

Calibrating process-based marine ecosystem models: an example case using Atlantis

Heidi Pethybridge^{1*}, Mariska Weijerman², Holly Perryman³, Asta Audzijonyte⁴, Javier Porobic¹,
Vidette McGregor⁵, Raphael Girardin⁶, Cathy Bulman¹, Kelly Ortega Cisneros⁷, Matteo
Sinerchia⁸, Trevor Hutton¹, Hector Lozano-Montes¹, Mao Mori⁴, Camilla Novaglio¹, Gavin
Fay⁹, Rebecca Gorton¹, Elizabeth Fulton¹

1. CSIRO Oceans and Atmosphere, GPO Box 1538, Hobart, Tasmania 7000, Australia
2. NOAA Fisheries, Pacific Islands Fisheries Science Centre, Honolulu, HI 96818, United States
3. Institute of Marine Research (IMR), Strandgaten 229, 5004, Bergen, Norway
4. Institute for Marine and Antarctic Studies, University of Tasmania, 20 Castray Esplanade, Battery Point, Hobart, TAS 7001, Australia.
5. National Institute of Water and Atmospheric Research (NIWA), 301 Evans Bay Parade, Greta Point, Wellington, New Zealand.
6. IFREMER, Channel and North Sea Fisheries Research Unit, 150 Quai Gambetta, BP 699, 62321, Boulogne-sur-mer, France.
7. Department of Ichthyology and Fisheries Science, Rhodes University, Grahamstown, South Africa
8. CNR-IAS Institute for the Anthropic impacts and Sustainability in marine environment. Torregrande, Oristano, Italy.
9. Department of Fisheries Oceanography, School for Marine Science and Technology, University of Massachusetts Dartmouth, 200 Mill Rd., Fairhaven, MA 02719, United States.

* corresponding author: Heidi.Pethybridge@csiro.au

Abstract

Calibration of complex, process-based ecosystem models is a timely task with modellers challenged by many parameters, multiple (emergent) outputs of interest and often a scarcity of empirical data. Incorrect calibration can lead to unrealistic ecological and socio-economic predictions with the modeller's experience and available knowledge of the modelled system largely determining the success of model calibration. Here we provide an overview of best practices when calibrating an Atlantis marine ecosystem model, a widely adopted framework that includes the parameters and processes included in many different ecosystem models. We highlight the importance of understanding the model structure and data sources of the modelled system. We then focus on several model outputs (biomass trajectories, age distributions, condition at age, realised diet proportions, and spatial maps) and describe diagnostic routines that can assist modellers to identify likely erroneous parameter values. We detail strategies to fine tune values of four groups of core parameters: growth, predator-prey interactions, recruitment and mortality. Additionally, we provide a pedigree routine to evaluate the uncertainty of an Atlantis ecosystem model based on data sources used. Describing best and current practices will better equip future modellers of complex, processed-based ecosystem models to provide a more reliable means of explaining and predicting the dynamics of marine ecosystems. Moreover, it promotes greater transparency between modellers and end-users, including resource managers.

Keywords: *best practices, model diagnostics, food web, pedigree, parameter estimation*

1.0 Introduction

Complex process-based (or conceptual) marine ecosystem models are increasingly used as strategic tools when considering how populations, communities and ecosystems respond to management or environmental changes (Larkin 1996, Garcia et al. 2003, Pikitch et al. 2004, Collie et al. 2016). Contrary to statistical models that seek to find associations between a minimum number of parameters to maximise fit to data, the goal of process-based models is to describe general functional relationships among several components of an ecosystem, thus providing greater mechanistic understanding of the complex dynamics of marine ecosystems (White and Marshall 2019). Many different process-based ecosystem models exist that are based on a range of theories and assumptions that vary from accepted to controversial (Fer et al. 2018). The breadth of this information (and uncertainty) has fostered large structural discrepancies between different ecosystem modelling frameworks that each balance complexity, realism and accuracy and are further categorised into species-based, size-based or agent (or individual)-based models (Pethybridge et al. 2018). Among the process-based ecosystem model types, there are a number of commonalities between the grounding theories and principles used (e.g. stoichiometry, food web, life-history, individual or population growth, and thermodynamics) meaning that they share similar data requirements. These data requirements are often unavailable for many components of the system of interest introducing bias from imprecise observation values (observation error), poor or oversimplified inclusion of processes (structural error) or inappropriate parameter values (parametric error) (Hill et al. 2007, Link et al. 2011). Unrealistic and biased model predictions can result from these uncertainties, making model calibration a fundamental process in ensuring appropriate parameter values are attained, so that model outputs agree with observed trends and most importantly have predictive power.

In relatively simple ecosystem models, automated optimization algorithms and statistical inferencing (Bayesian or frequentist) methods are routinely used during the calibration or parameter estimation phase to quantify uncertainty in model inputs and outputs based on probability distributions (Bolker

et al. 2009, Subbey 2018, Krapu and Borsuk 2019). For complex process-based models such analytical methods are often too computationally demanding meaning that initial parameter values are estimated before model simulation and that more labour-intensive (manual) methods are required to determine which sets of parameters need tuning and what magnitude of change might be required. The main reason is that these models have more complex structural elements, such as nonlinearities and multiple discontinuities, and have multiple outputs which lead to complex correlation structures of error. Often an iterative process, called pattern-oriented modelling, is needed to calibrate such models, adjusting parameter values to reproduce more ecologically reasonable dynamics and to fit historical observations (Grimm and Railsback 2012, Kramer-Schadt et al. 2007, Travers-Trolet et al. 2014).

The calibration process varies amongst process-based ecosystem models and primarily depends on the model's representation of the food web components and contributing biological and physical processes (Collie et al. 2016). Species-based ecosystem models typically aggregate species into functional groups – primarily based on similarities in diet, body size, life history characteristics, habitat dependencies, spatial patterns (e.g. depth distributions) and commercial or social significance. Size-based models characterize the food web based on body-size relationships and group together individuals (regardless of species) by size, age, or life-stage. Agent-based models track populations representing individual traits (size-at age, life-history, metabolism, and spatial distribution). Each functional, size, or agent group represented in a model requires parameterization, so the more groups included, the more intricate the calibration strategy. Relatively 'simple' ecosystem models that represent just a few groups (2-12), such as many size spectrum models (Blanchard et al. 2014), and models of intermediate complexity (MICE; Plagányi et al. 2014) have few parameters (typically no more than 20) that are either adjusted during calibration using standard statistical methods or fitted to data. The species-based model framework Ecopath with Ecosim (EwE; Polovina 1984, Christensen and Walters 2004), requires the calibration of 4-10 parameters per functional group after a balanced model has been achieved. General thermodynamics and

ecological principles can then be applied to assist in the checking of EwE parameter values (Link 2010).

At the most complex end of the ecosystem modelling spectrum are those models that seek to represent the entire trophic range separated by aggregated functional species groups and have multiple trait formulations and a number of coupled biophysical processes (ecological, oceanography, biogeochemistry, and those related to human activities). Such models include agent-based models (e.g., Fiechter et al. 2015, Shin and Cury 2004, Grüss et al. 2016) and hybrid species and size-based models (e.g., Melbourne-Thomas et al. 2011a,b, Fulton et al. 2004a, 2007, 2011, Blanchard et al. 2014). The number of parameters needing tuning in these models can be huge (from 6 to over 50 per functional group) reflecting an increase in the number of equations and time-steps required to emulate a large number of processes. The heightened complexity also increases the need for observational sources and theoretical understanding and as such models are often poorly constrained by available data which may result in potentially high degrees of uncertainty over specific parameter values (Collie et al. 2016, Weijerman et al. 2015a). While recent efforts have seen a number of best practice and review papers on how to calibrate an EwE mass balanced model (Essington 2007, Ainsworth and Walters 2015, Heymans et al. 2016), there is limited guidance available on calibration strategies for complex process-based ecosystem, or whole-of-system, models. Performance metrics for Atlantis ecosystem models have been reported (Kaplan and Marshall 2016, Olsen et al. 2016, McGregor et al. 2019), but primarily from the perspective of reviewing and validating finalized models.

This paper provides some pointers on how to best calibrate a complex process-based ecosystem model, using the Atlantis framework as an example. Models in the Atlantis framework typically include age-structured and biomass pool functional groups (species) in a spatially explicit framework and therefore include many parameters and processes used in other process-based ecosystem models. First, we detail the actions required to obtain an understanding of how the study ecosystem was conceptualised and what data were used for initial parameter estimations. Second,

we outline the types of model outputs needed to diagnose which parameter groups are most likely in need of adjusting. We describe how core parameters are commonly estimated and what magnitude of change could be expected. Lastly, we provide a simple set of guidelines for ecosystem modellers to assess uncertainty, based on a performance criterion. We focussed on aspects most commonly involved in model calibration, rather than the model development or construction. As such we are not addressing issues of structural model uncertainty (i.e. processes included and the way they are represented), but only the parameter uncertainty. We think that ecosystem model developers will find the assumptions and methods used to estimate values for core parameters highly useful, as will model end-users. Similarly, we primarily focussed on the biological parameters, rather than the parameters required by the physics, hydrodynamics or fisheries and assessment modules of Atlantis. The growing interests in process-based ecosystem models, such as Atlantis, and their use in considering ecosystem-based management options (Plagányi 2007, Grüss et al. 2017), makes this paper a timely and important contribution.

2.0 Conceptualization of the ecosystem in Atlantis

Atlantis (<https://research.csiro.au/atlantis/>) is a species-based and age-structured, spatially-explicit biophysical model that covers the full trophic spectrum, including ocean biogeochemistry and human interactions (Fulton et al. 2004a, Audzijonyte et al. 2019) (Fig. 1). The framework has been applied from polar to tropical systems (Weijerman et al. 2016) and is a commonly used tool to evaluate management strategies and to explore how ecological groups may respond to changes in the environment or human use (Fulton et al. 2011). Full details on the entire Atlantis framework can be found in the user manuals (Audzijonyte et al. 2017a, b).

One of the key features of Atlantis is that it allows users to select from a range of functional responses, stock-recruitment relationships, habitat effects and species interactions (see Audzijonyte et al. 2019). Yet, the core processes, and groups of parameters, that govern changes in the biomass of a particular functional group are not dissimilar to other physiologically structured models, and can be broadly calculated as:

$$\text{Biomass (N)} = \text{Growth} + \text{Recruitment} - \text{Predation} - \text{Additional mortality}$$

where biomass is tracked as nitrogen (N) and can be converted from carbon or dry weight using an assumed Redfield ratio (Redfield 1958). Descriptions and details of how to estimate these four groups of parameters are summarised in Table 1 and discussed in detail in section 5.0.

While not a strict requirement, lower trophic levels are generally treated as aggregate biomass pool groups (or with juvenile and adult stages) while high trophic invertebrate and vertebrate groups are age-structured (with a user-defined number of stages and cohorts). Recruitment is only explicitly tracked in the age-structured groups and determines new arrivals to the habitat/model domain (as eggs and larval stages are not explicitly modelled). Like in other two-pool models (e.g. de Roos and Persson 2001), age-structured groups in Atlantis are tracked through two nitrogen pools: structural (*SN*) representing irreversible mass (skeleton and organs) and reserve (*RN*) representing sufficient energy stores (adipose tissue, gonads, and body fat). This offers a powerful approach to model properties dependent on the condition and reserve dynamics (Kooijman 2010).

Many specific additional processes such as overwintering, symbiotic growth (e.g. zooxanthellae-host), mixotrophy, eddy strength, bleaching and bioerosion are also available (Audzijonyte et al. 2017a) but not discussed further here.

3.0 Transitioning from model development to calibration

The first place to start when calibrating any complex process-based ecosystem model is to understand how the ecosystem has been conceptualised during model development and what main data sources and approaches were used to estimate the core parameters of the model (Table 1) and the initial conditions (starting biomass and weights, environmental preferences and spatial distributions) (Table S1). This would include first obtaining clear descriptions of the biological components of the model (e.g. the taxonomic composition of the groups, number of stages, migration, habitat preference) and what functional response or relationship equations have been selected (for temperature, feeding, and life-history). Atlantis offers a range of options for users to

select from and it is important to understand the assumptions and data requirement for each of these options. The second step involves looking at how the functional groups interact through food web interactions. Due to a coarse spatial and temporal resolution, Atlantis does not aim to simulate accurate dynamics of the lowest trophic groups. Therefore, it is recommended to pool these groups into several larger categories and to ensure food web simple links (Fig. 2). Thirdly, it is crucial to look at the environmental factors (temperature, salinity, pH) included in the model, and their ranges throughout the model domain (e.g. maximum and minimum values and how they compare to the known tolerances or distribution ranges of functional groups). Recruitment, mortality and growth can be dependent on environmental conditions in the modelled spatial areas, which allows dynamically emergent responses of growth and distributions to environmental change. However, if not well represented, these factors can restrict a group's recruitment and spatial distribution or modify predator feeding rates and prey availability. Lastly, because Atlantis allows both horizontal and vertical spatial structure, effort should be made to check the assumptions and parameters behind the spatial distributions and movements (horizontal and vertical) of functional groups as they affect predator-prey interactions and density-dependent processes (Fulton et al. 2004b, Grüss et al. 2016).

A good practice is to clearly document and reference all the data sources used to estimate initial parameter values during the development phase and the time series data that will be used to assess model performance. This would ensure model transparency and transferability of information to other models and studies. This process will provide a first approximation of the uncertainty in the model and will ensure that the core model parameters are tuned to the most complete and current information. This will also assist in creating a pedigree of the model (as we propose below and outline in Table 2) that quantifies model uncertainty. Several approaches can be taken to source information for the various parameters (Table 1 and S1), with seeking guidance, advice and data from field experts an obvious first point of call. As new data are acquired, one should keep in mind potential caveats such as using data for a depleted or heavily fished stock, since very different relative biomass trajectories may be expected when conducting model projections for these groups compared to lightly or non-fished groups.

Lastly, the user should have a clear idea of how the many different model outputs will be assessed (using available tools) and how changes to parameters will be tracked over the course of the calibration process. The core Atlantis code does not have model output plotting functions and instead provides outputs in a range of different file formats (NetCDF, plain text and column delimited files). Many different tools and scripts to explore these outputs have been contributed by the Atlantis community over the last decade and are available on the public repository GitHub: <https://github.com/Atlantis-Ecosystem-Model/>. New modellers should also note that getting to the heart of anomalous behaviour in a complex process-based model may mean delving into the code. Atlantis is an open source package and its code is written in C/C++; helpful tips to diagnosing problems can be accessed through the Atlantis GitHub page. While these tools are tailored to Atlantis, the specific approaches to query and diagnose model behaviour might be of interest to a broader modelling community. The essential outputs to focus the assessment of any ecosystem model should include i) the biomass trends at ecosystem level; ii) box specific biomass; and iii) trophic structure at the ecosystem and box level.

3.1 Phases and goals of calibration

Setting clear objectives and performance criteria is crucial for the overall calibration strategy. There are typically three broad phases of calibration: 1. unfished system; 2. constant fishing; and 3. historical fishing and environmental drivers (Horne et al. 2010, Kaplan et al. 2010, Weijerman et al. 2015b). It is highly advisable to add complexity progressively for all functionalities - not just fishing. For example, it is advisable to initially turn off density dependent movement and set quadratic mortality parameters to zero or a very low value. In this paper, we discuss the calibration methods for the first phase, but the techniques should be revisited in the latter two calibration phases.

In the absence of fishing (phase one) the primary goal of the simulations is to produce stable biomass through time. That is: i) keep all groups from becoming extinct, and ii) obtain stable

biomasses that are not oscillating more than a certain percentage from their initial values, if they are set based on the observed ecosystem variables (abundance, biomass). What this percentage is depends on the system context and the degree of data uncertainty. For example, if initial model conditions reflect an unfished system then in this ‘unfished’ calibration phase the objective would be to keep all groups and age classes with $\pm 20\text{--}50\%$ of their initial values. For those who have good data (or for better known properties such as size at age) the aim would be to use the more restrictive $\pm 20\%$ target output band. Those with less data (or for more uncertain properties such as abundance of a poorly monitored group) would use a wider band, $30\text{--}50\%$ or potentially even higher for the most uncertain properties (such as the biomass of meiofauna). In contrast, if the model was started with initial conditions of a fished state and run forward in an unfished state the changes in realised biomass levels should reflect understanding of the level of depletion. Hence, in the absence of fishing all groups should reach carrying capacity with a biomass level similar to their ‘virgin’ (unfished) biomass (so for depleted groups abundance levels may be many times higher than the initial biomass). Ideally, these target ‘virgin’ biomass values would be taken from historical observations, but on occasion must be drawn from stock assessments. In such a case there should be close discussion with the assessment authors so all caveats and uncertainties pertaining to that assessment are understood.

In the second phase of calibration, the goal is to reproduce realistic responses to a range of fishing pressures. With an increase in fishing pressure, biomass should change (in most cases decrease) and do so more rapidly for species with low growth rates than for species with a high turn-over. However, the multispecies nature of these models means that counter-intuitive outcomes are possible, and may indicate incorrect assumptions about competitive, cannibalistic or predatory species interactions. If, for example, fishing decreases the abundance of a predator, and the abundance of its prey increases even if the latter is also fished, this may suggest a realistic response. Other responses to fishing may however be clearly unjustified (e.g. a small pelagic species going extinct when fished at low mortality levels).

For the third calibration phase, the goal is to reproduce historical trends in biomass when forcing the model with historical fishing and environmental drivers. Time series data should be compared with model trajectories and the trends and approximate magnitudes should be the same (and where possible the gross levels of variation).

Modellers should also think beyond gross biomass. Age (size or stage) distributions should be in line with observations or at least be plausible to field experts. Realised diets in the model are crucial for assessing the adequacy of species interactions, because it is possible to achieve correct biomasses based on completely incorrect diets (e.g. a top predator diet consisting of benthic invertebrates only, if too much access to this functional group was allowed in the model). Seasonal patterns of production or movement should also be assessed and for spatial models, the biomass distribution of all groups should be comparable to available species distribution or habitat maps.

Once the model is performing well and producing biomasses and size-at-age in the correct ballpark then the constraining assumptions (e.g. around density-dependent movement and additional mortality) can be relaxed and desired options put in place instead followed by calibration refinement.

3.2 Model time-steps and periodicity

The periodicity in which model fluxes are calculated and various model outputs are saved, is often variable based on user needs and the processes being modelled or diagnosed. For many ecosystem models, including EwE and size spectrum models, the time-step is rather coarse (seasonal or annual). Similar to other biogeochemical-based models, in Atlantis fluxes are calculated every 12 h to incorporate diurnal processes, such as vertical migration, although shorter (3 h or 6 h) time steps can be advised when interested in incorporating shorter time frame (e.g. tidal) processes. Longer time steps (of a day or even longer) can be used, but usually this is not done due to the importance of diurnal processes in most ecosystems. This, along with fine scale spatial structure, makes Atlantis model runs much more computationally demanding than many other ecosystem models. Adjusting

output reporting should balance the collection of information needed for calibration and the generation of output files as longer files slow down model run times. For instance, daily output may be needed to explore the effect of finer scale temporal dynamics (e.g. day of recruitment) and is commonly needed in the initial stages of model testing. Once the model is calibrated, output frequency is usually lowered – such as monthly, quarterly, seasonal or even annual. Annual output is often a useful setting for overall system metrics (such as gross biomass levels per group), with monthly or quarterly recommended for full spatial snapshots to ensure patterns, for both short and long-lived taxa, are well captured.

Every model run should have a sufficient ‘burn-in’ or ‘spin-up’ period to allow most groups to stabilise and to replace overly homogenous initial conditions with the true internal dynamics of the system. As a rule of thumb, the burn-in period should be long enough for the life cycles of all functional groups to be completed. Where longer (>10-20 year) burn-in periods are needed, outputs from a long run simulation can be used to create new initial conditions so that the model effectively has a stable state that captures age structures and 3D spatial distributions appropriate for the internal dynamics of the model.

4.0 Key ecosystem model outputs and diagnostic routines

After running an ecosystem model, a large and customizable range of outputs is available to assist in determining how the model is performing and which functional group parametrizations need further attention. For Atlantis output files can be produced for different purposes, with the main files including: (i) overall (system wide) biomass (tonnes) by functional group; (ii) spatial snapshots of biomass (tonnes), numbers, and *SN* and *RN* by functional group and age class; (iii) realised diet (proportion) for each age class of each functional group (and each model box); (iv) annual predation and fishing mortality for each functional and age group (per year); and (v) warnings, checks, and customised calibration messages. Other Atlantis output files are useful in reporting debug options to specifically look at the trends and patterns in stock structure, migration, movement, spawning, recruitment, aging, and fishing (Audzijonyte et al. 2017a).

Significant calibration time can be saved by thoughtful ‘pre-balance’ (PREBAL) diagnostics (Link et al. 2010). Essentially this includes checking if core model outputs span reasonable orders of magnitude or are trending in agreement with well accepted ecological principles (such as a 5-10% decline of biomass on \log_{10} scale with each increasing trophic level). Atlantis can output a number of integrative ecological indicators that can be used to check ecosystem condition (Fulton et al. 2005). These indicators are primarily intended for use in production runs to consider the implications of harvest strategies (Olsen et al. 2018) or to check the utility of the indicator themselves (Fulton et al. 2005, Shin et al. 2018). However, they can be also used to identify issues of model structure and data quality in accordance to PREBAL diagnostics that are regularly used in other ecological network models, such as EwE (Link 2010). For an Atlantis model the PREBAL diagnostic could include comparing biomass ratios of pelagic and demersal groups, piscivores and planktivores, and infauna and epifauna as indices of food web structure and integrity. Diversity (Kempton’s Q) and proportional habitat cover could also be considered along with the size and compositional indicators.

From the PREBAL checks and visualisation of temporal and spatial trends in the model outputs the modeller can compile a list of groups that are not performing well. Initial effort should go to adjusting parameters to improve the dynamics of troubled primary producer and consumer biomass groups as they influence the bottom-up ecosystem trophic dynamics that ultimately determine the production of higher trophic functional groups. A good place to start is to ensure a simple lower food web structure that includes linkages to detritus (Fig. 2). Gross “phytoplankton” or “zooplankton” groups rarely work when attempting to capture nutrient cycling and is best done by explicitly including both the microbial loop and the “classic” diatom-copepod pathways. Next, the modeller should carefully assess the degree of predation mortality through assessing realised diets of dominant predators. A very effective way of stopping a group from declining is to reduce the predation pressure, including degrees of cannibalism on the juveniles. Similarly, if the biomass of a group is increasing unrealistically, then one can increase the contribution of that group to the diet of

corresponding predators by modifying the predator-prey matrix (see below). Unlike other ecosystem models such as EwE, but similar to size-spectrum models, realised diets emerge dynamically in Atlantis as a function of prey availability, relative biomass and sizes. When confident on the food web structure, one should look to adjust the growth parameters as observational uncertainty of size-age data is likely to be large given the inherent variability of growth and the fact that functional groups often include a diverse set of species.

Once primary producer and consumer biomass groups are performing well, effort should go to diagnosing which parameters (Table 1 and detailed below) to adjust if higher order groups are crashing or increasing rapidly. Slightly different diagnostic routines are used for groups represented as a single biomass pool and those that are age-structured, though both heavily rely on model outputs that reflect realised predation and consumption levels (Fig. 3). This is because predation is typically the main source of mortality in Atlantis as growth is not set by a growth function but is based on realised consumption (explained more in section 5.1). For biomass pools, biomass can only be assessed through observing trends in total biomass because their reproduction is not specifically parameterized. For age-structured groups in Atlantis a larger number of model outputs (*SN*, *RN* and numbers at age) are produced giving much greater assistance in model assessment and diagnosis testing. For age-structured groups, the shape of the condition and numbers at age trajectories is a rapid way of determining what combination of parameters likely needs to be changed to produce stable size-at-age and abundance structures (Table S2). For many trajectories, a first place to start is to look at, and adjust, the parameters related to recruitment particularly if age-1 numbers are crashing within the first few years in the model run (indicating that adults may not have high enough reserves to support recruitment or that the number of recruits is far too low). Once these have been checked a similar sequence of parameter checking to biomass groups is routinely followed (Fig. 3).

Importantly, visualisation of spatial trends can further assist in diagnosing which parameters most likely need adjusting. For example, if age 1 numbers decline in only some boxes, it is unlikely that

the recruitment parameters will be the main cause but rather that parameters of predator-prey interactions need adjusting.

5.0 Core Atlantis parameters

Similar to other process-based ecosystem models that have a large number of parameters, in Atlantis many parameters should not need changing after the model development period. For example, it would be rare to change system level parameters related to bioturbation, bioirrigation, detrital breakdown, nitrification, light penetration or the Redfield ratio (particularly if the model has been developed in consultation with expert biogeochemists or experienced Atlantis modellers).

Depending on the level of model confidence for a particular functional group, often no or limited changes are needed for parameters reflecting general life-history and growth patterns (typically available on FishBase or SealeifeBase), number and biomass at age groups, spatial distributions, and environmental preferences (Table S1). Then there are a number of parameters that are central to model sensitivity (Pantus 2007), that are related to growth, predator-prey interactions, recruitment, and additional mortality (Table 1). Related to many of these parameters is that we still do not have a clear understanding of which biological mechanisms and drivers explain the variability that is observed in existing empirical databases. Here we describe best currently accepted practices in setting and adjusting these sets of parameters and in some cases provide parameter boundaries for broad ecological groups from existing, semi-validated Atlantis ecosystem models. Many of these parameters are also core to other ecosystem models, so best practices identified here should be transferrable between models. Conversely, efforts to estimate parameter values for other ecosystem modelling frameworks including web applications that use data derived from two global information systems (FishBase and SeaLifeBase)(Gruss, et al. 2019) should be utilised by Atlantis and other ecosystem modellers. For all the parameters detailed here, single parameter sensitivity simulations are the most common way of trying to test effects and determine the magnitude of change in a parameter value that is needed (Fulton et al. 2007, Ortega-Cisneros et al. 2017, Hansen et al. 2019).

For the estimation of all core parameters of an ecosystem model, it is essential to clearly establish how biomass measurements typically obtained in empirical studies, can be converted to match the currency of the model parameters. In Atlantis, the Redfield ratio is used to convert carbon to nitrogen while a wet weight (WW) to ash free dry weight (AFDW) ratio of 4 to 6 is commonly used. Where possible, species and regional specific conversion factors should be used, as values are considered to change between different vertebrates and invertebrates and over space and time. Structural weight (SN mg N) of a functional group is then considered to be 3.65 of the AFDW).

5.1 Growth and the feeding functional response

Functional feeding responses are included in ecosystem models to describe how prey availability affects consumer feeding rates and how food is then assimilated and converted to growth. Several different functional response models exist, but they all include parameters that represent food acquisition and processing such as prey assimilation efficiency, prey handling time, and predator clearance or ingestion rates. Direct measurements of these species-specific feeding and growth parameters are often poorly described, and thus ecosystem models use allometric assumptions in addition to more readily available observations (such as size-at-age, tagging, or stomach content data) to get a better feel for curve shape and bounds of parameter values (Table 1). Even the most meticulously estimated initial values for these parameters will likely need some adjustment as they represent the mean of several species in a functional group which dynamically change in space and time. In addition, prey selectivity, and clearance and ingestion rates are often influenced by water temperature and the size, morphology, chemical composition and behaviour of the prey.

Atlantis currently has twelve different options for modelling the feeding functional responses of predators, but most existing models default to using a modified Holling type II (Murray and Parslow 1997, Fulton et al. 2003) (Table S3). Type II functional responses assume that the consumer is limited by its capacity to process food, unless food is scarce in which case the searching rate parameter becomes more influential. Originally used for filter feeders, the modified version has been reformulated to minimise the number of parameters with a maximum growth (or feeding) rate

substituting the per-biomass handling time (Ht) parameter. In biomass pool and age-structured groups, the feeding response model generates, under constant prey biomass, different growth rates that are primarily determined by the maximum growth rate (mum) in addition to the clearance rate (C), both of which are specified for all age-cohorts. Growth can also be restricted by several other factors including light, nutrients, and spatial constraints. Both C and mum will limit a consumer's growth at low food densities if values are too low. Thus, if prey densities are perceived to be reasonable, the modeller should make efforts to test which parameter is limiting growth (and for which age classes), noting that both parameters may need some adjustment.

The clearance rate (C) defines the water volume filtered by a filter feeder or search volume of an active predator. There are different approaches to calculating initial values of C for vertebrates including using consumption rate estimates based on allometric relationships, data on the metabolic rate and dietary information (Table 1). As the type II functional response dictates that C increases with prey concentration, body-size relationship laws can be used to assume that for age-structured vertebrate groups' values of C will increase with age with rates asymptotically approaching a maximum value in adults (Kooijman 2010). A quick analysis of existing Atlantis ecosystem models suggest that values of C can be obtained by multiplying structural mass (SN) by a factor of 0.01 – 0.06 (Table S4).

The maximum growth rate (mum) of a predator can be related to the maximum ingestion rate ($Gmax$) as $mum = Gmax * E$, where E is the assimilation efficiency (Murray and Parslow 1997). The maximum ingestion rate is the inverse of the more familiar “handling time (Ht)” parameter in the traditional Holling type II relationship so that $mum = E/Ht$. Estimates of the maximum ingestion rates asymptotes with increased food density and are given by bioenergetics or ecophysiology models and can be described empirically through experimentation (Table 1). Assimilation efficiency is applied as a proportion for four different prey groups (live, plants, labile detritus and refractory detritus) and is used to convert food into assimilated energy. Assimilation efficiencies can be obtained from the experimental studies or from bioenergetics models with values typically higher in

live (primary) prey than in other (plant and detritus) food types (Table S4). With age, values of *mum* increase similarly to the von Bertalanffy growth function which is based on a bioenergetic expression of fish growth; therefore, size-at-age data can theoretically be used to estimate fish consumption rates (Essington et al. 2001). In addition, tagging data, stomach content analysis, and contaminant concentration can be used to refine parameters related to consumption. Based on existing empirical data, marine species have been shown to consume between 1-10% of their body mass per day when adult and up to 30% when juvenile (Essington et al. 2001). This typically corresponds to values of *mum* that are obtained when multiplying adult *SN* by a factor of 0.05 – 0.1 (Table S4). For most functional groups, we find that *mum* is higher than *C*, although for other groups (such as seabirds) we find that values for *C* and *mum* are similar or that *C* is slightly higher than *mum*.

5.2 Predator-prey interactions

Parameterization of the predator-prey matrix does have a major influences on the biomass trends of all functional groups. In Atlantis, these interactions are modelled in a similar way for both biomass pool and age-structured groups with the available biomass of the prey ($AvailB_{prey}$) to a predator given with the equation:

$$AvailB_{prey} = pPREY \cdot \delta_{overlap} \cdot \delta_{habitat} \cdot \delta_{size} \cdot B_{prey}$$

$pPREY$ is the stage-structured (juveniles and adults) predator-prey matrix to indicate the maximum proportion of prey biomass available to a consumer at a given point in space and time (Table 1), $\delta_{overlap}$ is the proportion of available prey based on the predator-prey spatial overlap, $\delta_{habitat}$ is the proportion of available prey based on the habitat refuge for prey, δ_{size} is the proportion of available prey based on the gape limitation of the predator (Section 5.2.1) and B_{prey} is the biomass of the prey. The concept is similar to the vulnerabilities used in the feeding arena theory (Walters and Kitchell 2001). The available food for each age class, in each spatial box, is then applied to a predator functional feeding response to determine the realised consumption or grazing of that prey (Table S3). If $pPrey$ is set to 0 predation will never occur, but a value of 1 does not guarantee that a prey will be present in the diet, as it will also depend on the spatial, temporal, habitat and size

overlap. Once the gross volume grazed is determined by the functional feeding response the consumption is pro-rated over the available prey biomasses. For example, if, after accounting for all predator-prey matrix, space and size limitations, the available prey includes 1000mg of clams and 1mg of fish, and the functional response determines that 10% of all available biomass can be grazed in that time step, the predator will ingest 100mg of clams and 0.1mg of fish. This has important implications for optimising the maximum available biomass of prey groups to predators, for example, one predator group can completely deplete the prey resource. Checking the realized diet is therefore a very important part of the tuning process.

Because of the large uncertainty typically associated with trophic interactions, the food web availability matrix is one of the most adjusted parameter when parameterizing ecosystem models including Atlantis. Realized diet values should, wherever possible, be compared to available dietary data from the study system within an appropriate time scale and, if needed, values in the predator-prey matrix should be adjusted (Table 1). Typically, p_{Prey} values >0.1 represent a strong interaction while values <0.001 signify weak interactions. Lower values are more common for vertebrate predation on biomass pool groups as there is no prey size limitation in these cases and these groups can be very abundant. High values can be common for predation on cephalopods, mesopelagic fishes and other fast growing prey groups. The availability of biomass pool groups should be low for predators that prefer to eat fish. Otherwise a predator, such as a seal, might entirely fill up on invertebrate prey and impose no top-down control on fishes. This can happen even if the availability of invertebrates to seals is as low as 0.001, as the final proportion in the diet is determined by availability and the biomass in the box. It is important to keep consumption of detritus to a minimum for all but true detritus specialists; and where it is included to make sure the associated assimilation efficiencies reflect how nutritious those food sources are to predators. This is because detritus pools can be quite extensive (particularly in the sediments) and artificially support groups that should not feed on them. In the worst case, a feedback can develop between the growth of a group (particularly a biomass pool) consuming detritus and production of detritus as waste by that group (which grows the pool of the consumers' food source allowing further growth).

5.2.1 Gape size

The gape size parameters determine the size selectivity (feeding window) and availability of age-structured prey to a predator. Prey outside the range have a refuge (narrowing the potential impact of a predator on prey abundances). In general, most predators consume prey that are relatively small (typically 10-20% of body size, Scharf et al. 2000) compared with their own body and gape size, however exceptions do exist (e.g. gelatinous zooplankton and cephalopods). In Atlantis the upper and lower gape limit parameters are applied to the structural size to look at the relative sizes of predators and prey, they do not consider mouth allometry. In most consumers mouth characteristics are strongly related to body length and feeding habits (Karachle and Stergiou 2011) while prey size is thought to be a linear function of a consumer's body length (Nilsson and Bronmark 2000). Based on these assumptions, a range of values can be considered for different ecological groups (Table S4).

5.3 Recruitment

Recruitment is the number of individuals that enter the system in the youngest age-class and is only explicitly modelled in age-structured groups and stage-structured biomass pools. Stock recruitment relationships are often essential to achieve co-existence in multi-species models, but are also among the most uncertain and hotly debated parameters. To determine the number of recruits, Atlantis determines the amount of spawn produced depending on the condition-at-age of the spawner biomass (the motivation for pulling apart structure and reserve weight) (Baretta et al. 1995). The proportion of weight-for-spawn represents the investment in reproductive structures (*FSP*, *KSPA*) used to translate available reserve weight-per-individual (*RN*) into the spawn they produce (Table 1). Fewer recruits are produced when the adults are underweight. Once the biomass pool of spawn has been determined, Atlantis calculates the number of recruits that will enter the model domain after a larval (or gestation) period.

There are several functional forms relating the number of recruits to the existing population that can be applied in Atlantis including, but not limited to, constant recruitment, linear recruitment, and Beverton-Holt (1957) and Ricker (1954) stock-recruitment models, as well as simple relationships relating recruits to planktonic or environmental conditions. These functional forms are supported by different parameters and equations (Table S5). Commonly, stock-recruitment models are only used for fish groups, or in groups that stock assessments have been undertaken with a constant or linear recruitment set for invertebrates and vertebrate groups that have fixed offspring (eggs or pups) including sharks, marine mammals, sea turtles and seabirds.

As Atlantis often does not include larval processes, many models use a density-dependent recruitment form consisting of a modified version of the Beverton-Holt spawning stock-recruit relationship that includes the condition of adults in the population. Atlantis requires knowledge of the three parameters used in the Beverton-Holt equation most commonly used in stock assessments.

$$\text{Number of recruits} = (\alpha * S) / (\beta + S)$$

where *alpha* is the asymptote for the number of recruits; *beta* is the spawning stock biomass (SSB_0 , mg WW) expected to produce recruitment of $\frac{1}{2}$ alpha. The relationship is often specified with respect to *steepness* - the proportion of virgin recruits produced by 20% of the virgin spawning stock (where high values indicative of a resilient population, which is robust to harvesting) (Mace and Doonan 1988, Francis 1992, Haddon 2001, Lee et al. 2012), virgin recruitment (R_0), and virgin spawning stock biomass (SSB_0). The initial values of the Atlantis *BHalpha* and *BHbeta* parameters can be estimated based on steepness, virgin spawning stock biomass and virgin recruitment (see example spreadsheet on the AtlantisWiki:

<https://confluence.csiro.au/pages/viewpage.action?pageId=238256130>) but likely need additional tuning to get ecological realistic values.

Due to the modifications to the Beverton-Holt spawning stock recruitment relationship that take into consideration the condition of potentially spawning adults, checking the realised SSR relationship is necessary during calibration. There may be remaining differences between the Atlantis version of

the spawning stock recruitment curves and those from corresponding stock assessments. In part, this may be due to ‘recruits’ in Atlantis entering the model at younger ages than ‘recruits’ in some stock assessments (i.e. the assessments may only track those recruiting into a fishery, which could be after many years post settlement within the ecosystem). Thus, they will have been subjective to mortality in the intervening years, hence, stock assessment estimates must be scaled to provide Atlantis numbers-of-recruits. Further, the dynamic nature of the ecosystem model will bring in fluctuations to the spawning stock recruitment relationship through adult condition that are not explicitly captured in stock assessment models.

5.4 Additional mortality

There are two main types of additional natural mortality terms: mL , the daily background linear mortality rate from sources not explicitly represented in the model; and, mQ , the daily quadratic (density dependent) mortality rate. Mortality (M) is calculated as:

$$M = (mS + mL + mQ * Quantity) * Quantity$$

where mS represents ‘special’ mortality based on environmental conditions (such as the local levels of oxygen), and $Quantity$ represents the metric for quantifying the functional group. For biomass-pooled groups this is biomass and for age-structured groups this is abundance. For age-structured groups M is computed per age class where mL and mQ are stage-specific (juvenile and adult).

Setting appropriate levels of the non-predation mortality parameters is extremely important, as their use (especially when mL or mQ is set to higher values such as $>1-e5$) can make the model insensitive to predator-prey interactions effectively turning a model into a set of single species models. It is for this reason that additional mortality settings should be used as an alternative after checking recruitment, predation pressure, and growth parameters (Fig. 3). In general, mL relates to the natural mortality term in population biology (as daily maximum value) and should be kept as low as possible (ideally zero unless there is compelling evidence to the contrary) to ensure that predation really is the dominant mortality source being represented. This should be used to fine tune the model or to represent not explicitly included processes such as senescence, parasite loads or

other stress factors (Walters and Christensen 2019). The values for mQ should also be kept low, though there is frequently the need to apply it to the largest (top) predators (e.g. mammals, sea birds) that have low natural predation levels once they are adults. In these cases, the mQ may be compensating for a lack of representation of behaviour or other ecological mechanisms that introduce density dependent controls on these populations (e.g. fine scale competition for breeding sites).

Additional ‘special’ mortality terms can be used to characterize starvation (for age-structured groups), smothering, bottom stress, loss due to hypoxia, disease, contaminant spills, fishing removals or extreme events. These forms of mortality are potentially influential but often rarely adjusted during calibration (smothering of macrophytes being an exception and the use of oxygen dependent mortality is growing). Starvation mortality is particularly underutilised. Typically, starvation mortality is set to 0.0001 for most vertebrate groups. One should investigate higher values of starvation-induced mortality, as well as testing a model run where all diets are set to 0 (with starvation expected to result). An Atlantis model typically allows for approximately 70% decrease in RN (reserve fat) before an organism is deemed to be starving which is in line with reported ranges of 60%-80% (Broekhuizen et al. 1994).

6.0 Evaluating model confidence and data quality

Uncertainty and variability are inherent with ecosystem modelling and should be appropriately explored and reported. Model development, including calibration, is a very data intensive task, making it difficult to track where exactly there is high degree of uncertainty. While those involved with model development will know which functional groups or parameters are most data limited, these are rarely reported consistently. At the end stages of model development and calibration, modellers should seek to relay details that reflect the complexity and quality of the model in a transparent manner. Standardizing the approaches taken to assessing model confidence will further allow for future ensemble model comparisons. Standardized ways of categorizing data sources in the development of EwE models has assisted their usage and uptake by allowing end-users (scientists

and fisheries managers) to assess model confidence. In a similar way, indicating the rigor of different aspects of the parameterisation in complex ecosystem models will greatly assist the calibration process. We propose a data quality approach, similar to the pedigree index used in EwE (Christensen et al. 2005), in which input and output parameters for modelled functional groups or a modelled system can be indexed (or ranked)(Table 2). Efforts should be made to combine this approach with detailed documentation of the main data source.

Based on the individual values for each functional group and group of parameters, a pedigree index for the overall model can be calculated; where the average is obtained across all the functional groups. Similarly to EwE or size-spectrum models, confidence intervals could also be assigned and then used as weightings or prior probability distributions when applying Bayesian approaches to parameter estimation (Aydin et al. 2005, Spence et al. 2015) or using other optimization algorithms, at least for a set of core model parameter. Some particularly promising techniques include approximate Bayesian computation (Elske van der Vaart et al. 2015) and Bayesian model emulation (Fer et al. 2018). Sensitivity analysis can then be targeted to the most uncertain influential parameters when evaluating the robustness of model predictions to small variations in input data. More recently, model skill assessment metrics have become a standard way of objectively assessing model performance in terms of how well it captures observed trends and variations on observed time series (Olsen et al. 2016).

7.0 Conclusions

Complex process-based ecosystem models are becoming more numerous and more accepted as a modelling approach to support strategic management of natural resources. Consequently, as the approach matures, so does the need for transparency in reporting the data sources and diagnostic approaches used to justify adjusting a potentially large number of parameter values during calibration. In this study we detailed the key steps and practices currently used to calibrate complex process-based ecosystem models, with Atlantis used as an exemplar (given its complexity and application across multiple system types). We recommend the following as best practices: (1) take

steps to understand how the system of interest is conceptualized in the model and ascertain what data sources and types have been used; (2) use available diagnostic routines to identify and report likely erroneous parameter values associated with recruitment, predator-prey interactions, growth, and additional mortality; and (3) create a pedigree based on data quality to define and document model confidence for each functional group. The core parameters discussed are similar to those used in other process-based ecosystem models, thus knowledge created here could assist other ecosystem modellers while also increasing understanding of the functionality of Atlantis ecosystem models. Moreover, the paper provides model end-users an overview of the thoroughness taken by modellers to ensure that the model is producing the most trustworthy outputs fitted to the most current and available data sources. Lastly, it provides the modelling community and end-users with access to information regarding the quality and history of the model.

Acknowledgements

Funded by a CSIRO Community of Practice internal investment project and contributions from NOAA, NIWA, IFREMER, and IMR's REDUS Project by allowing staff to contribute *in kind* via research time. KOC acknowledges support from the Claude Leon Foundation through a postdoctoral fellowship and the resources from the Cluster for High Performance Computing.

References

- Ainsworth, C.H., Walters, C.J., 2015. Ten common mistakes made in Ecopath with Ecosim modelling. *Ecological Modelling* 308, 14-17.
- Andersen, K.H., Jacobsen, N.S., Farnsworth, K.D., 2015. The theoretical foundations for size spectrum models of fish communities. *Canadian Journal of Fisheries and Aquatic Sciences* 73, 575-588.
- Audzijonyte, A., Gorton, R., Kaplan, I., Fulton, E.A., 2017a. Atlantis user's guide part I: general overview, physics & ecology. The Commonwealth Scientific and Industrial Research Organisation, Hobart, Australia
- Audzijonyte, A., Gorton, R., Kaplan, I., Fulton, E.A., 2017b. Atlantis user's guide part II: socio-economics. The Commonwealth Scientific and Industrial Research Organisation, Hobart, Australia.

- Audzijonyte, A., Pethybridge, H., Porobic, J., Gorton, R., Kaplan, I., Fulton, E.A., 2019. Atlantis: a spatially explicit end-to-end marine ecosystem model with dynamically integrated physics, ecology and socio-economics modules. *Methods in Ecology and Evolution*. doi: 10.1111/2041-210X.13272.
- Aydin, K.Y., McFarlane, G.A., King, J.R., Megrey, B.A., Myers, K.W., 2005. Linking oceanic food webs to coastal production and growth rates of Pacific salmon (*Oncorhynchus* spp.), using models on three scales. *Deep Sea Research Part II: Topical Studies in Oceanography* 52, 757-780.
- Baretta, J.W., Ebenhöf, W., Ruardij, P., 1995. The European regional seas ecosystem model, a complex marine ecosystem model. *Netherlands Journal of Sea Research* 33, 233-246.
- Barnes, C., Bethea, D.M., Brodeur, R.D., Spitz, J., Ridoux, V., Pusineri, C., Chase, B.C., Hunsicker, M.E., Juanes, F., Kellermann, A., Lancaster, J., 2008. Predator and prey body sizes in marine food webs. *Ecology* 89, 881-881. doi:10.1890/07-1551.1.
- Beverton, R.J.H., Ministry of Agriculture Fisheries and Food London (United Kingdom), Holt, S.J., 1957. On the dynamics of exploited fish populations. HMSO, London (UK) 533 p.
- Blanchard, J.L., Andersen, K.H., Scott, F., Hintzen, N.T., Piet, G., Jennings, S., 2014. Evaluating targets and trade-offs among fisheries and conservation objectives using a multispecies size spectrum model. *Journal of Applied Ecology* 51, 612-622.
- Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* 24, 127-135.
- Broekhuizen, N., Gurney, W.S.C., Jones, A., Bryant, A.D., 1994. Modeling compensatory growth. *Functional Ecology* 8, 770-782.
- Christensen, V., Walters, C.J., 2004. Ecopath with Ecosim: methods, capabilities and limitations. *Ecological Modelling* 172, 109-139.
- Christensen, V., Walters, C.J., Pauly, D., 2005. Ecopath with Ecosim: a user's guide, Fisheries Centre, University of British Columbia, Vancouver. November 2005 edition, 154 p. (available online at www.ecopath.org).
- Collie, J.S., Botsford, L.W., Hastings, A., Kaplan, I.C., Largier, J.L., Livingston, P.A., Plagányi, É., Rose, K.A., Wells, B.K., Werner, F.E., 2016. Ecosystem models for fisheries management: finding the sweet spot. *Fish and Fisheries* 17, 101-125.
- Cowie, G.L., Hedges, J.I., 1996. Digestion and alteration of the biochemical constituents of a diatom (*Thalassiosira weissflogii*) ingested by an herbivorous zooplankton (*Calanus pacificus*). *Limnology and Oceanography* 41, 581-594.
- De Roos, A.M., Persson, L., 2001. Physiologically structured models – from versatile technique to ecological theory. *Oikos* 94, 51-71.

- Essington, T.E., 2007. Evaluating the sensitivity of a trophic mass-balance model (Ecopath) to imprecise data inputs. *Canadian Journal of Fisheries and Aquatic Sciences* 64, 628-637.
- Essington, T.E., Kitchell, J.F., Walters, C.J., 2001. The von Bertalanffy growth function, bioenergetics, and the consumption rates of fish. *Canadian Journal of Fisheries and Aquatic Sciences* 58, 2129-2138.
- Fer, I., Kelly, R., Moorcroft, P.R., Richardson, A.D., Cowdery, E.M., Dietze, M.C., 2018. Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation. *Biogeosciences* 15, 5801-5830.
- Fiechter, J., Rose, K.A., Curchitser, E.N., Hedstrom, K.S., 2015. The role of environmental controls in determining sardine and anchovy population cycles in the California Current: analysis of an end-to-end model. *Progress in Oceanography* 138, 381-398.
- Francis, R.I.C.C., 1992. Use of risk analysis to assess fishery management strategies: a case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise, New Zealand. *Canadian Journal of Fisheries and Aquatic Sciences* 49, 922-930.
- Fulton, E.A., Smith, A.D.M., Johnson, C.R., 2003. Mortality and predation in ecosystem models: is it important how these are expressed? *Ecological Modelling* 169, 157-178.
- Fulton, E.A., Parslow, J.S., Smith, A.D.M., Johnson, C.R., 2004a. Biogeochemical marine ecosystem models II: the effect of physiological detail on model performance. *Ecological Modelling* 173, 371-406.
- Fulton, E.A., Smith, A.D.M., Johnson, C.R., 2004b. Effects of spatial resolution on the performance and interpretation of marine ecosystem models. *Ecological Modelling* 176, 27-42.
- Fulton, E.A., Smith, A.D.M., Punt, A.E., 2005. Which ecological indicators can robustly detect effects of fishing? *ICES Journal of Marine Science* 62, 540-551.
- Fulton, E.A., Smith, A.D.M., Smith, D.C., 2007. Alternative management strategies for southeast Australian commonwealth fisheries: stage 2 - quantitative management strategy evaluation. Commonwealth Scientific and Industrial Research Organisation (CSIRO) Hobart.
- Fulton, E.A., Link, J.S., Kaplan, I.C., Savina-Rolland, M., Johnson, P., Ainsworth, C., Horne, P., Gorton, R., Gamble, R.J., Smith, A.D.M., Smith, D.C., 2011. Lessons in modelling and management of marine ecosystems: the Atlantis experience. *Fish and Fisheries* 12, 171-188.
- Fulton, E.A., Bulman, C.M., Pethybridge, H., Goldsworthy, S.D., 2018. Modelling the Great Australian Bight ecosystem. *Deep-Sea Research Part II-Topical Studies in Oceanography* 157, 211-235.
- Garcia, S.M., Zerbi, A., Aliaume, C., Do Chi, T., Lasserre, G., 2003. The ecosystem approach to fisheries: issues, terminology, principles, institutional foundations, implementation and outlook. United Nations. Food and Agriculture Organization. Rome, Italy. FAO Fisheries Technical Paper No. 443. 71 p.

- Garcia, S.M., Cochrane, K.L., 2005. Ecosystem approach to fisheries: a review of implementation guidelines. *ICES Journal of Marine Science* 62, 311-318.
- Grimm, V., Railsback, S.F., 2012. Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367, 298-310.
- Grüss, A., Schirripa, M.J., Chagaris, D., Velez, L., Shin, Y.-J., Verley, P., Oliveros-Ramos, R., Ainsworth, C.H., 2016. Estimating natural mortality rates and simulating fishing scenarios for Gulf of Mexico red grouper (*Epinephelus morio*) using the ecosystem model OSMOSE-WFS. *Journal of Marine Systems* 154, 264-279.
- Grüss, A., Rose, K.A., Simons, J., Ainsworth, C.H., Babcock, E.A., Chagaris, D.D., De Mutsert, K., Froeschke, J., Himchak, P., Kaplan, I.C., O'Farrell, H., Rejon, M.J.Z., 2017. Recommendations on the use of ecosystem modeling for informing ecosystem-based fisheries management and restoration outcomes in the Gulf of Mexico. *Marine and Coastal Fisheries* 9, 281-295.
- Grüss, A., Palomares, M.L.D., Poelen, J.H., Barile, J.R., Aldemita, C.D., Ortiz, S.R., Barrier, N., Shin, Y.-J., Simons, J., Pauly, D., 2019. Building bridges between global information systems on marine organisms and ecosystem models. *Ecological Modelling* 398, 1-19.
- Gulland, J.A., 1970. The fish resources of the ocean. United Nations. Food and Agriculture Organization. Rome, Italy. FAO Fisheries Technical Paper No. 97. 255 p.
- Haddon, M., 2001. Modelling and quantitative methods in fisheries. Chapman & Hall, Boca Raton, FL (USA) 406 p.
- Hansen, C., Drinkwater, K.F., Jähkel, A., Fulton, E.A., Gorton, R., Skern-Mauritzen, M., 2019. Sensitivity of the Norwegian and Barents Sea Atlantis end-to-end ecosystem model to parameter perturbations of key species. *Plos One* 14, e0210419.
- Heymans, J.J., Coll, M., Link, J.S., Mackinson, S., Steenbeek, J., Walters, C., Christensen, V., 2016. Best practice in Ecopath with Ecosim food-web models for ecosystem-based management. *Ecological Modelling* 331, 173-184.
- Hill, S.L., Watters, G.M., Punt, A.E., McAllister, M.K., Le Quéré, C., Turner, J., 2007. Model uncertainty in the ecosystem approach to fisheries. *Fish and Fisheries* 8, 315-336.
- Holling, C.S., 1966. The functional response of invertebrate predators to prey density. *Memoirs of the Entomological Society of Canada* 98, 5-86.
- Horne, P.J., Kaplan, I.C., Marshall, K.N., Levin, P.S., Harvey, C.J., Hermann, A.J., Fulton, E.A., 2010. Design and parameterization of a spatially explicit ecosystem model of the Central California Current. U.S. Dept. Commer., NOAA Technical Memorandum NMFS-NWFSC-104. 140 p.
- Kaplan, I.C., Levin, P.S., Burden, M., Fulton, E.A., 2010. Fishing catch shares in the face of global change: a framework for integrating cumulative impacts and single species management. *Canadian Journal of Fisheries and Aquatic Sciences* 67, 1968-1982.

- Kaplan, I.C., Horne, P.J., Levin, P.S., 2012. Screening California Current fishery management scenarios using the Atlantis end-to-end ecosystem model. *Progress in Oceanography* 102, 5-18.
- Kaplan, I.C., Marshall, K.N., 2016. A guinea pig's tale: learning to review end-to-end marine ecosystem models for management applications. *ICES Journal of Marine Science* 73, 1715-1724.
- Karachle, P.K., Stergiou, K.I., 2011. Mouth allometry and feeding habits of some Mediterranean fishes. *Acta Ichthyologica Et Piscatoria* 41, 265-275.
- Kooijman, S.A.L.M., 2000. *Dynamic energy and mass budgets in biological systems*. Cambridge University Press.
- Kooijman, S.A.L.M., 2010. *Dynamic energy budget theory for metabolic organisation*. Cambridge University Press.
- Kramer-Schadt, S., Revilla, E., Wiegand, T., Grimm, V., 2007. Patterns for parameters in simulation models. *Ecological Modelling* 204, 553-556.
- Krapu, C., Borsuk, M., 2019. Probabilistic programming: A review for environmental modellers. *Environmental Modelling & Software* 114, 40-48.
- Larkin, P.A., 1996. Concepts and issues in marine ecosystem management. *Reviews in Fish Biology and Fisheries* 6, 139-164.
- Lee, H.-H., Maunder, M.N., Piner, K.R., Methot, R.D., 2012. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? *Fisheries Research* 125-126, 254-261.
- Link, J., 2002. Does food web theory work for marine ecosystems? *Marine Ecology Progress Series* 230, 1-9.
- Link, J.S., 2004. A general model of selectivity for fish feeding: a rank proportion algorithm. *Transactions of the American Fisheries Society* 133, 655-673.
- Link, J.S., 2010. Adding rigor to ecological network models by evaluating a set of pre-balance diagnostics: a plea for PREBAL. *Ecological Modelling* 221, 1580-1591.
- Link, J.S., Fulton, E.A., Gamble, R.J., 2010. The northeast US application of Atlantis: a full system model exploring marine ecosystem dynamics in a living marine resource management context. *Progress in Oceanography* 87, 214-234.
- Link, J.S., Gamble, R.J., Fulton, E.A., 2011. NEUS--Atlantis construction, calibration, and application of an ecosystem model with ecological interactions, physiographic conditions, and fleet behavior. U.S. Dept. Commer., NOAA Technical Memorandum NMFS-NE-218. 247 p.
- Link, J.S., Ihde, T.F., Harvey, C.J., Gaichas, S.K., Field, J.C., Brodziak, J.K.T., Townsend, H.M., Peterman, R.M., 2012. Dealing with uncertainty in ecosystem models: the paradox of use for living marine resource management. *Progress in Oceanography* 102, 102-114.

- Livingston, P.A., Aydin, K., Buckley, T.W., Lang, G.M., Yang, M.-S., Miller, B.S., 2017. Quantifying food web interactions in the North Pacific – a data-based approach. *Environmental Biology of Fishes* 100, 443-470.
- Mace, P.M., Doonan, I.J., 1988. A generalized bioeconomic simulation model for fish population dynamics. MAFFish, N.Z. Ministry of Agriculture and Fisheries.
- McGregor, V.L., Horn, P.L., Fulton, E.A., Dunn, M.R., 2019. From data compilation to model validation: a comprehensive analysis of a full deep-sea ecosystem model of the Chatham Rise. *PeerJ* 7, e6517. <https://doi.org/10.7717/peerj.6517>.
- Melbourne-Thomas, J., Johnson, C.R., Fulton, E.A., 2011a. Regional-scale scenario analysis for the Meso-American Reef system: modelling coral reef futures under multiple stressors. *Ecological Modelling* 222, 1756-1770.
- Melbourne-Thomas, J., Johnson, C.R., Perez, P., Eustache, J., Fulton, E.A., Cleland, D., 2011b. Coupling biophysical and socioeconomic models for coral reef systems in Quintana Roo, Mexican Caribbean. *Ecology and Society* 16.
- Murray, A.G., Parslow, J.S., 1997. Port Phillip Bay integrated model: final report. Port Phillip Bay Environmental Study. Technical Report No.44. CSIRO Environmental Projects Office, Canberra, ACT 2601, GPO Box 1666.
- National Marine Fisheries Service, 1999. Ecosystem-based fishery management: a report to Congress by the Ecosystem Principles Advisory Panel. United States. National Marine Fisheries Service. Ecosystem Principles Advisory Panel, Washington, D.C.
- Nilsson, P.A., Brönmark, C., 2000. Prey vulnerability to a gape-size limited predator: behavioural and morphological impacts on northern pike piscivory. *Oikos* 88, 539-546.
- Olsen, E., Fay, G., Gaichas, S., Gamble, R., Lucey, S., Link, J.S., 2016. Ecosystem model skill assessment. Yes we can!. *Plos One* 11.
- Olsen, E., Kaplan, I.C., Ainsworth, C., Fay, G., Gaichas, S., Gamble, R., Girardin, R., Eide, C.H., Ihde, T.F., Morzaria-Luna, H.N., Johnson, K.F., Savina-Rolland, M., Townsend, H., Weijerman, M., Fulton, E.A., Link, J.S., 2018. Ocean futures under ocean acidification, marine protection, and changing fishing pressures explored using a worldwide suite of ecosystem models. *Frontiers in Marine Science* 5.
- Ortega-Cisneros, K., Cochrane, K., Fulton, E.A., 2017. An Atlantis model of the southern Benguela upwelling system: validation, sensitivity analysis and insights into ecosystem functioning. *Ecological Modelling* 355, 49-63.
- Pantus, F.J., 2007. Sensitivity analysis for complex ecosystem models, School of Physical Sciences. University of Queensland, Brisbane, Australia.
- Pethybridge, H.R., Choy, C.A., Polovina, J.J., Fulton, E.A., 2018. Improving Marine Ecosystem Models with Biochemical Tracers. *Annual Review of Marine Science* 10, 199-228.

- Pikitch, E.K., Santora, C., Babcock, E.A., Bakun, A., Bonfil, R., Conover, D.O., Dayton, P., Doukakis, P., Fluharty, D., Heneman, B., Houde, E.D., Link, J., Livingston, P.A., Mangel, M., McAllister, M.K., Pope, J., Sainsbury, K.J., 2004. Ecosystem-Based Fishery Management. *Science* 305, 346.
- Pinnegar, J.K., 2014. DAPSTOM - an integrated database & portal for fish stomach records. (Version 4.7). Centre for Environment, Fisheries & Aquaculture Science. 39 p, Lowestoft, UK.
- Plagányi, É.E., 2007. Models for an ecosystem approach to fisheries. United Nations. Food and Agriculture Organization. Rome, Italy. FAO Fisheries Technical Paper No. 477. 108 p.
- Plagányi, É.E., Punt, A.E., Hillary, R., Morello, E.B., Thébaud, O., Hutton, T., Pillans, R.D., Thorson, J.T., Fulton, E.A., Smith, A.D.M., Smith, F., Bayliss, P., Haywood, M., Lyne, V., Rothlisberg, P.C., 2014. Multispecies fisheries management and conservation: tactical applications using models of intermediate complexity. *Fish and Fisheries* 15, 1-22.
- Polovina, J.J., 1984. Model of a coral reef ecosystem. *Coral Reefs* 3, 1-11.
- Porobic, J., Fulton, E., Frusher, S., Parada, C., Haward, M., Ernst, B., Stram, D., 2018. Implementing Ecosystem-based Fisheries Management: Lessons from Chile's experience. *Marine Policy* 97, 82-90.
- Porobic, J., Fulton, E., Parada, C., Frusher, S., Ernst, B., Manriquez, P., 2019. The impact of fishing on a highly vulnerable ecosystem, the case of Juan Fernandez Ridge ecosystem. *Plos One* 14(2): e0212485. <https://doi.org/10.1371/journal.pone.0212485>.
- Redfield, A.C., 1958. The biological control of chemical factors in the environment. *American Scientist* 46, 205-221.
- Rice, J., Gislason, H., 1996. Patterns of change in the size spectra of numbers and diversity of the North Sea fish assemblage, as reflected in surveys and models. *ICES Journal of Marine Science* 53, 1214-1225.
- Ricker, W.E., 1954. Stock and Recruitment. *Journal of the Fisheries Research Board of Canada* 11, 559-623.
- Scharf, F.S., Juanes, F., Rountree, R.A., 2000. Predator size-prey size relationships of marine fish predators: interspecific variation and effects of ontogeny and body size on trophic-niche breadth. *Marine Ecology Progress Series* 208, 229-248.
- Sheldon, R.W., Parsons, T.R., 1967. A continuous size spectrum for particulate matter in the sea. *Journal of the Fisheries Research Board of Canada* 24, 909-915.
- Shin, Y.J., Cury, P., 2004. Using an individual-based model of fish assemblages to study the response of size spectra to changes in fishing. *Canadian Journal of Fisheries and Aquatic Sciences* 61, 414-431.
- Shin, Y.J., Houle, J.E., Akoglu, E., Blanchard, J.L., Bundy, A., Coll, M., Demarcq, H., Fu, C.H., Fulton, E.A., Heymans, J.J., Salihoglu, B., Shannon, L., Sporcic, M., Velez, L., 2018. The

- specificity of marine ecological indicators to fishing in the face of environmental change: a multi-model evaluation. *Ecological Indicators* 89, 317-326.
- Spence, M.A., Blackwell, P.G., Blanchard, J.L., 2015. Parameter uncertainty of a dynamic multispecies size spectrum model. *Canadian Journal of Fisheries and Aquatic Sciences* 73, 589-597.
- Sturludottir, E., 2018. Exploring the effects of discarding using the Atlantis ecosystem model for Icelandic waters. *Scientia Marina* 82, 51-62.
- Subbey, S., 2018. Parameter estimation in stock assessment modelling: caveats with gradient-based algorithms. *Ices Journal of Marine Science* 75, 1553-1559.
- Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L., Martinez, N.D., Rantala, H., Romanuk, T.N., Stouffer, D.B., Tylianakis, J.M., 2012. Food webs: reconciling the structure and function of biodiversity. *Trends in Ecology & Evolution* 27, 689-697.
- Travers-Trolet, M., Shin, Y.J., Field, J.G., 2014. An end-to-end coupled model ROMS-N2P2Z2D2-OSMOSE of the southern Benguela foodweb: parameterisation, calibration and pattern-oriented validation. *African Journal of Marine Science* 36, 11-29.
- van der Vaart, E., Beaumont, M.A., Johnston, A.S.A., Sibly, R.M., 2015. Calibration and evaluation of individual-based models using Approximate Bayesian Computation. *Ecological Modelling* 312, 182-190.
- Walters, C., Christensen, V., 2019. Effect of non-additivity in mortality rates on predictions of potential yield of forage fishes. *Ecological Modelling* 410, e108776. <https://doi.org/10.1016/j.ecolmodel.2019.108776>.
- Walters, C., Kitchell, J.F., 2001. Cultivation/depensation effects on juvenile survival and recruitment: implications for the theory of fishing. *Canadian Journal of Fisheries and Aquatic Sciences* 58, 39-50.
- Weijerman, M., Fulton, E.A., Janssen, A.B.G., Kuiper, J.J., Leemans, R., Robson, B.J., van de Leemput, I.A., Mooij, W.M., 2015a. How models can support ecosystem-based management of coral reefs. *Progress in Oceanography* 138, 559-570.
- Weijerman, M., Fulton, E.A., Kaplan, I.C., Gorton, R., Leemans, R., Mooij, W.M., Brainard, R.E., 2015b. An Integrated Coral Reef Ecosystem Model to Support Resource Management under a Changing Climate. *Plos One* 10, e0144165.
- Weijerman, M., Link, J.S., Fulton, E.A., Olsen, E., Townsend, H., Gaichas, S., Hansen, C., Skern-Mauritzen, M., Kaplan, I.C., Gamble, R., Fay, G., Savina, M., Ainsworth, C., van Putten, I., Gorton, R., Brainard, R., Larsen, K., Hutton, T., 2016. Atlantis Ecosystem Model Summit: Report from a workshop. *Ecological Modelling* 335, 35-38.
- Welch, H.E., 1968. Relationships between assimilation efficiencies and growth efficiencies for aquatic consumers. *Ecology* 49, 755-759.

White, C.R., Marshall, D.J., 2019. Should we care if models are phenomenological or mechanistic?
Trends in Ecology & Evolution 34, 276-278.

Tables and Figures

Table 1. Core biological parameters of an Atlantis ecosystem model that often need optimising during the calibration phase. Parameters marked with a * indicate that they effect processes influenced by a temperature (e.g. Q10).

Process	Parameter	T	Units	Description and use	Preferred data sources ^
Growth	Maximum growth efficiency (<i>mum</i>)	*	mgN day ⁻¹ or mgN ind ⁻¹ day ⁻¹	Included in a grazing (feeding) equation for all groups it relates to the maximum ingestion rate. Biomass pools have 1 entry, while age-structured groups have 1 per age class.	<ol style="list-style-type: none"> 1. Inferred by equations for consumption, and an assumption about assimilation or conversion efficiency (see further details in Audzijonyte et al. 2017a). 2. Related to the maximum ingestion rate (<i>Gmax - units</i>) as $mum = Gmax * E$, where <i>E</i> is the assimilation efficiency (Murray and Parslow 1997). 3. consumption rates based on stomach content data, bioenergetics models or regression models (such as von Bertalanffy growth model)
	Clearance rates (<i>C</i>)	*	m ³ mg N ⁻¹ day ⁻¹ or m ³ ind ⁻¹ day ⁻¹	Defines the water volume filtered by a filter feeder, or search volume of a consumer. It is directly related to the maximum consumption rate. Biomass pools have 1 entry, while age-structured groups have 1 per age class.	<ol style="list-style-type: none"> 1. Daily ration estimates based on stomach content data 2. Literature on gape size, swimming speed of a predator, and search time per food, to calculate the amount of water filtered (Porobic et al. 2019). 3. Allometric relationships e.g. $Ci = a \cdot (RNi + SNi) 0.7$ for mammals and birds and $Ci = b \cdot ((RNi + SNi) - (RNi-I + SN i-I))$ for fish where <i>Ci</i> is consumption for age group <i>i</i>, <i>a</i> is a constant, <i>b</i> is a conversion efficiency of ~10, and <i>RN</i> and <i>SN</i> are weight-at-age in terms of reserve and structure (Kaplan et al. 2012, Weijerman et al. 2015b, Sturludottir et al. 2018). 4. Based on estimates of the allometric slope of consumption rates (Essington et al. 2001), when size-at-age is taken from von Bertalanffy growth relationships. 5. Ecopath QB and PB parameters converted to a daily rate and multiplying by a factor (i.e. 1.0 – 1.3) to account for representation of maximum consumption and growth rates.
Food web	Prey availability or predation pressure (<i>pPrey</i>)		Scaled (0-1)	Stage-structured (juveniles and adults) diet connection matrix (or predator-prey matrix) to indicate the maximum proportion of prey biomass availability to a consumer at a given point in space and time. If 0, predation will never occur.	<ol style="list-style-type: none"> 1. Diet composition studies based on stomach content analysis, stable isotopes and fatty acids from the study system (Pethybridge et al. 2018) 2. Developing a general model of fish feeding selectivity (Link 2004, Link et al. 2011) 3. Data from similar ecosystems; open access databases (Pinnegar 2014, Barnes et al. 2008, Livingston et al. 2017, Link 2002, Thompson et al 2012)
	Assimilation efficiency (<i>E</i> , <i>Eplant</i> , <i>EDR</i> , <i>EDL</i>)	*	Scaled (0-1)	Efficiency in which consumed food translates to realised growth, which differs for different food types (live, plant, labile detritus, and refractory detritus).	<ol style="list-style-type: none"> 1. Actual values derived from the literature with nitrogen assimilation at a higher efficiency than carbon, especially in food with lower protein (N) content (Cowie and Hedges 1996). 2. Higher assimilation efficiencies expected for predators than for herbivores (Welch 1968). 3. Assimilation efficiency = $(I-E)/I \times 100\%$, where <i>I</i> is the ingestion rate and <i>E</i> is the egestion (defecation) rate.

	Gape limits (<i>KLP</i> , <i>KUP</i>)		Scaled (0-1)	Determines the upper and lower limits of the size selectivity (or refuge) of age-structured prey to a predator.	1. Actual values derived from the literature (Karachle and Stergiou 2011, Nilsson and Brönmark 2000) 2. Assuming a linear function of consumer body length, allometric relationships.
Recruitment	Numbers of recruits (<i>KDENR</i> , <i>BHalpha</i> , <i>BHbeta</i>)		# (constant) # (per adult) mg N	Recruitment is typically set as constant for invertebrates, turtles, mammals and birds, or based on the Beverton-Holt (1959) or Ricker (1969) spawning stock - recruit relationships for fishes.	1. Stock assessment models; convert quantities derived from stock assessment to Atlantis Beverton-Holt parameters. 2. Single-species life history parameters (such as fecundity) featured in papers, reports or Fishbase/Seabase references 3. Observed recruitment patterns
	Adult investment of reserves (<i>FSP</i> , <i>KSPA</i>)		Scaled (0-1) & Constant	Translates available reserve weight per individual into the spawn they produce.	1. Physiology and biochemical indices (Proximate composition or calorimetry measures of reproduction and structural tissues) 2. Energy allocation models that link reproductive outputs to growth rates. 3. Life history theory (small short-lived species allocate more energy relative to size to reproduction)
Additional mortality	Linear (mL), Quadratic (mQ), Starvation (starve)		day ⁻¹	Background linear and quadratic natural mortality not represented in the model via predation or fisheries.	1. Physiological studies that look at starvation effects. 2. Stock assessments often calculate natural mortality values, which then needs to be transferred into the different types of mortality

^ Where data cannot be sourced, field experts should always be consulted.

Table 2. A proposed pedigree routine for core input parameters and output-data comparisons used in an Atlantis model with higher values indicating better quality of data sources used. This routine should be followed for all functional groups and core parameter groups. Adapted from Christensen et al. (2005) for the EwE ecosystem model framework. Examples of data sources for the different parameters are presented in Table S2.

Parameter	Data type used	Index
Predator-prey matrix (pPrey)	From existing models of the same system (e.g. EwE)	1
	Qualitative knowledge of similar groups and systems	2
	Expert knowledge of the same group and system	3
	Quantitative knowledge of same group / different systems	4
	Quantitative knowledge of same group / same system	5
Weights (SN, RN), Growth curves, Recruitment	Approximate or indirect methods	1
	Taken from another model of the same group / system	2
	Expert knowledge	3
	Empirical relationships (including from stock assessments)	4
	Local data taken from the literature or surveys	5
Initial biomass, Fishing mortality	Approximate or indirect methods	1
	Taken from another model of the same group / system	2
	Local data taken from the literature	3
	Local stock assessments	4
	Sampling based (e.g. independent surveys)	5
Outputs		
Biomass or other variable trends	Approximate or indirect methods	1
	Relative local estimates (some idea of local trends)	2
	Fit to patchy or incomplete local time-series data (<5 years)	3
	Fit to complete time-series data (>5 years)	4
	Fit to complete spatial time series data	5

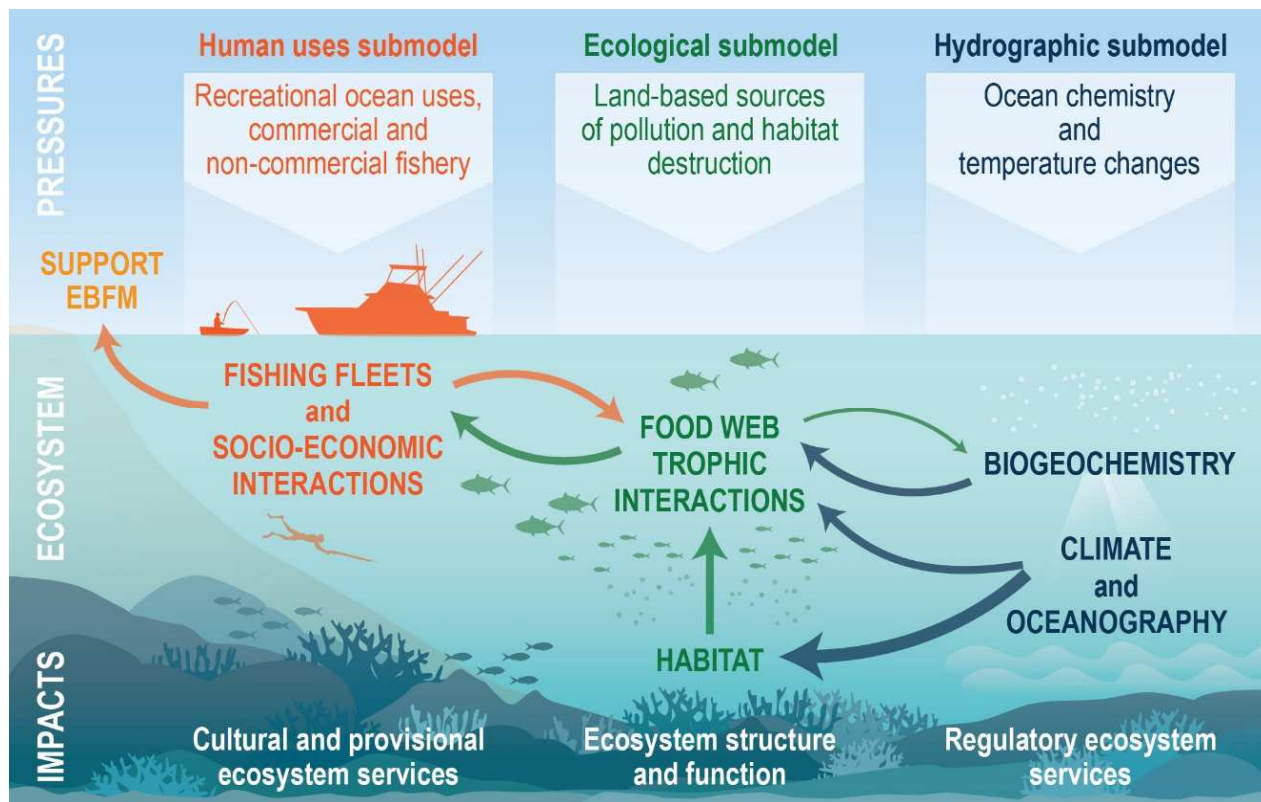


Figure 1. Schematic representation of the main processes defined in various sub-models included in an Atlantis ecosystem model. The pressures are included as forcing files, the model simulates the changes in the ecosystem (biophysical and socio-economic) condition which has an impact on the ecosystem services it provides. Figure created by Amanda Dillon, NOAA PIFSC.

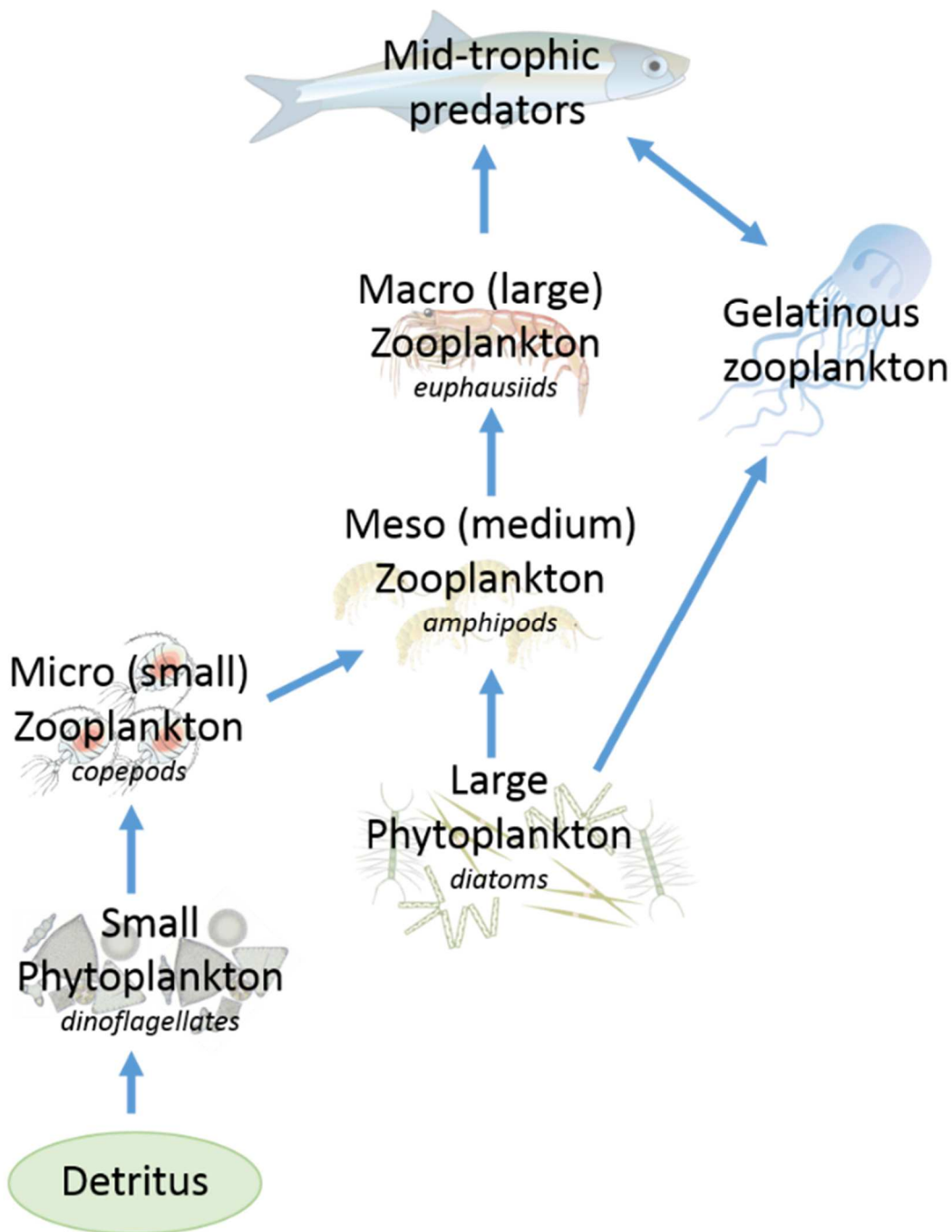


Figure 2. Representation of the lower trophic food web including linkages to the detrital-microbial food web. Images from Tracy Saxby IAN image library.

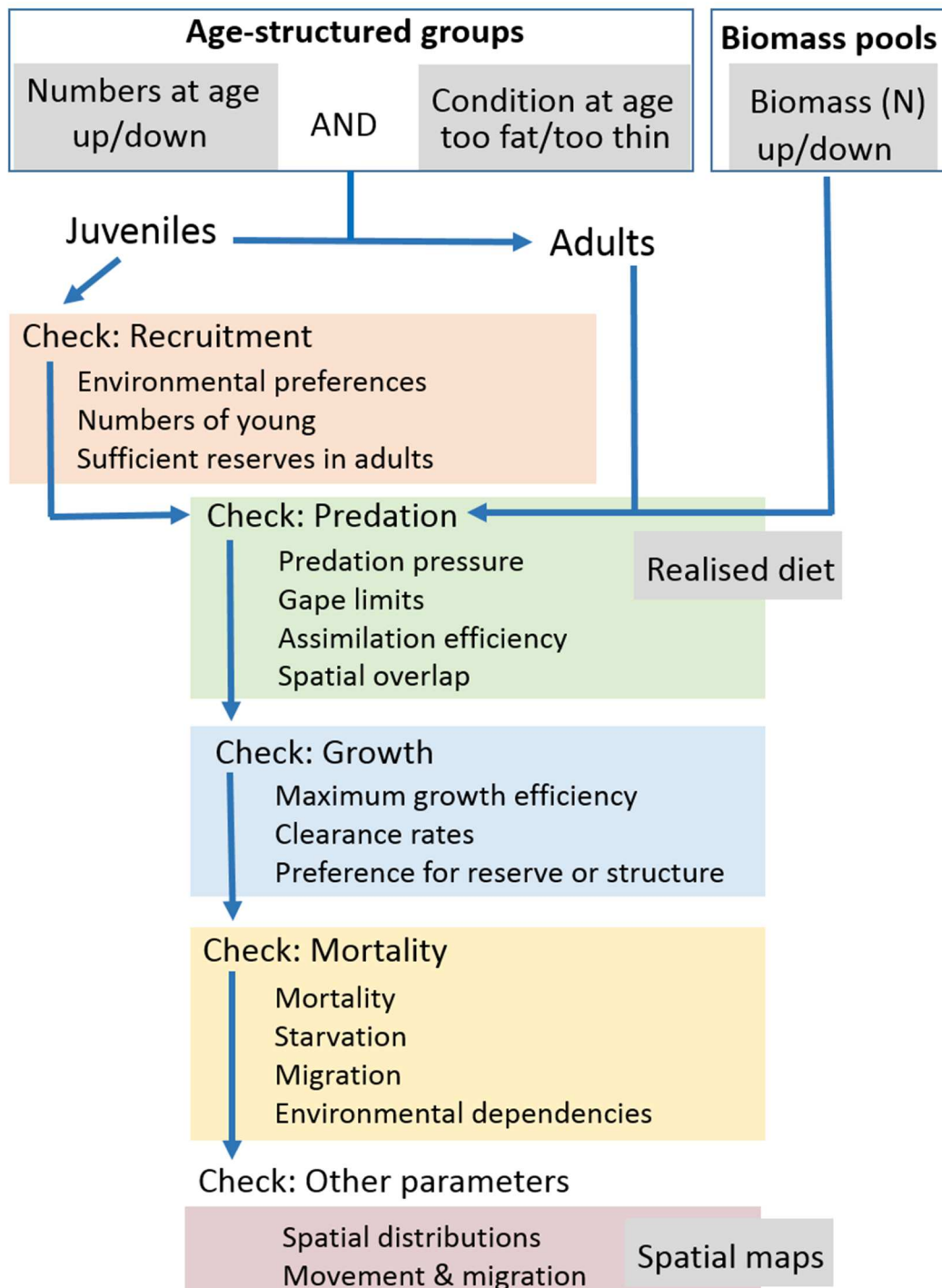


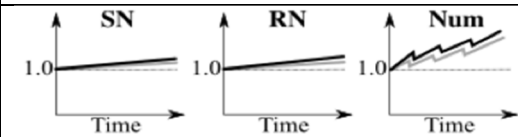
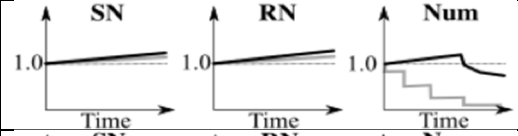
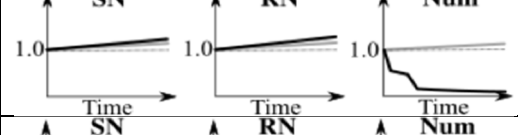
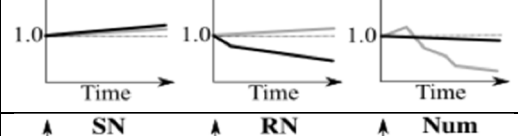
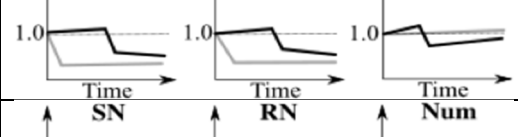
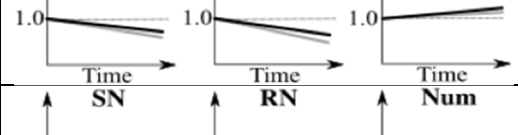
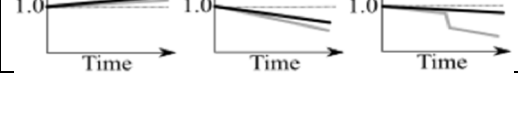
Figure 3. Calibration flow chart indicating the sequence in which core parameters for biomass pool and age-structured functional groups should be checked in Atlantis. Grey boxes indicate essential model outputs that can assist in diagnosing which parameters need adjusting.

Supplementary Information

Supplementary Table S1. Main sources of empirical or model base data to assist different biological input and output parameters of an Atlantis ecosystem model (of which often don't require lengthy optimization during the model calibration phase, depending on data type and quality; Table 2).

Parameters or trends	Data sources
Nutrient and primary productivity dynamics	<ul style="list-style-type: none"> – Hydrodynamics (nutrient composition) observational data – Biogeochemical or NPZ model – Remote sensing satellite data
Population size (biomass/abundance)	<ul style="list-style-type: none"> – Observational-based field surveys (counts and binary data) – Mark-recapture studies including close-kin genetic tagging – Acoustic video or sonar-based assessment methods
Habitat cover	<ul style="list-style-type: none"> – Fisheries (catch-effort) data – Bioregionalisation/GIS approaches and associated mapping (with data from numerous and diverse sources) – Habitat assessment and classification surveys (diving or video/quadrant or transect)
Spatial distributions and habitat associations (vertical and horizontal)	<ul style="list-style-type: none"> – Tagging or acoustic telemetry data – Bioregionalisation/GIS approaches and associated mapping of species occurrence records – Acoustic or sonar-based observations – eDNA data (species presence/absence) – Occurrence records (Fisheries, survey or citizen science data) – Species distribution or state-space models that make inferences based on known habitat and environmental preferences
Movement patterns and migration	<ul style="list-style-type: none"> – Acoustic and remote sensing telemetry – Fisheries spatiotemporal data – Mark and recapture studies
General life-history and growth (size-at-age/maturity)	<ul style="list-style-type: none"> – Morphological-based or chronological (e.g. otolith count) assessment – Published values on FishBase, SeafBase, or other publicly available databases – Growth models used in stock assessments
Environmental tolerances (Temperature, salinity, pH)	<ul style="list-style-type: none"> – Ecophysiological (field or experimental) studies – Inferred from spatiotemporal information on species distribution and depth range

Supplementary Table S2. Diagnostic plots to assist calibration of age-structured functional groups in an Atlantis ecosystem model. Tracking the relative changes in structural biomass (SN; hard parts of an organism, bones), residual biomass (RN; soft parts of an organism, fat) and density (Num) can effectively diagnose improper parameterizations. The following table summarizes common symptoms of improper parameterizations, how to observe the symptom, and the actions to take to re-parameterize the group to improve calibration. Model parameters that are key to model calibration (detailed in Table 1) include: the predator-prey matrix ($pPrey$), maximum growth rates (mum), clearance rates (C), stage-specific mortality rates (mL and mQ), number of recruits ($KDENR$ and $BHalpha$), the size of recruits ($KWRR$ and $KWSR$), and the preference for growing RN rather than SN (pR).

Symptom	Observation		Action
Too many of them	SN and RN are acceptable. Problematic rapid increase in number of juveniles and adults.		Reduce number of juveniles and adults - Check diet: increase predation on the group or decrease predation on the prey ($pPrey$) - Check growth: decrease maximum growth rates (mum) and/or clearance rates (C) - Check mortality: increase stage-specific mortality rates (mL and/or mQ)
Too few babies	SN and RN are acceptable. Problematic crash in number of juveniles.		Increase number of juveniles - Check recruitment: increase recruitment ($KDENR$ and/or $BHalpha$) - Check diet: decrease the predation on juveniles ($pPrey$)
Too few adults	SN and RN are acceptable. Problematic decrease in number of adults.		Increase number of adults - Check diet: decrease predation on adults ($pPrey$) - Check mortality: decrease stage-specific mortality rates (mL and/or mQ) for adults
Adults too skinny	SN is acceptable. Problematic continuous decrease in RN of adults, which causes poor recruitment (i.e., crash of juvenile numbers)		Increase RN of adults - Check recruitment: decrease recruitment ($KDENR$ and/or $BHalpha$) - Check growth: increase preference for laying down RN rather than SN (pR) - Check diet: increase predation by adults ($pPrey$) - Check clearance rates: increase clearance rates (C) of adults
Size-at-age	Stable recruitment. Problematic sudden drop of SN and RN for both juveniles and adults.		Stabilize SN and RN - Check recruitment: make sure size or recruits ($KWRR$ and/or $KWSR$) is correct - Check growth: increase maximum growth rates (mum) and/or clearance rates (C)
Not growing enough	Numbers are acceptable. Problematic continuous decrease in SN and RN of both juveniles and adults.		Increase SN and RN - Check diet: increase predation by the group at all stages ($pPrey$) - Check growth: increase maximum growth rates (mum)
Not getting fat	SN is acceptable. Problematic continuous decrease in RN of both juveniles and adults and crashing number of juveniles.		Increase RN and improve recruitment - Check growth: increase preference for laying down RN rather than SN (pR) - Check reproduction: change age class of reproduction ($FSPB$)

Supplementary Table S3. Description of the Holling type II feeding functional responses in Atlantis that calculate the amount of prey biomass consumed or grazed on by a predator (Gr_{prey}). See Table 4 for meanings and descriptions of the Atlantis parameters. In both equations B is the biomass of a consumer (predator) while $AvailB_{prey}$ is the biomass of all prey available to the consumer. Other feeding functional response equations available in Atlantis are described in the Atlantis manual, Table 10.5 (Audzijonyte et al. 2017a).

Feeding functional response flag (<i>predcase_XX</i>)	Description	Atlantis parameters	Equation
0 = Modified Holling type II	Adopted from PPBIM ¹	$pPrey$ mum C E	$Gr_{prey} = \frac{B C AvailB_{prey}}{1 + \frac{C \sum_i (E_i AvailB_{prey,i})}{mum}}$
8 = standard Holling type II	As traditionally used and described by Holling 1966, and is characterized by a decelerating intake rate, which follows from the assumption that the consumer is limited by its capacity to process food.	C Handling time (ht in biomass pools and both hta and htb in age-structured groups).	$Gr_{prey} = \frac{B C AvailB_{prey}}{1 + C HT \sum_i AvailB_{prey,i}}$ <p>where</p> $HT = \delta_{starve} hta (SN XCN)^{-htb}$ <p>where δ_{starve} is a scaler that reduces handling time (HT), SN is the structural mass, and XCN is the C:N Redfield Ratio.</p>

¹ PPBIM: Port Phillip Bay Integrated Model.

Supplementary Table S4. Estimated boundaries for gape size and assimilation parameters for different ecological groups based on existing Atlantis ecosystem models. These parameter ranges represent an enormous amount of species and study systems and thus should only be used as a guideline where absolutely no data is available in the model system.

Ecological group	Growth		Gape size		Assimilation Efficiency			
	Clearance rate (C)	Maximum growth rate (mum)	Lower limit (KLP_X)	Upper limit (KUP_X)	Live food and carrion (E_X)	Plants (EPlant_X)	labile detritus (EDL_X)	refractory detritus (EDR_X)
Units	m ³ ind ⁻¹ day ⁻¹	mgN ⁻¹ day ⁻¹	Scaler, 0-1 (small to large)		Scaler, 0-1 (low – high)			
Fish, sharks, whales	SN ^x x 0.01-0.06	SN ^x x 0.05-0.1	0.0001; 1.0e-10*	0.4-0.45; 0.000002*	0.07 - 0.9	0.001 - 0.2	0.001 - 0.2	0.001 - 0.1
Cephalopods	0.01 - 0.2	0.02 - 0.4	0.005	0.2 - 0.4	0.25 - 0.5	0.25 - 0.5	0.001 - 0.2	0.001 - 0.1
Macrobenthos	0.0002 - 0.004	0.001 - 0.03	0.001-0.00001	0.2 - 0.7	0.3 - 0.5	0.01 – 0.6	0.1 - 0.2	0.001 - 0.1
Deposit feeders and meiobenthos	0.0002 - 0.005	0.005 - 0.07	NA	NA	0.01 - 0.6	0.01 - 0.5	0.002-0.3	0.001 - 0.15
Infaunal carnivores and benthic grazer	0.0001 - 0.03	0.001 - 0.03	NA	NA	0.5	0.5	0.002 - 0.3	0 - 0.1
Filter feeders	0.0005 - 0.02	0.01-0.08	NA	NA	0.3 - 0.9	0.3 - 0.9	0.1 - 0.45	0 - 0.1
Zooplankton	0.01-0.1	0.05-0.4	0.001-0.00001	0.2 - 0.7	0.25 - 0.5	0.25 - 0.6	0 - 0.2	0.001 - 0.1
Phytoplankton	NA	0.05-1.0	NA	NA	NA	NA	NA	NA
Macrophytes	NA	0.02-0.1	NA	NA	NA	NA	NA	NA

^ for an age structured group. * filter feeding vertebrates

Supplementary Table S5. Description of the most commonly used recruitment options in Atlantis. See Table 4 for Atlantis parameters meanings and descriptions. B = biomass

Recruitment flag (flagrecruit)	Description	Atlantis parameters	Equation to calculate number of recruits (R)
1 = Constant	A constant number of recruits enters the population at each recruitment episode	<i>KDENR</i>	$R = K DENR$
3 = Atlantis-BH	Similar to Beverton-Holt Spawning stock-recruitment relationship, but includes the condition of adults in the population such that less recruits are produced when the adults are underweight	<i>BHalpha</i> (α), <i>BHbeta</i> (β), <i>FSPB</i> , <i>KSPA</i> , <i>FSP</i> , <i>FSPB</i>	$R = \frac{\alpha L}{\beta + B}$ <p>Where</p> $L = \sum_{a=ageclass} FSPB_a \times S_a \times N_a$ <p>where</p> <p><i>FSPB_a</i> proportion at age that are mature, <i>S_a</i> spawn produced per individual in age class <i>a</i> and is a function of <i>KSPA</i>, <i>FSP</i>, <i>RN</i>, and <i>SN</i>, <i>N_a</i> number in age class <i>a</i></p>
10 = BH	Almost identical to Beverton-Holt SSR except that the entire biomass is used rather than only those mature	<i>BHalpha</i> (α), <i>BHbeta</i> (β)	$R = \frac{\alpha B}{\beta + B}$
12 = Linear	A constant number of recruits per adult (male and female are not separated and this needs to be taken into account if values are used from the literature based on births per adult female)	<i>KDENR</i> <i>FSPB</i>	$R = K DENR \times N_{mature}$ where <i>N_{mature}</i> is the number of mature individuals determined by the <i>FSPB</i>