GROWTH MODELS FOR FISHERIES: THE EFFECT OF UNBALANCED SAMPLING ERROR ON MODEL SELECTION, PARAMETER ESTIMATION, AND BIOLOGICAL PREDICTIONS

FAY HELIDONIOTIS[1,2]* AND MALCOLM HADDON[2]

[1]*Institute for Marine and Antarctic Studies, University of Tasmania, Private Bag 129, Hobart, Tasmania 7001, Australia;* [2]*CSIRO Marine & Atmospheric Research, GPO Box 1538, Hobart, Tasmania, 7001, Australia*

ABSTRACT Field studies that attempt to estimate the mean growth rates of individuals in a population usually yield unbalanced data and various forms of sampling error. If these data are a poor representation of a population, then the estimates of growth rates may also be unrepresentative of the population. We present an assessment of the performance and uncertainty in 4 growth models—the von Bertalanffy, Gompertz, inverse logistic, and Schnute models—fitted to data with various scenarios of sampling error. The performance of each model was determined by comparing highest likelihoods outcomes to known case outcomes. A Monte Carlo simulation framework was used to generate data consisting of 8 typical scenarios of sampling error common in tag–recapture data. Each growth model was evaluated according to 2 metrics: the error rate (i.e., a metric for model uncertainty) and the prediction error (i.e., the accuracy of biological predictions such as age). Results indicate that an inadequate size range in the data (i.e., a lack of juvenile size classes) would often lead to a high error rate. When negative growth increment data are included, the K parameter of the von Bertalanffy model increased, and the L_∞ decreased. The inverse logistic model sometimes produced absurd parameter estimates but nevertheless generated the lowest prediction errors. A high prediction error can potentially have far more serious implications to fishery stock assessments than is currently appreciated. Given widespread use of the von Bertalanffy and Gompertz models, selected solely on the basis of model selection criteria, it is clear that greater care and scrutiny are warranted in the selection of growth models in the presence of sampling error.

KEY WORDS: growth, von Bertalanffy, Gompertz, inverse logistic, *Haliotis rubra*, Australia

INTRODUCTION

Careful estimates of growth are essential for many studies of marine populations and aspects such as productivity (Summerfelt 1987), fish ecology (Neuheimer & Taggart 2007), life history characteristics (Grover 2005), and the assessment and management of fish stocks (Hilborn & Walters 1992). These aspects may vary spatially and temporally between spatially segregated populations, and may ultimately influence decisions in fisheries management (Hilborn & Walters 1992).

Growth parameters are often used to discriminate between fast and slow-growing populations, and for determining the spatial scale of stocks (Worthington et al. 1995). However, claims of fine spatial scale heterogeneity in growth would be questionable if variability in growth parameters was the result of artifacts of sampling error in the data used. In Tasmania, where the blacklip abalone fishery (*Haliotis rubra*, Leach) accounts for ~25% of the global abalone wild catch (FAO 2006), careful estimates of growth are important for effective management (Haddon et al. 2008, Haddon 2011).

The Tasmanian abalone fishery consists of hundreds of spatially discrete populations that may be ecologically dissimilar at fine spatial scales (tens or hundreds of meters) (Prince et al. 1987, Nash 1992, Miller et al. 2008). Ideally, different populations would be sampled consistently, but this is rarely possible when there are numerous populations, particularly when samples are taken underwater and under variable conditions. In the case of tag–recapture data, there is usually an inconsistent number of observations per size class between samples, which gives rise to unbalanced sampling error in the data and it is not known how it may influence parameter

estimates. This presents a particular problem when interpopulation comparisons are important to the scientific understanding and management of marine populations. Inconsistent sampling may obscure biological findings and thwart comparisons between populations. It is therefore important to consider the effect of unbalanced sampling error and measurement error on model uncertainty (Francis & Shotton 1997).

Uncertainty in growth model selection often derives from a lack of data on the growth of juveniles, which is the size range with the greatest discriminating power between candidate models (Urban 2002, Rogers-Bennett et al. 2007). Unfortunately, for many species, including abalone, tagging smaller juveniles (<50 mm in the case of *Haliotis rubra*) is not a viable option because the process of tagging is likely to affect growth (Day & Fleming 1992, Wang 1998). Clearly, it is important to specify minimum data requirements when analyses lead to management decisions (Francis & Shotton 1997, Punt 2006), and therefore it is important to determine the extent to which a paucity of data from juvenile size classes affects model uncertainty.

Another source of model uncertainty arises from a thin choice in candidate models (Katsanevakis & Maravelias 2008). This may lead to "retrospective regret" in the selected model because a larger range of plausible models was not considered (Draper 1995, Hamilton et al. 2007, Katsanevakis & Maravelias 2008). With the exception of 1 study (Rogers-Bennett et al. 2007), the selection of growth models for abalone has generally been limited to 2 models—namely the von Bertalanffy and the Gompertz models (Day & Fleming 1992). Historically, there has been strong reliance on the von Bertalanffy model and, if the von Bertalanffy was not the best fitting model, then the Gompertz was selected, effectively by default despite alternative models being available (Rogers-Bennett et al. 2007). The inverse logistic model was recently developed and was found to be statistically optimum for the majority of blacklip abalone

populations in Tasmania (Haddon et al. 2008, Helidoniotis et al. 2011). The Schnute model (Schnute 1981, Francis 1995) is another alternative that has been used for its flexibility, and it offers the prospect of accurately describing different growth trajectories (Worthington et al. 1995); however, with abalone, the Schnute model tends to assume a growth trajectory similar to either the von Bertalanffy or Gompertz models.

Aspects of sampling error that can affect model uncertainty include sample size, the size range of observations (Rogers-Bennett et al. 2007), and the relative density of samples within a size range (Wang 1998). Typical tag–recapture data consist of relatively fewer small and large specimens, with more specimens in the midsize range (Wang et al. 1995). In addition, observations of apparent negative growth increments are common in tagging studies despite being biologically absurd. It is unknown whether growth models are robust to these kinds of data characteristics, and model uncertainty may arise if there is incomplete information on the population and community dynamics in the data (Francis & Shotton 1997). Therefore the likelihood and consequences of model uncertainty should be considered.

Model uncertainty can lead to misleading outcomes and invalid conclusions (Cox 2002). Nevertheless, key decisions in fisheries management must be made, usually on an annual basis, and rely on information from models. In practice, the true growth trajectory is usually not known, and stochastic versions have been formulated instead. For example, a well-known disadvantage of the deterministic von Bertalanffy is that parameter estimates are biased if the growth variation of individuals is ignored (Sainsbury 1980, Wang & Thomas 1995, Eveson et al. 2007). A solution is to use probability density functions (pdf's) around the K and L_∞ parameters (Troynikov et al. 1998, Eveson et al. 2007). The deterministic von Bertalanffy and Gompertz models require pdf's around the L_∞ to avoid predicting negative increments for larger size classes. However the inverse logistic model achieves this without requiring a pdf around its parameters (Helidoniotis et al. 2011). One approach toward evaluating the implications of model uncertainty in the presence of unbalanced sampling error is to use simulation testing based on Monte Carlo techniques. In the current work, the simulation framework consists of an operating model that represents a "known" outcome. This was used to generate simulated data, including the typical scenarios of different sampling errors. The operating model can take the form of any growth model being considered and then candidate growth models are fitted to the simulated data. The highest likelihood outcomes are then compared with the known case outcomes (Marcot 2012). To study the performance of each candidate growth model we selected 2 criteria: model-fitting criteria and biological criteria. For each criterion, a metric was developed to quantify performance. For the model selection criterion, the metric developed was the error rate addressing model uncertainty (i.e., the probability that an incorrect model was selected as the "best" statistical model). For the biological criterion, the metric was the prediction error (i.e., the accuracy of biological predictions such as age). Blacklip abalone from a site in Tasmania, Australia, were used as a case study, and the performance of 4 growth models are tested against 8 plausible scenarios of sampling error. The 8 scenarios combined address unbalanced data, sampling errors, and different degrees of bias. This study examines differences between the best-fitting candidate model and the operating model under different scenarios of simulated sampling error.

MATERIALS AND METHODS

The Growth Models

Assessments were made of 4 deterministic candidate growth models used to describe growth in abalone: the von Bertalanffy, Gompertz, inverse logistic, and Schnute models. The candidate growth models include a reparameterized analog of the von Bertalanffy model for tag–recapture data used for estimating expected length increments following specific time increments (Fabens 1965, Haddon 2011).

The standard von Bertalanffy model is given as

$$\Delta\hat{L} = (L_\infty - L_i)(1 - e^{-K\Delta t}) + \varepsilon \tag{1}$$

The reparameterized Gompertz (Troynikov et al. 1998) model for estimating length increments following time increments is

$$\Delta\hat{L} = L_\infty \left(\frac{L_i}{L_\infty}\right)^{\exp(-g\Delta t)} - L_i + \varepsilon \tag{2}$$

The inverse logistic model (Haddon et al. 2008) is

$$\Delta\hat{L} = \frac{Max\Delta L}{1 + e^{Ln(19)\frac{(L_i - L_{50}^m)}{(L_{95}^m - L_{50}^m)}}} + \varepsilon \tag{3}$$

A length-based analog of the Schnute model (Francis 1995) is

$$\Delta\hat{L} = [L_i^b e^{-a\Delta t} + c(1 - e^{-a\Delta t})]^{1/b} - L_i + \varepsilon \quad \text{if} \quad a \neq 0, b \neq 0 \tag{4}$$

where $\Delta\hat{L}$ is the expected length increment, L_∞ is the shell size when the mean length increment is 0 (von Bertalanffy and Gompertz), L_i is the initial length when first tagged and released, K is the rate of change in length increment with increasing shell size, g is the exponential decrease in growth increment and is less than 0, Δt is the time at liberty (as a fraction of a year), $Max\Delta L$ is the maximum growth increment, L_{50}^m is the initial length that produces a growth increment of $0.5Max\Delta L$, L_{95}^m is the initial length that produces a growth increment of $0.05Max\Delta L$, b describes curvature in the Schnute model, a "has no simple biological meaning" (Francis 1995) but scales the rate at which the curve approaches an asymptote if there is one, c is the growth curve shape parameter, and ε represents the independent, additive, normal random residual errors.

The Simulation Testing Framework

The simulation testing framework consisted of a training data set, testing data sets (simulated), operating models, and candidate models. The starting parameters for the operating models were derived by fitting the models to a training data set. Simulation testing involved systematically selecting a particular growth model as the operating model and then generating simulated testing data sets using Monte Carlo methods. The 4 candidate growth models were fitted, in turn, to the simulated testing data sets. The operating models represented the underlying biological reality, which defined the correct model, and the simulated testing data sets, generated from the operating

models, represented a population with known growth parameters. The simulated growth data was assumed to incorporate both stochastic process and observation uncertainty.

Training Data Set

To select realistic parameters for each operating model, a training data set of tag–recapture data was selected from a field survey at Black Island, Tasmania, Australia (42.86° S, 148.00° E; Fig. 1). The selection of the training data set was independent of the growth models and was representative of growth increments for blacklip abalone in some populations of Tasmania's blacklip abalone fishery. This data set was made up of yearly length increments from juvenile and mature abalone whose size at tagging ranged from 57–171 mm. Growth increments that were more negative than –3 mm were removed (Fig. 1). Note that –3 mm was selected because it was assumed to allow for measurement error applicable to all size classes that is not necessarily related to shell chipping; this is common practice in treating growth increment data in abalone (see Naylor et al. (2003)). The final sample consisted of 116 observations, which is at the low end of a typical sample size of tag–recapture data for populations around Tasmania.

Operating Model and Starting Parameters

Operating model parameters were obtained by fitting 3 of the growth models (von Bertalanffy, Gompertz, and inverse logistic) to the training data. Despite its flexible nature, the Schnute was often similar to either the von Bertalanffy or the Gompertz when fitted to data. The Schnute model, therefore, was not suitable as an operating model because it does not have a unique form. Only the von Bertalanffy, Gompertz, and inverse logistic models were used as operating models, but all 4, including the Schnute, were used as candidate models.

Simulating Testing Data Sets

A Monte Carlo simulation framework was used to generate data sets from each operating model. For each of the 8 data scenarios, 500 data sets were generated from each of the 3 operating models. A sample of initial lengths was selected randomly from the size range of the selected scenario. Normal random errors were added to the predicted length increments produced by each operating model in each simulated data set, and the variance was equal for each operating model. This gave all 3 operating models the same process and observation errors (Zhou 2007) and reflects the typical distributions observed in the size
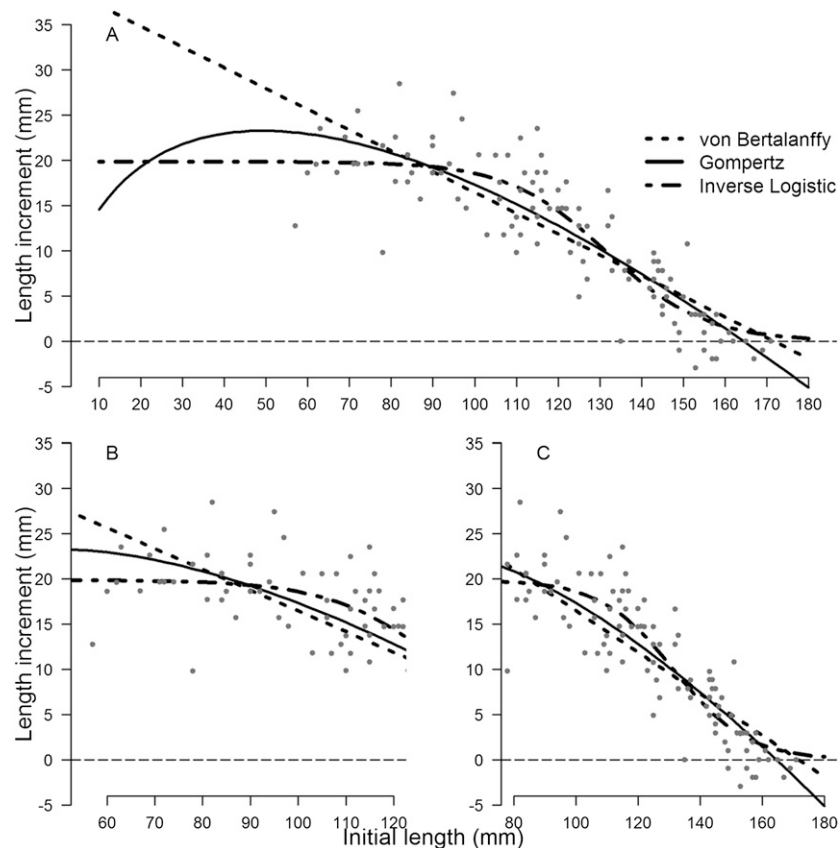

Figure 1. (A–C) The von Bertalanffy, Gompertz, and inverse logistic growth models fitted to tag–recapture data that were used as the training data set. The data are from Black Island (42.9687° S, 145.4924° E) and were the best example of tag–recapture data in terms of sample size and initial size range (A). In this data set, the size range most underrepresented is in the juvenile size range (i.e., in which the initial length (Lt) is smaller than 60 mm. (B) The relationship of the model with respect to smaller abalone (Lt, 55–120 mm) in which most of the discriminating power takes place, although the high degree of scatter in this region may reduce the discriminating power between models. (C) The relationship of the models with respect to larger abalone (Lt, 80–170 mm), which is adequately represented in typical tag–recapture data. The Schnute model is not represented here because it assumes either a Gompertz or von Bertalanffy growth trajectory.

distributions of younger cohorts (Helidoniotis & Haddon 2012). The alternative of using a nonconstant variance was not used because that would have led to the 4 growth model forms receiving unequal treatment with respect to residuals. Furthermore, normal random errors were used because this type of error structure is the most common assumption used in fitting growth models to tag–recapture data (Hilborn & Walters 1992). However, if a different probability density function (other than the normal distribution) provides a better fit for some other species or population, then the expectation would be to implement that error structure. In fitting growth models to numerous abalone populations, the estimated SD of normal errors around the model may range from 2.5–6.8 ($n = 27$ from samples collected around Tasmania) (Helidoniotis et al. 2011). An SD of 3.69 was therefore selected as a reasonable estimate for all operating models in the simulations.

Scenarios of Unbalanced Sampling Error

Two categories of test data sets were considered—namely, ideal data sets and nonideal data sets, the latter of which consisted of the 8 scenarios of sampling error. The properties of the ideal data sets were determined from considering what characteristics would constitute ideal field data (Table 1). These included a large sample size ($n = 540$) and an initial size range that extended over both the juvenile and adult size ranges (50–180 mm). The ideal data set was used for 2 purposes: first, to act as a positive control to test the capacity of the candidate models to recover the correct underlying growth model when available data are ideal, and second, to examine the effect of negative growth increments on parameter estimation. Negative increments are commonplace in tag–recapture data (Quinn & Deriso 1999) and reflect measurement error or, in the case of abalone, occasional shell damage (Francis & Shotton 1997). Adding normal random error to simulated growth increment data can also result in negative increments, especially for the larger initial lengths. Using the ideal data set, the effect of normal random measurement error on parameter estimates, arising from negative increments, was examined in 3 ways: (1) by removing all negative increments, (2) by removing data with increments that are more negative than –3 mm, and (3) no removal.

For the nonideal data sets, the various forms of sampling error were characterized by considering typical scenarios of sampling error in tag–recapture data from 20 populations around Tasmania. The characteristics of sampling error were grouped into 3 main categories: (1) inadequate sample size, (2) small size range consisting of small or large individuals, and/or (3) variable relative density of observations taken across size classes. Two levels were selected within each of the 3 categories, which, when considered in factorial combination, yielded 8 realistic scenarios of sampling error (Table 1). For sample size, the 2 levels were $n = 100$ and $n = 500$, encompassing the lower and upper limits of the 20 available empirical data sets (where $n = 100$–250 is typical). For size range, observations indicate that the minimum initial lengths at release from 20 populations ranged from 47–82 mm. Thus2two different initial size ranges at tagging were used (55–120 mm and 80–170 mm) to reflect a lack of large animals and small animals respectively. The 55–120-mm size range represents predominantly juvenile animals (minimum change in length increment, $\Delta L > 10$ mm), whereas the size range 80–170 mm includes late juveniles to late-stage adults with negligible growth (minimum $\Delta L = 0$). This larger size range is more common in tag–recapture data. For the relative density of observations, 2 different density ratios for mid- and large-size animals (0.3 and 0.6, and 0.7 and 0.2, respectively) were used, and there was a constant density of 0.1 for the smaller animals. Relative density is associated with the relative probability of recovering abalone of different size classes (Wang 1999). Growth increments more negative than –3 mm were omitted from all simulated data sets to reflect typical practice involving tag–recapture data (Haddon et al. 2008).

Fitting Candidate Models to Data

With the exception of the ideal test data sets, all 4 candidate growth models were fitted to each of the 500 testing data sets simulated by the 3 operating models. The Schnute model was not fitted to the ideal testing data set because it was not used as

---

TABLE 1.

A description of the sampling scenarios for 2 categories of simulated data sets: ideal and nonideal.

Scenario	Sample Size (n)	Sampling Density, P; Low–Mid–Large	Length Range, R (mm)	Increments Included
Ideal data	540	0.33–0.33–0.33	50–180	All negative increments
	540	0.33–0.33–0.34	50–181	More than –3 mm
	540	0.33–0.33–0.35	50–182	Positive increments only
Nonideal data				
A	100	0.1–0.3–0.6	55–120	More than –3 mm
B	100	0.1–0.3–0.6	80–170	More than –3 mm
C	100	0.1–0.7–0.2	55–120	More than –3 mm
D	100	0.1–0.7–0.2	80–170	More than –3 mm
E	500	0.1–0.3–0.6	55–120	More than –3 mm
F	500	0.1–0.3–0.6	80–170	More than –3 mm
G	500	0.1–0.7–0.2	55–120	More than –3 mm
H	500	0.1–0.7–0.2	80–170	More than –3 mm

For the nonideal data set 8 scenarios (A–H) of unbalanced sampling error are described that are typical of abalone tag–recapture surveys in Tasmania. Three aspects of sampling error are sample size, sampling density, and the initial lengths tagged. For sampling density, Low corresponds to the size range that is in the lower 25% of the initial length range, Mid represents 25%–75% of the initial length range, and Large is more than 75% the initial length range.

an operating model and therefore it had no utility as a possible control or no use in tests for effects of negative increments. To improve the efficiency of the fitting process, the parameters of the operating models (and the Schnute model) were used as the starting parameters for the candidate models when fitted to simulated testing data (Table 2). Each candidate model was fitted to each of the testing data sets using maximum likelihood methods (described later).

Assessing Model Performance

The most standard approach for determining which model is appropriate is to use model selection based on information criteria such as Akaike's information criteria (AIC). However, as the use of growth models increases, it becomes important to evaluate their performance and uncertainty in their application to real-world problems (Marcot 2012) in addition to standard model selection. Real-world problems that are biological in nature (e.g., age) are useful to both fisheries and conservation, and therefore a biological criterion was included. An example of a biological criterion is the time taken for a population to reach legal size and enter the fishery (Rogers-Bennett et al. 2007). This is equivalent to the average minimum age of the population when it can be harvested legally. The performance of the models was therefore evaluated according to 2 criteria: model-fitting criteria and biological criteria. Metrics were developed to quantify performance within each criterion. For the model selection criterion, the metric used was error rate, and the diagnostics used was the distribution of parameter estimates relative to the "known" values of the operating model. For the biological criterion, the metric used was prediction error (e.g., differences in

time-to-fishery estimates between the operating model and the candidate models) as a result of unknowingly selecting the incorrect model. In addition, the "correct" candidate model was compared with the related operating model by calculating the coefficient of variation (CV) in time-to-fishery estimates (Rogers-Bennett et al. 2007). Therefore, the CVs measure variation only under the related operating model, (e.g., the CVs of the von Bertalanffy time-to-fishery estimates are for data generated under the von Bertalanffy operating model), and a CV was calculated for the 500 simulations under each scenario in turn; hence, a range of CV estimates was thus obtained (i.e., there were 8 data scenarios, and therefore 8 CVs were obtained for each candidate model).

Criterion 1: Model Selection and Error Rate

The optimum statistical model was selected on the basis of AIC, which is defined as AIC $= -2\Sigma LL_m + 2K_m$, where K_m is the total number of parameters for model m (including σ_m, the SD for model m), and $-\Sigma LL_m$ is given by

$$-\Sigma LL_m = -\sum_{Lt=1}^{n} Ln\left(\frac{1}{\sigma_m\sqrt{2\pi}}e^{-\left[\frac{(\Delta L_{Lt}-\Delta\hat{L}_{Lt})^2}{2\sigma_m^2}\right]}\right) \qquad (5)$$

where ΔL_{Lt} is the observed length increment for each of the n initial lengths at tagging, Lt; and $\Delta\hat{L}_{Lt}$ is the expected length increment for initial length Lt. The negative log-likelihood for model m, $-\Sigma LL_m$, was minimized using the optim function in R (R Development Core Team 2008). The SD of the residuals was estimated and stored along with model parameter values and related model fits at each run. In each case, the candidate model that was related mathematically to the operating model was termed the correct model, whereas models that were unrelated to the operating model were termed the incorrect models (Zhou 2007).

The candidate model with the smallest AIC was selected as the "best," or statistically optimum model, regardless of whether it was a correct or incorrect model. The probability of selecting the best model was calculated as the proportion of the 500 simulations, for each of the 8 scenarios of data quality, for which the correct candidate model had a lower AIC than other candidate models (Table 3). If the 3 main candidate models, ignoring the Schnute, were to perform equally well under the 8 data scenarios (Table 1), the expectation would be that the probability of the correct candidate model being selected as the statistically best model should be $P = 0.33$. The Schnute model was omitted because it was found invariably to mimic either the von Bertalanffy model or the Gompertz model. A probability gradient between $P = 0.33$–1.00 was used to determine which was the best candidate model, where $P < 0.33$ indicates ambiguity and $P \geq 0.95$ indicates an unambiguous result.

The effect of negative increments on the variability of parameter estimates across each of the 500 simulated data sets was examined by plotting pairs of parameters from the sets generated by each candidate model. If the available data adequately represent the growth trajectory for a given candidate model, then the expectation was that estimated parameter values of the candidate model would be distributed approximately normally about the known parameter values from the operating model, and would be centered around this parameter value (Hilborn &

TABLE 2.

Starting parameter values for the operating and candidate models used in simulations.

	Parameters	Starting Values	
von Bertalanffy	L_∞	172	
	K	0.261	
	SD	3.69	
Gompertz	L_∞	165	
	g	0.387	
	SD	3.69	
inverse logistic	$Max\Delta L$	20	
	L_{50}	132	
	L_{95}	166	
	SD	3.69	
Schnute		**Gz-Schnute**	**VB-Schnute**
	b	0.002	0.854
	a	0.387	0.278
	c	1.012	80.7
	SD	3.69	3.69

Starting parameters were obtained from a training set of tag–recapture data from a population of blacklip abalone in Tasmania. Three operating models were used: the von Bertalanffy (VB), Gompertz (Gz), and inverse logistic. The Schnute model was used only as an candidate model to the 3 other operating models. Two combinations of starting parameters were necessary for the Schnute candidate model; Gz-Schnute simulates a Gompertz model, whereas VB-Schnute simulates a von Bertalanffy model.

A summary of the probabilities of the candidate model being selected as the best statistical model, as determined by the Akaike's information criteria, when fitted to simulated data from the 3 operating models.

Operating Model		von Bertalanffy				Type I Error	Gompertz				Type I Error	Inverse logistic				Type I Error
Candidate Model		VB	Gz	IL	Sch		VB	Gz	IL	Sch		VB	Gz	IL	Sch	
Ideal data, $n = 540$																
All negative increments		**1**	0	0	0	0.00	0	**1**	0	0	0.00	0	0	**1**	0	0.00
Negative increments > –3 mm		**0.9**	0	0	0	0.14	0	**1**	0	0	0.00	0	0	**1**	0	0.00
Positive increments (>0 mm)		**0.1**	0.0	0.9	0	0.89	0	**1**	0	0	0.00	0	0	**1**	0	0.00
Nonideal data (sampling error)																
$n = 100$	A	**0.65**	0.28	0.07	0	0.35	0.25	**0.69**	0.06	0	0.31	0.07	0.56	**0.37**	0	0.63
	B	**0.59**	0.09	0.32	0	0.41	0.41	**0.35**	0.24	0	0.65	0.02	0.01	**0.97**	0	0.03
	C	**0.69**	0.26	0.05	0	0.31	0.28	**0.69**	0.03	0	0.31	0.15	0.68	**0.17**	0	0.83
	D	**0.62**	0.17	0.21	0	0.38	0.35	**0.47**	0.18	0	0.53	0.01	0.03	**0.96**	0	0.04
$n = 500$	E	**0.80**	0.07	0.13	0	0.20	0.11	**0.82**	0.07	0	0.18	0.00	0.13	**0.87**	0	0.13
	F	**0.67**	0.01	0.31	0	0.33	0.55	**0.25**	0.20	0	0.75	0.00	0.00	**1.00**	0	0.00
	G	**0.82**	0.06	0.12	0	0.18	0.08	**0.85**	0.07	0	0.15	0.01	0.22	**0.77**	0	0.23
	H	**0.72**	0.00	0.28	0	0.28	0.40	**0.43**	0.17	0	0.57	0.00	0.00	**1.00**	0	0.00

All 4our candidate models were fitted to each operating model for both ideal and nonideal data sets that consisted of 8 scenarios of sampling error (A–H, Table 1). The 4 candidate models are von Bertalanffy (VB), Gompertz (Gz), inverse logistic (IL), and Schnute (Sch). The ideal data are included as a positive control for the simulation against which the 8 data scenarios are compared. For each operating model, the "correct" candidate models is identified in bold type. The probability of type I error is shown for each scenario under each operating model.

Mangel 1997). If the available data were not representative of the population, then the expectation was that the scatter of the candidate model parameter estimates around the known parameters, from the operating model, would be skewed or otherwise distorted. A robust model was considered to be a candidate model with a relatively narrow CV for its parameter estimates, and a model that accurately recovers the parameters of the related operating model.

Criterion 2: Biological Criteria and Prediction Error

The robustness of the growth models was also assessed using biological criteria. A biological criterion used in other studies of abalone growth was the time taken for abalone to recruit to the fishery (Rogers-Bennett et al. 2007). In Tasmania, abalone recruit to the fishery after they reach a legal minimum shell length (LML). The LML differs across regions around Tasmania, and the appropriate LML at Black Island (the site of the training data set) is 140 mm. Because the juvenile stage for abalone is considered to commence at the formation of the first respiratory pore at about 2 mm in shell length (Cropp 1989), the time to fishery was determined by the number of years taken for an abalone to grow from 2–140 mm. Time-to-fishery estimates for the candidate models were compared with the equivalent estimates from the operating model (which represents the known time-to-fishery estimate).

The von Bertalanffy and Gompertz models can be rearranged with respect to Δt to estimate the time taken (in years) for abalone to recruit to the fishery: Time to fishery for von Bertalanffy was

$$\Delta t = \frac{-\ln\left(1 - \left(\dfrac{\Delta L}{L_\infty - L_t}\right)\right)}{k} \qquad (6)$$

Time to fishery for Gompertz was

$$\Delta t = \ln\left(\ln\frac{\Delta L + L_t}{L_\infty} \times \left(\ln\frac{L_t}{L_\infty}\right)^{-1}\right) \times g^{-1} \qquad (7)$$

For the inverse logistic, estimating time to fishery was more involved. Growth has to be simulated iteratively to generate length at age in estimating the time taken to reach the LML (i.e., to enter the fishery). To estimate the time taken to reach the LML, the process starts with $L_t = 2.0$ and is iterated with time steps of 1 wk until $L_{t+1} \geq 140$ mm using

$$L_{t+1} = L_t + \frac{Max\Delta L(\Delta t + C\sin(2\pi(t_R - p)) - C\sin(2\pi(t_T - p)))}{1 + e^{Ln(19)\left(\frac{L_t - L_{50}}{L_{95} - L_{50}}\right)}} \qquad (8)$$

where C is the amplitude of the seasonality effect for ΔL; t_R and t_T are the dates of recapture and tagging, respectively (as fractions of a year; for example, June 30 = 0.5; $t_R = t_T + \Delta t$); and p is the date at which the seasonal growth rate was maximum (as a fraction of a year).

Time to fishery for Schnute was

$$\Delta t = -\ln\frac{(\Delta L + L_t)^b - c}{L_t^b - c} \times a^{-1} \qquad (9)$$

The growth model equations were rearranged according to Eqs (6)–(8) to obtain time-to-fishery estimates (Δt) for each of the 500 simulations under 8 data scenarios. The parameters returned by the 500 simulations were substituted into Eqs (6)–(8) to obtain time-to-fishery estimates for each of the 500 simulations of the 8 data scenarios. Fixed estimates of L_t and ΔL were used with $L_t = 2$ mm, and $\Delta L = 138$ mm (i.e., 140 – 2 mm, being the length increment between the commencement of the juvenile phase and the size when individuals enter the fishery). A robust candidate model was considered to be one with the relatively

smallest difference in time-to-fishery estimates, between the candidate and operating models, across all scenarios of sampling error.

In reality, the true form of the growth model was unknown and the goal of this analysis was to quantify the possible consequence of unknowingly selecting the incorrect model. Growth models have many uses, and so it was important to acknowledge that the estimate of the prediction error used in this study (i.e., to estimate time to fishery) will not be similar if calculated for another purpose. The equation for calculating the prediction error involved 2 components: (1) the biological accuracy determined as the difference between the mean estimates of the incorrect candidate models and the unrelated operating model, and (2) the variance in time-to-fishery estimates for each incorrect candidate model across all data scenarios combined (Table 4). Relatively high variance is an indication of a high prediction error.

$$\text{Prediction error} = \frac{\left(\left(\sum_{o=1}^{o_t}\sum_{d=1}^{8}(T_o - T_m)_{o,d}^2\right) + \text{var}\right)}{o_t} \quad (10)$$

where $T_o - T_m$ was the sum of the difference in time to fishery (i.e., the mean difference in time-to-fishery estimates between the incorrect candidate model m and the unrelated operating model o for each data scenario d, where the mean refers to the mean over all 500 simulations), var was the variance (for the incorrect candidate model m under the (unrelated) operating model o for all data scenarios combined), o was the unrelated operating model (von Bertalanffy, Gompertz, or inverse logistic), o_t was the number of unrelated operating models (i.e., $o_t = 2$ for the incorrect von Bertalanffy, Gompertz, and inverse logistic candidate models; $o_t = 3$ for the incorrect Schnute candidate model) and d was the data quality scenario (Table 1).

The higher the prediction error for a given candidate model, the higher the potential risk in using that growth model for fishery management. A weighting was applied by squaring the difference (between the candidate and operating models) to help differentiate between a high-difference/low-variance score and a low-difference/high-variance score. This weighting reflects that a high-difference/low-variance score was more risky than a low-difference/high-variance score.

RESULTS

The von Bertalanffy, Gompertz, and inverse logistic growth models differ in their description of growth, particularly in the juvenile sizes (<80 mm in the fits of Fig. 1), indicating that juvenile size classes are necessary for minimizing model uncertainty and optimizing model selection. Typical tag–recapture data in abalone do not adequately represent the full size range of juveniles, and where there was a (characteristic) lack of data on juveniles, it was expected that the ability to discriminate between growth models would be weak (Fig. 1).

The Effect of Negative Increments on Model Uncertainty

Using ideal data in which sample size is large ($n = 540$) and when growth increments less than –3 mm were removed, the probability of obtaining the correct von Bertalanffy, Gompertz, and inverse logistic models were all high ($P > 0.9$). This result was expected and confirmed that the simulation framework used in this study can produce an expected outcome, and therefore any unexpected outcomes are likely to be real and not the result of unknown idiosyncrasies of the methodology used. When using a similar sample size ($n = 500$) but in the presence of sampling error (scenarios E–H) the probability of selecting the "correct" von Bertalanffy and Gompertz models decreased substantially relative to the ideal data, with P ranging from 0.59–0.82 for the von Bertalanffy and 0.25–0.85 for the Gompertz. This isolates the effect that size range and proportion sampled may be having on model selection.

The removal of all negative increments dramatically reduced the probability of the von Bertalanffy model being selected as the best statistical model ($P = 0.1$, Table 3); instead, the "incorrect" inverse logistic model had a high probability ($P = 0.9$, Table 3) of being selected under a von Bertalanffy operating model. The other candidate models (i.e., the Gompertz and inverse logistic) were selected in all instances under their corresponding operating

TABLE 4.

The prediction error (Eq (10)) for quantifying the implications of selecting the incorrect candidate model.

Candidate Model	von Bertalanffy		Gompertz		Inverse logistic		Schnute		
Operating Model	Gz	IL	VB	IL	VB	Gz	VB	Gz	IL
A	–2.12	–1.42	2.08	1.10	0.72	–1.18	2.07	0.01	1.10
B	–2.45	–2.33	3.11	0.30	1.60	–0.38	0.10	–2.44	–2.33
C	–2.17	–1.41	2.29	1.16	0.68	–1.27	0.68	–1.27	–0.13
D	–2.49	–2.59	2.98	–0.18	1.53	–0.51	1.32	–1.36	–1.55
E	–2.13	–1.40	2.06	1.10	0.78	–1.10	2.04	–0.02	1.08
F	–2.46	–2.38	3.11	0.24	1.70	–0.31	0.56	–2.03	–1.97
G	–2.20	–1.45	2.21	1.13	0.74	–1.08	2.25	0.04	1.18
H	–2.52	–2.62	3.11	–0.11	1.67	–0.37	0.33	–2.29	–2.41
$\Sigma(\text{difference}^2)$	75.85		61.82		18.74		56.16		
Variance	0.21		1.37		1.21		2.40		
Meanprediction error	38		32		10		20		

Mean residuals in time-to-fishery estimates between the incorrect candidate model and operating models from 500 simulations are presented. The 3 operating models were von Bertalanffy (VB), Gompertz (Gz), and inverse logistic (IL). The prediction error for each candidate model (von Bertalanffy, Gompertz, inverse logistic, and Schnute) are shown for 8 data quality scenarios (A–H; refer to Table 1 for description of scenarios).

model and were therefore unaffected by removing negative increments (Table 3, Fig. 2). Removing all negative increments had a marked affect on the parameters of the von Bertalanffy, resulting in a bias in the L_∞ and K parameters (Fig. 2). In contrast, parameters of the Gompertz and inverse logistic candidate models were not biased by the removal of negative increments (Fig. 2).

Criterion 1: Model Selection and the Effect of Unbalanced Sampling on Model Uncertainty

von Bertalanffy

Using nonideal data, the probability of the correct von Bertalanffy candidate model being selected statistically under its related operating model was relatively low ($P = 0.59$–0.82 across all scenarios), indicating that the von Bertalanffy was not statistically robust to all scenarios of unbalanced sampling error (Table 3). The 2 data scenarios in which the correct von Bertalanffy had the greatest probability of being selected were E and G ($P = 0.80$–0.82), in which sample size was high ($n = 500$) and the initial size range consisted of smaller animals (50–120 mm). Nevertheless these probabilities were still below the (arbitrary) unambiguous probability of $P \geq 0.95$. Scenarios B and D resulted in the lowest probability of the correct von Bertalanffy being selected as the best statistical model. This suggests that a low sample size combined with a lack of smaller animals was the weakest data scenario for a von Bertalanffy model.

Gompertz

The probability of the correct Gompertz candidate model being selected for any scenario under its related operating model ranged between 0.25–0.85 (Table 3). The correct Gompertz candidate model occurred when sample size was high ($n = 500$) and initial sizes were low (50–120 mm; probability, 0.82 and 0.85; scenarios E and G; Table 3). Nevertheless, these probabilities remain below the unambiguous probability of $P \geq 0.95$. For the Gompertz model, the biggest disadvantage was data lacking smaller juveniles (<80 mm; scenarios B, D, F, and H) as there was high ambiguity in the selection of the best model, and an increase in sample size to $n = 500$ did not help increase the probability of the correct Gompertz ($P = 0.25$–0.47) (Table 3). As with the von Bertalanffy model, a lack of smaller animals in the juvenile stages was the weakest scenario for fitting a Gompertz curve.

Inverse Logistic

Unlike the correct von Bertalanffy or Gompertz, the correct inverse logistic candidate model had the highest probability (>0.95; scenarios B, D, F, and H) of being unambiguously the best model under its related operating model (Table 3). The correct inverse logistic also had the greatest range of probabilities, ranging from 0.17–1.00. In contrast to the von Bertalanffy and Gompertz models, the data quality criterion that had the biggest advantage on robustness was when the length of initial sizes was in the largest size classes: (i.e., 80–170 mm; scenarios B, D, F, and H). For the inverse logistic model, a large sample size
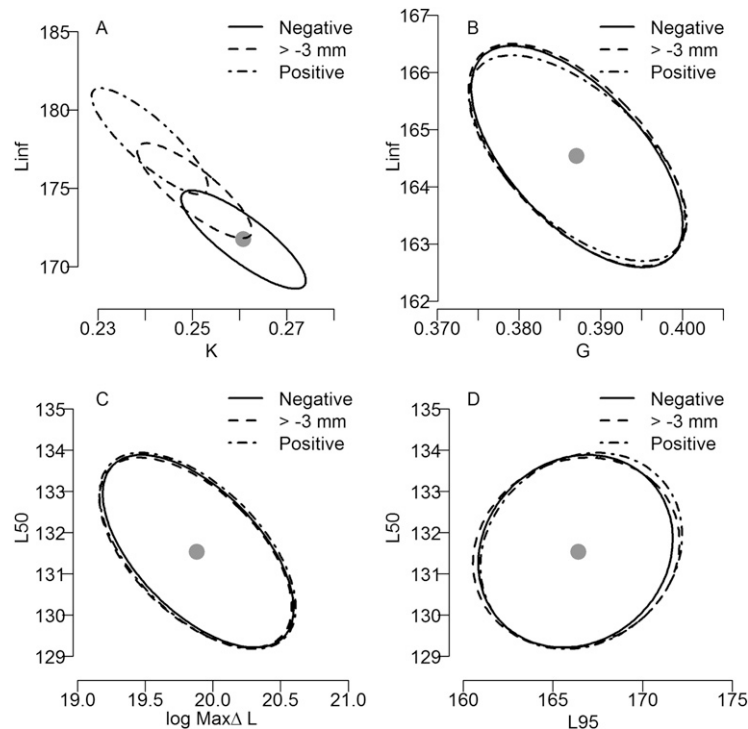
Figure 2. (A–D) The effect of removing negative growth increments on the distribution of parameter estimates for 3 candidate models: von Bertalanffy, Gompertz, and inverse logistic. The distribution of parameter estimates are shown for 500 simulations of ideal data adjusted by the incremental removal of negative increments: negative, no removal of negative increments; –3 mm, removal of increments less than –3 mm; positive, all negative increments removed. The ellipses represent 95% confidence limits around the distribution of candidate model parameters estimates. The stronger the parameter correlation, the more oval the 95% confidence limits. In each plot, the point represents the parameter values used in the operating model. The greater the overlap between ellipses, the lesser the effect of removing negative increments on the distribution of parameters.

($n = 500$) greatly improves its robustness to any other data scenario, and this was not always the case for the von Bertalanffy or Gompertz. A size range encompassing only smaller animals reduces the robustness of the inverse logistic model, probably because the available size range of the data (<120 mm) was not sufficient. For example, the estimate of the L_{50} parameter from the operating model ($L_{50} = 132$ mm) was greater than the maximum initial shell length of 120 mm. When the initial size range was small (55–120 mm; particularly for scenarios A and C, where $n = 100$), some parameter estimates of the inverse logistic model were highly skewed (Fig. 3). Increasing the sample size to $n = 500$ (scenarios E and G) greatly reduced the skew in parameter estimates and improved the performance of the inverse logistic model (Fig. 3).

Schnute

The Schnute model was never selected as the best model (Table 3). The flexibility conferred by the many submodel forms of the Schnute model, through the a and b parameters, enables it to assume either the Gompertz or the von Bertalanffy form, depending on the underlying data (Table 2). However, the Schnute cannot characterize the growth of juveniles that was found to be effectively constant annually (but with seasonal variation in juvenile abalone), and such growth was best represented by the inverse logistic growth model (Helidoniotis et al. 2011). Previous studies suggest that the growth rates of the smallest juveniles increased with initial size for blacklip and greenlip abalone (Shepherd & Hearn 1983, Troynikov et al. 1998). Of the 8 alternative shapes that the Schnute model can adopt (Francis 1995), none are similar to an inverse logistic model. In addition, fitting the Schnute model to data can be difficult because the shape of the predicted growth trajectory is very sensitive to the selection of starting parameters. For blacklip abalone, the starting parameters selected for the Schnute model greatly influenced whether it assumed a von Bertalanffy- or Gompertz-like growth pattern. When the starting parameters of the Schnute simulated a Gompertz–like model (Table 2), the Schnute converged on a false minimum and always simulated a Gompertz model regardless of the operating model (the Gompertz model in such cases would exhibit the same maximum likelihood value as that for the Schnute model). When the starting parameters of the Schnute resembled a von Bertalanffy-like growth trajectory (Table 2), the Schnute became flexible enough to simulate either a von Bertalanffy- or Gompertz-like growth trajectory, depending on the operating model. When fitted to data from the inverse logistic model the Schnute could not assume an inverse logistic-like growth trajectory and would reproduce either von Bertalanffy- or Gompertz-like growth.
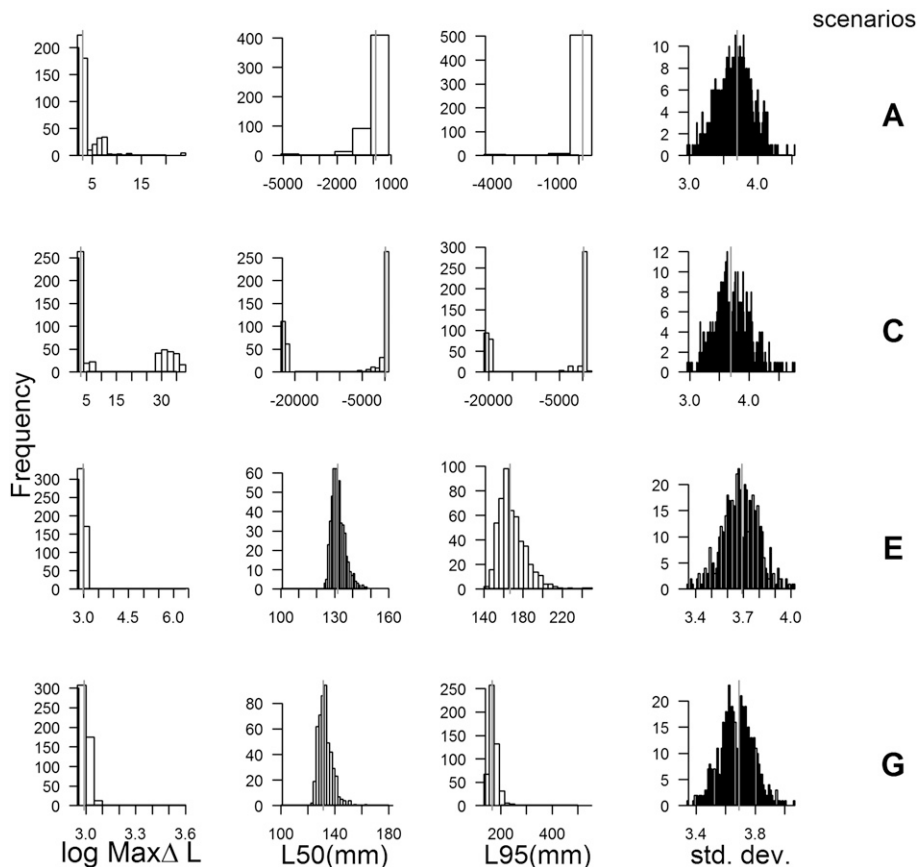

Figure 3. Parameter estimates for the inverse logistic model under the inverse logistic operating model for data scenarios A, C, E, and G in which the initial size range was 55–120 mm (refer to Table 1 for a description of the scenarios). Row 1 is scenario A, row 2 is scenario C, row 3 is scenario E, and row 4 is scenario G. Scenarios A and C illustrate the skewed distribution of parameter estimates of the candidate model fits. The estimates for $Max\Delta L$ (log $Max\Delta L$) were often so highly skewed that a natural log transformation was required. The vertical line represents the "known" estimate of the inverse logistic operating model parameter.

The Schnute had a similar likelihood to the von Bertalanffy or Gompertz; but, having more parameters than either of these growth models, the Schnute model always had a larger AIC. It would need to fit the growth data rather more closely than either of these other models to be selected as the best model according to AIC.

Summary

The data scenarios in which the von Bertalanffy and Gompertz were most robust were when smaller size classes (55–120 mm in initial shell length) were well represented in the data. An increase in sample size to $n = 500$ only improved the robustness of the von Bertalanffy. However, this data scenario was the least common in tag–recapture data. Conversely, the inverse logistic model performed best under an initial size range consisting of larger animals (80–170 mm), in which the minimum shell length increment was 0 mm. This data scenario was more typical of tag–recapture surveys. Therefore, the inverse logistic model had a high probability of recovering the known underlying growth based on data with the most typical characteristics of sampling error.

Criterion 2: Biological Criteria and Prediction Error

When each candidate model was the correct model, it produced accurate time-to-fishery estimates (Fig. 4). The correct von Bertalanffy model generated the shortest estimate (0.10 y; range, 6.4–6.5 y), the correct inverse logistic model generated an intermediate duration estimate (0.23 y; range, 7.8–8.1 y), and the correct Gompertz model generated the longest estimate (0.3 y; range, 8.5–8.8 y). Also, the correct von Bertalanffy and Gompertz models generated the most precise time-to-fishery

estimates (CV = 0.01–0.05; Table 5), whereas the inverse logistic model was slightly less precise (CV = 0.03–0.08; Table 5).

The performance of the incorrect models in each case reflected the differences between their general estimates of time to fishery. Thus, the von Bertalanffy and Gompertz performed worst as the candidate model when the other was the operating model. The inverse logistic model performed in an intermediate fashion but always came second to the correct model. In all cases, the Schnute model generated the most variable estimates of time to fishery and was never the optimum model (Fig. 4, Table 5).

In reality, the true growth trajectory of abalone populations may be obscured by sampling error. It has been common practice among research programs to select a growth model indiscriminately, which may be unknowingly incorrect, and fit it to data with an unknown degree of sampling error. This may ultimately result in a growth model that is not representative of the population. It is therefore important to assess the prediction error (i.e., the biological inaccuracy of unknowingly using the incorrect growth model). Time-to-fishery estimates were used to measure prediction error and to assess the implications of selecting an incorrect model, and estimates were compared among the 4 candidate models (Fig. 4). The difference in time-to-fishery estimates was calculated between each correct candidate model and the unrelated operating models. Relatively large differences indicated a relatively high prediction error associated with that candidate model. Of all 4 candidate models, the incorrect Gompertz model was the most inaccurate in terms of departures from known time-to-fishery estimates, deviating a maximum of 3.1 y under the von Bertalanffy operating model. The incorrect von Bertalanffy model was the second most inaccurate model, with the second highest departures from known time-to-fishery estimates, with an absolute maximum of 2.7 y under the inverse logistic operating model. The incorrect Schnute model was the third most inaccurate, with an absolute maximum deviation of 2.4 y under the inverse logistic operating model. The incorrect inverse logistic model was the most accurate of all the models in estimating time to fishery under a different operating model with a maximum deviation of 1.69 y. The von Bertalanffy model had the greatest mean prediction error (38; Table 4), followed by the Gompertz
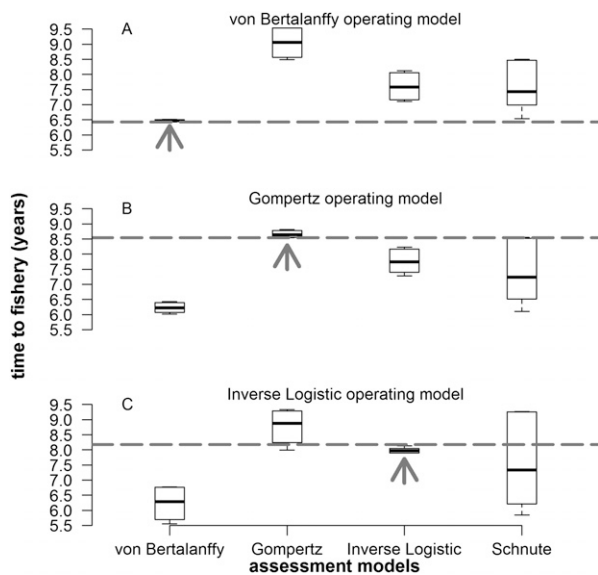
Figure 4. (A–C) Mean values for time-to-fishery estimates are presented from 500 simulations generated from 3 operating models: von Bertalanffy (A), Gompertz (B), and inverse logistic (C). The horizontal dashed line represents the expected known time-to-fishery estimate as determined by the operating model in each case. Time-to-fishery estimates for 4 candidate models (von Bertalanffy, Gompertz, inverse logistic, and Schnute) are shown for all 8 data quality scenarios combined. The arrows indicate the correct candidate model within each plot.

TABLE 5.

Range of coefficient of variation (CV) estimates from all data quality scenarios combined.

	Time-to-Fishery (CV)	Model Parameter (CV)		
		L_∞	K	
von Bertalanffy	0.01–0.05	0.7–5.1	3.4–11.9	
		L_∞	g	
Gompertz	0.01–0.05	0.4–3.7	2.8–7.5	
		$Max\Delta L$	L_{50}	L_{95}
Inverse logistic	0.03–0.08	3.4–1019	1.0–600	1.0–821

Estimates are presented for biological (time to fishery) and statistical (parameter values) criteria. The CV range is for each candidate model under its related operating model. For example, the CV of the von Bertalanffy parameters/time-to-fishery estimates is for data generated under the von Bertalanffy operating model).

TABLE 6.

A summary of the robustness of 4 growth models according to 3 criteria.

	Criteria*	von Bertalanffy	Gompertz	Inverse Logistic	Schnute
1	Parameter estimates	Good	Good	Bad	NA
2	Biological estimates	Poor	Poor	Good	Medium
3	Risk	More risky	More risky	Less risky	Medium

* Criterion 1: Statistical criteria and the distribution of candidate model parameters around the operating model parameters. Good indicates a relatively small distribution; Poor indicates a relatively wide distribution. Criterion 2: Biological criteria (i.e., the accuracy in time-to-fishery estimates between the incorrect candidate model and operating models). Poor indicates relatively large inaccuracy; Good indicates relatively low inaccuracy. Criterion 3: The biological risk of selecting an incorrect model as determined by the prediction error, see Eq (10). NA, not applicable.

(32), then the Schnute (20); the model with the lowest mean prediction error was the inverse logistic (10).

Comparing Statistical and Biological Criteria

Absurd parameter estimates did not necessarily lead to absurd biological estimates particularly for the inverse logistic model (Table 6). The results of the time-to-fishery estimates for the inverse logistic were more accurate than expected given the highly skewed and excessively wide distribution of parameter estimates (Fig. 3). The range of the CV for parameter estimates was greater than the range of CV for the biological estimates (i.e., time-to-fishery estimates), particularly for the inverse logistic model (Table 5). For example, the range of the CV for the von Bertalanffy (CV = 0.7–5.1 and CV = 3.4–11.9 for L_∞ and K, respectively) was more than 10 times the range of CV for its time-to-fishery estimates (CV = 0.01–0.05). The range of the CV for the inverse logistic model (CV = 3.4–1019 and CV = 1.00–821 for $Max\Delta L$ and L95, respectively) was at least 100–10,000 times the range of CV for its time-to-fishery estimates (CV = 0.03–0.08).

DISCUSSION

The Effect of Sampling Error on Model Selection

There are competing hypotheses as to which model describes most appropriately the growth of abalone (Day & Fleming 1992), and, given the variation exhibited by different abalone populations, this is not surprising. For the most part, the error rate (type I error) in model outcome was consistently higher (0.3–0.75; Table 3) under the most typical scenario of sampling error (i.e., when tag recapture data consisted of the adult size ranges, 80–170 mm; scenarios B, D, F, and H, Table 1). This sampling error is ubiquitous in field sampling programs given the difficulty in collecting cryptic juvenile abalone (<80 mm for *Haliotis rubra* in Tasmania). According to model selection criteria, the inverse logistic candidate model had a higher probability of recovering its own operating model than the von Bertalanffy or Gompertz had of recovering their related operating models (Table 3). Therefore, under the most typical scenarios of sampling error (i.e., when tag–recapture data consists of the adult size ranges) the inverse logistic model would be the more robust model.

The statistical selection of the best-fitting model can also be influenced by measurement error, such as negative increments (Francis 1988). With ideal data (i.e., large sample size, large size range), the removal of negative increments had no effect on the selection of the correct Gompertz and inverse logistic candidate models, but clearly affected the robustness of the von Bertalanffy model.

To avoid fitting to negative increments, the L_∞ parameter of the von Bertalanffy and Gompertz models have been reinterpreted using stochastic modeling by truncating the distribution of the L_∞ parameter (Sainsbury 1980, Troynikov et al. 1998). When measurement error is properly removed from the data, it is possible that the robustness of the stochastic von Bertalanffy model may be improved. Although negative increments are conspicuous in the larger size classes, they nevertheless imply that uncertainties in shell length could occur across the entire initial size range, and this needs to be investigated further.

The Effect of Sampling Error on Variability in Parameter Estimates

Parameters of the von Bertalanffy model are often used to compare populations. For example, in a study of 3 blacklip populations from South Australia, K ranged from 0.32–0.41/yr and L_∞ ranged from 139–144 mm (Shepherd & Hearn 1983). Similarly, in a study of 3 populations from Victoria, K ranged from 0.15–0.37/yr (difference, 0.22) and L_∞ ranged from 121–152 mm (difference, 31 mm). As a result of these differences, and the fact that the geographical distance between the South Australia and Victorian populations is approximately 700 km (McShane et al. 1988, McShane & Smith 1992), it is reasonable to assume that these populations are biologically different. However, the differences in the range of parameter estimates with this magnitude of geographical separation are approximately within the difference in range observed in the current study for a single population in Tasmania under typical scenarios of sampling error. The parameter estimates for the von Bertalanffy ranged from 0.19–0.34/yr for K (difference, 0.15) and from 150–194 mm for L_∞ (difference, 44 mm) as a result of sampling error. It is important to consider these ranges when making inter-population comparisons, and this result raises the question of what proportion of the fine-scale biological differences reported among wild populations is an artifact of sampling error in the data.

The Effect of Sampling Error on Biological Predictions

In most cases, model selection stops at statistical goodness-of-fit tests such as the AIC. However, biological criteria are also important when assessing how well models perform in recovering the known growth characteristics (Rogers-Bennett et al. 2007). Ultimately, this is important for improving the management of fished stocks.

All candidate models performed well under their related operating model forms, accurately recovering the known time-to-fishery estimates regardless of the data scenario (Fig. 4). Although all models indicated clear discrepancies between the correct candidate and operating models for statistical and parameter properties, such seemingly large discrepancies were less apparent in biological time-to-fishery estimates (Fig. 4). For example, some parameter estimates were absurd for the inverse logistic model but did not lead to similar differences in time-to-fishery estimates (Table 5).

Growth models, however, have many uses apart from time-to-fishery estimates. As the use of growth models increases, the selection of the most appropriate growth model would depend on its performance and uncertainty in other real-world problems, and not just time-to-fishery estimates. The growth of the juveniles vary, with some studies demonstrating that growth is constant over a range of initial sizes (Prince et al. 1988, Helidoniotis & Haddon 2012), which is consistent with the inverse logistic model (Haddon et al. 2008), or that the growth rate of juveniles increases over the juvenile sizes, which is consistent with the Gompertz model (Shepherd & Hearn 1983, Troynikov et al. 1998)

Minimum Data Requirements

For parameter estimates to be reliable, data need to be sufficient to define the parameters of interest (Haddon 2011). For that to occur, field surveys need to be model driven (Scandol 2004). A large range of initial size classes is more important for model robustness than large sample size alone. Typical data sets for abalone, however, consist mostly of large animals, and only the inverse logistic model was robust to this type of data scenario. Statistically, the inverse logistic model has a requirement for data to be available across size ranges that are extensive enough to allow the estimation of at least 2 parameter values—a combination of either $Max\Delta L$ and L_{50}, $Max\Delta L$ and L_{95}, or L_{50} and L_{95}. The data typical of tag–recapture studies tends to encompass the size ranges predicted by the latter 2 parameters (i.e., L_{50} and L_{95}), consisting of medium to large abalone, thereby meeting the requirements of the inverse logistic model.

Implications of Unknowingly Selecting the Incorrect Model

In practice, the true form of growth in any population is likely to be unknown, and it is valuable to assess the consequences of making inaccurate biological predictions as a result of unknowingly using the incorrect model. Of the 4 candidate models considered, the incorrect inverse logistic model provided the most accurate time-to-fishery estimates relative to the incorrect von Bertalanffy, Gompertz, or Schnute models. The low AIC of the inverse logistic model meant it was the appropriate model; however, parameter estimates were absurd. If the AIC was ignored, the inverse logistic model may have been rejected after the model-fitting stage because of the generation of absurd parameter estimates if there was sampling error. This restriction means that sampling error can lead back to one of the original problems, which is having a limited choice in model selection.

Implications of Different Models to Fisheries Management

The choice of growth model can have significant implications for fisheries management. For example, growth models are used in Tasmania when setting the LML for harvesting and protecting spawning stock. The Gompertz model implies slower initial growth, which would imply lower productivity and may lead to more conservative fishing policies (Rogers-Bennett et al. 2007), but may result in larger lost opportunity costs to the fishery. However, if the Gompertz model was used incorrectly to generate the LML, it would underestimate the LML for an intended level of protection, thereby underprotecting the spawning stock and increasing the risk of recruitment overfishing. The von Bertalanffy model would indicate more rapid initial growth than may actually occur, suggesting that the stock is more productive than in reality, and so this model could not be considered as being risk averse in fisheries policy decisions (Rogers-Bennett et al. 2007). If the von Bertalanffy is used in stock assessments, surplus stock production may be overestimated, leading to risky harvest strategies. On the other hand, if the von Bertalanffy model was used incorrectly to generate the size-related LML, it would overestimate an LML, thereby overprotecting spawning stock. This may result in lost opportunity costs to the fishery, which is risky to the economy but conservative to stock protection. Improving the techniques that inform the management of the blacklip abalone fishery in Tasmania is important because this is the largest, single, managed wild abalone fishery in the world and it makes a major contribution to the local economy (~AUD$100M pa before processing). The von Bertalanffy and Gompertz models are used commonly in Australia to describe the growth of abalone based on tag–recapture data. Their widespread use could potentially have serious implications in biasing fishery stock assessments.

LITERATURE CITED

Cox, L. A. 2002. Risk analysis: foundations, models, and methods. New York: Springer. 556 pp.

Cropp, R. 1989. Abalone culture in Tasmania. *Technical Report.* Hobart: Department of Sea Fisheries, Tasmania.

Day, R. W. & A. E. Fleming. 1992. The determinants and measurement of abalone growth. In: S. A. Shepherd, M. J. Tegner & S. A. Guzmán del Próo, editors. Abalone of the world: biology, fisheries and culture. Oxford: Blackwell. pp. 141–164.

Draper, D. 1995. Assessment and propagation of model uncertainty. *J. R. Stat. Soc. B* 57:45–97.

Eveson, J. P., T. Polacheck & G. M. Laslett. 2007. Consequences of assuming an incorrect error structure in von Bertalanffy models: a simulation study. *Can. J. Fish. Aquat. Sci.,* 64:602–617.

Fabens, A. J. 1965. Properties and fitting of the von Bertalanffy growth curve. *Growth* 29:265–289.

FAO 2006. FAO Fishery Information, Data and Statistics Unit. 2006. Total production 1950-2004. Universal software for fishery statistical time series [online or CD-ROM]. Statistics and Information Service, editors. Food and Agriculture Organization of the United Nations. Rome. (http://www.fao.org/fishery/statistics/software/fishstat/en)

Francis, R. I. C. C. 1988. Maximum likelihood estimation of growth and growth variability from tagging data. *N. Z. J. Mar. Freshw. Res.* 22:42–51.

Francis, R. I. C. C. 1995. An alternative mark–recapture analogue of Schnute's growth model. *Fish. Res.* 23:95–111.

Francis, R. I. C. C. & R. Shotton. 1997. "Risk" in fisheries management: a review. *Can. J. Fish. Aquat. Sci.* 54:1699–1715.

Grover, M. C. 2005. Changes in size and age at maturity in a population of kokanee *Oncorhynchus nerka* during a period of declining growth conditions. *J. Fish Biol.* 66:122–134.

Haddon, M. 2011. Modeling and quantitative methods in fisheries. Boca Raton, FL: Chapman and Hall/CRC. 449 pp.

Haddon, M., C. Mundy & D. Tarbath. 2008. Using an inverse-logistic model to describe growth increments of blacklip abalone (*Haliotis rubra*) in Tasmania. *Fish Bull.* 106:58–71.

Hamilton, G., R. McVinish & K. Mengersen. 2007. Using a Bayesian net to aid in the selection of candidate variables In: 2nd Annual Conference for the SRA-ANZ (Society of Risk Analysis Australian & New Zealand Regional Organisation).

Helidoniotis, F. & M. Haddon. 2012. Growth model selection for juvenile blacklip abalone (*Haliotis rubra*): assessing statistical and biological validity. *Mar. Freshw. Res.* 63:23–33.

Helidoniotis, F., M. Haddon, G. Tuck & D. Tarbath. 2011. The relative suitability of the von Bertalanffy, Gompertz and inverse logistic models for describing growth in blacklip abalone populations (*Haliotis rubra*) in Tasmania, Australia. *Fish. Res.* 112:13–21.

Hilborn, R. & M. Mangel. 1997. The ecological detective: confronting models with data. Princeton, NJ: Princeton University Press. 315 pp.

Hilborn, R. & C. J. Walters. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. London: Chapman and Hall. 570 pp.

Katsanevakis, S. & D. Maravelias. 2008. Modelling fish growth: multi model inference as a better alternative to *a priori* using von Bertalanffy equation. *Fish Fish.* 9:178–187.

Marcot, B. G. 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecol. Modell.* 230:50–62.

McShane, P. E. & M. G. Smith. 1992. Shell growth checks are unreliable indicators of age of the abalone *Haliotis rubra* (Mollusca, Gastropoda). *Aust. J. Mar. Freshw. Res.* 43:1215–1219.

McShane, P. E., M. G. Smith & K. Beinssen. 1988. Growth and morphometry in abalone (*Haliotis rubra* Leach) from Victoria. *Aust. J. Mar. Freshw. Res.* 39:161–166.

Miller, K. J., B. T. Maynard & C. N. Mundy. 2008. Genetic diversity and gene flow in collapsed and healthy abalone fisheries. *Mol. Ecol.* 18:200–211.

Nash, W. J. 1992. An evaluation of egg-per-recruit analysis as a means of assessing size limits for blacklip abalone (*Haliotis rubra*) in Tasmania. In: S. A. Shepherd, M. J. Tegner & S. A. Guzmán del Próo, editors. Abalone of the world: biology, fisheries and culture. Oxford: Blackwell Scientific. pp. 318–340.

Naylor, J. R., N. L. Andrew & S. Kim. 2003. Fishery independent surveys of the relative abundance, size-structure, and growth of paua (*Haliotis iris*) in PAU 4. In: New Zealand fisheries assessment report 2003/8. Wellington: Ministry of Fisheries. p. 16.

Neuheimer, A. B. & C. T. Taggart. 2007. The growing degree-day and fish size-at-age: the overlooked metric. *Can. J. Fish. Aquat. Sci.* 64:375–385.

Prince, J. D., T. L. Sellers, W. B. Ford & S. R. Talbot. 1987. Experimental evidence for limited dispersal of haliotid larvae (genus *Haliotis*: Mollusca: Gastropoda). *J. Exp. Mar. Biol. Ecol.* 106:243–263.

Prince, J. D., T. L. Sellers, W. B. Ford & S. R. Talbot. 1988. A method for ageing the abalone *Haliotis rubra* (Mollusca: Gastropoda). *Aust. J. Mar. Freshw. Res.* 39:167–175.

Punt, A. E. 2006. The FAO precautionary approach after almost 10 years: have we progressed towards implementing simulation-tested feedback–control management systems for fisheries management? *Nat. Resour. Model.* 19:441–464.

Quinn, T. J., II & R. B. Deriso. 1999. Quantitative fish dynamics. New York, Oxford University Press. 542 pp.

R Development Core Team. 2008. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rogers-Bennett, L., D. Rogers & S. A. Schultz. 2007. Modeling growth and mortality of red abalone (*Haliotis rufescens*) in northern California. *J. Shellfish Res.* 26:719–727.

Sainsbury, K. J. 1980. Effect of individual variability on the von Bertalanffy growth equation. *Can. J. Fish. Aquat. Sci.* 37:241–247.

Scandol, J. P. 2004. A framework for the assessment of harvested fish resources in NSW. In: Fisheries resource assessment series. Sydney: NSW Department of Primary Industries. 96 pp.

Schnute, J. 1981. A versatile growth model with statistically stable parameters. *Can. J. Fish. Aquat. Sci.* 38:1128–1140.

Shepherd, S. A. & W. S. Hearn. 1983. Studies on southern Australian abalone (genus *Haliotis*) IV: growth of *H. laevigata* and *H. rubra*. *Aust. J. Mar. Freshw. Res.* 34:461–475.

Summerfelt, R. C. 1987. Age and growth in fishes. In: Summerfelt, R. C. & G. E. Hall, eds. Papers presented at the International Symposium on Age and Growth of Fish. June 9–12, 1985, Des Moines, Iowa. Iowa State University Press, p 544.

Troynikov, V. S., R. W. Day & A. Leorke. 1998. Estimation of seasonal growth parameters using a stochastic Gompertz model for tagging data. *J. Shellfish Res.* 17:833–838.

Urban, H. J. 2002. Modeling growth of different developmental stages in bivalves. *Mar. Ecol. Prog. Ser.* 238:109–114.

Wang, Y. G. 1998. Growth curves with explanatory variables and estimation of the effect of tagging. *Aust. N. Z. J. Stat.* 40:299–304.

Wang, Y. G. 1999. Estimating equations for parameters in stochastic growth models from tag–recapture data. *Biometrics* 55:900–903.

Wang, Y.-G. & M. R. Thomas. 1995. Accounting for individual variability in the von Bertalanffy growth model. *Canadian Journal of Fisheries and Aquatic Sciences,* 52:1368–1375.

Wang, Y.-G., M. R. Thomas & I. F. Somers. 1995. A maximum likelihood approach for estimating growth from tag–recapture data. *Can. J. Fish. Aquat. Sci.* 52:252–259.

Worthington, D. G., N. L. Andrew & G. Hamer. 1995. Covariation between growth and morphology suggests alternative size limits for the abalone, *Haliotis rubra*, in NSW, Australia. *Fish. Bull. (Wash. DC)* 93:551–561.

Zhou, S. 2007. Discriminating alternative stock–recruitment models and evaluating uncertainty in model structure. *Fish. Res.* 86:268–279.