Deliverable No. 4.4

Project acronym:
MareFrame

Project title:
“Co-creating Ecosystem-based Fisheries Management Solutions"

Grant agreement No: **613571**
Project co-funded by the European Commission within the Seventh Framework Programme

Start date of project: **1st January 2014**
Duration: **48 months**

Due date of deliverable:	31/12/2016
Submission date:	29/12/2016
File Name:	D4.4 MAREFRAME_Comparison of the performance of two EMs with known (simulated) data
Revision number:	2
Document status:	Final[1]
Dissemination Level:	PU[2]

Revision Control

Role	Name	Organisation	Date	File suffix[3]
Author	Erla Sturludóttir	UI		ES
Author	Paul Frater	UI		PNF
WP leader	Paul Fernandes	UNIABDN		PF
Coordinator	Anna K. Daníelsdóttir	MATIS	2912/2016	AKD
Administrative Coordinator	Oddur M. Gunnarsson & Kristinn Ólafsson	MATIS	28/12/2016	OMG, KÓ
Scientific Coordinator	Gunnar Stefánsson	UI	22/12/2016	GS

[1] Document will be a draft until it was approved by the coordinator
[2] PU: Public, PP: Restricted to other programme participants (including the Commission Services), RE: Restricted to a group specified by the consortium (including the Commission Services), CO: Confidential, only for members of the consortium (including the Commission Services)
[3] The initials of the revising individual in capital letters

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

Deliverable D4.4

Comparison of the performance of two EMs with known (simulated) data

December 2016

Contents

Executive summary

This report is the 4th deliverable in WP4 and studies the performance of two models commonly used in this WP, Ecopath with Ecosim (EwE) and Gadget. An Atlantis model was constructed for Icelandic waters and used as an operating model, i.e. simulated data from Atlantis was fed into EwE and Gadget. Thus, Atlantis was considered as the "true" ecosystem and EwE and Gadget studied to see if they were able to replicate the truth. Such a comparison is not possible e.g. using survey data as such data are not the true population but measurements of one. A methodology was developed to automatically extract data from Atlantis and import to the other models. Also, balancing and fitting routines were written for EwE to make the modelling process more automatic and less subjective. This study was considered data rich as the best possible knowledge of the ecosystem was known, e.g. the true biomass was known. The EwE model had 42 functional groups and time-series fitting was done for the 25 vertebrate groups. A single-species Gadget model was constructed for cod and haddock. The EwE model was able to simulate the magnitude and the trends for both biomass and catches for most of the vertebrate groups. It did however overestimate the biomass of the juvenile stanza groups. The Gadget model did replicate the trends in biomass very accurately but did overestimate the biomass for the both groups. A further study is needed to investigate why the Gadget model overestimated the biomass. Also, further study is needed to test the performance of the model when the information of the ecosystem is limited but that will be addressed in the 7th deliverable in this WP.

1 Introduction

Single species models have been used for decades in fisheries management to estimate stock size and to investigate effects of fishing pressure. However, they only focus on one species at a time and ignore all species interactions. Multispecies and ecosystem models include species interactions and some even included interactions between environmental variables and species. The more complex ecosystem models like Ecopath with Ecosim (EwE; Christensen & Walters, 2004) and Atlantis (Fulton et al., 2004) can improve understanding of the ecosystem and can be useful for strategic decision making while single species or simple multispecies models like Gadget (Begley and Howell, 2004) may be more appropriate for tactical decisions.

Single species models are not enough for the ecosystem approach to fisheries management (EAFM) and the addition of ecosystem models are necessary. They for example allow investigating the effects on a predator when its prey is being fished. The most complex models such as Atlantis can incorporate a socio-economic model which allow for exploring not only the ecosystem but also the effects the ecosystem has on the economy and vice versa.

Before an ecosystem model can be used for the EAFM it is vital to know whether it is reliable. This has usually been done by comparing the model to historic data which is often limited (Olson et al, 2016). Output from different models have also been compared to see if they give similar results (Smith , 2011; Forrest et al., 2015).

Another way to test how reliable models are is to feed simulated data into these models and test their performance. Producing simulated data for a whole ecosystem is not an easy task but this can be done using the Atlantis framework. The Atlantis model is a whole-of-an-ecosystem model. It is spatially explicit and considers the oceanography and follows the flow of nutrients through the food web. The vertebrate groups in the model have age structure and the model tracks number and size of each age group.

In this study the Atlantis model, constructed for Icelandic waters, was used to produce simulated data. These data were used to build an EwE model and a single species Gadget model. The performance of these two models were tested and compared. This study is considered data rich, i.e. the best possible information is available to construct the models.

2 Methods

An Atlantis model has been constructed for Icelandic waters, (see detailed description of the Atlantis model in Deliverable 4.6), simulating the entire marine ecosystem. The Atlantis model is used as an operating model and known simulated data from the model is used to construct an EwE and a Gadget model. The modelling process of these two models will be described in this section.

2.1 Ecopath with Ecosim

Ecopath with Ecosim (EwE) (Christensen & Walters, 2004) was chosen to test the performance of a commonly used ecosystem model. An independent R-based (R core team, 2014) version of EwE, Rpath (Aydin, 2016), was used for the modelling process. The EwE model is based on mass balancing equations. The production (*P*) of group *i* is described as follows:

$$P_i = Y_i + M2_i * B_i + E_i + BA_i + MO_i * B_i$$

Where the predation mortality (M2) is:

$$M2_i = \sum_{j=1}^{n} \frac{Q_j * DC_{ij}}{B_i}$$

and other mortality (MO) is:

$$M0_i = \frac{P_i(1 - EE_i)}{B_i},$$

where Q_j is the total consumption of predator *j* and *DC* is the proportion of group *i* in the diet of predator *j*, Y_i is the catch of group *i*, B_i is the biomass, E_i the net migration, BA_i is the bioaccumulation. EE_i is the ecotophic efficiency and indicates the proportion of the production that is explained by the model. In a balanced model *EE* is between 0 and 1.

The growth rate $\left(\frac{\partial B}{\partial t}\right)$ in Ecosim is defined as:

$$\frac{\partial B_i}{\partial t} = g_i \sum_{j}^{n} c_{ji} - \sum_{j}^{n} c_{ij} + E_i - (M0_i + F_i)B_i$$

where g_i is the growth efficiency of group *i*, and *F* is the harvest rate.

The c_{ij} is:

$$c_{ij} = Q_{ij} * \frac{V_{ij} * Y_j}{V_{ij} - 1 + Y_j} * \frac{D_{ij} * Y_i}{D_{ij} - 1 + Y_i} \quad (1)$$

where c_{ij} is the consumption of predator *j* on prey *i*, Q_{ij} is the total consumption of predator *j* on prey *i* from the Ecopath model. *V* is the vulnerability parameter and has a default value of 2, *D* represents the handling time and has a default value of 1000. *Y* is the proportional change in biomass of predator and prey since the start of the simulation.

The term *V* is the parameter that is usually tuned in the predator-prey interaction. At the default value the total consumption of the predator on its prey will increase by 1/3 when the predator biomass (Y_j) doubles in size (Figure 1). If the predator biomass decreases to half its size the consumption decreases by 1/3. If the V is greater the changes will be larger until they become linear when V = 100. If V approaches 1 there will be little change in consumption if there is a change in the predator biomass. If however, there is a change in prey biomass (Y_i) the consumption changes in same way regardless of the *V* parameter (Figure 2).

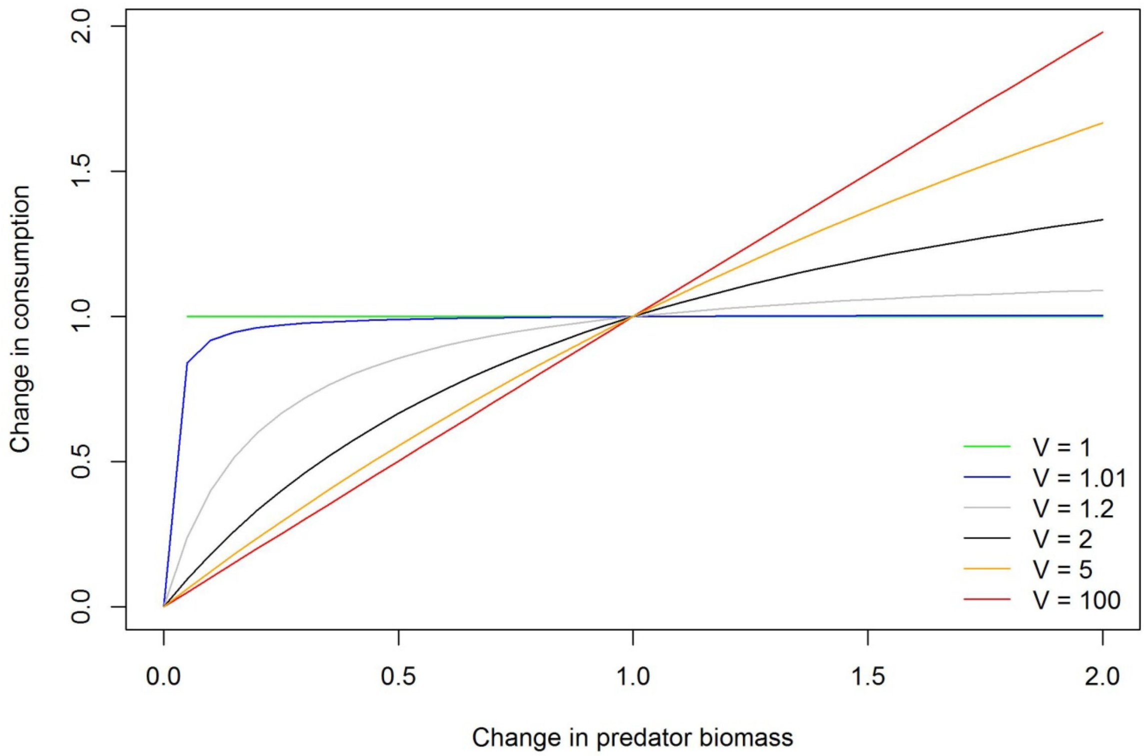

Figure 1. The predator-prey interaction described in Equation 1 for changing predator biomass by six different vulnerability parameters.

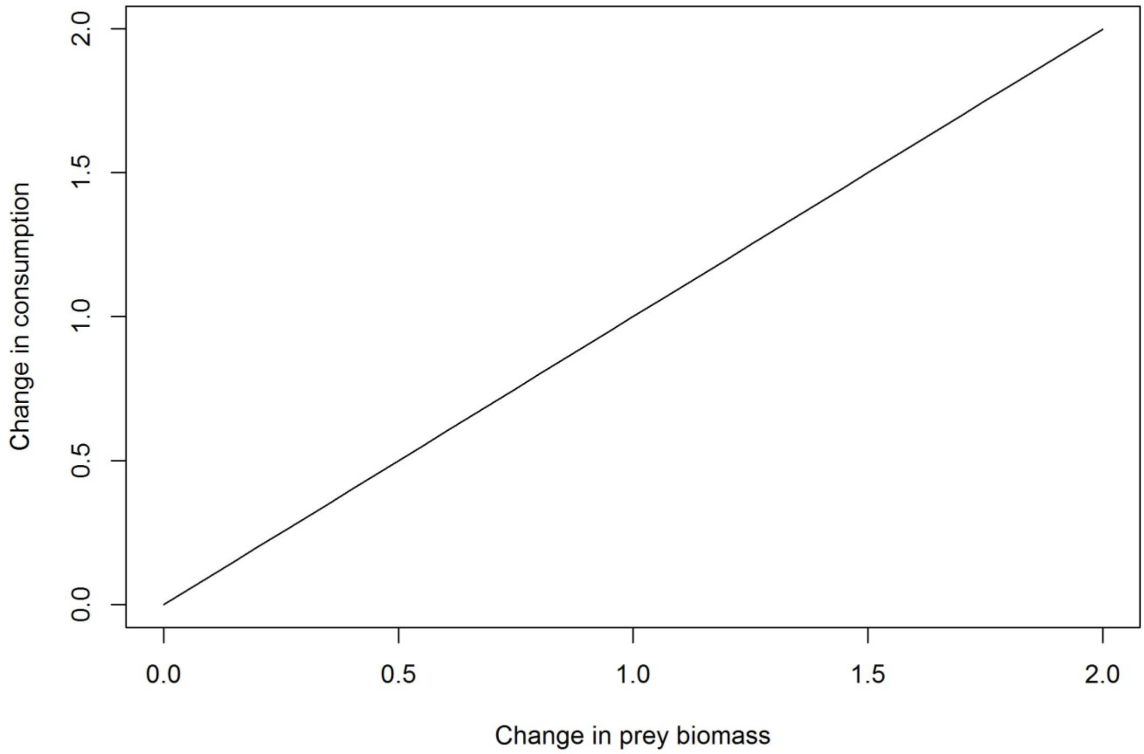

Figure 2. The predator-prey interaction described in Equation 1 for changing prey biomass.

2.1.1 Functional groups

The Atlatnic model of Icelandic waters was used as an operating model, i.e. simulated data from Atlantis was imported into Rpath. The Atlantis model includes 52 functional groups (Table 1) and the vertebrate groups have up to ten age classes. This was simplified before importing the data into Rpath. The functional groups pelagic bacteria, sediment bacteria, refractory detritus and labile detritus were considered as one detritus group in Rpath and dinoflagellates was combined with the diatom group. The benthic invertebrate groups were also reduced: scallop and quahog were merged with the filter feeders and cucumbers were combined with the deposit feeders, lobster and megazoobenthos were combined and, seagrass, macroalgea and microphytobenthos were included into one algea group. The Ecopath model had therefore 42 functional groups (Table 4). The age classes were also simplified, instead of having ten age classes for the vertebrates only four functional groups had age classes and were divided into juveniles (0-4 years old) and adults (4+ years old).

Table 1. Functional groups in the Atlantis model.

Vertebrates	Invertebrates and other groups
Cod (*Gadus morhua*)	Cephalopod
Haddock (*Melanogrammus aeglefinus*)	Shrimp
Saithe (*Pollachius virens*)	Microzooplankton
Redfish (*Sebastes* sp)	Mesozooplankton
Greenland Halibut (*Reinhardtius hippoglossoides*)	Macrozooplankton
Flatfish	Gelatinous zooplankton
Herring (*Clupea harengus*)	Norway Lobster
Capelin (*Mallotus villosus*)	Other Megazoobenthos
Migratory pelagic	Iceland Scallop
Other Codfish	Ocean Quahog
Other Demersal Commerical	Cucumbers
Other Demersal Fish	Deposit Feeder
Sandeel Fish	Other Benthic Filter Feeders
Long Lived Demersal	Benthic Grazer
Large Pelagic Fish	Benthic Carnivore
Small Pelagic Fish	Meiobenthos
Skates Rays	Diatom
Small sharks	Pico-phytoplankton
Large Sharks	Macroalgae
Seabird	Microphytobenthos
Pinniped	Seagrass
Minke Whale (*Balaenoptera acutorostrata*)	Dinoflagellates
Whale Baleen	Pelagic Bacteria
Whale Tooth	Sediment Bacteria
Whale Tooth Other	Labile detritus
	Refractory detritus
	Carrion

2.1.2 Biomass and landings

The tenth year of the Atlantis simulation was chosen as a starting point for the EwE model to avoid unbalance in age structure and diet composition in the beginning of the Atlantis simulation. Biomass and landings from the starting year were imported to Rpath. The biomass and landings of the vertebrates in Atlantis are separated for the age classes so it can easily be divided between the age groups of the multi stanza groups in the Ecopath model. The biomass of the migratory groups migratory pelagic, seabird, minke whale and baleen whales were scaled to represent their time inside the modelled area. The bioaccumulation parameter (*BA*) was set as the difference in biomass between the first and the second year.

The biomass was considered unknown for the invertebrates and the primary producers. In these cases, the biomass is estimated by the Ecopath model by setting EE to a certain value but this is a common practise in Ecopath models. The EE was set to 0.95 except for the primary producers the EE was set to 0.5 as recommended in Heymans et al. (2016).

2.1.3 Production

The production to biomass ratio (P/B) for the fish groups is assumed to be equal to total mortality (Z) in Ecopath (Haymans et al., 2016). When the Atlantis model was set up the age distributions of the vertebrate groups were acquired from natural mortality (M) based on maximum age (Hoenig, 1983). The M was calculated from the maximum age (t_{max}) as follows:

$$M = -\ln(0.01)/t_{max}$$

The same natural mortality was assumed for the Ecopath model. The fishing mortality (F) was calculated as C/B where C is the catch and B is biomass. If a group was a stanza group the juveniles were assumed to have 1.5 times higher natural mortality than the adults. The P/B then becomes Z = M + F.

For the invertebrate and primary producers, the P/B is taken from an Ecopath model for the Norwegian and the Barents Sea (Dommasnes et al., 2001).

2.1.4 Consumption

Approximate consumption (*Q*) can be calculated from Atlantis model for the vertebrate groups. How much age class *a* of group *g* needs to grow in a year was calculated. From that it is possible to calculate how much it needs to eat to maintain growth as the proportion of food that goes into growth (assimilation efficiency) is known in the Atlantis model.

$$Q_a = \frac{growth_a}{assimilation\ efficiency}$$

The total consumption (*Q*) of the group was then calculated based on the biomass in each age group as follows:

$$Q = \frac{\sum_{i=1}^{A} Q_a * B_a}{\sum_{i=1}^{A} B_a}$$

where Q_a is the consumption for age group *a*, B_a is the biomass of age group *a* and *A* is the number of age groups.

Consumption for the other groups was taken from Dommasnes et al. (2001).

2.1.5 Diet composition

The diet output from Atlantis is a snapshot of the diet composition at that time-step. The diet output was taken in the end of the starting year (year 10 in the Atlantis simulation). The diet data from Atlantis gives the proportion of each prey by the age class of the predator. The diet composition was weighted by the consumption of each age class to allow the age class that feeds the most to have the highest influence on the diet composition of the group.

The age of the prey is not given, which is usually the case with real diet data (stomach content). When a prey belongs to a multi-stanza group it needs to be allocated to either of the age groups or divided into both of them. How to do this is not trivial but size of the predators and prey can be used as a guideline. In Atlantis the prey needs to be less than 40% of the weight of the predator to be eaten. The weight of the predators and preys were used to allocate them into age groups. If the maximum weight of a prey was less than 40% of the weight of the predator the prey was divided into the stanza groups based on its biomass.

2.1.6 Parameters for multi-stanza groups

There are special parameters needed only for the multi-stanza groups. The age in months are needed to define the groups. The four multi-stanza groups where split when harvesting begins at age 48 months. The von Bertalanffy parameter (K) was estimated from the size of each age group in the Atlantis model. The model also needs the ratio of weight at maturity and weight at infinity (W_{mat}/W_{inf}). These were also obtained from the Atlantis model.

2.1.7 Balancing

The Ecopath model estimates the biomass for the groups that had biomass missing by setting the EE set to 0.95 or 0.50. It estimates the EE for groups that had all parameters set except the EE. If one or more groups have EE > 1 the model is considered unbalanced. EE > 1 means that a group is not producing enough to compensate for consumption on it and is therefore not in balance. Not all groups should be in balance, e.g. a group that is decreasing in biomass over time. This can however be corrected in the model by setting the bioaccumulation parameters to a value < 0. This has been done in this study. There are six options available to balance the model: 1) Increase the biomass of a group with EE > 1. 2) Decrease the biomass of a predator feeding on a group with EE > 1 3) Increase P/B of a group with EE > 1. 4) Decrease the Q/B of a predator feeding on a group with EE > 1. 5) Adjust the diet composition of a predator that is feeding on a group with EE > 1 to make it feed less on that group but more on other groups. 6) Decrease the BA value. Groups can also have EE < 0 when BA < 0. There are three options to balance such a group: a) Increase the BA value. b) Increase consumption of the group by increasing the Q/B of its predator or c) by adjusting the diet compositions to make it feed more on that group but less on other groups. It can be very difficult to say which of these five options should be carried out.

Kavanagh et al. (2004) suggested a balancing routine that changes the biomass and the diet composition iteratively. In the present study options 1, 2, 6 and, a were not considered because the biomass and BA from the Atlantis model were considered accurate. The other three options were studied to see what effect they had on the groups in the model. Increasing the P/B has only an effect on the group with that P/B. However, changing that parameter may not be the correct thing to do and may prevent a change in parameters that should be changed instead. It also changes the ratio between P/B and Q/B, the gross food conversion efficiency (P/Q = GE). Decreasing the Q/B of a predator has influence on all groups that the predator consumes and the predator itself. The last option of adjusting the diet composition causes less consumption of groups with EE > 1 and higher consumption for groups with EE < 1. The effect on other groups is however less than in option 4 and the consumption rate of the predator is not altered.

In this study an automated routine was written in R to balance the model and only options 3, 5 and, c were used as they have the least impact on the groups in the model. First, the diet of groups with EE < 0 where adjusted to make EE = 0. Next, the P/B of groups with E > 1 was increased, so that the groups would have EE < 1, but the increase was constrained to 25%. When the P/B is increased the GE will consequently also increase. GE is normally between 0.1 and 0.3 but can be found to be as high as 0.5 (Darwall, 2010). Therefore, if the change in P/B caused GE to become larger than 0.5 the Q/B was scaled to constrain GE to 0.5. If the model was still unbalanced the next step was to adjust the diet matrix. The proportion of a group with EE > 1 in the diet of its predators was lowered so the EE < 1 but the change was restricted to 25%. The proportions of other groups in the diet of the predators were scaled for groups that had EE <= 0.95 so that the diet proportions would sum to one. If these two last steps did not result in a balanced model, they were repeated until balance was achieved (Fig. 3).

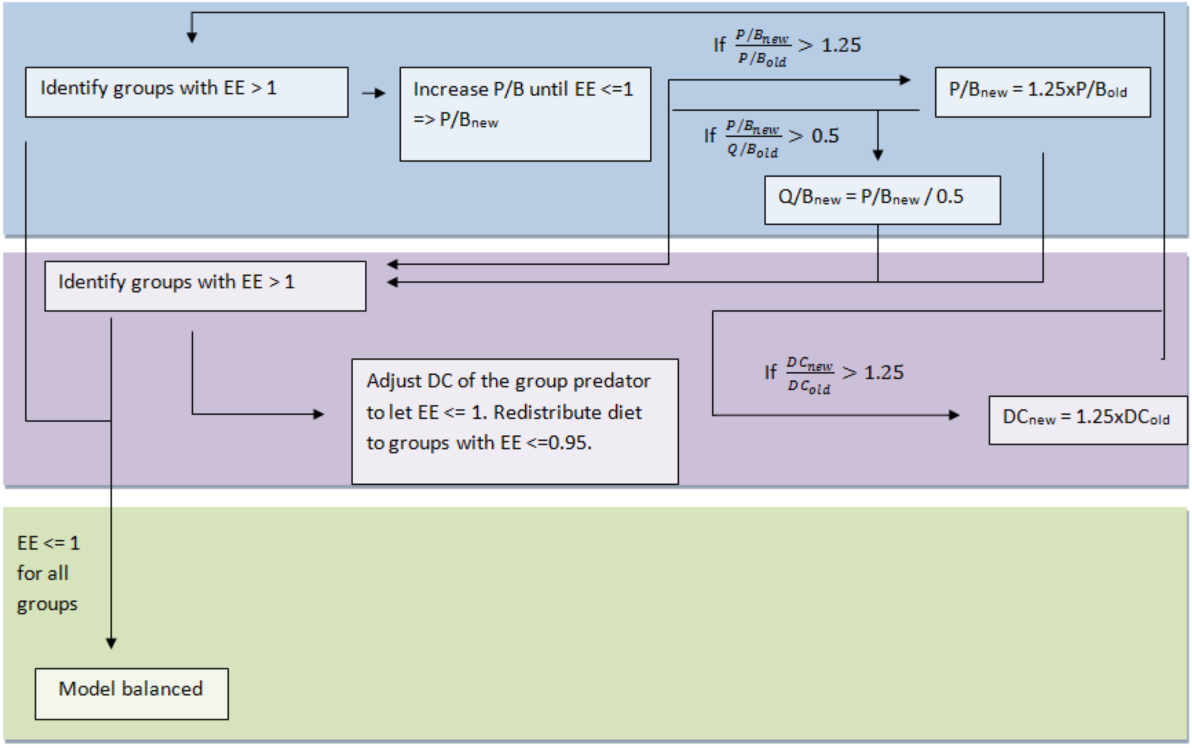

Figure 3. The balancing process of Ecopath.

2.1.8 Simulation in Ecosim

Annual harvesting rate was calculated from the catch and biomass data from Atlantis and imported into Rpath. The parameters from the balanced Ecopath model were used for simulation of biomass and catches for 56 years in Ecosim.

The vulnerability parameters (V) are equal to 2 in the Ecosim model but can be decreased or increased to represent bottom-up or top-down effects. The V in each predator-prey interactions was estimated by minimizing the sum of squares (SS) of the biomass and catch data from the vertebrate groups, resulting in 49 time-series. The predator-prey interactions in the model were 671 which results in 671 V. It was not possible to estimate all these parameters simultaneously and therefore they were estimate one at a time. The V was constrained to be between 1.01 and 100. The V of the interaction that gave the most improvement was set to its estimated value and another round of estimations was carried out. This was then repeated 29 times or until there was no improvement in AIC.

$$AIC_{SS} = 2k + n \ln SS$$

where k is the number of parameters, n is the number of observations and

$$SS = \sum_{i=1}^{2} \sum_{g=1}^{G} \sum_{t=1}^{T} \left(y_{gti} - \hat{y}_{gti}\right)^2$$

where $i = 1$ is biomass and $i = 2$ is catches, G is number of groups, T is the length of the simulation, y_{gt} are the observations from Atlantis for group g at time t and $\hat{y}_{gt}$ are the simulated values from Ecosim.

A second round of estimation was carried out where the V parameters were at its estimated value and the previous estimated parameters were estimated again.

2.2 Gadget

Gadget (Globally applicable Area Disaggregated Ecosystem Simulator) was also used as a test model to compare performance using Atlantis as a known operating system. Contrary to EwE, Gadget is often used as a more simplistic version of an ecosystem simulator. While EwE attempts to balance the trophic dynamics of a system, Gadget runs a simulation of each population in the model and uses statistical procedures to fit itself to actual survey and landings data. This fitting process can be time and computer intensive, and creating a Gadget model to compare the entire Atlantis simulated ecosystem would be impractical if not impossible. In addition, Gadget is often used by research institutes to perform single species stock assessments. Given this we compared Gadget to Atlantis using two species of economic importance in Icelandic waters, cod and haddock.

2.2.1 Gadget Background and Data Simulation

2.2.1.1 Background of the Gadget Model

Gadget simulates age-structured population(s) by starting with an initial number of individuals, which can be either specified or estimated by the model. Although Gadget is capable of setting up multi-area models, we parameterized all models as single area models. Therefore, explanations of the model and presented results should be assumed as coming from a single area. Within each year of the simulated model run a certain proportion of each age class is removed due to natural mortality, removed from fishing, and a new year class is added via either a renewal option or by using one of several available spawning functions (e.g. Ricker, Beverton-Holt). Additionally, the age of each year class is increased by one at the end of the last timestep of the year. Natural mortality is given by the equation, $N_{a+1,t+1} = N_{a,t}e^{-M}$, where $N_{a,t}$ is the number of individuals in each age class at each year and M is a mortality variable that can either be specified or estimated. The number of individuals removed due to fishing is calculated based on actual landed biomass data and using length, weight, and age data collected from commercial catches. This results in each age class declining across time. Renewal of the population happens either by a renewal option or a spawning function. The renewal option simply adds a certain number of recruits for each year of the model run. This annual renewal value can either be specified directly or designated as a variable to be estimated. Alternatively, a spawning function can be used which adds a certain number of recruits at each designated spawning event based on the spawning stock biomass of each population in the model at that time.

In addition to the simulation described above, Gadget also uses various likelihood components to fit the simulated population(s) to actual data. These likelihood components can range from basic biological data on species (e.g. age, length) to predatory data to survey indices and migration data. The most common likelihood parameters are age and length from standardized surveys and commercial catch surveys as well as basic survey indices (i.e. in number of fish caught per year from a standardized survey). Likelihood components are used to compare output simulated by Gadget to observed data using some likelihood function. The most common likelihood function used is a basic sum-of-squares, which adds the squared distance of model output data to observed data at each timestep for each likelihood component. Scores from these likelihood components are added across timesteps to produce a final likelihood score for the model. Gadget then uses this summed likelihood score in an optimizing run and attempts to find parameters that minimize this value.

2.2.1.2 Data Simulation

In order to produce the data used for likelihood components we took Atlantis output and simulated mock surveys similar to those performed in Iceland by the Marine Research Institute, which are done in March and October of every year. We also took the landings data created by Atlantis, and simulated a commercial catch sample based on the proportion of fish caught in each age, area, and year of the Atlantis model. Atlantis only computes age, structural nitrogen (structural N) and reserve nitrogen (reserve N) of organisms, and Gadget uses length distributions as one of its likelihood parameters. Therefore, we needed to calculate length based off the weight provided by Atlantis. To compute length from weight consistently we used the known structural N biomass of fish from Atlantis output and calculated weight based off of the length-weight relationship of both cod and haddock for each of those respective models. To calculate length of fish from Atlantis output we used structural N data from Atlantis output and computed length based on this weight. Structural N is less sensitive to seasonal changes in the Atlantis model (i.e. biomass alterations due to poor diet, lack of food, and spawning would be observed primarily in reserve N) and therefore provides a more consistent proxy for weight. In addition, biomass of each individual is initially calculated in Atlantis using the relationship $BM = 2.65 structualN + structuralN$, where BM is biomass of an individual fish. We used this same relationship to compute the length of a given fish from its structural N by reconfiguring the length-weight relationship formula $w = al^b$, where w is weight, l is length, and a and b are parameters specific to each species of fish. We derived the a and b parameters for the length-weight relationship from actual length and weight data for cod and haddock collected during the standardized spring and autumn bottom trawl surveys performed by the Icelandic Marine Research Institute. These parameters were as follows: cod – a=0.008249352, b=3.026918; haddock – a=0.00587903, b=3.116172. However, since we had the weight instead of length from Atlantis output, we rearranged this formula as such:

$$l = (w / a)^{1/b}$$

to compute length from weight, where l is the length of fish in cm, w is the biomass as calculated using the equation for BM above, and a and b are the same parameters in the length-weight relationship above. After computing lengths we then had a total count of fish for each specific age, weight, and length combination in every area, year, and month used in the Atlantis model.

As an effort to more accurately mimic the variability of lengths found in nature we distributed the number of fish in each age, weight and length combination based on the probability density function of a normal distribution using the respective length of each of the above combinations as the mean and standard deviation specific to age for both cod and haddock, which we calculated from actual Iceland Groundfish Survey data.

From this we mimicked a survey using following the selectivity curve:

$$Selection = 3\text{e-}05 / (1 + e^{-0.046(l - 49)})$$

Where l is the length and *Selection* is the proportion of fish "sampled" by the selectivity curve. The various constants were chosen to most accurately mimic selection of fish to the Icelandic Groundfish Surveys both in terms of size and number of fish caught in the surveys. One survey was performed in the third month of each year to mimic a spring survey, and one in the tenth month of each year to mimic an autumn survey (i.e. we pulled a sample of Atlantis data from each area and year within the specified month using the above curve).

To sample commercial catches for age and length we took the number of fish caught in each age and weight combination with calculated length and took 0.1444% of each combination. We sampled this particular percentage as it resulted in a similar number of fish to what was recorded in actual commercial catch samples by the Icelandic Marine Research Institute.

Samples were simulated in R using the package mfdbatlantis (Lentin, 2016) to read Atlantis output and compute surveys. Simulated survey data and actual Atlantis data were imported to a local MareFrame database using the R package mfdb (Lentin, 2014).

2.2.2 Setting up the Gadget Models

We set up single area Gadget models to attempt to simulate the cod and haddock populations produced by Atlantis. Each model was run on quarterly timesteps and included a single stock that renewed annually in the first timestep of each year. The number of years included in the model was exactly the same as those in Atlantis (1948-2013). Both models were set up to have nearly every variable estimated; therefore, initial values for each age, annual recruitment values, natural mortality, growth function parameters, and catch and survey selectivity curve parameters were all estimated in each model. The only parameters not estimated were those for the length-weight relationship and the variation in length-at-age of each respective species (which were calculated based on the simulated survey data from Atlantis output-see *2.2.1.2*)

The Gadget models were run using an iterative re-weighting procedure designed to reduce the variance of each single likelihood component in succession and using the estimated parameters to produce a final model that best explains all components (Stefansson, 2003). As part of this procedure each estimated parameter is allowed to vary and a simulation is run Model outputs

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

from the simulation are compared to sampled data for each likelihood component and a sums-of-squares is produced. This is performed iteratively within an optimization algorithm until a minimum sums-of-squares is reached. For initial parameter values, upper and lower bounds, and final parameter values for each model see Table 2 and Table 3.

Table 2. Parameters and their initial values, upper and lower bounds, and final values for cod Gadget model.

Parameter	Function	Initial Value	Lower Bound	Upper Bound	Final Value
Linf	Growth	170	150	250	172.39
Kappa	Growth	0.2	0.01	0.3	0.1
Beta-binomial	Growth Implementation	6	1.00E-08	100	0.46
M	Natural Mortality	0.2	0.15	0.4	0.22
Area multiplier	Initial Values	5	1.00E-05	10	6.79
Area initial	Initial Values	10	1.00E-05	100	29.43
Age 1 initial	Initial Values	20	1.00E-05	100	73
Age 2 initial	Initial Values	20	1.00E-05	100	100
Age 3 initial	Initial Values	20	1.00E-05	100	98.4
Age 4 initial	Initial Values	20	1.00E-05	100	72.03
Age 5 initial	Initial Values	20	1.00E-05	100	55.44
Age 6 initial	Initial Values	20	1.00E-05	100	56.18
Age 7 initial	Initial Values	20	1.00E-05	100	17.06
Age 8 initial	Initial Values	20	1.00E-05	100	34.13
Age 9 initial	Initial Values	20	1.00E-05	100	5.05
Age 10 initial	Initial Values	20	1.00E-05	100	17.01
Age 11 initial	Initial Values	20	1.00E-05	100	8.5
Age 12 initial	Initial Values	20	1.00E-05	100	0.92
Age 13 initial	Initial Values	20	1.00E-05	100	<0.01
Age 14 initial	Initial Values	20	1.00E-05	100	<0.01
Age 15 initial	Initial Values	20	1.00E-05	100	<0.01
Age 16 initial	Initial Values	20	1.00E-05	100	<0.01
Age 17 initial	Initial Values	20	1.00E-05	100	7.96
Age 18 initial	Initial Values	20	1.00E-05	100	24.86
Age 19 initial	Initial Values	20	1.00E-05	100	30.06
Recruitment Length	Recruitment	20	5	40	12.99
Area multiplier	Recruitment	100	1.00E-05	500	50.14
Recruitment per year	Recruitment	1	0.001	1000	varies
Recruit length standard dev.	Recruitment	2	0.1	10	4.47
Commercial alpha	Commercial Fleet Selectivity	0.5	0.01	3	0.1
Commercial l50	Commercial Fleet Selectivity	65	5	200	55.23
Spring Survey alpha	Spring Survey Selectivity	0.05	0.001	3	0.05
Spring Survey l50	Spring Survey Selectivity	50	5	55	55
Autumn Survey alpha	Autumn Survey Selectivity	0.05	0.001	3	0.02
Autumn Survey l50	Autumn Survey Selectivity	50	5	55	55

Table 3. Parameters and their intial values, upper and lower bounds, and final values for haddock Gadget model.

Parameter	Function	Initial Value	Lower Bound	Upper Bound	Final Value
Linf	Growth	100	75	150	100.9
Kappa	Growth	0.2	0.01	0.3	0.22
Beta-binomial	Growth Implementation	6	1.00E-08	100	0.74
M	Natural Mortality	0.2	0.15	0.4	0.38
Area multiplier	Initial Values	5	1.00E-05	10	1.83
Area initial	Initial Values	10	1.00E-05	100	75.41
Age 1 initial	Initial Values	20	1.00E-05	100	40.57
Age 2 initial	Initial Values	20	1.00E-05	100	91.7
Age 3 initial	Initial Values	20	1.00E-05	100	48.36
Age 4 initial	Initial Values	20	1.00E-05	100	78.89
Age 5 initial	Initial Values	20	1.00E-05	100	21.29
Age 6 initial	Initial Values	20	1.00E-05	100	26.92
Age 7 initial	Initial Values	20	1.00E-05	100	<0.01
Age 8 initial	Initial Values	20	1.00E-05	100	1
Age 9 initial	Initial Values	20	1.00E-05	100	0.32
Age 10 initial	Initial Values	20	1.00E-05	100	0.25
Age 11 initial	Initial Values	20	1.00E-05	100	<0.01
Age 12 initial	Initial Values	20	1.00E-05	100	0.1
Age 13 initial	Initial Values	20	1.00E-05	100	0.01
Age 14 initial	Initial Values	20	1.00E-05	100	0.05
Age 15 initial	Initial Values	20	1.00E-05	100	0.03
Age 16 initial	Initial Values	20	1.00E-05	100	0.05
Age 17 initial	Initial Values	20	1.00E-05	100	<0.01
Age 18 initial	Initial Values	20	1.00E-05	100	<0.01
Age 19 initial	Initial Values	20	1.00E-05	100	<0.01
Recruitment Length	Recruitment	20	1	40	1
Area multiplier	Recruitment	100	1.00E-05	500	97.38
Recruitment per year	Recruitment	1	0.001	1000	115.44
Recruit length standard dev.	Recruitment	2	0.1	10	0.16
Commercial alpha	Commercial Fleet Selectivity	0.05	0.01	3	0.14
Commercial l50	Commercial Fleet Selectivity	65	5	200	42.29
Spring Survey alpha	Spring Survey Selectivity	0.05	0.001	3	0.04
Spring Survey l50	Spring Survey Selectivity	40	5	50	50
Autumn Survey alpha	Autumn Survey Selectivity	0.05	0.001	3	0.05
Autumn Survey l50	Autumn Survey Selectivity	40	5	50	37.83

2.3 Comparisons

The simulated biomass and catches from EwE and estimated biomass from Gadget were compared to the true biomass and catches from the Atlantis model. They were compared using Pearson's correlation, model reliability (RI) and model efficiency (MEF) as suggested by Stow et al. (2009). The Pearson's correlation (Eq. 1) shows if the model can catch the trends of the true time-series but it does not show if the model is able to simulate the correct magnitude. It is therefore useful to look at more than one metrics when evaluating model skill. RI is a metric that measures how far the model is from the truth on average. MEF measures if the model is better or worse than the average of the observations. These three metrics are as follows:

$$r = \frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2 \sum_{i=1}^{n}(P_i - \bar{P})^2}} \qquad (1)$$

$$RI = exp \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{O_i}{P_i}\right)^2} \qquad (2)$$

$$MEF = \frac{\sum_{i=1}^{n}(O_i - \bar{O})^2 - \sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \qquad (3)$$

where O_i is observation i of n, $\bar{O}$ is the average of the observations, P_i is prediction i of n and $\bar{P}$ is the average of the predictions. Ideally, all these metric should be close to one. If RI is equal to one then the model and the observations are in the same level on average but if RI is e.g. 3 it means that the model differs by a factor of 3 from the observations on average. The model is as good as the average of predicting the truth if MEF is equal to zero. If it is higher than zero it is better than the straight line but if it is lower than zero it is actually worse than a straight line. However, a model can have a negative MEF but still get correlation close to one. To study which model, Gadget or EwE, performed better, i.e. which were able to simulate the true biomass more accurately, these three metrics were compared.

3 Results

3.1 Balancing Ecopath

The Ecopath model was parameterized and 11 groups had EE outside of the interval 0 to 1 (Table 4). The Seabird group had EE < 0 but redfish, Greenland halibut, both stanza-groups of herring, demersal commercial, small and large pelagic, small and large sharks and pinniped had EE ranging from 1.29 to 3.34. In step 1 of the balancing process the diet is adjusted the group with EE < 0 to make EE > 0. After step 3 there were only four groups left with EE > 1 and maximum EE had gone from 3.34 to 2.06. After step 5 there was one group left, the large shark, and it needs two additional steps to make EE < 1 (Figure 4).

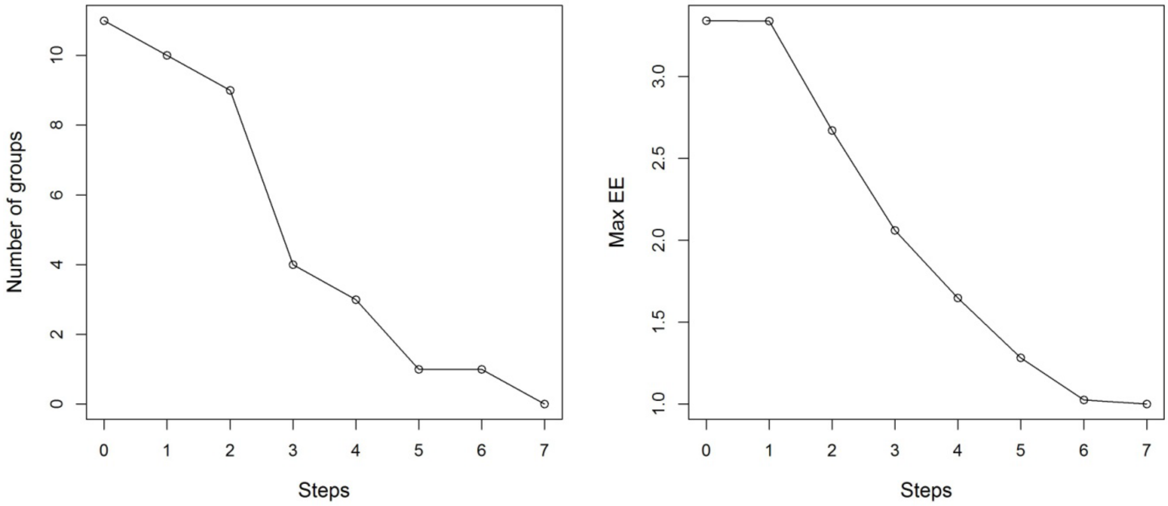

Figure 4. Number of groups with EE < 0 or EE > 1 in each step and the maximum EE in each step.

The changes the balancing process had on predation mortality and P/B and Q/B can be seen in Table 5. The predation mortality was reduced by 3-45% for the groups that had EE > 1. This consequently led to an increase in predation mortality for other groups but that increase was only 0-12% except the seabird group needed 60% increase in predation mortality. Six group needed an increase of P/B by 25% but the P/B for large shark had to be almost doubled, going from P/B = 0.05 to P/B = 0.09.The Q/B parameter was only increased for the herring to maintain GE ≤ 0.5. That increase also resulted in an increase of Q/B for the juvenile herring as that is based on the Q/B for the adults. The balanced model can be seen in Table 6.

Table 4. The Ecopath model before balancing. EE < 0 and EE > 1 are shown in red.

Group	type	TL	Biomass	PB	QB	EE	GE	Removals
Cod 0-4	0	4.07	366493	0.37	4.44	0.59	0.08	6537
Cod 4+	0	4.28	1523327	0.49	2.01	0.61	0.25	448843
Haddock 0-4	0	3.38	107812	0.58	3.33	0.51	0.18	8672
Haddock 4+	0	3.35	80137	0.92	1.93	0.74	0.48	50415
Saithejuv 0-4	0	4.17	82634	0.38	3.11	0.74	0.12	1361
Saithe 4+	0	4.26	371033	0.37	1.49	0.78	0.25	53280
Redfish	0	3.91	1836558	0.09	0.97	2.15	0.10	0
Greenland Halibut	0	4.27	571364	0.16	1.89	1.47	0.09	6105
Flatfish	0	2.88	225305	0.30	1.74	0.22	0.17	19981
Herring 0-4	0	3.69	417900	0.51	1.30	1.42	0.40	4749
Herring 4+	0	3.69	471015	0.39	0.87	1.53	0.45	29929
Capelin	0	3.50	5899716	1.17	3.03	0.61	0.39	121793
Migratory pelagic	0	3.53	1253964	0.51	1.71	0.56	0.30	0
Other Codfish	0	3.89	115588	0.47	1.88	0.94	0.25	19333
Demersal Commerical	0	3.72	255543	0.31	1.90	1.40	0.16	19572
Other Demersal Fish	0	3.46	534144	0.58	1.79	0.32	0.32	0
Sandeel Fish	0	3.47	1273289	0.58	3.22	0.55	0.18	0
Long Lived Demersal	0	4.42	115273	0.15	1.31	0.85	0.12	0
Large Pelagic Fish	0	3.95	87526	0.15	1.33	1.54	0.12	0
Small Pelagic Fish	0	3.61	106630	0.51	2.39	2.05	0.21	0
Small Sharks	0	4.50	117525	0.09	1.06	1.29	0.08	0
Skates	0	4.06	61269	0.15	1.12	0.53	0.14	0
Large Sharks	0	4.60	111533	0.05	0.95	3.34	0.05	0
Seabird	0	4.30	29786	0.11	1.38	-0.03	0.08	0
Pinniped	0	4.67	1835	0.13	1.48	2.61	0.09	0
Minke Whale	0	4.09	69106	0.10	1.58	0.11	0.06	0
Whale Baleen	0	3.64	389033	0.08	0.82	0.29	0.10	15025
Whale Tooth	0	4.82	408143	0.06	1.85	0.17	0.03	1414
Whale Tooth Other	0	4.69	11323	0.16	0.45	0.15	0.35	0
Cephalopod	0	3.67	159023	2.44	12.00	0.95	0.20	0
Shrimp	0	2.03	669659	1.25	5.00	0.95	0.25	502
Mesozooplankton	0	2.46	12608290	4.00	15.00	0.95	0.27	0
Microzooplankton	0	2.00	9389383	10.00	25.00	0.95	0.40	0
Macrozooplankton	0	2.75	3437459	2.50	15.00	0.95	0.17	0
Gelatinous Zoo	0	3.45	18154	2.50	15.00	0.95	0.17	0
Megazoobenthos	0	3.32	1990241	1.50	9.75	0.95	0.15	312
Deposit Feeders	0	2.00	17626020	1.50	9.75	0.95	0.15	0
Benthic Filter Feeders	0	2.04	5966701	1.50	9.75	0.95	0.15	0
Benthic Grazers	0	2.00	778138	1.50	9.75	0.95	0.15	0
Benthic Carnivore	0	3.06	5158090	1.50	9.75	0.95	0.15	0
Meiobenthos	0	2.00	18304910	1.50	9.75	0.95	0.15	0
Diatom	1	1.00	5921743	117.73	0.00	0.50	0.00	0
Pico-phytoplankton	1	1.00	188725	117.73	0.00	0.50	0.00	0
Macroalgae	1	1.00	5636284	0.65	0.00	0.50	0.00	0
Labile Detritus	2	1.00	1186180000	0.49	0.00	0.73	0.00	0
Carrion	2	1.00	311	2.60	0.00	0.08	0.00	0
Fleet	3	5.00	0	0.00	0.00	0.00	0.00	0

Table 5. Change in predation mortality, PB and QB of the groups in the model due to the balancing process.

Group	Predation mortality change (%)	PB change (%)	QB change (%)
Cod 0-4	8	0	0
Cod 4+	12	0	0
Haddock 0-4	4	0	0
Haddock 4+	6	0	0
Saithejuv 0-4	8	0	0
Saithe 4+	12	0	0
Redfish	-33	56	0
Greenland Halibut	-22	25	0
Flatfish	2	0	0
Herring 0-4	3	25	16
Herring 4+	-19	25	13
Capelin	2	0	0
Migratory pelagic	2	0	0
Other Codfish	3	0	0
Demersal Commerical	-16	25	0
Other Demersal Fish	3	0	0
Sandeel Fish	2	0	0
Long Lived Demersal	11	0	0
Large Pelagic Fish	-19	25	0
Small Pelagic Fish	-25	53	0
Small Sharks	-3	25	0
Skates	10	0	0
Large Sharks	-45	95	0
Seabird	60	0	0
Pinniped	-42	56	0
Minke Whale	0	0	0
Whale Baleen	10	0	0
Whale Tooth		0	0
Whale Tooth Other	10	0	0
Cephalopod	7	0	0
Shrimp	1	0	0
Mesozooplankton	4	0	0
Microzooplankton	4	0	0
Macrozooplankton	6	0	0
Gelatinous Zoo	1	0	0
Megazoobenthos	2	0	0
Deposit Feeders	2	0	0
Benthic Filter Feeders	2	0	0
Benthic Grazers	1	0	0
Benthic Carnivore	2	0	0
Meiobenthos	2	0	0
Diatom	4	0	
Pico-phytoplankton	4	0	
Macroalgae	1	0	
Labile Detritus	2	4	
Carrion	3	0	

Table 6. The balanced Ecopath model. Parameters estimated by the model are shown in blue.

Group	type	TL	Biomass	PB	QB	EE	GE	Removals
Cod 0-4	0	4.07	366493	0.37	4.44	0.63	0.08	6537
Cod 4+	0	4.27	1523327	0.49	2.01	0.63	0.25	448843
Haddock 0-4	0	3.38	107812	0.58	3.33	0.53	0.18	8672
Haddock 4+	0	3.35	80137	0.92	1.93	0.74	0.48	50415
Saithejuv 0-4	0	4.17	82634	0.38	3.11	0.79	0.12	1361
Saithe 4+	0	4.25	371033	0.37	1.49	0.82	0.25	53280
Redfish	0	3.90	1836558	0.14	0.97	1.00	0.15	0
Greenland Halibut	0	4.27	571364	0.21	1.89	1.00	0.11	6105
Flatfish	0	2.88	225305	0.30	1.74	0.22	0.17	19981
Herring 0-4	0	3.69	649451	0.64	1.50	0.75	0.43	4749
Herring 4+	0	3.69	471015	0.49	0.98	1.00	0.50	29929
Capelin	0	3.50	5899716	1.17	3.03	0.62	0.39	121793
Migratory pelagic	0	3.53	1253964	0.51	1.71	0.57	0.30	0
Other Codfish	0	3.82	115588	0.47	1.88	0.96	0.25	19333
Demersal Commerical	0	3.71	255543	0.38	1.90	1.00	0.20	19572
Other Demersal Fish	0	3.46	534144	0.58	1.79	0.33	0.32	0
Sandeel Fish	0	3.47	1273289	0.58	3.22	0.55	0.18	0
Long Lived Demersal	0	4.39	115273	0.15	1.31	0.96	0.12	0
Large Pelagic Fish	0	3.94	87526	0.19	1.33	1.00	0.15	0
Small Pelagic Fish	0	3.61	106630	0.79	2.39	1.00	0.33	0
Small Sharks	0	4.44	117525	0.11	1.06	1.00	0.11	0
Skates	0	4.05	61269	0.15	1.12	0.59	0.14	0
Large Sharks	0	4.56	111533	0.09	0.95	1.00	0.09	0
Seabird	0	4.30	29786	0.11	1.38	0.01	0.08	0
Pinniped	0	4.65	1835	0.20	1.48	1.00	0.13	0
Minke Whale	0	4.08	69106	0.10	1.58	0.11	0.06	0
Whale Baleen	0	3.64	389033	0.08	0.82	0.29	0.10	15025
Whale Tooth	0	4.78	408143	0.06	1.85	0.17	0.03	1414
Whale Tooth Other	0	4.68	11323	0.16	0.45	0.16	0.35	0
Cephalopod	0	3.67	170101	2.44	12.00	0.95	0.20	0
Shrimp	0	2.03	678088	1.25	5.00	0.95	0.25	502
Mesozooplankton	0	2.46	13066970	4.00	15.00	0.95	0.27	0
Microzooplankton	0	2.00	9728470	10.00	25.00	0.95	0.40	0
Macrozooplankton	0	2.75	3645231	2.50	15.00	0.95	0.17	0
Gelatinous Zoo	0	3.45	18425	2.50	15.00	0.95	0.17	0
Megazoobenthos	0	3.32	2032894	1.50	9.75	0.95	0.15	312
Deposit Feeders	0	2.00	17981720	1.50	9.75	0.95	0.15	0
Benthic Filter Feeders	0	2.04	6098121	1.50	9.75	0.95	0.15	0
Benthic Grazers	0	2.00	788204	1.50	9.75	0.95	0.15	0
Benthic Carnivore	0	3.06	5262121	1.50	9.75	0.95	0.15	0
Meiobenthos	0	2.00	18674470	1.50	9.75	0.95	0.15	0
Diatom	1	1.00	6146199	117.73	0.00	0.50	0.00	0
Pico-phytoplankton	1	1.00	195560	117.73	0.00	0.50	0.00	0
Macroalgae	1	1.00	5709199	0.65	0.00	0.50	0.00	0
Labile Detritus	2	1.00	1186180000	0.50	0.00	0.72	0.00	0
Carrion	2	1.00	311	2.60	0.00	0.09	0.00	0
Fleet	3	4.99	0	0.00	0.00	0.00	0.00	0

3.2 Ecosim

3.2.1 Simulation without time-series fitting

The Ecosim model was run for 57 years based on the parameters from the Ecopath model. The model simulates biomass and catches and the comparison of the Ecosim results are compared to the Atlantis output for the vertebrate groups in Figure 5 - Figure 7.

The metrics for comparisons were calculated for both the biomass and catch and can be seen in Table 7. The simulated biomass from Ecosim and the biomass from Atlantis had positive correlation for 24 out of 29 vertebrate groups, 14 of them had correlation above 0.75 and nine of those groups had correlation higher than 0.9. There were however five groups which had a negative correlation but these were all groups of top predators. Of those 29 vertebrate groups 20 were harvested and all of them had positive correlation with catches from the Atlantis model.

The RI metric measures how far on average the simulated biomass and catches are from the Atlantis output. The biomass was in 11 out of 29 cases within 50% of the true biomass but in five cases the difference between the simulated and true biomass was greater than 100%. For the catches, six out of 20 groups harvested had the difference between the simulated catches and the Atlantis catches less the 50% but seven groups had a difference of more than 100% (Table 7).

The MEF metric measures if the simulated results are better or worse than the overall average, i.e. a straight line. If MEF < 0 the model is worse than a straight line. This was the case for 19 groups for the biomass and nine for the catches (Table 7). Only four groups had MEF > 0.5 for the biomass and seven for the catches. MEF came out worst for the juvenile groups, groups with no harvest and the mammal and seabird groups. Groups can have high correlation and RI close to 1 but still have MEF < 0. This was e.g. the case with other demersal, sandeel and large pelagic fish.

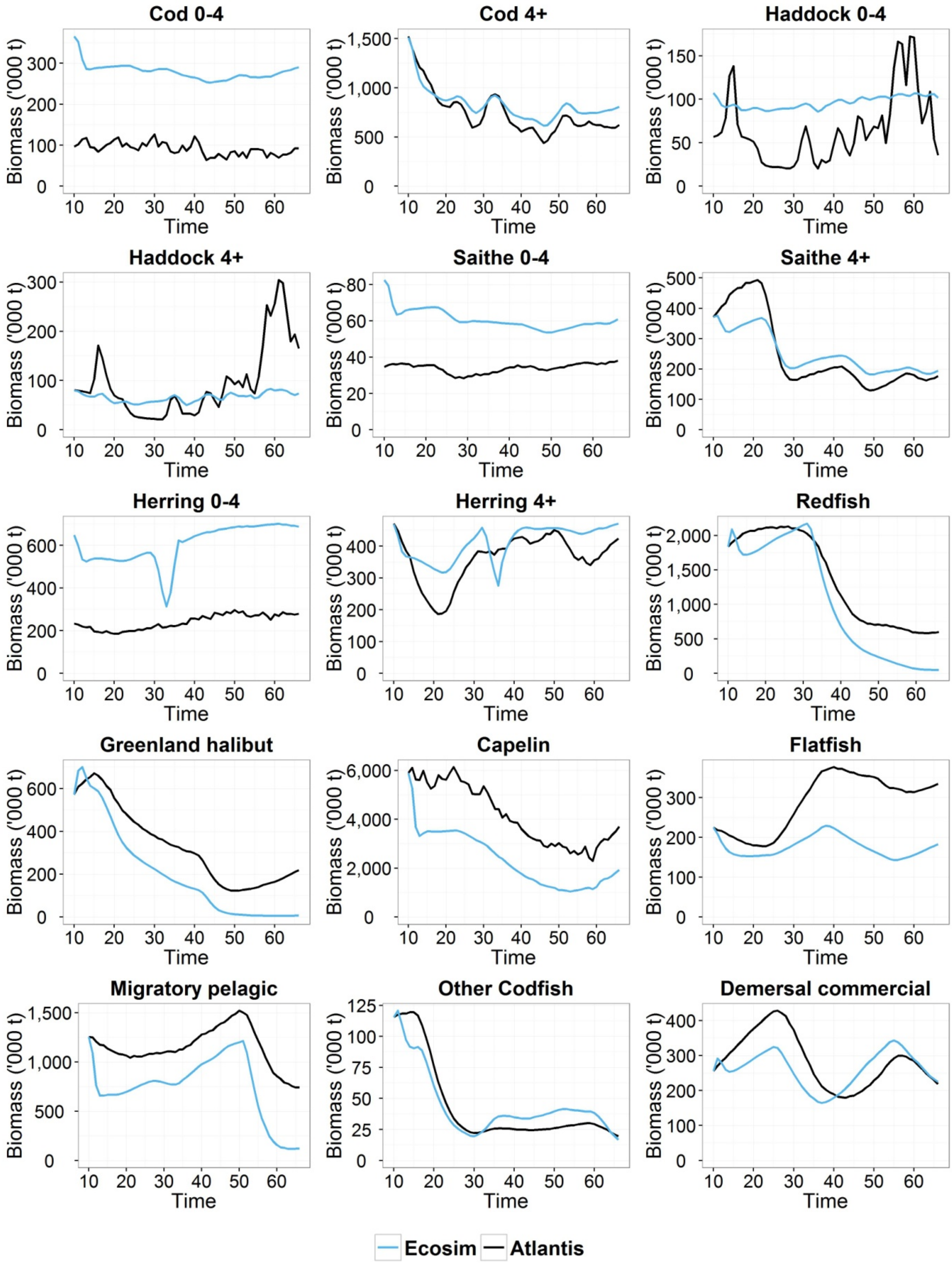

Figure 5. Simulated biomass of vertebrates from Ecosim before fitting to time-series, compared to the Atlantis output.

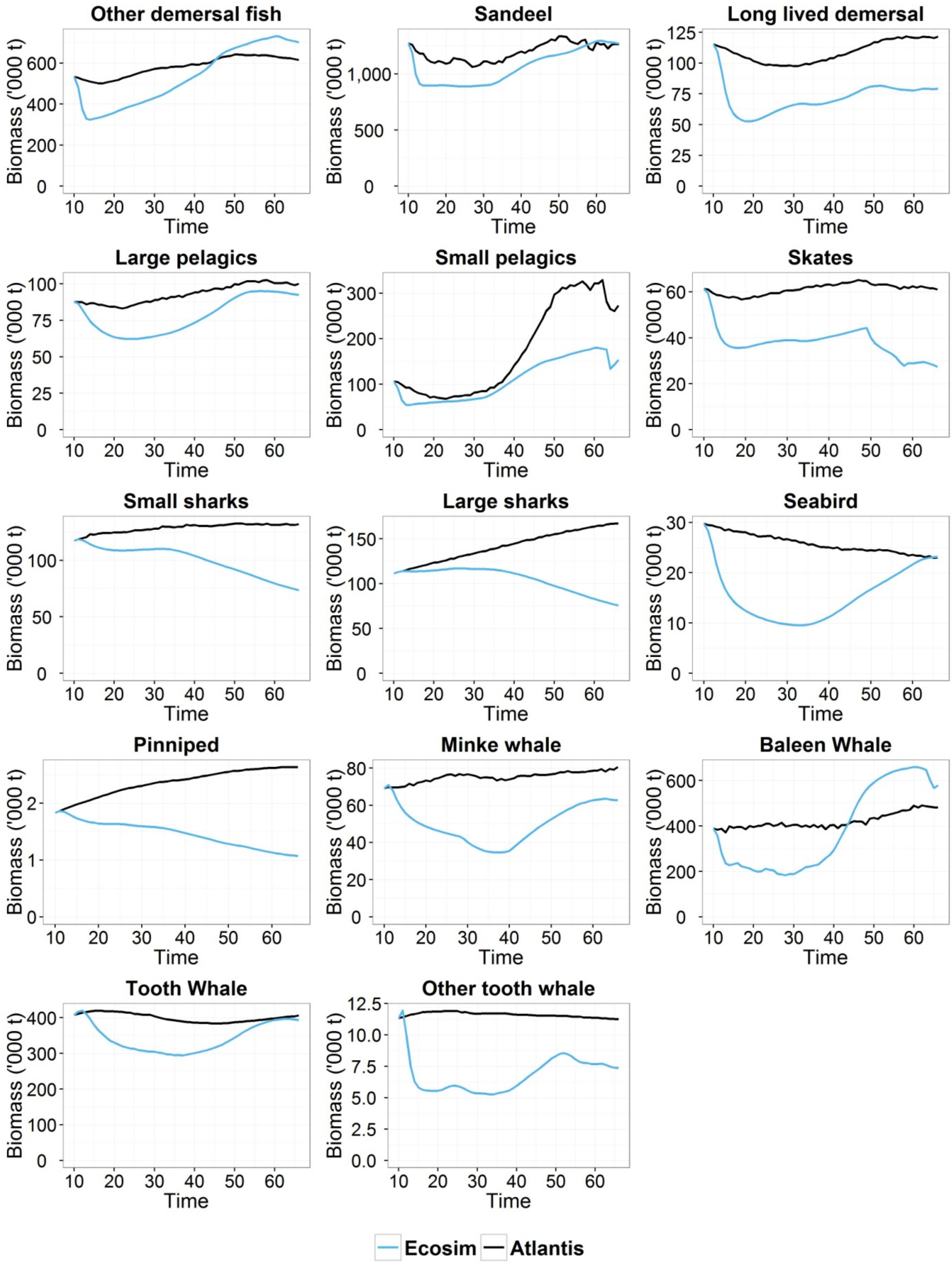

Figure 6. Simulated biomass of vertebrates from Ecosim before fitting to time-series, compared to the Atlantis output.

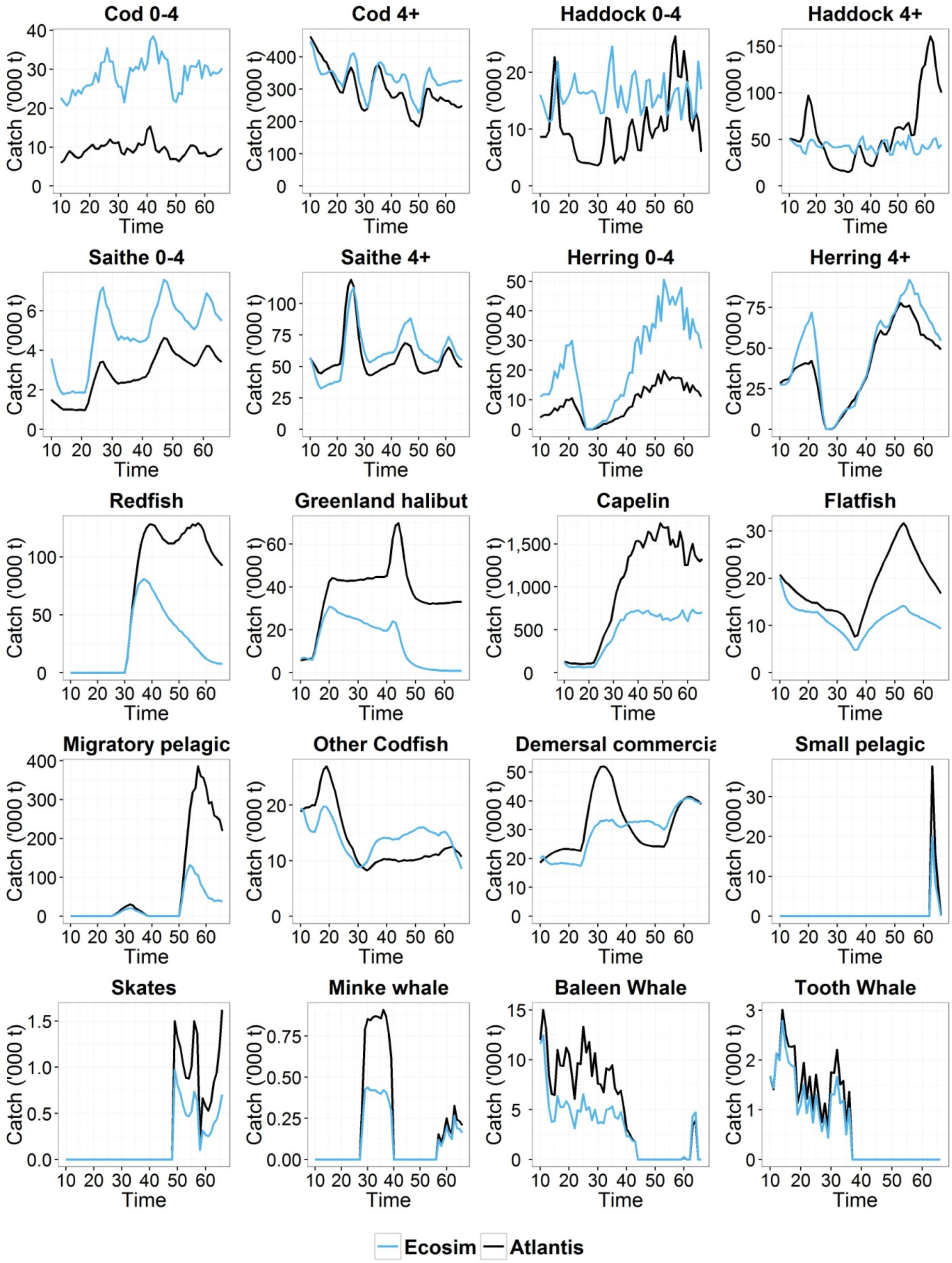

Figure 7. Simulated catches of vertebrates from Ecosim before fitting to time-series, compared to the Atlantis output.

Table 7. EwE skill assessment: Pearson's correlation (r), model reliability (RI) and model efficiency for simulated biomass and catches from EwE before time-series fitting.

Group	Biomass			Catches		
	r	RI	MEF	r	RI	MEF
Cod 0-4	0.45	3.14	-155.6	0.63	3.13	-117.2
Cod 4+	0.97	1.18	0.75	0.83	1.19	0.38
Haddock 0-4	0.65	2.23	-0.4	0.01	2.23	-1.28
Haddock 4+	0.79	1.89	0.02	0.14	1.89	-0.13
Saithejuv 0-4	0.13	1.8	-118.75	0.96	1.80	-2.97
Saithe 4+	0.96	1.24	0.74	0.79	1.24	0.51
Redfish	0.99	3.06	0.66	0.70	4.22	-0.15
Greenland Halibut	0.97	6.54	0.37	0.48	6.71	-2.58
Flatfish	0.45	1.69	-2.37	0.60	1.69	-1.01
Herring 0-4	0.79	2.52	-113.43	0.99	2.52	-7.03
Herring 4+	0.74	1.25	0.12	0.96	1.25	0.78
Capelin	0.95	1.96	-1.29	0.97	1.97	-0.08
Migratory pelagic	0.91	2.21	-3.94	0.86	2.97	0.22
Other Codfish	0.95	1.30	0.87	0.67	1.33	0.41
Demersal Commerical	0.56	1.26	0.19	0.63	1.28	0.31
Other Demersal Fish	0.95	1.26	-4.44			
Sandeel Fish	0.88	1.17	-2.9			
Long Lived Demersal	0.68	1.56	-18.79			
Large Pelagic Fish	0.88	1.24	-5.94			
Small Pelagic Fish	0.98	1.52	0.29	1.00	1.82	0.77
Small Sharks	-0.76	1.36	-69.01			
Skates	0.08	1.68	-114.41	0.99	2.01	0.67
Large Sharks	-0.86	1.48	-6.74			
Seabird	-0.13	1.93	-36.9			
Pinniped	-0.94	1.74	-17.56			
Minke Whale	0.05	1.61	-93.66	0.99	1.73	0.65
Whale Baleen	0.85	1.61	-22.92	0.93	1.73	0.59
Whale Tooth	0.30	1.21	-31.61	0.99	1.27	0.94
Whale Tooth Other	-0.72	1.8	-737.05			

3.2.2 Simulation with time-series fitting

The biomass and catches from Atlantis and Ecosim were used to calculate SS. Without any fitting the AIC was 91212 but was reduced by setting $V = 1.06$ for the Diatoms (Table 8). This resulted in higher productivity for the diatoms when their biomass decreased. This higher productivity provides more food for the whole food web. The estimation usually resulted in lowering the V parameter which means that the groups have higher consumption at lower biomass level (see Figure 1). The best model of these 22 was the one that included estimates for all the interaction listed in Table 8.

The result of the time-series fitting can be seen in Figure 8 for the simulated biomass of harvested vertebrate groups and in Figure 9 for other vertebrate groups and the simulated catches in Figure 10.

The vertebrate groups had positive correlation for biomass for 24 of the 29 groups, 15 of these groups had higher correlation than 0.75 and nine groups had correlation above 0.9 (

Table 9). 15 of the harvested groups had correlation higher than 0.75 but one group had negative correlation. Overall there was improvement in correlation after the time-series fitting but it was the groups with the worst correlation that improved the most, e.g. the correlation for pinnipeds went from -0.94 to 0.87. For some groups the correlation got worse, e.g. for the minke whale and herring groups.

Table 8. Estimates of vulnerability (V) for predator-prey interactions, number of V parameters estimated and the AIC. Each model contains estimates of V for the predator-prey interactions in the rows above.

model	predator	prey	V	#Param	AIC
1				0	91212
2	Diatom	Outside	1.06	1	88926
3	Greenland halibut	Cephalopod	1.01	2	88995
4	Redfish	Macrozooplankton	1.01	3	88767
5	Migratory pelagic	Mesozooplankton	1.01	4	88823
6	Other codfish	Benthic Filter Feeders	100.00	5	88829
7	Capelin	Macrozooplankton	1.20	6	87518
8	Greenland halibut	Macrozooplankton	1.01	7	87530
9	Macrozooplankton	Mesozooplankton	100.00	8	85642
10	Whale baleen	Macrozooplankton	1.01	9	85554
11	Benthic Filter Feeders	Labile Detritus	100.00	10	85546
12	Greenland halibut	Cod 0-4	1.01	11	85511
13	Minke whale	Macrozooplankton	1.01	12	85498
14	Whale tooth	Pinniped	100.00	13	85500
15	Macrozooplankton	Diatom	5.83	14	84607
16	Megazoobenthos	Benthic Filter Feeders	1.26	15	84634
17	Redfish	Small pelagic	100.00	16	84629
18	Whale tooth other	Cephalopod	1.01	17	84631
19	Whale baleen	Mesozooplankton	1.01	18	84619
20	Cephalopod	Macrozooplankton	1.01	19	84605
21	Seabird	Sandeel fish	1.01	20	84577
22	Whale tooth	Cod 0-4	1.01	21	84553

There was an overall improvement in RI after time-series fitting. The average RI for biomass went from 1.9 to 1.4 and improved the most for Redfish and Greenland halibut. Groups with RI less than 1.5 for biomass increased from 11 before fitting to 23 after fitting. However, there were few groups where RI increased, e.g. the juvenile cod and saithe groups and also small pelagic (Table 7 and Table 9).

Even though the average value of MEF increased, groups with MEF < 0 for biomass stayed the same after time-series fitting but went from nine to six for the catches. The mammal and other non-harvested top predator improved the most in MEF but still had MEF < 0. The juvenile cod and saithe groups that already had large negative value got even worse after time-series fitting.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

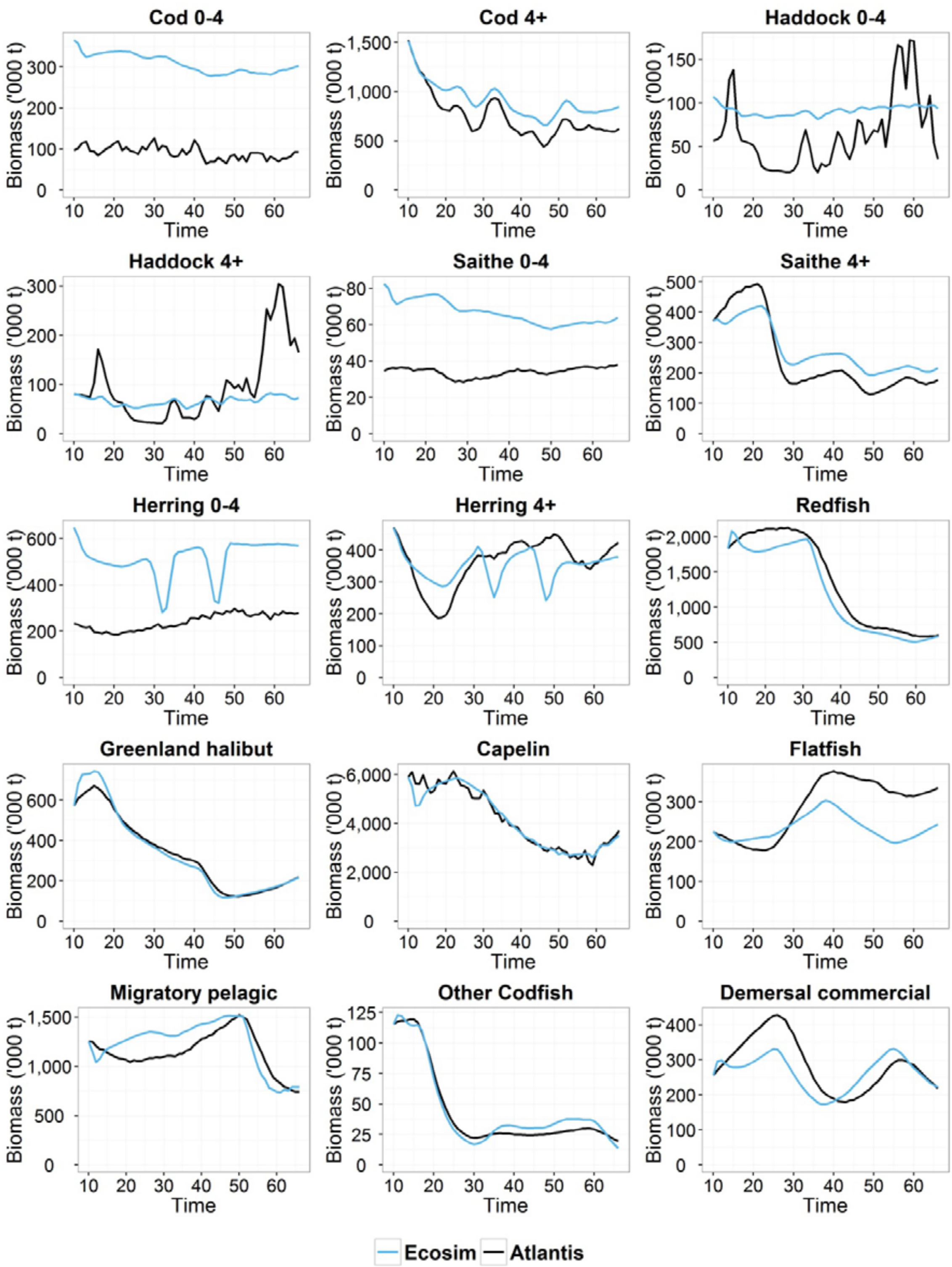

Figure 8. Simulated biomass of vertebrates from Ecosim after fitting to time-series, compared to the Atlantis output.

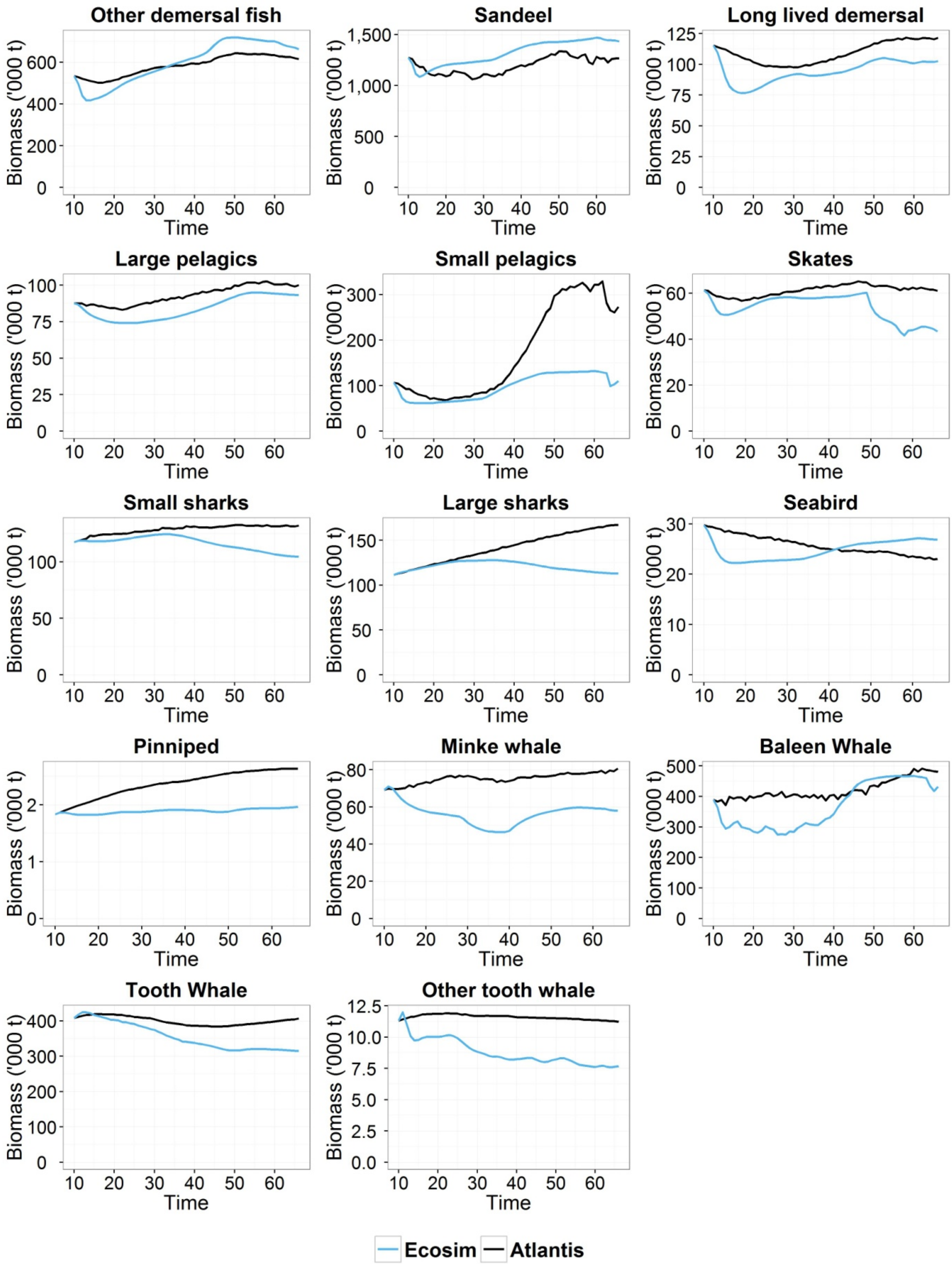

Figure 9. Simulated biomass of vertebrates from Ecosim after fitting to time-series, compared to the Atlantis output.

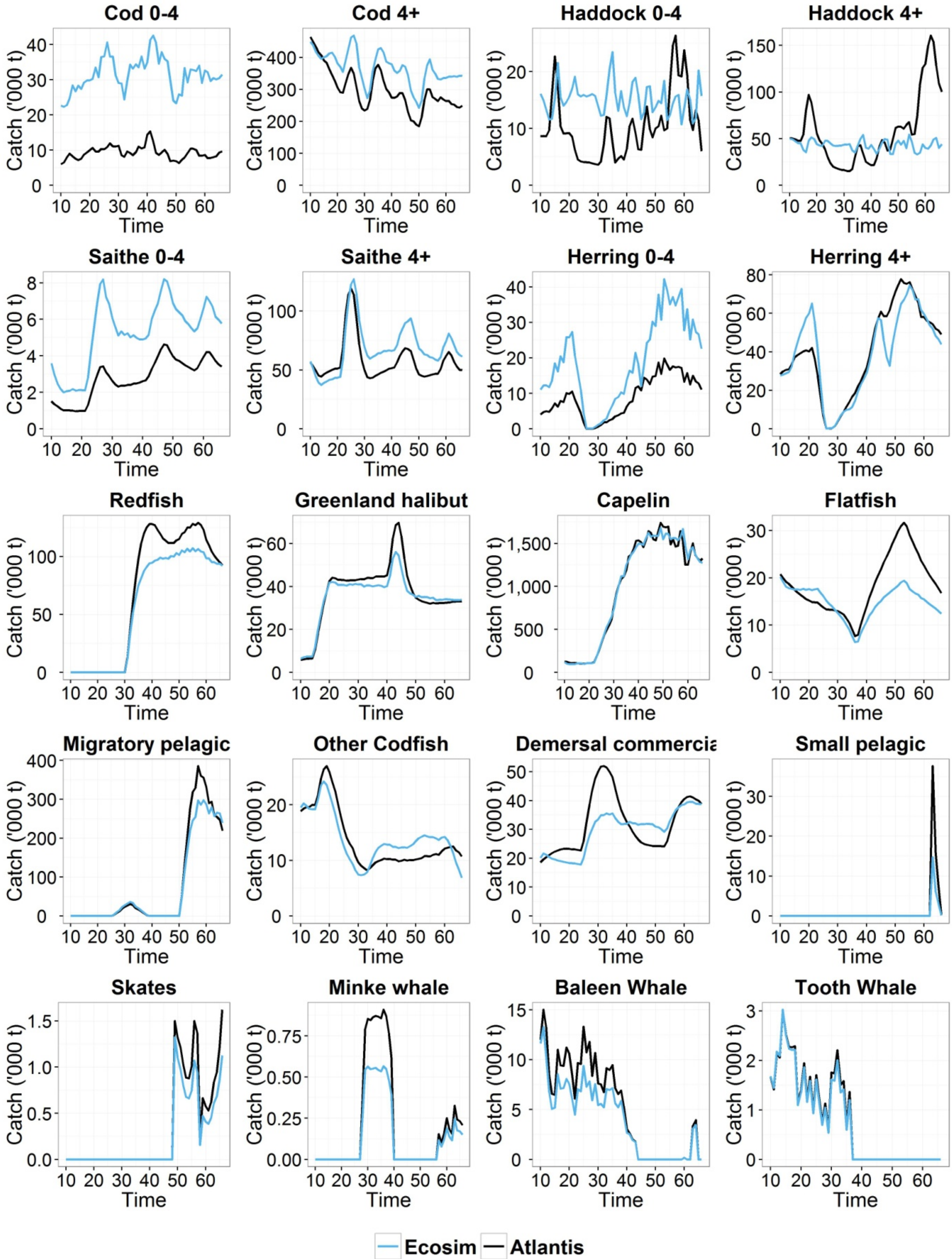

Figure 10. Simulated Catch of vertebrates from Ecosim after fitting to time-series compared to the Atlantis output.

Table 9. EwE skill assessment: Pearson's correlation (r), model reliability (RI) and model efficiency for simulated biomass and catches from EwE after time-series fitting.

	Biomass			Catch		
Group	r	RI	MEF	r	RI	MEF
Cod 0-4	0.64	3.45	-206.9	0.70	3.45	-155.9
Cod 4+	0.98	1.27	0.4	0.85	1.27	-0.4
Haddock 0-4	0.63	2.16	-0.27	-0.03	2.16	-1.02
Haddock 4+	0.76	1.90	0.02	0.07	1.90	-0.14
Saithejuv 0-4	-0.08	1.98	-177.26	0.94	1.97	-4.9
Saithe 4+	0.98	1.29	0.78	0.83	1.28	0.22
Redfish	0.99	1.16	0.92	0.99	1.20	0.92
Greenland Halibut	0.99	1.08	0.97	0.97	1.11	0.89
Flatfish	0.60	1.33	-0.29	0.69	1.34	0.12
Herring 0-4	0.33	2.18	-66.63	0.97	2.18	-3.18
Herring 4+	0.47	1.24	0.18	0.91	1.24	0.82
Capelin	0.97	1.06	0.95	1.00	1.06	0.99
Migratory pelagic	0.79	1.15	0.29	0.99	1.19	0.96
Other Codfish	0.99	1.21	0.97	0.87	1.23	0.76
Demersal Commerical	0.67	1.22	0.36	0.72	1.24	0.45
Other Demersal Fish	0.99	1.10	-0.36			
Sandeel Fish	0.76	1.12	-2.33			
Long Lived Demersal	0.73	1.18	-2.58			
Large Pelagic Fish	0.92	1.12	-1.26			
Small Pelagic Fish	0.92	1.74	-0.14	1.00	2.51	0.61
Small Sharks	-0.46	1.13	-13.44			
Skates	0.09	1.2	-17.11	1.00	1.36	0.91
Large Sharks	-0.29	1.23	-1.8			
Seabird	-0.51	1.15	-2.37			
Pinniped	0.87	1.26	-3.78			
Minke Whale	-0.27	1.37	-51.26	1.00	1.46	0.82
Whale Baleen	0.78	1.24	-4.04	0.99	1.30	0.87
Whale Tooth	0.84	1.16	-18.8	1.00	1.07	1.00
Whale Tooth Other	0.54	1.35	-256.16			

3.3 Gadget

Gadget reliably modelled the trends for both cod biomass (Figure 11) and haddock biomass (Figure 12). This is reflected in the high Pearson's correlation for both species (cod: $r = 0.92$; haddock: $r = 0.97$). In terms of absolute numbers Gadget modelled the number of fish in the Atlantis model marginally well for cod (Figure 13, $r = 0.78$) and quite well for haddock (Figure 14; $r = 0.99$). Although Gadget reliably represented the patterns modelled by Atlantis it consistently overestimated biomass for both species of fish. This is reflected in the other two model reliability metrics (RI and MEF, see

Table 10).

The RI for cod biomass was 1.62 and for haddock biomass was 1.45, which means that, on average, Gadget overestimated cod biomass by 62% and haddock biomass by 45%. Visually, Gadget seemed to overestimate biomass of both cod and haddock by a greater factor when the observed biomass and number of fish were greater. Cod biomass was greatest in the Atlantis model near the beginning of the model run. This is also when Gadget overestimated biomass by the greatest amount, nearly three times as much. However, as the Atlantis model progressed and cod biomass declined the Gadget estimate for biomass came closer to that of Atlantis. When RI is calculated for cod biomass without using data from the first 20 years of the model the RI drops to 1.45. For haddock, the magnitude of biomass in Gadget approached that of Atlantis in the model years from about 1970 to 1980, when haddock biomass was lowest. RI of haddock biomass from using data from only the model years 1969-1979 was 1.16, or a 16% overestimation of biomass. However, in the model years of the mid-2000s when haddock biomass was high in the Atlantis model, then the Gadget model overestimated biomass of fish by a factor approaching two.

MEF for both cod and haddock biomass was <0 meaning that, overall, the average of Atlantis output values for biomass of each respective species would be better predictors than the output from the Gadget models for each respective species. MEF for cod biomass was -5.06 and for haddock biomass was -0.24 indicating that Gadget was better at predicting biomass for haddock than for cod, but still not a better predictor than simply the average of the haddock biomass from the Atlantis model.

Gadget was less consistent when looking at the absolute numbers of fish. Even though Gadget overestimated the biomass of both cod and haddock it underestimated the absolute number of cod and overestimated the absolute number of haddock. RI for the absolute number of cod was 1.31, but absolute number of fish in the Gadget model were lower than that found in Atlantis. Therefore, Gadget estimated, on average, the inverse of 1.31, which is about 76% of the number of cod that were found in the Atlantis model. The RI for total number of haddock was 1.59, which means that Gadget estimated 159% of the number of haddock in the Atlantis model.

The MEF for absolute number of cod indicated that Gadget more closely estimated cod numbers than cod biomass. MEF for absolute number of cod simulated by Gadget was -0.10. While this is still less than 0 it is closer than biomass to being as good of a predictor as the average of the Atlantis values. MEF for haddock numbers was -1.78, which was lower than the MEF for haddock biomass. Regardless, all MEF metrics for biomass and numbers for both cod and haddock were <0 indicating that the average Atlantis values would have been a better predictor of the Atlantis model according to this metric.

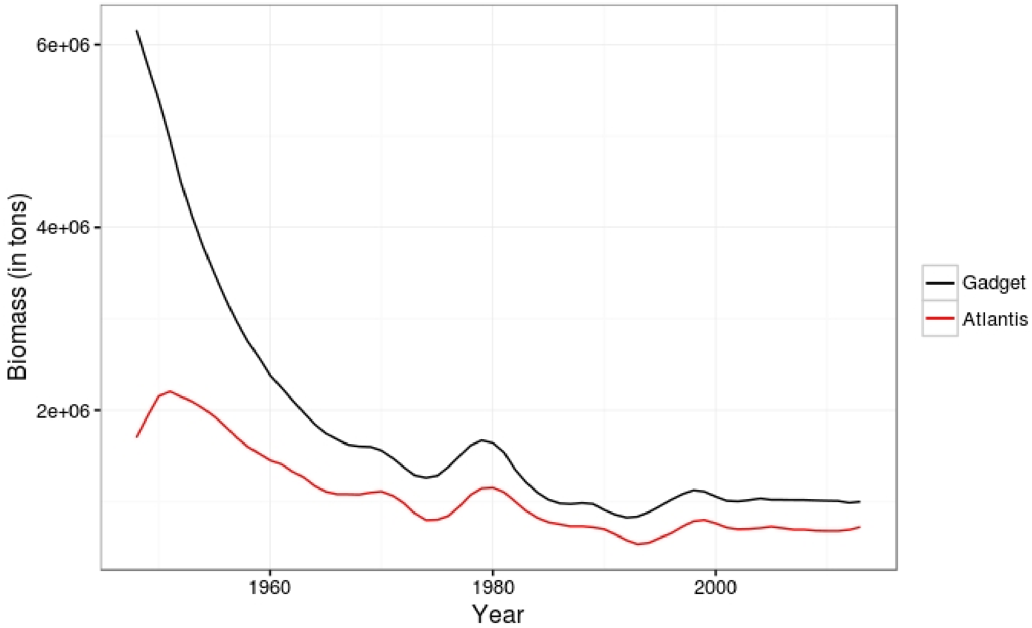

Figure 11. Annual cod biomass in tons for both Atlantis (red) and Gadget (black) across all years of the models.

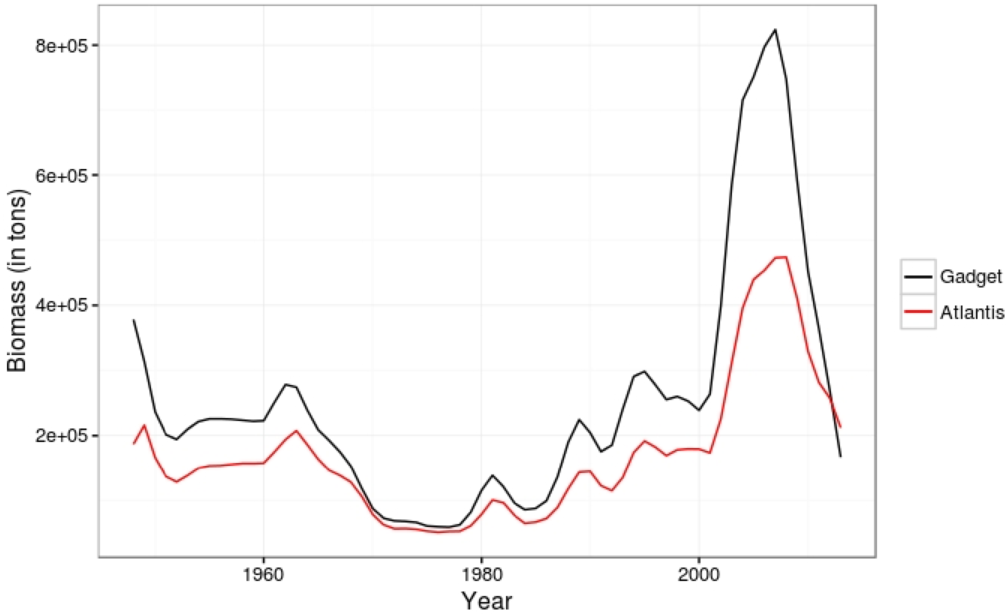

Figure 12. Annual haddock biomass in tons for both Atlantis (red) and Gadget (black) across all years of the models.

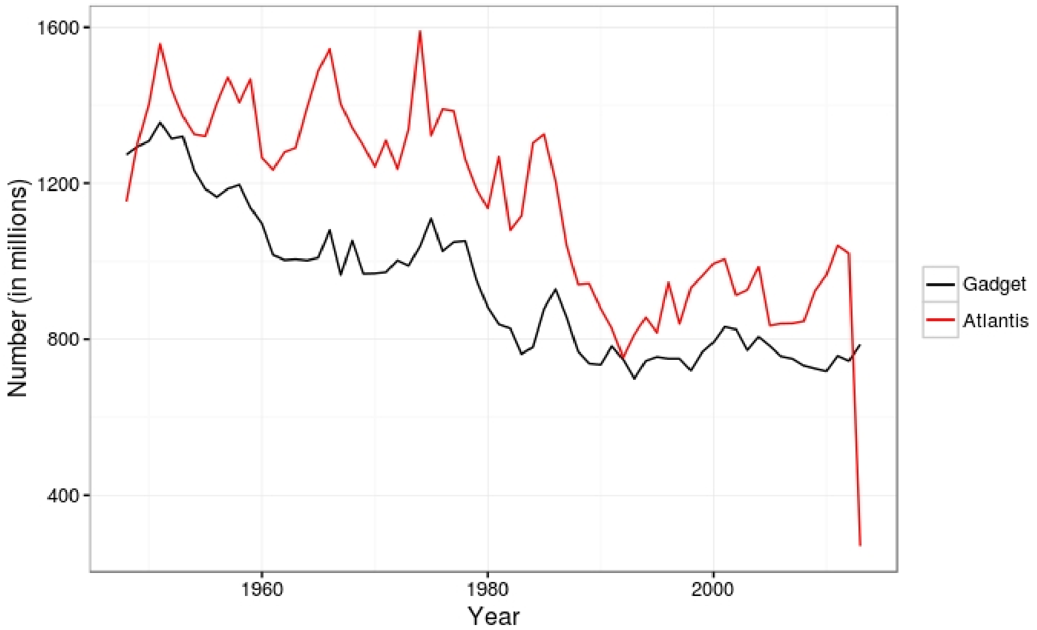

Figure 13. Annual number of cod in millions of fish for both Atlantis (red) and Gadget (black) across all years of the models.

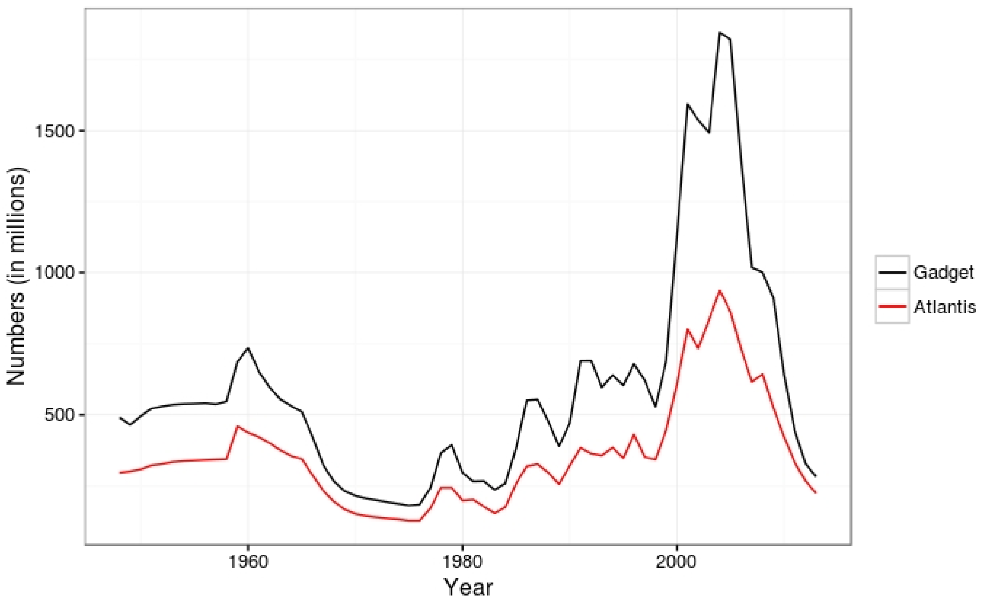

Figure 14. Annual number of haddock in millions of fish for both Atlantis (red) and Gadget (black) across all years of the models.

Table 10. Gadget skill assessment: Pearson's correlation (r), model reliability (RI), and model efficiency (MEF) for simulated biomass and number of fish from Gadget.

Species	Biomass			Numbers		
	r	RI	MEF	r	RI	MEF
Cod	0.92	1.62	-5.07	0.78	1.31	-0.10
Haddock	0.97	1.45	-0.24	0.99	1.59	-1.78

3.4 Comparison between Gadget and EwE

The simulated biomass of cod and haddock (4+ groups in EwE) from the Gadget and the EwE model were compared. The Gadget model is better at simulating the changes in biomass. The haddock has recruitment spikes that control the fluctuations in biomass and the Gadget model was able to estimate that whereas the EwE model was not. For a group with no recruitment spikes such as cod, the EwE does a better job at simulating biomass before the time series fitting. The Gadget model had higher correlation for haddock but the EwE had higher correlation for cod before time-series fitting but much lower after time series fitting when the correlation had decreased to 0.7. The RI was better for the haddock in the Gadget model but worse for the cod. The Gadget model overestimated the biomass for both groups which resulted in high RI and MEF < 0. Before time-series fitting the cod had very high MEF but after time series fitting it was similar as for the Gadget model.

4 Discussion

4.1 The modelling process

The best option when estimating parameters in a model is to do that simultaneously. This was however not possible in this study as there were 671 parameters that could be estimated. It can be seen in Figure 15 that the major improvements occur in steps. The first major improvement takes place with the V parameter estimated in the "consumption" for diatoms. The next one is when the V is estimated in the capelin-macrozooplankton interaction but this interaction does not cause improvement until the V in the codfish- filter feeder interaction has been changed making the codfish feed less on the filter feeders. It is not always easy to trace the effects through the food web but it seems that less feeding on filter feeders causes more food for macrozooplankton. The capelin can then feed more on the macrozooplankton without having too much negative effect on the fit of the other groups. The method used for time-series fitting in this study can therefore miss out of improvements because it is not done simultaneously. The fit of the model would probably improve a lot if the parameters from the Ecopath model, e.g. the Q/B parameters were included into the model fitting. The Q/B was calculated from Atlantis and is below 1 for some groups, e.g. the other tooth whale and the fit of this group would improve if that parameter would be increased.

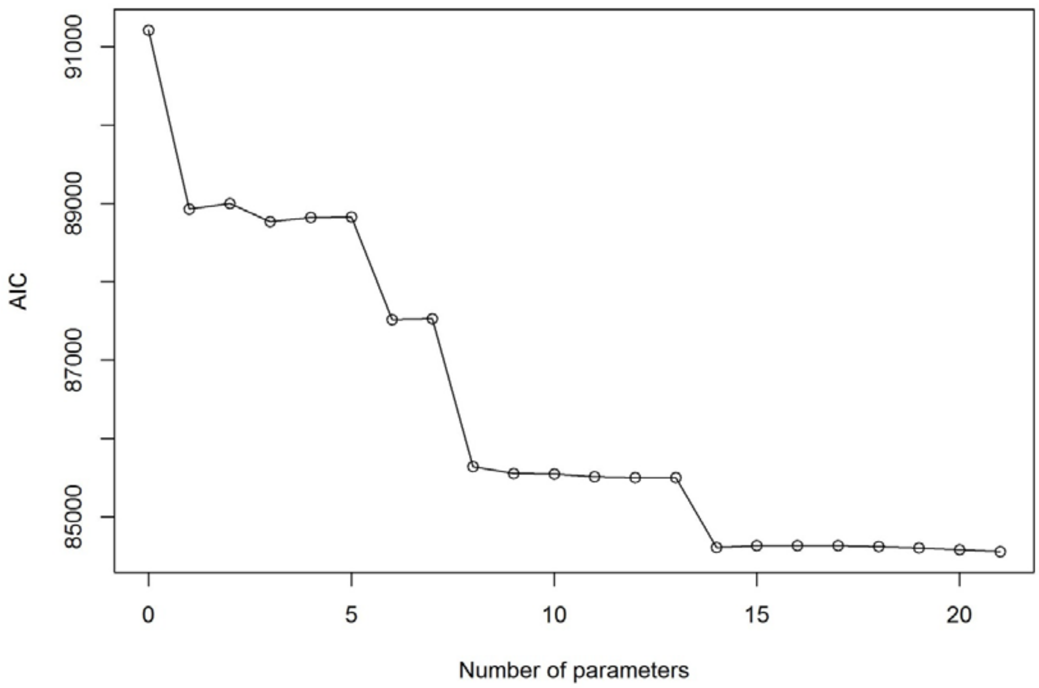

Figure 15. AIC for each model in Table 8.

The consumption calculated from the Atlantis model is not always realistic and this has to do with how the groups are set up in Atlantis. If the age groups have similar size the groups do not need to grow to be able to fit into the size of the next age group. This will result in zero consumption. This is a problem with some of the groups, e.g. the other tooth whales and long lived species such as baleen whales, large sharks and redfish. These groups grow quickly into their adult size and grow very slowly after that. They do however lose weight due to reproduction which they need to compensate for leading to some consumption. This is something that could be improved in the Atlantis model by either let them lose more weight during reproduction, lowering the assimilation efficiency parameter or turning on the respiration model.

4.2 The model performance

Both the EwE and the Gadget model were able to capture the trends of the biomass for the vertebrate groups. In this study the precision of information was very high. The biomass was known but that is not the case in real life. Therefore, this study shows that these models can theoretically simulate biomass from an ecosystem but not how they would perform when fed with real data. It is possible to try to mimic real data by adding error or bias to the data. This is a material for further research on model performance.

The EwE model did fit well to the true values for most of the vertebrate groups but the forecasting ability of the model was not tested. The estimation of the V parameter can result in unrealistic behaviour of the groups. When V approaches 1 the production of the groups can become unrealistically high at low biomass which would lead to high tolerance of high fishing pressure. This is also something that is worth investigating further.

Three metrics were used to test the performance and to compare the models. It is very important that the models have high correlation with the true values. In scenarios testing, it is important that the groups respond in a correct way to changes in fishing pressure and the correlation is an indicator on how well they may forecast changes. A group that has negative correlation with the true values does not have a reliable forecast. The second metric, RI, measures how far the simulated biomass is from the true biomass. It depends on what purpose of the model is how important it is to get the right magnitude. If the purpose is to forecast the effect of changing fishing pressure the magnitude might not be important if the trends are correct. The magnitude might be more important when the purpose of the model is to provide advice on how much should be fished but the Gadget model is often used for that purpose. The Gadget model consistently overestimated the biomass of the two groups that were tested with that model. It is not clear what caused the overestimation. It is possible that this could be a result of the methodology used in this study but a further research on this is needed. If this is something inherent in the model itself a further research is needed on the effect this may have on the advice.

References

Aydin, K., Lu999cey, S., and Gaichas, S. (2016). Rpath: R implementation of Ecopath with Ecosim. R package version 0.0.0.9003.

Begley, J., and D. Howell. 2004. "An Overview of Gadget, the Globally Applicable Area-Disaggregated General Ecosystem Toolbox." ICES CM, 1–15. http://www.hafro.is/gadget/.

Christensen, V., & Walters, C. J. (2004). Ecopath with Ecosim: methods, capabilities and limitations. Ecological modelling, 172(2), 109-139.

Darwall, W. R., Allison, E. H., Turner, G. F., & Irvine, K. (2010). Lake of flies, or lake of fish? A trophic model of Lake Malawi. Ecological Modelling, 221(4), 713-727.

Elvarsson, B. T. (2015). Rgadget. R package version. 1.0.

Forrest, R. E., Savina, M., Fulton, E. A., & Pitcher, T. J. (2015). Do marine ecosystem models give consistent policy evaluations? A comparison of Atlantis and Ecosim. Fisheries Research, 167, 293-312.

Fulton, E. A., Parslow, J. S., Smith, A. D., & Johnson, C. R. (2004). Biogeochemical marine ecosystem models II: the effect of physiological detail on model performance. Ecological Modelling, 173(4), 371-406.

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

Hoenig, J. M. (1983). Empirical use of longevity data to estimate mortality rates. Fish. Bull, 82(1), 898-903.

Kavanagh, P., Newlands, N., Christensen, V., & Pauly, D. (2004). Automated parameter optimization for Ecopath ecosystem models. Ecological Modelling, 172(2), 141-149.

Lentin, J. (2014). mfdb: MareFrame DB querying library. R package version 1.0.

Lentin, J. (2016). mfdbatlantis: MareFrame Atlantis Interface. R package version 4.0-0.

Olsen, E., Fay, G., Gaichas, S., Gamble, R., Lucey, S., & Link, J. S. (2016). Ecosystem Model Skill Assessment. Yes We Can!. PloS one, 11(1), e0146467.

Palomares, M. L. D., & Pauly, D. (1998). Predicting food consumption of fish populations as functions of mortality, food type, morphometrics, temperature and salinity. Marine and freshwater research, 49(5), 447-453.

Pauly, D. (1980). On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. Journal du Conseil, 39(2), 175-192.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Smith, A. D. et al. (2011). Impacts of fishing low–trophic level species on marine ecosystems. Science, 333(6046), 1147-1150.

Stefansson, G. (2003). Issues in multispecies models. Natural Resource Modeling, 16:415-437.

Stow, C. A. et al. (2009). Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems, 76(1), 4-15.