Canadian Journal of Fisheries and Aquatic Sciences
Journal canadien des sciences halieutiques et aquatiques

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

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2017-0165.R2
Manuscript Type:	Article
Date Submitted by the Author:	13-Aug-2017
Complete List of Authors:	Carruthers, Thomas; University of British Columbia Kell, Laurence; ICCAT Secretariat, Palma, Carlos; ICCAT Secretariat
Is the invited manuscript for consideration in a Special Issue? :	N/A
Keyword:	STOCK ASSESSMENT < General, MARINE FISHERIES < General, TUNA < Organisms, SAMPLING < General, STATISTICAL ANALYSIS < General

SCHOLARONE™
Manuscripts

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

3 Thomas Carruthers¹, Laurence Kell², and Carlos Palma².

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

6 2. International Commission for the Conservation of Atlantic Tunas. Corazón de María, 8.
7 28002 Madrid, Spain. Laurie.kell@iccat.int; Carlos.palma@iccat.int

8

9

10 Corresponding Author

11 Thomas R. Carruthers

12 Tel: (+1) 604 822-6903

13 Fax: (+1) 604 822-8934

14 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.

34 Introduction

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

49

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

58 mortality-at-age using catch per unit effort (CPUE) from the fisheries as a proxy for relative
59 stock abundance.

60

61 Stock assessments of the Atlantic tropical tunas, bigeye tuna (*Thunnus obesus*) and yellowfin
62 tuna (*Katsuwonus pelamis*), by The International Commission for the Conservation of
63 Atlantic Tunas (ICCAT) use a range of modelling approaches to evaluate fishery status
64 relative to reference points. Since the 1990's these have included VPAs (e.g. XSA Shepherd
65 1999; ADAPT-VPA, Gavaris 1988), biomass dynamic models (Prager, 2002), and statistical
66 catch-at-size models (ICCAT 2015; 2016). The VPA approach is still applied since it
67 provides continuity, is fast and simple to apply, has established estimation properties and is
68 suited to the complex vulnerability dynamics of these stocks that may be difficult to
69 approximate using statistical catch-at-age models, for example.

70

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

74

75 There are two principal data processing steps in the production of the total CAA dataset: (1)
76 the assignment of size to all catches to derive total CAS (a size for every fish that was caught)
77 (2) the aging of total CAS to get total CAA (an age for every fish that was caught). In the
78 majority of highly migratory pelagic fisheries only a small fraction of fish that are caught are
79 measured. Since all fish must be assigned a size in step 1 above, this can be seen as a missing
80 data or 'imputation' problem.

81

82 In the case of the tropical tunas of the Atlantic, the total catch-at-size dataset is constructed
83 from three main types of data that are reported to ICCAT: ('Task I' data) annual catches
84 recorded by region, gear, nation and species, ('Task II' data) catches and effort by area, gear,
85 nation, species and month and sampled size frequencies by area, gear, flag, species and
86 month (and by sex if possible). Where total catch-at-size data are not reported, size frequency
87 samples from that flag and gear must be raised to sum to the total catch in weight reported.
88 The current approach to this is complex, difficult to reproduce and relies on ad-hoc rules and
89 assumptions (Kell et al. 2003). The primary objective of this paper is to establish an
90 alternative approach for constructing this total CAS dataset that is empirical, reproducible
91 and can account for uncertainty.

92

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

104

105 Despite doubt over the reliability of the total CAA dataset (e.g. Fromentin et al. 2014) VPAs
106 continue to be used in the stock assessment of Atlantic billfish and tunas. In many cases data

107 preparation and population dynamics modelling are carried out as separate, discrete stages of
108 the stock assessment process. In the case of VPA assessments of Atlantic tropical tunas the
109 uncertainty in parameter estimates (i.e. model predictions) is normally calculated by
110 bootstrapping the time series of catch per unit effort (CPUE) which is assumed to be a proxy
111 for relative stock abundance and used to calibrate the VPA (e.g. Yamaguchi and Matsuishi
112 2007). This accounts for observation error and not the additional uncertainty and potential
113 biases that may be incurred from the processing of the CAA data. To our knowledge there
114 has been no attempt to rigorously account for the uncertainties of data processing in the
115 outputs of VPA stock assessments.

116

117 In this paper empirically-derived size imputation algorithms are investigated that generate
118 multiple total catch-at-size datasets. We undertake a comparative evaluation of the impact of
119 size imputation relative to typical axes of uncertainty in stock assessments. The hypothesis is
120 evaluated that variability among catch-at-size imputations leads to a greater degree of
121 variability in estimates of management reference points than traditional axes of uncertainty
122 such as natural mortality rate and choice of relative abundance index (e.g. Hillary et al.
123 2015).

124

125 **Methods**

126 *Background to imputation models and their assumptions*

127 We assume that size data are Missing At Random (MAR, Rubin 1976) a missing data
128 condition that assumes: (1) the sizes of missing data can be predicted from observable
129 covariates such as location and time of capture and (2) the probability that an observation is
130 missing is unrelated to the size of the fish itself. For the MAR condition to be violated, size
131 data must be missing due to size itself without any other predictor of the missing size of the

132 fish. For example, a bias towards the measuring of larger fish. In such a case the data will not
133 be MAR and the mechanism is considered ‘non-ignorable’. If however size data are missing
134 due to an observable covariate such as the strength of the fishing gear or size of hook, then
135 these covariates may be used to impute the size of the fish and satisfy the MAR assumptions.
136 In the tuna case studies considered in this paper, the principal reason for a lack of size data is
137 the absence of sampling protocols after capture. If sampling was undertaken all fish were
138 measured regardless of length, which indicates that the MAR assumption is likely to be
139 satisfied in these cases.

140

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

150

151 With this problem in mind, Rubin (1987) developed a procedure known as Multiple
152 Imputation (MI). In essence, MI is EM with replication in which the imputation model is
153 repeatedly fitted to a sample of the full dataset in a similar way to bootstrapping. Typically
154 MI produces between 4 and 10 sets of imputed values. When the mean of the replicate
155 imputations for each data point is used and the standard error incorporated into the final
156 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).

161

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).

168

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).

179

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:

$$(1) \quad P(O_j|M_i) \propto W_{i,j} = P(S, Y, G) = \frac{\exp(-0.5v/(C_{SY}^2 + C_{SG}^2 + C_{YG}^2 - 2C_{SY}C_{SG}C_{YG} - 1))}{2\sqrt{2}\pi^{3/2} \sqrt{1 - (C_{SY}^2 + C_{SG}^2 + C_{YG}^2) + 2C_{SY}C_{SG}C_{YG}}}$$

where

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

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 σ_Y (in years):

$$(3) \quad Y_{i,j} = \frac{|t_i - t_j|}{\sigma_Y}$$

206

207 where time t is converted to a real number (e.g. February 6th 1995 = 1995.11).

208

209 Similarly, standardized geographic distance G is calculated by

210

$$(4) \quad G_{i,j} = \frac{\tan^{-1}(\sqrt{\varphi_{i,j}}\sqrt{1-\varphi_{i,j}})K/(\pi)}{\sigma_G}$$

212

213 where σ_G is the standard deviation (in km), K is the earth's mean circumference (40 041km)

214 and ϕ is given by

215

$$(5) \quad \varphi_{i,j} = \sin\left(\frac{lat_j - lat_i}{2}\right)^2 + \cos(lat_i) \cdot \cos(lat_j) \cdot \sin\left(\frac{lon_j - lon_i}{2}\right)^2$$

217

218 where latitude (lat) and longitude (lon) are expressed in radians.

219

220 Standardized seasonal distance S is calculated from the day of the year (e.g. for January 7th, d

221 = 7). In order to make these 'wrap' between years (so that day 365 and day 1 are adjacent) we

222 used a sine – arcsine transformation:

223

$$(6) \quad S_{i,j} = \frac{360\sin^{-1}(\gamma_{i,j})/\pi}{\sigma_S}$$

225

226 where σ_S is the standard deviation (in days) and

227

$$(7) \quad \gamma_{i,j} = \sin(|\pi(d_i - d_j)/360|)$$

229

230 In some situations catches for which size data are missing were reported in weight rather than
231 the numbers of individual fish. In these cases we sampled a sufficient number of length
232 observations to sum to the reported weight of the catch using the deterministic growth curve.
233 While subsequent VPA analyses involve various growth curves, these are not critical at the
234 imputation stage. This is because alternative growth curves only affect the number of
235 imputations required to meet a target catch weight, rather than the imputed sizes themselves,
236 and the estimated assessment quantities were unrelated to stock magnitude. At the imputation
237 level, alternative growth curves varied overall results less than half of a percent and therefore
238 for the sake of simplicity, a single growth curve was assumed for each stock (based on the
239 first factor level, see below).

240

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

243

244 ***Empirically fitting the Bayesian length imputation model***

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

249

250 To ensure that the imputation models were credible we fitted the correlation coefficients (C_{SY} ,
251 C_{SG} , C_{YG}) and the standard deviations (σ_S , σ_G , σ_Y) to the length composition of 15 fleets that
252 reported length observations with sufficient spatio-temporal coverage (these flag and gear
253 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 (σ).

256

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

260

$$(8) \quad \hat{z}_i = E(z_i) = \sum_j W_{i,j} z_j / \sum_j W_{i,j}$$

262

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 $\hat{z}$ with the real mean length $\bar{z}$ that was observed at the same time and location:

268

$$(9) \quad P(\bar{z}_i | \theta) = \frac{1}{\bar{z}_i s \sqrt{2\pi}} \exp \left(-\frac{(\ln(\bar{z}_i) - \ln(\hat{z}_i))^2}{2s^2} \right)$$

270

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

273

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).

278

279 The BMI approach was implemented using Markov Chain Monte Carlo (MCMC via the
280 Metropolis Hastings algorithm) to numerically approximate the posterior distribution of the
281 correlation and standard deviation parameters. In this way we sampled credible combinations
282 of model parameters (σ , C) for the imputation models and created multiple datasets from
283 them. Since it was possible for some fleets to have length sample data with weak contrast
284 with respect to one of the three dimensions G , S and Y , we prescribed weak prior
285 distributions. The correlation parameters C , were assigned a normal prior with mean zero and
286 standard deviation 0.5. The standard deviation parameters were assigned lognormal prior
287 means of 500km (σ_G), 2 years (σ_Y) and 30 days (σ_Y) all of which had a coefficient of variation
288 of 50%. The model posterior distributions were very similar given alternative priors. For
289 example, lognormal prior means of 2000km, 5 years and 60 days respectively lead to less
290 than 5% changes in mean posterior estimates of the core model parameters. MCMC chains
291 showed rapid mixing and according to trace plots Gelman and Rubin (1992) and Geweke
292 (1992) diagnostics, convergence could not be rejected. Parameter posteriors were sampled
293 after a 'burn-in' of 5000 iterations.

294

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

302

303 *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

329 imputation drives VPA assessments, for each length group size we imputed 20 replicate total-
330 catch-at-length datasets. Imputations were carried out for the time period 1968-2013 and
331 totalled around 950 million yellowfin tuna and 450 million bigeye tuna, the majority of which
332 were imputed at the spatial resolution of 1 degree ocean cell.

333

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

342

343 Both aging approaches require a growth model which is also used by subsequent VPA
344 assessments to convert predicted numbers-at-age to predicted biomass-at-age. Factor 5
345 investigated two levels for the growth model. The first level for both stocks was a von
346 Bertalanffy growth model used in the most recent stock assessments (ICCAT 2015; 2016).
347 The second level was a Gascuel et al. (1992) growth model for yellowfin tuna and a Richards
348 growth curve for bigeye tuna (Hallier et al. 2005) (Figure 3). To ensure theoretical
349 consistency, the same growth curve used to create the total catch-at-age dataset was also used
350 in the subsequent VPA stock assessment. The VPA assessments, which predict numbers at
351 age, require a growth model to calculate spawning biomass for recruitment modelling, and
352 vulnerable biomass for calibration to relative abundance indices.

353

354 In addition to the total catch-at-age datasets, the tuned VPA assessments applied to these
355 stocks (Porch 2013) are calibrated using a time-series of relative abundance, typically a
356 standardized catch-per-unit-effort (CPUE) index. It is common for numerous CPUE indices
357 to be considered that infer conflicting trends in relative abundance which in turn may result in
358 poor fits to these data and bias in assessment model estimates. Therefore in the real stock
359 assessment setting, ICCAT working groups first look at the relative abundance trends
360 inferred by the individual CPUE series and select groups of CPUE indices that represent
361 common hypotheses about trends in stock abundance (ICCAT 2016; 2017). For factor 6, this
362 procedure was used to define two levels for CPUE trend in abundance, one for each group
363 CPUE indices (CPUE I and CPUE II, Figure 3).

364

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

371

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

379

380 *Calculating variables of management interest*

381 The 8 factors of the VPA analysis chain (Figure 2) resulted in a total of 3840 VPA
382 assessments. For each assessment we calculated current spawning biomass relative to
383 biomass at maximum sustainable yield (B/B_{MSY}) and current fishing mortality rate relative to
384 maximum sustainable yield (F/F_{MSY}) as the two principal variables of management interest
385 (Sissenwine and Shepherd 1987). These two variables offer a concise fishery assessment
386 summary because they represent stock status (B/B_{MSY} is commonly used to delineate
387 overfished and underfished stock status) and exploitation rate (F/F_{MSY} is commonly used to
388 delineate overfishing from underfishing). To evaluate the impact of the various VPA analysis
389 stages, assessment results can be plotted as points in the space of B/B_{MSY} and F/F_{MSY} (a
390 ‘Kobe’ plot) which is a standard output in a wide range of fishery assessment settings
391 including the advice framework of the tuna Regional Management Organizations (Kell et al.
392 2016).

393

394 Maximum sustainable yield reference points are often calculated by characterizing the
395 relationship between spawning biomass and recruitment. However in many stock assessments
396 (of which Atlantic yellowfin tuna and bigeye tuna are examples) there is little evidence for
397 such a relationship, and past recruitment varies independent of stock size (Szuwalski et al.
398 2015). Therefore biomass (B_{MSY}) and fishing mortality rate (F_{MSY}) reference points were
399 calculated from mean model estimated recruitment levels and the mean vulnerability-at-age
400 over the most recent three years.

401

402 **Results**403 *Predictive capacity of the Bayesian imputation model*

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

415

416 ***The relative important of assessment factors in determining fishery reference points***

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

425

426 The degree of uncertainty introduced from replicate imputations of the total catch at length
427 data set (factor 3) was inconsistent among the stocks and estimated management reference
428 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_{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_{MSY} and F/F_{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_{MSY} for yellowfin tuna but was pivotal in determining F/F_{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_{MSY}) for yellowfin. For both stocks, by far the most important source of variance in VPA estimates of stock status (B/B_{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_{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_{MSY} for yellowfin.

Uncertainty due to data processing

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

461

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

474

475 Discussion

476 Previous attempts to characterize uncertainty in VPA assessments have focused on
477 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 importance 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).

509

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

528 uncertainty from data processing and can approximate complex historical age vulnerability
529 dynamics.

530

531 The BMI-VPA approach has a number of weaknesses of which high computational
532 requirements are among the most significant. In contrast to other methods for characterizing
533 uncertainty in VPA assessments that do not attempt to deal with data processing (e.g. Hillary
534 2011) our approach can take several hours to generate a suitable number (e.g. 20) multiple
535 imputed datasets (note this is automated and still rapid compared with current manual
536 approach to building the total CAS dataset). The computational demand is not easily solved
537 with parallel computing due to the relatively high memory requirements (typically greater
538 than 10GB per imputation). The spatio-temporal imputation model used to sample sizes is
539 also rather naïve. For example it does not discriminate between longitudinal and latitudinal
540 distances which may be questionable for tropical tuna species that are less likely to inhabit
541 distant latitudes but may mix throughout equatorial longitudes. Similarly, the imputation
542 method disregards coastal and high-seas location and does not make use of spatial covariates
543 that could predict missing length data according to known spawning and rearing areas.

544

545 Acknowledging these weaknesses we describe and demonstrate a novel approach for
546 accounting for uncertainties in data processing in stock assessments. The BMI-VPA method
547 offers substantial benefits over the status quo: it is empirically derived, reproducible and
548 performance can be evaluated by conventional diagnostics of model fit. Imputation models of
549 this type are flexible and may be improved by the inclusion of other covariates relating to
550 life-history, migration and distribution. A future priority is simulation testing to identify the
551 circumstances where the BMI-VPA approach is likely to be most suitable.

552

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.

578

579 **References**

- 580 Agnew, D.J., Pearce, J., Pramod, G., Peatman, T., Watson, R., Beddington, J.R., Pitcher, T.J.
581 2009. Estimating the worldwide extent of illegal fishing. PLoS ONE. **4**(2): e4570.
582 <https://doi.org/10.1371/journal.pone.0004570>.
- 583 Andridge, R. R., Little, R. J. A. 2010. A Review of Hot Deck Imputation for Survey Non-
584 response. Int. Stat. Rev. **78**(1): 40–64. doi: 10.1111/j.1751-5823.2010.00103.x
- 585 Berger, A.M., Grandin, C.J., Taylor, I.G., Edwards, A.M., Cox, S. 2017. Status of the Pacific
586 Hake (whiting) stock in U.S. and Canadian waters in 2017. Prepared by the Joint
587 Technical Committee of the U.S. and Canada Pacific Hake/Whiting Agreement, National
588 Marine Fisheries Service and Fisheries and Oceans Canada. 202 p.
- 589 Butterworth, D.S. and Punt, A.E., 1999. Experiences in the evaluation and implementation of
590 management procedures. ICES J. Mar. Sci. **56**: 985–998,
591 <http://dx.doi.org/10.1006/jmsc.1999.0532>.
- 592 Dempster, A.P., Laird, N.M., Rubin, D.B. 1977. Maximum likelihood via the EM algorithm.
593 J. Royal. Stat. Soc. Ser. B. **39**: 1-22.
- 594 Fonteneau, A., Chassot, E., Bodin, N. 2013. Global spatio-temporal patterns in tropical tuna
595 purse seine fisheries on drifting fish aggregating devices (DFADs): Taking a historical
596 perspective to inform current challenges. Aquat. Living Resour. **26**: 37–48.
- 597 Fournier, D.A., Sibert, J.R., Majkowski, J. and Hampton, J. 1990. MULTIFAN: a likelihood-
598 based method for estimating growth parameters and age composition from multiple length
599 frequency data sets illustrated using data for southern bluefin tuna (*Thunnus maccoyii*).
600 Can J. Fish. Aqua. Sci. **47**: 301-317.

- 601 Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen,
602 A., and Sibert, J. 2012. AD Model Builder: using automatic differentiation for statistical
603 inference of highly parameterized nonlinear models. *Optim. Method. Softw.* **27**: 233-249.
- 604 Fraile, I., Arrizabalaga, H., Rooker, J.R. 2015. Origin of Atlantic bluefin tuna (*Thunnus*
605 *thynnus*) in the Bay of Biscay. *ICES J. Mar. Sci.* **72** (2): 625-635.
- 606 Fromentin, J.M. Bonhommeau, S., Arrizabalaga, H., Kell, L.T. 2014. The spectre of
607 uncertainty in management of exploited fish stocks: The illustrative case of Atlantic
608 bluefin tuna. *Mar. Pol.* **47**: 8-14.
- 609 Gascuel D., Fonteneau A., Capisano A. 1992. A two-stanza growth model for the yellowfin
610 tuna (*Thunnus albacares*) in the eastern Atlantic. *Aquat. Living Resour.* **5** (3): 155-172.
- 611 Gavaris, S. 1988. An adaptive framework for the estimation of population size. Canadian
612 Atlantic Fisheries Scientific Advisory Committee, Research Document, 88/29. 12 p.
- 613 Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences.
614 *Stat. Sci.* **7**: 457-472.
- 615 Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating
616 posterior moments, in *Bayesian Statistics 4*. Bernardo JM, Berger JO, Dawid AP, Smith
617 AFM (eds). Clarendon Press, Oxford, UK.
- 618 Gislason, H., Pope, J.G., Rice, J.C. and Daan, N., 2008. Coexistence in North Sea fish
619 communities: implications for growth and natural mortality. *ICES J. Mar. Sci.* **65**(4): 514-
620 530.
- 621 Goodyear, P. 1997. Fish age determined from length: an evaluation of three methods using
622 simulated red snapper data. *Fish. Bull.* **95**: 39-46.
- 623 Hallier, J.P., Stequert, B., Maury, O., Bard, F-X. 2005. Growth of bigeye tuna (*Thunnus*
624 *obesus*) in the eastern Atlantic Ocean from tagging-recapture data and otolith readings.
625 *Col. Vol. Sci. Pap. ICCAT.* **57**(1): 181-194.

- 626 Hillary, R.M. 2011. Assessing components of uncertainty in VPA abundance and mortality
627 estimates using an alternative exploitation rate-based algorithm. *Nat. Resour. Model.*
628 **25**(4): 574-598.
- 629 Hillary, R.M., Preece, A.L., Davies, C.R., Kurota, H., Sakai, O., Itoh, T., Parma, A.M.,
630 Butterworth, D.S., Ianelli, J. and Branch, T.A., 2015. A scientific alternative to moratoria
631 for rebuilding depleted international tuna stocks. *Fish Fish.* **17**: 469-482.
- 632 ICCAT. 2014. Report of the 2014 Atlantic Bluefin Tuna Stock Assessment Session.
633 International Commission for the Conservation of Atlantic Tunas (ICCAT). Available at:
634 https://www.iccat.int/Documents/Meetings/Docs/2014_BFT_ASSESS-ENG.pdf [accessed
635 June 5th 2017].
- 636 ICCAT. 2015. Report of the 2015 ICCAT bigeye tuna stock assessment session. International
637 Commission for the Conservation of Atlantic Tunas. Col. Vol. Sci. Pap. **72** (1): 86-183.
638 Available at: http://www.iccat.int/Documents/CVSP/CV072_2016/n_1/CV072010086.pdf
639 [accessed April 2017]
- 640 ICCAT. 2016. Report of the 2016 ICCAT yellowfin tuna stock assessment session.
641 International Commission for the Conservation of Atlantic Tunas. San Sebastian. Col. Vol.
642 Sci. Pap. **73** (1): 76-227. Available at:
643 http://www.iccat.int/Documents/CVSP/CV073_2017/n_1/CV073010076.pdf [accessed
644 April 2017]
- 645 ICCAT. 2017. Access to ICCAT statistical databases. Available at:
646 <http://www.iccat.int/en/accesingdb.htm> <http://www.iccat.int/en/accesingdb.htm> [accessed
647 March 25th 2017]
- 648 Kimura, D. 1977. Statistical assessment of the Age-Length Key. *J. Fish. Res. Board Canada.*
649 **43**(3): 317-324.

- Kell, L.T., Die, D.J., Restrepo, V.R., Fromentin, J.M., Ortiz de Zarate, V. and Pallares, P.,
2003. An evaluation of management strategies for Atlantic tuna stocks. *Scientia
Marina*, **67**(1): 353-370.
- Kell, L. T. and Kell. A. 2011. A comparison of age slicing and statistical age estimation for
Mediterranean swordfish (*Xiphias gladius*). *Col. Vol. Sci. Pap. ICCAT*. **66** (2011): 1522-
1534.
- Little, R. J. A. and Rubin, D.B. 1987. *Statistical Analysis with Missing Data*. Wiley, New
York.
- Liu, Q., Pitcher, T., Al-Hossaini, M., 1989. Ageing with fisheries length-frequency data,
using information about growth. *J. Fish Biol.* **35**: 169-177.
- Mason, A., Richardson, S., Plewis, I., Best, N. 2012. Strategy for modelling non-random
missing data mechanisms in observational studies using Bayesian methods. *J. Off. Stat.*
28(2): 279–302.
- Medina, A., Abascal, F. J., Megina, C., Garcia, A. 2002. Stereological assessment of the
reproductive status of female Atlantic northern bluefin tuna during migration to
Mediterranean spawning grounds through the Strait of Gibraltar. *J. Fish Biol.* **60**: 230–217.
- Methot, R.D. and Wetzel, C.R. (2012) Stock synthesis: a biological and statistical framework
for fish stock assessment and fishery management. *Fish. Res.* **142**: 86-99.
- Mishra, G. D., and A. J. Dobson. 2004. Multiple imputation for body mass index: lessons
from the Australian Longitudinal Study on Women's Health. *Stat. Med.* **23**: 3077-3087.
- Miyake, P.M. 2004. A brief history of the tuna fisheries of the world. In *Management of tuna
fishing capacity: conservation and socio-economics*. Second meeting of the Technical
Advisory Committee of the FAO project. Edited by W.H. Bayliff, J.I. de Leiva Moreno
and J. Majkowski. Food and Agriculture Organization of the United Nations, Madrid, pp.
23-50.

- Pinheiro, J. C. and Bates, D. M. (1996). Unconstrained parameterizations for variance-covariance matrices. *Stat. Comp.* **6**(3): 289-296.
- Porch, C.E., 2003. VPA-2BOX Version 3.01 User's guide. *Sustainable Fisheries Division 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.
- Punt, A.E. 1997. The performance of VPA-based management. *Fish. Res.* **29**: 217-243.
- Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016. Management strategy evaluation: best practices. *Fish Fish.* **17**, 303–334, <http://dx.doi.org/10.1111/faf.12104>.
- R Core Team. 2017. R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-9. <http://www.R-project.org>.
- Rose, C., Smith, M.D. 1996. The Multivariate Normal Distribution, *Mathematica Journal.* **6**: 32 – 37.
- Rubin, D. B. 1976. Inference and Missing Data. *Biometrika* **63**: 581-592.
- Rubin, D. B. 1987. Multiple Imputations for Nonresponse in Surveys. John Wiley and Sons, New York.
- Shepherd, J.G. 1999. Extended survivors analysis: an improved method for the analysis of catch-at-age data and abundance indices. *ICES J. Mar. Sci.* **56**: 584-591.
- Sissenwine, M. and Shepherd, J., 1987. An alternative perspective on recruitment over fishing and biological reference points. *Can. J. Fish. Aquat. Sci.* **44** (4): 913–918.
- Szuwalski, C.S., Vert-Pre, K.A., Punt, A.E., Branch, T.A., Hilborn, R. 2014. Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. *Fish Fish.* doi: 10.1111/faf.12083.

- 700 Taylor, N.G., McAllister, M.K., Lawson, G. L. Carruthers, T.R., Block, B.A. 2011. Atlantic
701 bluefin tuna: a novel multistock spatial model for assessment of population biomass. PLoS
702 ONE. 6(12):1:10.
- 703 TTSG. 2012. Alternative virtual population analyses of yellowfin tuna (*Thunnus albacares*),
704 1970-2010. International Commission for the Conservation of Atlantic Tunas, Tropical
705 Tunas Species Group. Col. Vol. Sci. Pap. **68**(3): 1044-1059.
- 706 Van Poorten, B.T., Carruthers, T.R., Ward, H.G.M., Varkey, D.A. 2015. Imputing recreational
707 angling effort from time-lapse cameras using an hierarchical Bayesian model. Fish. Res.,
708 **172**: 265-273.
- 709 WCPFC. 2014. Stock assessment of bluefin tuna in the Pacific Ocean in 2014. Report of the
710 Pacific Bluefin Tuna Working Group. Western and Central Pacific Fisheries Commission,
711 Majuro. WCPFC-SC0-2014/SA-WP-11. Available at: <https://www.wcpfc.int/node/19201>
712 [accessed June 2017]
- 713 Williams, A.J., Farley, J.H., Hoyle, S., Davies, C.R., Nichol, S.J. 2012. Spatial and sex-
714 specific variation in growth of albacore tuna (*Thunnus alalunga*) across the South Pacific
715 Ocean. PLoS ONE. 7(6): e39318. <https://doi.org/10.1371/journal.pone.0039318>
- 716 Yamaguchi, H. and Matsuishi, T. 2007. Effects of sampling errors on abundance estimates
717 from virtual population analysis for walleye Pollock in northern waters of Sea of Japan.
718 Fish. Sci. **73**: 1061-1069.

719 **Tables**

720 Table 1. Posterior mean estimates of imputation model parameters for fleets where a length

721 imputation model was fitted. The coefficient of determination (R^2) provides a measure of the

722 explanatory power of two models: one that estimates just standard deviation parameters (σ ,

723 assumes no correlation structure) and another that also estimates the correlation parameters

724 C . The difference in R^2 value between models with and without the correlation structure is

725 included in the column ‘Delta’. Fleets not included in this table had insufficient spatio-

726 temporal coverage of length composition to fit an imputation model. In these cases length

727 data were still imputed that were observed by these fleets but using imputation model

728 parameters from the corresponding Japanese fleet of the same gear type.

Species	Flag	Gear	No correlation structure				Correlation structure estimated								Delta R ²
			σ_Y	σ_S	σ_G	R ²	σ_Y	σ_S	σ_G	C _{YS}	C _{YG}	C _{YS}	R ²		
Bigeye	Purse seine	USA	0.97	21	352	0.48	1.75	33	555	-0.85	-0.78	0.97	0.69	0.21	
Bigeye	Longline	Japan	9.01	61	1132	0.31	2.60	59	1161	0.65	-0.83	-0.54	0.33	0.03	
Bigeye	Bait boat	Japan	0.45	15	477	0.73	0.99	28	489	0.94	-0.63	-0.57	0.78	0.05	
Bigeye	Purse seine	Japan	9.66	31	8071	0.23	2.75	36	890	0.90	0.91	0.86	0.28	0.05	
Bigeye	Purse seine	Spain	0.45	25	351	0.75	0.66	26	757	-0.02	0.16	0.51	0.82	0.07	
Bigeye	Longline	Brazil	0.48	14	278	0.67	0.59	14	144	-0.72	0.20	0.42	0.71	0.04	
Yellowfin	Purse seine	USA	0.63	42	311	0.29	2.37	73	391	-0.97	-0.55	0.59	0.32	0.03	
Yellowfin	Longline	Japan	2.74	62	1624	0.51	3.65	91	1392	-0.71	-0.32	-0.35	0.60	0.09	
Yellowfin	Bait boat	Japan	0.45	12	625	0.68	0.93	19	512	-0.30	-0.55	0.18	0.74	0.06	
Yellowfin	Purse seine	Japan	0.45	603	220	0.40	1.62	148	1276	0.46	-0.96	-0.48	0.43	0.03	
Yellowfin	Longline	Spain	0.56	18	401	0.49	0.90	21	457	0.57	0.13	0.64	0.51	0.02	
Yellowfin	Bait boat	Spain	0.45	23	200	0.57	0.76	19	115	0.15	0.20	0.21	0.68	0.11	
Yellowfin	Purse seine	Spain	31.88	22	407	0.48	9.47	42	557	0.80	0.30	-0.19	0.48	0.00	
Yellowfin	Longline	Brazil	5.96	49	841	0.39	0.94	67	618	0.95	0.87	0.96	0.71	0.33	
Yellowfin	Bait boat	Brazil	5.34	21	406	0.41	4.30	39	592	0.92	-0.72	-0.76	0.43	0.02	

731 Table 2. The factorial design of the Bayesian Multiple Imputation – Virtual Population
 732 Analysis experiments.

Factor	Level
1. Stock	Bigeye tuna, yellowfin tuna
2. Imputation length group size (n)	100, 1000, 10000
3. Imputation replicates	20 replicates
4. Aging method	Slicing, mixed distributions
5. Growth Model	von Bertalanffy, Richards (bigeye) / Gascuel (yellowfin)
6. CPUE trend in abundance	CPUE I, CPUE II
7. Natural mortality	Working group (WG), Gislason
8. F-ratio (fishing mortality rate in plus group relative to last true age)	Fixed at 1, estimated

733

Draft

734 Table 3. The reduction in residual deviance attributed to each factor of the VPA analysis
735 chain for estimates of F/F_{MSY} and B/B_{MSY} . In this analysis, replicate imputations (factor 3) are
736 treated as a random effect (labelled ‘RE’) and are equal to the residual deviance after
737 accounting for all the remaining, fixed factors. In each case (e.g. estimates of F/F_{MSY} for
738 bigeye tuna) the various factors of the VPA analysis chain are ranked in order of those
739 associated with the greatest variability in the reference point (F/F_{MSY} or B/B_{MSY}).

Factor in VPA analysis chain	Degrees of freedom	Deviance reduction
Bigeye tuna: $\log(F/F_{MSY})$		
3. Replicate imputations	RE	217
6. CPUE trend in abundance	1	191
7. Natural mortality rate	1	96
4. Aging method	1	32
8. FRatio	1	25
5. Growth model	1	6
2. Length group size	2	4
Bigeye tuna: $\log(B/B_{MSY})$		
7. Natural mortality rate	1	1222
3. Replicate imputations	RE	189
6. CPUE trend in abundance	1	64
8. FRatio	1	62
4. Aging method	1	27
2. Length group size	2	8
5. Growth model	1	2
Yellowfin tuna: $\log(F/F_{MSY})$		
8. FRatio	1	224
4. Aging method	1	211
5. Growth model	1	189
3. Replicate imputations	RE	186
7. Natural mortality rate	1	19
2. Length group size	2	7
6. CPUE trend in abundance	1	1
Yellowfin tuna: $\log(B/B_{MSY})$		
7. Natural mortality rate	1	2482
5. Growth model	1	789
8. FRatio	1	642
3. Replicate imputations	RE	554
2. Length group size	2	49
4. Aging method	1	35
6. CPUE trend in abundance	1	2

740

741 **Figures**

742

743 Figure 1. The geographical distribution of total catch in numbers by 10 degree ocean cell
 744 ('CATDIS', ICCAT 2017) for which size must be imputed, panels a and c) and the size data
 745 that are available (panels b and d) from 1970 onwards. The area of the plotted points is
 746 proportional to numbers that are available (maximum point size is represents 20 million fish).

747

748
749 Figure 2. Convergence in the estimates of deviance reduction due to factors of the VPA
750 analysis chain with increasing replicate imputations n . The lines correspond with the values
751 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.

755
756 Figure 4. The fit of the imputation model to mean size data of six fleets fishing bigeye tuna.
757 Plotted points represent mean size observations (x-axis) and the corresponding predicted
758 mean size by the imputation model. Each plot includes estimates of the coefficient of
759 determination (R^2 , explaining the fraction of variability in observations explained by the
760 imputation model) and the parameters of the trivariate normal distribution.

761

762 Figure 5. The fit of the imputation model to mean size data of nine fleets fishing yellowfin

763 tuna. Plotted points represent mean size observations (x-axis) and the corresponding

764 predicted mean size by the imputation model. Each plot includes estimates of the coefficient

765 of determination (R^2 , explaining the fraction of variability in observations explained by the

766 imputation model) and the parameters of the trivariate normal distribution.

767
768 Figure 6. Plot of spawning stock biomass and fishing mortality rate relative to MSY levels
769 (B/B_{MSY} , F/F_{MSY} respectively) for Atlantic bigeye tuna.

770

771 Figure 7. Plot of spawning stock biomass and fishing mortality rate relative to MSY levels

772 (B/B_{MSY} , F/F_{MSY} respectively) for Atlantic yellowfin tuna.

773
774

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.

779 Figure 9. Plot of VPA estimates of spawning stock biomass and fishing mortality rate
780 relative to MSY levels for Atlantic yellowfin tuna. The ellipses encompass 95% of the VPA
781 assessment results for the multiple imputed datasets and the black point is the grand mean.

Draft