical catches are assigned an age). The population dynamics model of the VPA leads with an estimate of the numbers remaining in the oldest age class. The entire cohort is then reconstructed back in time by accounting for natural mortality rate while adding the reported catch of younger age classes. VPA assumes that catches and natural mortality are known without error, and therefore vulnerability-at-age does not require approximation by a model. The implementation of VPA that is used to inform management of Atlantic tunas, ADAPT-VPA (Conser 1993), estimates the terminal numbers or fishing
mortality-at-age using catch per unit effort (CPUE) from the fisheries as a proxy for relative stock abundance.

Stock assessments of the Atlantic tropical tunas, bigeye tuna (*Thunnus obesus*) and yellowfin tuna (*Katsuwonus pelamis*), by the International Commission for the Conservation of Atlantic Tunas (ICCAT) use a range of modelling approaches to evaluate fishery status relative to reference points. Since the 1990’s these have included VPAs (e.g. XSA Shepherd 1999; ADAPT-VPA, Gavaris 1988), biomass dynamic models (Prager, 2002), and statistical catch-at-size models (ICCAT 2015; 2016). The VPA approach is still applied since it provides continuity, is fast and simple to apply, has established estimation properties and is suited to the complex vulnerability dynamics of these stocks that may be difficult to approximate using statistical catch-at-age models, for example.

In this paper we address two principal limitations of the VPA assessment approach: (1) a lack of transparency and reproducibility in the processing of the total CAA dataset; (2) the inability to account for uncertainty associated with the processing of the total CAA dataset.

There are two principal data processing steps in the production of the total CAA dataset: (1) the assignment of size to all catches to derive total CAS (a size for every fish that was caught) (2) the aging of total CAS to get total CAA (an age for every fish that was caught). In the majority of highly migratory pelagic fisheries only a small fraction of fish that are caught are measured. Since all fish must be assigned a size in step 1 above, this can be seen as a missing data or ‘imputation’ problem.
In the case of the tropical tunas of the Atlantic, the total catch-at-size dataset is constructed from three main types of data that are reported to ICCAT: (‘Task I’ data) annual catches recorded by region, gear, nation and species, (‘Task II’ data) catches and effort by area, gear, nation, species and month and sampled size frequencies by area, gear, flag, species and month (and by sex if possible). Where total catch-at-size data are not reported, size frequency samples from that flag and gear must be raised to sum to the total catch in weight reported.

The current approach to this is complex, difficult to reproduce and relies on ad-hoc rules and assumptions (Kell et al. 2003). The primary objective of this paper is to establish an alternative approach for constructing this total CAS dataset that is empirical, reproducible and can account for uncertainty.

In the case of Atlantic bigeye tuna and yellowfin tuna assessments (ICCAT 2015; 2016) the total catch-at-size dataset is then converted to the total CAA dataset by either cohort slicing or statistical age-at-size methods (e.g. Kimura 1977; Goodyear 1997). For Atlantic bigeye and yellowfin tuna, cohort slicing is generally used, where a fish of a given size is aged deterministically using a growth curve. This approach assumes that a single age class corresponds to each size of fish. Alternatively, statistical age-at-size methods may be used that sample ages stochastically from an appropriate age spectra for a particular size class. Statistical age-at-size methods often make use of a pre-calculated age-length key that determines the probability of assigning an age (conditional on length) based on age frequency observations or from percentiles of an error distribution superimposed over a fitted growth curve (Kell and Kell 2011).

Despite doubt over the reliability of the total CAA dataset (e.g. Fromentin et al. 2014) VPAs continue to be used in the stock assessment of Atlantic billfish and tunas. In many cases data
preparation and population dynamics modelling are carried out as separate, discrete stages of
the stock assessment process. In the case of VPA assessments of Atlantic tropical tunas the
uncertainty in parameter estimates (i.e. model predictions) is normally calculated by
bootstrapping the time series of catch per unit effort (CPUE) which is assumed to be a proxy
for relative stock abundance and used to calibrate the VPA (e.g. Yamaguchi and Matsuishi
2007). This accounts for observation error and not the additional uncertainty and potential
biases that may be incurred from the processing of the CAA data. To our knowledge there
has been no attempt to rigorously account for the uncertainties of data processing in the
outputs of VPA stock assessments.

In this paper empirically-derived size imputation algorithms are investigated that generate
multiple total catch-at-size datasets. We undertake a comparative evaluation of the impact of
size imputation relative to typical axes of uncertainty in stock assessments. The hypothesis is
evaluated that variability among catch-at-size imputations leads to a greater degree of
variability in estimates of management reference points than traditional axes of uncertainty
such as natural mortality rate and choice of relative abundance index (e.g. Hillary et al.
2015).

Methods

Background to imputation models and their assumptions

We assume that size data are Missing At Random (MAR, Rubin 1976) a missing data
condition that assumes: (1) the sizes of missing data can be predicted from observable
covariates such as location and time of capture and (2) the probability that an observation is
missing is unrelated to the size of the fish itself. For the MAR condition to be violated, size
data must be missing due to size itself without any other predictor of the missing size of the
fish. For example, a bias towards the measuring of larger fish. In such a case the data will not
be MAR and the mechanism is considered ‘non-ignorable’. If however size data are missing
due to an observable covariate such as the strength of the fishing gear or size of hook, then
these covariates may be used to impute the size of the fish and satisfy the MAR assumptions.
In the tuna case studies considered in this paper, the principal reason for a lack of size data is
the absence of sampling protocols after capture. If sampling was undertaken all fish were
measured regardless of length, which indicates that the MAR assumption is likely to be
satisfied in these cases.

The simplest model-based method of imputing MAR data is Expectation Maximisation (EM,
Dempster et al. 1977). EM is iterative and starts by imputing missing data. A model is then
fitted to the (now complete) dataset, and is used to re-impute missing values. This continues
until convergence in the imputed values is achieved. Expectation maximisation is one of the
most widely used of the ‘more principled’ methods (Little and Rubin 1987). The product of
EM is a single imputed data set corresponding to the best fit of the model to the observed
data. However, EM provides no statistically rigorous manner with which to estimate standard
errors even in instances where the imputation model fits the observed data poorly and there is
considerable uncertainty associated with imputation.

With this problem in mind, Rubin (1987) developed a procedure known as Multiple
Imputation (MI). In essence, MI is EM with replication in which the imputation model is
repeatedly fitted to a sample of the full dataset in a similar way to bootstrapping. Typically
MI produces between 4 and 10 sets of imputed values. When the mean of the replicate
imputations for each data point is used and the standard error incorporated into the final
analysis as a measure of uncertainty, the MI method is termed ‘repeated-imputation
inference’. Bayesian multiple imputation (BMI) is version of MI in which the imputation model is fitted in a Bayesian framework and posterior estimates of model parameters are sampled to generate multiple imputed datasets (e.g. Mason et al. 2012; Van Poorten et al. 2015).

An alternative to EM is hot-deck imputation in which missing values are imputed with observed values that are likely to be similar (Andridge and Little 2010). Hot-deck methods may define similarity by any relevant criterion. For example, 'Hot-deck next case' simply imputes the required data from the next complete case and 'hot-deck nearest neighbour' calculates of the 'distance' between incomplete case and donor case (e.g. nearby geographically, observed in the same day).

In this research we describe a novel approach to imputing size data that is a hybrid of BMI and hot-deck nearest neighbour approaches. We specify a distance model in which sampling weights are assigned to every observed length datum in relation to each missing length datum; essentially a matrix with a row for each location in space and time in which fish are caught and a column for each location in space and time of a size observation. The approach then randomly samples a length datum from the observed length records in proportion to their assigned sampling weight. These lengths may be assigned an age using either deterministic or statistical aging methods. Our approach allows for repeated-imputation inference by generating multiple datasets that can be used in multiple corresponding VPA assessments (a BMI-VPA approach).

The length imputation model
For any missing datum $M_i$, the sample weight $W_{i,j}$ of a length observation $O_j$ is calculated from a standard trivariate normal distribution (Rose and Smith 1996) based on the distance in terms of season $S$, year $Y$ and geographic location $G$. Temporal distances were intended to capture changes in size structure due to exploitation (predicted by stock assessments, e.g. ICCAT 2015; 2016) while spatial distances were included to recognise spatial size structuring of tunas (e.g. Taylor et al. 2011; Williams et al. 2012). The sample weight $W$ can then be used in a multinomial model to sample individual length observations in proportion to their relative weight given the location of a missing length datum: $P(O_j|M_i)$. Expressed fully this is:

\[
P(O_j|M_i) \propto W_{i,j} = P(S, Y, G) = \frac{\exp\left(-0.5v/(c_{SG}^2 + c_{YG}^2 + c_{GY}^2 + 2c_{SG}c_{SG}c_{YG} - 1)\right)}{2\sqrt{2\pi}^{3/2} \sqrt{1 - (c_{SG}^2 + c_{YG}^2 + c_{GY}^2 + 2c_{SG}c_{SG}c_{YG})}}
\]

where

\[
v = S^2(c_{SG}^2 - 1) + Y^2(c_{YG}^2 - 1) + G^2(c_{GY}^2 - 1) + 2[SY(c_{SY} - c_{SG}c_{YG}) + SG(c_{SG} - c_{SY}c_{YG}) + YG(c_{YG} - c_{SY}c_{YG})]
\]

and $C$ is the correlation among variables (e.g. $C_{YG}$ is the correlation between year and geographic location).

Standardized temporal distance in years $Y$ is calculated as the absolute difference between the time of the missing observation $i$ and each length record $j$ divided by a standard deviation $\sigma_Y$ (in years):

\[\text{Standardized Temporal Distance} = \frac{|T_i - T_j|}{\sigma_Y}\]
where time $t$ is converted to a real number (e.g. February 6$^{th}$ 1995 = 1995.11).

Similarly, standardized geographic distance $G$ is calculated by

\[
G_{i,j} = \frac{\tan^{-1}(\sqrt{\frac{\varphi_{i,j}-\varphi_{j,i}}{1-\varphi_{j,i}}} K/\pi)}{\sigma_G}
\]

where $\sigma_G$ is the standard deviation (in km), $K$ is the earth’s mean circumference (40 041km) and $\phi$ is given by

\[
\varphi_{i,j} = \sin\left(\frac{\text{lat}_i-\text{lat}_j}{2}\right)^2 + \cos(\text{lat}_i) \cdot \cos(\text{lat}_j) \cdot \sin\left(\frac{\text{lon}_i-\text{lon}_j}{2}\right)^2
\]

where latitude $(\text{lat})$ and longitude $(\text{lon})$ are expressed in radians.

Standardized seasonal distance $S$ is calculated from the day of the year (e.g. for January 7$^{th}$, $d = 7$). In order to make these ‘wrap’ between years (so that day 365 and day 1 are adjacent) we used a sine – arcsine transformation:

\[
S_{i,j} = \frac{360 \sin^{-1}(\gamma_{i,j})/\pi}{\sigma_S}
\]

where $\sigma_S$ is the standard deviation (in days) and

\[
\gamma_{i,j} = \sin\left(\left[\pi(d_i - d_j)/360\right]\right)
\]
In some situations catches for which size data are missing were reported in weight rather than the numbers of individual fish. In these cases we sampled a sufficient number of length observations to sum to the reported weight of the catch using the deterministic growth curve. While subsequent VPA analyses involve various growth curves, these are not critical at the imputation stage. This is because alternative growth curves only affect the number of imputations required to meet a target catch weight, rather than the imputed sizes themselves, and the estimated assessment quantities were unrelated to stock magnitude. At the imputation level, alternative growth curves varied overall results less than half of a percent and therefore for the sake of simplicity, a single growth curve was assumed for each stock (based on the first factor level, see below).

The length imputation model was implemented in the statistical environment R (3.3.2, 64bit; R core team 2017).

**Empirically fitting the Bayesian length imputation model**

The imputation models used the ‘Task II’ length dataset to impute spatial catches of the Task II catch dataset (ICCAT 2017). Task II catches are not complete and were scaled to match total landings reported by each flag (e.g. U.S.A.) and gear type (e.g. longline). The spatial range and magnitude of these data are illustrated in Figure 1.

To ensure that the imputation models were credible we fitted the correlation coefficients ($C_{SY}$, $C_{SG}$, $C_{YG}$) and the standard deviations ($\sigma_S$, $\sigma_G$, $\sigma_Y$) to the length composition of 15 fleets that reported length observations with sufficient spatio-temporal coverage (these flag and gear combinations are listed in Table 1). To determine whether the estimation of trivariate normal
correlation coefficients ($C$) was worthwhile we also fitted a model that assumed no
correlation requiring estimation of only the standard deviation parameters ($\sigma$).

Given the multinomial probability model used to select lengths (Eqn 1), the expected imputed
length $\bar{z}$ for a given position can be predicted as the average of the available length
observations weighted by $W$:

$$
\bar{z}_i = E(z_i) = \frac{\sum_j W_{i,j} z_j}{\sum_j W_{i,j}}
$$

where $W_{i,j} = 0$ when $i=j$ to avoid the length samples of a particular mean length observation
from contributing to its own calculation. We used the model to predict a set of mean length
samples for each fleet, $\bar{z}$. Using a log-normal likelihood function we could then compare our
predicted mean length estimates $\bar{z}$ with the real mean length $\bar{z}$ that was observed at the same
time and location:

$$
P(\bar{z}_i | \theta) = \frac{1}{\bar{z}_i \sqrt{2\pi}} \exp \left( -\frac{(\ln(\bar{z}_i) - \ln(\bar{z}_j))^2}{2s^2} \right)
$$

The lognormal standard deviation $s$ was assumed to be proportional to the observed mean
length with a fixed coefficient of variation of 10%.

The length imputation model fitting was undertaken using AD Model Builder (Fournier et al.
2012). In order to ensure that the variance-covariance matrix for the multivariate normal
distribution (Eqn. 1) was positive semi-definite and allow for unconstrained estimation we
used the Cholesky parameterization (Pinheiro and Bates 1996).
The BMI approach was implemented using Markov Chain Monte Carlo (MCMC via the
Metropolis Hastings algorithm) to numerically approximate the posterior distribution of the
correlation and standard deviation parameters. In this way we sampled credible combinations
of model parameters ($\sigma, C$) for the imputation models and created multiple datasets from
them. Since it was possible for some fleets to have length sample data with weak contrast
with respect to one of the three dimensions $G$, $S$ and $Y$, we prescribed weak prior
distributions. The correlation parameters $C$, were assigned a normal prior with mean zero and
standard deviation 0.5. The standard deviation parameters were assigned lognormal prior
means of 500km ($\sigma_G$), 2 years ($\sigma_Y$) and 30 days ($\sigma_Y$) all of which had a coefficient of variation
of 50%. The model posterior distributions were very similar given alternative priors. For
example, lognormal prior means of 2000km, 5 years and 60 days respectively lead to less
than 5% changes in mean posterior estimates of the core model parameters. MCMC chains
showed rapid mixing and according to trace plots Gelman and Rubin (1992) and Geweke
(1992) diagnostics, convergence could not be rejected. Parameter posteriors were sampled
after a ‘burn-in’ of 5000 iterations.

For most fleets there was insufficient length sample data to fit an imputation model (i.e. not
one of the 15 flag and gear combinations in Table 1). Fleets that have size samples and fitted
imputation models represented around 35% of bigeye catches and 15% of yellowfin catches.
In these cases missing length observations were imputed from the length data of the correct
fleet but imputation model parameters were borrowed from the Japanese fleet of the same
gear type. The Japanese fleets were used since their length observations were among the most
numerous and had the widest spatio-temporal coverage.

Accounting for uncertainty in assessments
VPA assessments involve a chain of analysis; a set of sequential steps with associated uncertainties (Kell et al. 2003). Several of these steps represent data processing rather than stock assessment modelling. To quantify the relative importance of data processing in determining the outputs stock assessments, tuned VPAs (Porch 2003) were conducted over a range of scenarios representing the conventional axes of assessment uncertainty in addition to uncertainty from data-processing (among imputations). A factorial design was adopted in which the factors corresponded with 8 stages in the VPA chain of analyses (Table 2, Figure 2). Some of the factors correspond to decisions about VPA assessment specification (e.g. the functional form of the growth model or the choice of relative abundance index). In these cases wherever possible the levels reflected base-case and alternative scenarios for the most recent stock assessments.

The first factor corresponded to stock with levels for Atlantic bigeye tuna and Atlantic yellowfin tuna. These stocks were chosen for their economic significance and availability of a recent, fully documented VPA assessment (ICCAT 2015; 2016 for bigeye tuna and yellowfin tuna respectively). More than one stock was analysed to help distinguish between case-specific and more general findings.

Factors 2 and 3 relate to the specification of the length imputation model described above. Both bigeye tuna and yellowfin tuna aggregate in shoals of similar sized fish. It is therefore inappropriate to impute many different sizes for fish that were caught together. To address this, we imputed identical lengths for groups of fish. This is equivalent to sampling a length from Eqn. 1 and replicating this value. To understand how this might affect VPA analyses we identified three levels for the factor ‘length-group size’, where size imputations were carried out in groups of 100, 1000, and 10 000 individuals. To understand how uncertainty from
imputation drives VPA assessments, for each length group size we imputed 20 replicate total-
catch-at-length datasets. Imputations were carried out for the time period 1968-2013 and
totalled around 950 million yellowfin tuna and 450 million bigeye tuna, the majority of which
were imputed at the spatial resolution of 1 degree ocean cell.

Factors 4 and 5 of the VPA analysis chain convert the imputed total catch-at-length dataset to
the total catch-at-age dataset. In factor 4, two levels were considered for the aging method:
cohort slicing (deterministic prediction of age from inverse growth curve) that is the current
default method of the most recent stock assessments and an alternative statistical aging
approach using an age-length key (conditional probability of age given length) derived from
mixed distributions (Kell and Kell 2011). The approach of Kell and Kell (2011) is an
extension of Liu et al. (1989) and uses additional biological information to characterize the
age-length key.

Both aging approaches require a growth model which is also used by subsequent VPA
assessments to convert predicted numbers-at-age to predicted biomass-at-age. Factor 5
investigated two levels for the growth model. The first level for both stocks was a von
Bertalanffy growth model used in the most recent stock assessments (ICCAT 2015; 2016).
The second level was a Gascuel et al. (1992) growth model for yellowfin tuna and a Richards
growth curve for bigeye tuna (Hallier et al. 2005) (Figure 3). To ensure theoretical
consistency, the same growth curve used to create the total catch-at-age dataset was also used
in the subsequent VPA stock assessment. The VPA assessments, which predict numbers at
age, require a growth model to calculate spawning biomass for recruitment modelling, and
vulnerable biomass for calibration to relative abundance indices.
In addition to the total catch-at-age datasets, the tuned VPA assessments applied to these stocks (Porch 2013) are calibrated using a time-series of relative abundance, typically a standardized catch-per-unit-effort (CPUE) index. It is common for numerous CPUE indices to be considered that infer conflicting trends in relative abundance which in turn may result in poor fits to these data and bias in assessment model estimates. Therefore in the real stock assessment setting, ICCAT working groups first look at the relative abundance trends inferred by the individual CPUE series and select groups of CPUE indices that represent common hypotheses about trends in stock abundance (ICCAT 2016; 2017). For factor 6, this procedure was used to define two levels for CPUE trend in abundance, one for each group CPUE indices (CPUE I and CPUE II, Figure 3).

In many stock assessment models the data that are typically available are not sufficiently informative to reliably estimate various parameters. A common example is the instantaneous natural mortality rate ($M$) which in yellowfin tuna and bigeye tuna VPAs is an age-specific vector that is fixed. Factor 7 included two levels for natural mortality rate at age, the first was the base-case vector established by the respective ICCAT species working group (WG) the second was based on the ogive derived by the meta-analysis Gislason et al. (2008) (Figure 3).

In the VPAs used to assessment these tuna species, a value for fishing mortality in the oldest age class is required to estimate the terminal numbers-at-age in each cohort from catches. This is parameterized such that fishing mortality rate of the oldest age class is a fraction of the second oldest age class, referred to as an F-ratio. In factor 8 of the analysis chain, we evaluated two levels for VPA assessment assumptions: level 1 in which the F-ratio was fixed at 1, and level 2 where F-ratio was estimated during model fitting to the relative abundance index.
Calculating variables of management interest

The 8 factors of the VPA analysis chain (Figure 2) resulted in a total of 3840 VPA assessments. For each assessment we calculated current spawning biomass relative to biomass at maximum sustainable yield ($B/B_{MSY}$) and current fishing mortality rate relative to maximum sustainable yield ($F/F_{MSY}$) as the two principal variables of management interest (Sissenwine and Shepherd 1987). These two variables offer a concise fishery assessment summary because they represent stock status ($B/B_{MSY}$ is commonly used to delineate overfished and underfished stock status) and exploitation rate ($F/F_{MSY}$ is commonly used to delineate overfishing from underfishing). To evaluate the impact of the various VPA analysis stages, assessment results can be plotted as points in the space of $B/B_{MSY}$ and $F/F_{MSY}$ (a ‘Kobe’ plot) which is a standard output in a wide range of fishery assessment settings including the advice framework of the tuna Regional Management Organizations (Kell et al. 2016).

Maximum sustainable yield reference points are often calculated by characterizing the relationship between spawning biomass and recruitment. However in many stock assessments (of which Atlantic yellowfin tuna and bigeye tuna are examples) there is little evidence for such a relationship, and past recruitment varies independent of stock size (Szuwalski et al. 2015). Therefore biomass ($B_{MSY}$) and fishing mortality rate ($F_{MSY}$) reference points were calculated from mean model estimated recruitment levels and the mean vulnerability-at-age over the most recent three years.

Results

Predictive capacity of the Bayesian imputation model
The trivariate normal distribution (location, year and season) was capable of providing reasonable to good predictive capacity for mean size (Table 1, Figures 4 and 5). The coefficient of determination ($R^2$ value) for the fitted models fell in the range of 30-80% but were often over 50%. Estimating the additional correlation parameters generally did not lead to significant improvements in predictive capacity for most fleets (Table 1). Two notable exceptions were USA purse seiners fishing bigeye tuna and Brazilian longliners fishing yellowfin tuna. Since the observed size data are a product of numerous processes including fishery targeting and the spatio-temporal pattern of both fishing and the size composition of the stock, the explanation for correlation among parameters is not clear. The results presented below for the BMI-VPA analysis used the imputation model with correlated parameters for generating the total catch-at-length data set.

The relative important of assessment factors in determining fishery reference points

The various factor combinations for the VPA analysis chain led to a high degree of variability among assessment estimates of $B/B_{MSY}$ and $F/F_{MSY}$ (Figures 6 and 7). An ANOVA was performed to partition the variance in fishery reference points by factor to evaluate their relative importance in the VPA analysis chain. These results are framed in terms of the reduction in residual deviance that can be attributed to each factor. Convergence diagnostics indicate that the partitioning of this variance is reasonably stable after 15 replicate imputations (Figure 2) and that the 20 replicates were sufficient to draw robust conclusions about the relative importance of the various factors.

The degree of uncertainty introduced from replicate imputations of the total catch at length data set (factor 3) was inconsistent among the stocks and estimated management reference points (Table 3). In the case of the bigeye tuna stock, estimates of $F/F_{MSY}$ varied more
strongly among imputations than any of the other factors (although the reduction in residual
deviance was similar to alternative scenarios of CPUE trend in abundance). In contrast
imputation was substantially less influential than natural mortality in determining $B/B_{\text{MSY}}$ for
the same species. In the case of yellowfin tuna, while replicate imputations were ranked
lower (fourth in terms of the reduction in deviance for both $B/B_{\text{MSY}}$ and $F/F_{\text{MSY}}$ estimates) the
degree of uncertainty generated by imputations was comparable to other factors such as the
level of growth model.

The relative importance of the various factors was highly inconsistent among the VPA
estimates for the two stocks. For example the choice of CPUE relative abundance trend
(factor 6) was inconsequential in determining $F/F_{\text{MSY}}$ for yellowfin tuna but was pivotal in
determining $F/F_{\text{MSY}}$ estimates for bigeye tuna (Table 3, Figures 8 and 9). CPUE relative
abundance trend was also not critical in determining stock status estimates ($B/B_{\text{MSY}}$) for
yellowfin. For both stocks, by far the most important source of variance in VPA estimates of
stock status ($B/B_{\text{MSY}}$) was the choice of natural mortality rate vector (factor 7). Although
substantially less important, replicate imputations were amongst the largest contributors to
variance in $B/B_{\text{MSY}}$ after natural mortality rate.

In general, the length group size of the imputations (the number of fish imputed with identical
sizes) did not strongly determine reference points for either stock. Similarly aging method
(factor 4) was relatively inconsequential with the exception of estimates of $F/F_{\text{MSY}}$ for
yellowfin.

Uncertainty due to data processing
In Figures 8 and 9 the VPA assessment results were disaggregated into five factors (aging method, growth model, trend in abundance, natural mortality rate, F-ratio approach) to highlight variance in model predictions of F/F_{MSY} and B/B_{MSY} among total catch-at-length imputations (‘imputation variance’). Two things are striking about these results, namely that the results were case-specific and that some factors affected F/F_{MSY} and others B/B_{MSY}. For example when the F-ratio was estimated rather than fixed at 1, in the case of bigeye tuna this resulted in a smaller variance for model predictions whereas the reverse was seen for yellowfin and the F-ratio affected estimates of B/B_{MSY} more than estimates of F/F_{MSY}.

Depending on the combination of other factor levels, imputation variance could be very low or very high. When imputation variance was very high (e.g. Figure 8 panel u) it was around half of the total variance of all VPA scenarios, with multiple imputations providing model predictions over a wide range of B/B_{MSY} and F/F_{MSY}. The magnitude of imputation variance was inconsistent among the two stocks and the other factor levels (as illustrated by the inconsistent patterns in plotted ellipses among the panels of Figures 8 and 9). For example, there was a critical interaction between stock (factor 1) and the assumption about F-ratio estimation (factor 8). In Figures 8 and 9 this is illustrated by clear patterns among the rows of the figure panels. For bigeye tuna the estimated F-ratio scenarios (e.g. top two panel rows of Figure 8) imputation variance was relatively low. In contrast the same scenarios for yellowfin tuna (e.g. top two panel rows of Figure 98) generated amongst the greatest imputation variance.

Discussion

Previous attempts to characterize uncertainty in VPA assessments have focused on observation error. For example bootstrapping of the total catch-at-age data prior to VPA
analysis (Yamaguchi and Matsuishi 2007) a method that is also applied in Atlantic tuna
assessments (e.g. TTSG 2012). Hillary (2011) provided a useful analytical solution to
evaluate the role of error in catch-at-age data without having to undertake the
computationally intensive Monte-Carlo approaches of a bootstrap. These methods consider
how to deal with error after the construction of the total catch-at-age dataset. In our BMI-
VPA approach we focus on additional uncertainties of processing the catch-at-age dataset in
the first instance, which may be substantial and otherwise ignored.

In contrast to current methods for constructing the total catch-at-length datasets for Atlantic
yellowfin tuna and Atlantic bigeye tuna which are ad-hoc and not easily reproduced, we
describe and implement an empirical approach to formulating these datasets. In many cases
the fitted imputation models that include covariates for distance in time, season and location
provided reasonable predictive capacity. This finding is supported by previous work that has
identified spatio-temporal size structuring in yellowfin tuna and bigeye tuna (Fonteneau et al.
2013).

It proved difficult to reject or accept the hypothesis that VPA assessment outputs are more
strongly determined by imputation of the catch-at-length dataset than traditional axes of
uncertainty. The relative important of imputation varied among estimated reference points
and the stock in question. However in general, imputation was often of comparable or greater
importance in determining stock status relative to reference points as at least one other factor
of the VPA analysis chain that is typically considered as important and used in assessment
sensitivity testing (e.g. natural mortality rate, CPUE relative abundance index, growth
model). This result suggests that alternative plausible constructions of the total catch-at-
length dataset could provide very different estimates of stock status and current exploitation
rate. The results highlight that the relative importance of new information varies by stocks and management reference points. For example, for both bigeye tuna and yellowfin tuna uncertainty about natural mortality rate has the biggest impact on $B/B_{MSY}$ while for $F/F_{MSY}$ the imputation replicate has the biggest impact for bigeye tuna and F-ratio for yellowfin tuna.

The relative importance of natural mortality rate corroborates findings of previous investigations of VPA assessments (Punt 1997).

In many settings statistical catch-at-age (SCA) assessments are viewed favourably and are seen as possible replacements for VPAs. A central advantage of the SCA approach is that it requires only a sample of age composition data (or length composition data) and can account for observation error in these data. A principal disadvantage of SCA assessments is that they typically model age vulnerability using parametric models (selectivity curves) that can struggle to approximate complex vulnerability dynamics arising from cohort targeting (e.g. Pacific Hake, Berger et al. 2017), species targeting (e.g. global shift to targeting tuna in deeper waters, Miyake 2004) and inconsistent spatio-temporal overlap of fishing on the stock (e.g. Atlantic bluefin tuna, ICCAT 2014). An additional and critical disadvantage of SCA assessments is that they can be numerically unstable due to the inclusion of many nuisance parameters relating to complex historical vulnerability dynamics (e.g. stock assessment of Pacific bluefin tuna, WCPFC 2014). Indeed, numerical stability is one of central reasons why VPA assessments are still used to provide management advice for Atlantic bigeye tuna and yellowfin tuna. There are parallels between SCA assessments and the BMI-VPA approach.

SCA assessments essentially scale-up total observed catches to the total catch-at-age data set: in a sense total catch-at-age is also imputed in an SCA and interpreted as fishing mortality rate-at-age by similar population dynamics assumptions as a VPA. BMI-VPA analysis may offer an attractive alternative to SCA since in addition to observation error, it can account for
uncertainty from data processing and can approximate complex historical age vulnerability
dynamics.

The BMI-VPA approach has a number of weaknesses of which high computational
requirements are among the most significant. In contrast to other methods for characterizing
uncertainty in VPA assessments that do not attempt to deal with data processing (e.g. Hillary
2011) our approach can take several hours to generate a suitable number (e.g. 20) multiple
imputed datasets (note this is automated and still rapid compared with current manual
approach to building the total CAS dataset). The computational demand is not easily solved
with parallel computing due to the relatively high memory requirements (typically greater
than 10GB per imputation). The spatio-temporal imputation model used to sample sizes is
also rather naïve. For example it does not discriminate between longitudinal and latitudinal
distances which may be questionable for tropical tuna species that are less likely to inhabit
distant latitudes but may mix throughout equatorial longitudes. Similarly, the imputation
method disregards coastal and high-seas location and does not make use of spatial covariates
that could predict missing length data according to known spawning and rearing areas.

Acknowledging these weaknesses we describe and demonstrate a novel approach for
accounting for uncertainties in data processing in stock assessments. The BMI-VPA method
offers substantial benefits over the status quo: it is empirically derived, reproducible and
performance can be evaluated by conventional diagnostics of model fit. Imputation models of
this type are flexible and may be improved by the inclusion of other covariates relating to
life-history, migration and distribution. A future priority is simulation testing to identify the
circumstances where the BMI-VPA approach is likely to be most suitable.
Our analyses focus on the specific case of the VPA analysis chain. However this example demonstrates how repeated imputation inference could be used more generally for other fisheries modelling problems that are sequential in nature and rely on processed data that are themselves the product of models. Other possible applications include accounting for historical illegal, unreported and unregulated fishing (Agnew et al. 2009), changes in growth or condition factor due to shifting ecological or oceanographic conditions (Medina et al. 2002), designation of stock-of-origin in multi-stock fisheries (e.g. Atlantic bluefin tuna, Fraile et al. 2015), taxonomic classification error (in our analyses many of the yellowfin tuna that were measured were probably bigeye tuna for example) and construction of CPUE indices of abundance where targeting is uncertain (i.e. there is uncertainty over the measure of fishing effort). Since the full analysis chain presented in this paper can account for a very wide range of assessment uncertainty, it could be used as a basis for defining operating models for management strategy evaluation (MSE) (Butterworth and Punt 1999; Punt et al. 2016).

Depending on other factors of the VPA analysis chain, imputation variance could be very high or very low. Importantly however, this magnitude was not consistent over stocks. This suggests that it may be difficult to anticipate assessment situations where ignoring data processing is inconsequential from situations where it leads to a large compression of uncertainty. A recommendation arising from this finding is that regardless of whether BMI is used, VPA stock assessments should be fitted to alternative plausible total catch-at-age datasets to better characterize assessment uncertainty.

Acknowledgements

This manuscript benefitted from the helpful comments of Murdoch McAllister and Carl Walters.
References


Contribution SFD-01/02–151.

Prager, M. H. 2002. Comparison of logistic and generalized surplus-production models
applied to swordfish, Xiphias gladius, in the north Atlantic Ocean." Fish. Res. 58: 41-57.


Management strategy evaluation: best practices. Fish Fish. 17, 303–334,
http://dx.doi.org/10.1111/faf.12104.

project.org.

Rose, C., Smith, M.D. 1996. The Multivariate Normal Distribution, Mathematica Journal. 6:
32 – 37.


New York.

Shepherd, J.G. 1999. Extended survivors analysis: an improved method for the analysis of


common assumptions about recruitment: a meta-analysis of recruitment dynamics for
worldwide marine fisheries. Fish Fish. doi: 10.1111/faf.12083.


Tables

Table 1. Posterior mean estimates of imputation model parameters for fleets where a length imputation model was fitted. The coefficient of determination ($R^2$) provides a measure of the explanatory power of two models: one that estimates just standard deviation parameters ($\sigma$, assumes no correlation structure) and another that also estimates the correlation parameter $C$. The difference in $R^2$ value between models with and without the correlation structure is included in the column ‘Delta’. Fleets not included in this table had insufficient spatio-temporal coverage of length composition to fit an imputation model. In these cases length data were still imputed that were observed by these fleets but using imputation model parameters from the corresponding Japanese fleet of the same gear type.

<table>
<thead>
<tr>
<th>Species</th>
<th>Flag</th>
<th>Gear</th>
<th>No correlation structure</th>
<th>Correlation structure estimated</th>
<th>Delta</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\sigma_Y$</td>
<td>$\sigma_S$</td>
<td>$\sigma_G$</td>
<td>$\sigma_Y$</td>
</tr>
<tr>
<td>Bigeye</td>
<td>Purse seine</td>
<td>USA</td>
<td>0.97</td>
<td>21</td>
<td>352</td>
<td>1.75</td>
</tr>
<tr>
<td>Bigeye</td>
<td>Longline</td>
<td>Japan</td>
<td>9.01</td>
<td>61</td>
<td>1132</td>
<td>2.60</td>
</tr>
<tr>
<td>Bigeye</td>
<td>Bait boat</td>
<td>Japan</td>
<td>0.45</td>
<td>15</td>
<td>477</td>
<td>0.99</td>
</tr>
<tr>
<td>Bigeye</td>
<td>Purse seine</td>
<td>Japan</td>
<td>9.66</td>
<td>31</td>
<td>8071</td>
<td>2.75</td>
</tr>
<tr>
<td>Bigeye</td>
<td>Purse seine</td>
<td>Spain</td>
<td>0.45</td>
<td>25</td>
<td>351</td>
<td>0.66</td>
</tr>
<tr>
<td>Bigeye</td>
<td>Longline</td>
<td>Brazil</td>
<td>0.48</td>
<td>14</td>
<td>278</td>
<td>0.59</td>
</tr>
<tr>
<td>Yellowfin</td>
<td>Purse seine</td>
<td>USA</td>
<td>0.63</td>
<td>42</td>
<td>311</td>
<td>2.37</td>
</tr>
<tr>
<td>Yellowfin</td>
<td>Longline</td>
<td>Japan</td>
<td>2.74</td>
<td>62</td>
<td>1624</td>
<td>3.65</td>
</tr>
<tr>
<td>Yellowfin</td>
<td>Bait boat</td>
<td>Japan</td>
<td>0.45</td>
<td>12</td>
<td>625</td>
<td>0.93</td>
</tr>
<tr>
<td>Yellowfin</td>
<td>Purse seine</td>
<td>Japan</td>
<td>0.45</td>
<td>603</td>
<td>220</td>
<td>1.62</td>
</tr>
<tr>
<td>Yellowfin</td>
<td>Longline</td>
<td>Spain</td>
<td>0.56</td>
<td>18</td>
<td>401</td>
<td>0.90</td>
</tr>
<tr>
<td>Yellowfin</td>
<td>Bait boat</td>
<td>Spain</td>
<td>0.45</td>
<td>23</td>
<td>200</td>
<td>0.76</td>
</tr>
<tr>
<td>Yellowfin</td>
<td>Purse seine</td>
<td>Spain</td>
<td>31.88</td>
<td>22</td>
<td>407</td>
<td>9.47</td>
</tr>
<tr>
<td>Yellowfin</td>
<td>Longline</td>
<td>Brazil</td>
<td>5.96</td>
<td>49</td>
<td>841</td>
<td>0.94</td>
</tr>
<tr>
<td>Yellowfin</td>
<td>Bait boat</td>
<td>Brazil</td>
<td>5.34</td>
<td>21</td>
<td>406</td>
<td>4.30</td>
</tr>
</tbody>
</table>
Table 2. The factorial design of the Bayesian Multiple Imputation – Virtual Population Analysis experiments.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stock</td>
<td>Bigeye tuna, yellowfin tuna</td>
</tr>
<tr>
<td>2. Imputation length group size (n)</td>
<td>100, 1000, 10000</td>
</tr>
<tr>
<td>3. Imputation replicates</td>
<td>20 replicates</td>
</tr>
<tr>
<td>4. Aging method</td>
<td>Slicing, mixed distributions</td>
</tr>
<tr>
<td>5. Growth Model</td>
<td>von Bertalanffy, Richards (bigeye) / Gascuel (yellowfin)</td>
</tr>
<tr>
<td>6. CPUE trend in abundance</td>
<td>CPUE I, CPUE II</td>
</tr>
<tr>
<td>7. Natural mortality</td>
<td>Working group (WG), Gislason</td>
</tr>
</tbody>
</table>
| 8. F-ratio (fishing mortality rate in plus group relative to last true age) | Fixed at 1, estimated
Table 3. The reduction in residual deviance attributed to each factor of the VPA analysis chain for estimates of $F/F_{MSY}$ and $B/B_{MSY}$. In this analysis, replicate imputations (factor 3) are treated as a random effect (labelled ‘RE’) and are equal to the residual deviance after accounting for all the remaining, fixed factors. In each case (e.g. estimates of $F/F_{MSY}$ for bigeye tuna) the various factors of the VPA analysis chain are ranked in order of those associated with the greatest variability in the reference point ($F/F_{MSY}$ or $B/B_{MSY}$).

<table>
<thead>
<tr>
<th>Factor in VPA analysis chain</th>
<th>Degrees of freedom</th>
<th>Deviance reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bigeye tuna: $\log(F/F_{MSY})$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Replicate imputations</td>
<td>RE</td>
<td>217</td>
</tr>
<tr>
<td>6. CPUE trend in abundance</td>
<td>1</td>
<td>191</td>
</tr>
<tr>
<td>7. Natural mortality rate</td>
<td>1</td>
<td>96</td>
</tr>
<tr>
<td>4. Aging method</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>8. FRatio</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>5. Growth model</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2. Length group size</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><strong>Bigeye tuna: $\log(B/B_{MSY})$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Natural mortality rate</td>
<td>1</td>
<td>1222</td>
</tr>
<tr>
<td>3. Replicate imputations</td>
<td>RE</td>
<td>189</td>
</tr>
<tr>
<td>6. CPUE trend in abundance</td>
<td>1</td>
<td>64</td>
</tr>
<tr>
<td>8. FRatio</td>
<td>1</td>
<td>62</td>
</tr>
<tr>
<td>4. Aging method</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>2. Length group size</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>5. Growth model</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Yellowfin tuna: $\log(F/F_{MSY})$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. FRatio</td>
<td>1</td>
<td>224</td>
</tr>
<tr>
<td>4. Aging method</td>
<td>1</td>
<td>211</td>
</tr>
<tr>
<td>5. Growth model</td>
<td>1</td>
<td>189</td>
</tr>
<tr>
<td>3. Replicate imputations</td>
<td>RE</td>
<td>186</td>
</tr>
<tr>
<td>7. Natural mortality rate</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>2. Length group size</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>6. CPUE trend in abundance</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Yellowfin tuna: $\log(B/B_{MSY})$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Natural mortality rate</td>
<td>1</td>
<td>2482</td>
</tr>
<tr>
<td>5. Growth model</td>
<td>1</td>
<td>789</td>
</tr>
<tr>
<td>8. FRatio</td>
<td>1</td>
<td>642</td>
</tr>
<tr>
<td>3. Replicate imputations</td>
<td>RE</td>
<td>554</td>
</tr>
<tr>
<td>2. Length group size</td>
<td>2</td>
<td>49</td>
</tr>
<tr>
<td>4. Aging method</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>6. CPUE trend in abundance</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 1. The geographical distribution of total catch in numbers by 10 degree ocean cell (‘CATDIS’, ICCAT 2017) for which size must be imputed, panels a and c) and the size data that are available (panels b and d) from 1970 onwards. The area of the plotted points is proportional to numbers that are available (maximum point size is represents 20 million fish).
Figure 2. Convergence in the estimates of deviance reduction due to factors of the VPA analysis chain with increasing replicate imputations \( n \). The lines correspond with the values of Table 1 but are adjusted for the number of imputation replicates.
Figure 3. The levels for three factors by species: natural mortality rate, growth model and CPUE trend in abundance.
Figure 4. The fit of the imputation model to mean size data of six fleets fishing bigeye tuna.

Plotted points represent mean size observations (x-axis) and the corresponding predicted mean size by the imputation model. Each plot includes estimates of the coefficient of determination ($R^2$, explaining the fraction of variability in observations explained by the imputation model) and the parameters of the trivariate normal distribution.
Figure 5. The fit of the imputation model to mean size data of nine fleets fishing yellowfin tuna. Plotted points represent mean size observations (x-axis) and the corresponding predicted mean size by the imputation model. Each plot includes estimates of the coefficient of determination ($R^2$, explaining the fraction of variability in observations explained by the imputation model) and the parameters of the trivariate normal distribution.
Figure 6. Plot of spawning stock biomass and fishing mortality rate relative to MSY levels $(B/B_{MSY}, F/F_{MSY}$ respectively) for Atlantic bigeye tuna.
Figure 7. Plot of spawning stock biomass and fishing mortality rate relative to MSY levels (\(B/B_{\text{MSY}}, F/F_{\text{MSY}}\) respectively) for Atlantic yellowfin tuna.
Figure 8. Plot of VPA estimates of spawning stock biomass and fishing mortality rate relative to MSY levels for Atlantic bigeye tuna. The ellipses encompass 95% of the VPA assessment results for the multiple imputed datasets and the black point is the grand mean.
Figure 9. Plot of VPA estimates of spawning stock biomass and fishing mortality rate relative to MSY levels for Atlantic yellowfin tuna. The ellipses encompass 95% of the VPA assessment results for the multiple imputed datasets and the black point is the grand mean.