A Stepwise Fitting Procedure for automated fitting of Ecopath with Ecosim models

Erin Scott [a],[*], Natalia Serpetti [a], Jeroen Steenbeek [b], Johanna Jacomina Heymans [a]

[a] *Scottish Association for Marine Science, Oban, Argyll, PA37 1QA, UK*
[b] *Ecopath International Initiative Research Association, Barcelona, Spain*

Abstract

The Stepwise Fitting Procedure automates testing of alternative hypotheses used for fitting Ecopath with Ecosim (EwE) models to observation reference data (Mackinson et al. 2009). The calibration of EwE model predictions to observed data is important to evaluate any model that will be used for ecosystem based management. Thus far, the model fitting procedure in EwE has been carried out manually: a repetitive task involving setting >1000 specific individual searches to find the statistically 'best fit' model. The novel fitting procedure automates the manual procedure therefore producing accurate results and lets the modeller concentrate on investigating the 'best fit' model for ecological accuracy.

Code Metadata

Current code version	*V1*
Permanent link to code/repository used of this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-15-00065
Legal Code License	*GNU General Public License*
Code versioning system used	*svn*
Software code languages, tools, and services used	*Visual Basic.Net*
Compilation requirements, operating environments & dependencies	*OS needs to be Microsoft Windows XP service pack 3, Windows Vista, Windows 7, Windows 8.X, or Windows 10 or newer with.NET Framework 4 Full installed. If the computer does not run Microsoft Office 2007 or newer will need to install Microsoft Access 2007 database drivers.*
If available Link to developer documentation/manual	https://github.com/ElsevierSoftwareX/SOFTX-D-15-00065
Support email for questions	ewedevteam@gmail.com

1. Motivation and significance

Ecopath with Ecosim (EwE) is an ecological modelling software suite that is freely accessible with over 6000 registered users from 164 countries [1,2]. It is used to build models of marine and freshwater as well as terrestrial ecosystems. Ecopath describes energy flows through a trophic web of functional groups. Ecosim provides a time-dynamic simulation of the ecosystem, with key initial parameters inherited from the base Ecopath model [1].

The calibration of Ecosim models by fitting model predictions to time series reference data (observations) is important, as it evaluates the assumptions made when parameterising the Ecopath model [3]. The model fitting procedure tests a

* Correspondence to: Computing Science and Mathematics, University of Stirling, Stirling, FK9 4LA, UK.
E-mail addresses: egs@cs.stir.ac.uk (E. Scott), Natalia.Serpetti@sams.ac.uk (N. Serpetti), jeroen.steenbeek@gmail.com (J. Steenbeek), Sheila.Heymans@sams.ac.uk (J.J. Heymans).

Table 1
The eight alternative hypotheses and changing factors parameterised and compared by the Stepwise Fitting Procedure. ✓and
✗ indicate if a factor is included and not included respectively.

Hypothesis	Fishing data	Trophic effects (Vulnerabilities)	PP anomaly
Baseline	✗	✗	✗
Baseline and trophic effects	✗	✓	✗
Baseline and PP anomaly	✗	✗	✓
Baseline, trophic effects and PP anomaly	✗	✓	✓
Fishing	✓	✗	✗
Fishing and trophic effects	✓	✓	✗
Fishing and PP anomaly	✓	✗	✓
Fishing, trophic effects and PP anomaly	✓	✓	✓

combination of hypotheses, to assess the top-down and bottom-up controls on the time series dynamics. Vulnerability parameters, indicating the strength of the interaction between predators and prey, and/or an anomaly primary production (PP) function that represent an unknown function that can play a bottom-up control on the ecosystem [4]. Vulnerability parameters and a PP function are estimated using an optimisation search routine in Ecosim which reduces the sum of squares difference between the predicted and observed data. The effects of top-down controls (e.g. fishing) and bottom-up controls (e.g. primary productivity) depend on how the various model ecosystem components interact [3].

The Ecosim fitting procedure has previously been a manual process using the 'fit to time series' search interface within EwE [1]. This involves setting up individual searches where specific conditions apply, including selecting specific time series to be enabled or disabled (e.g. with and without fishing), thus searching for best fits with and without fishing. It further involves selecting which search to carry out; vulnerability, PP anomaly or both. The number of parameters to be estimated in the search also needs to be set, and a sensitivity search is needed to determine which vulnerability parameters must be estimated. The set-up of one search therefore needs to be repeated depending on the number of parameters that can be estimated and hypotheses being tested. This repetitive task increases the chances of human error. In a large model such as the model of the West Coast of Scotland [5], which contains 48 time series of catch or biomass that can be used to estimate fitting parameters, either through vulnerability, PP anomaly or vulnerability plus PP anomaly searches, the number of searches to be run becomes substantial. The fitting procedure assumes that the 48 time series are independent, but that each data point in the time series are not, thus we can estimate 47 parameters: 2–23 PP spline points (for smoothing the primary production anomaly over the 23 years of the time series) or 47 vulnerabilities, which leads to 1658 possible combinations to be tested. Previously, manual fitting of the West Coast of Scotland model varied only each 5th combination (i.e. 5, 10, 15 vulnerabilities, 5, 10, 15 spline points etc.) in [5] to reduce the repetitive nature of the task. The fitting procedure developed here gives ecosystem modellers the opportunity to test all fitting combinations with accuracy and speed, and thus allows focus on the analyses of results.

2. Software description

The novel Stepwise Fitting Procedure presented here automates the Ecosim model fitting procedure described in [3] and used by [4,5]. As [3] states it is convention and good practice to explore different search strategies for parameterisations that best explain the observation data to find the best-fit model. The automated procedure focuses on the impact of fishing, changes in the predator–prey dynamics (predation vulnerability settings), changes in primary production (PP), and all of the above [3]. A high level view of the stepwise fitting procedure is shown in Fig. 1. Eight alternative hypotheses are parameterised and compared. A more detailed description of the eight types of hypotheses can be found in [5] and are outlined for convenience in Table 1 which presents the factors which are included in each hypothesis.

The fitting procedure determines all the alternative hypotheses that may describe the observation data in equal measure and finds the 'best-fit' model. The alternative hypotheses are calculated by the fitting procedure from the maximum number of parameters that can be estimated ($m - 1$) [6]. The maximum number of estimated parameters m is calculated by the procedure from the total number of loaded time series of type: relative biomass, total mortality, catch and average weight [6]. Here we assume that all the time series data are independent but that each data point in the time series are not independent from the other data points in that series. The hypotheses are listed as fitting iterations in the user interface of the stand-alone version (Fig. 2).

The fitting procedure calculates the total number of iterations for each hypothesis by calculating all combinations of number of parameters estimated (K) based on m and/or the maximum number of spline points. For example, if $m = 3$ then for the hypothesis "Baseline and trophic effects" the procedure would calculate 3 iterations where the value of K will be 1, 2 and 3 in each of the iterations. All forcing data including absolute biomass, forced biomass, time forcing data, effort data by gear type, fishing mortality, forced total mortality and forced catch time series are disabled in baseline iterations [6]. There is the option to include and enable absolute biomass time series to the calculation of m and to include these time series in baseline iterations. If a forcing function is applied to the PP in the model, anomaly and vulnerability plus anomaly iteration searches will be calculated and shown in the iteration list of the fitting procedure user interface. The spline point step size can
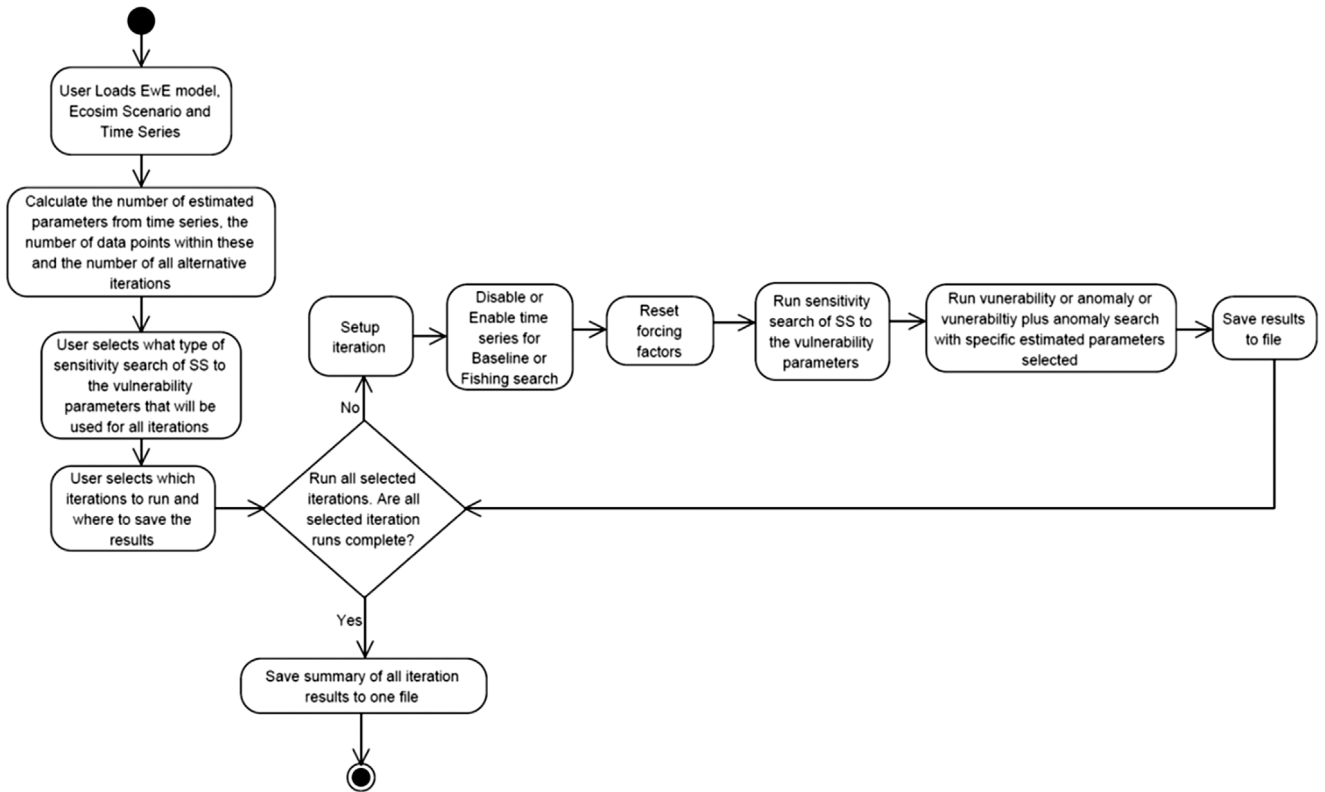
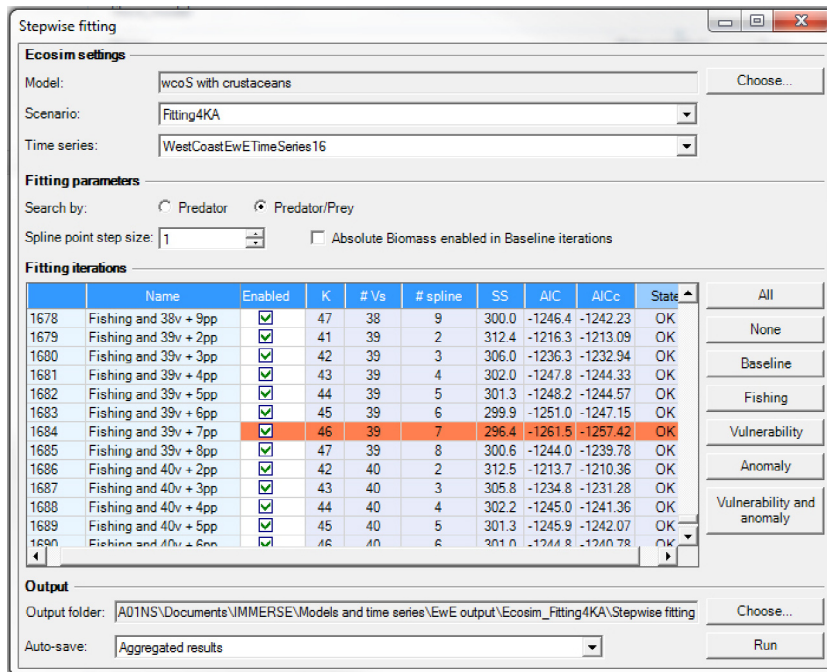Fig. 1. Activity diagram of the workflow of the Stepwise Fitting Procedure.

Fig. 2. The Stepwise Fitting Procedure user interface highlights the best fit iteration of the West of Scotland model.

also be set, with the default value set at 1. Before each iteration search is started the vulnerabilities and PP anomaly are reset by the procedure.

The fitting procedure was created to automate and assist in finding a 'best-fit' model. The 'best-fit' model is found

and determined by the minimum difference between model predictions to time series observations using the weighted sum of squared differences (SS) and the Akaike Information Criterion (AIC) [7], which penalises for fitting too many parameters based on the number of time series available for estimating

the SS. The formula used for the AIC is defined as:

$$AIC = n \cdot \log\left(minSS/n\right) + 2K \qquad (1)$$

where n is calculated by the fitting procedure as the total number of observations, or time series values, from the loaded times series of type: relative biomass, total mortality, catch and average weight. minSS is the minimum sum of squares calculated by the algorithm, and K is the number of parameters estimated. In addition the procedure also calculates the AICc [8], to address small sample size. The AICc is defined as:

$$AICc = AIC + 2K \cdot (K - 1)/(n - K - 1). \qquad (2)$$

The fitting procedure calculates these values for each iteration and identifies the best-fit by the iteration which has the lowest AICc value.

The number of possible vulnerabilities that can be estimated are often significantly more than that of the data available to constrain the model, thus it is imperative to test the vulnerabilities that affect the most detectable change [3]. Before each iteration search is started a sensitivity search of sum of squares (SS) to the vulnerability parameters is carried out to find which of the vulnerability parameters creates the largest change on the SS, and therefore the most sensitive vulnerability parameters are used in the iteration search. The fitting procedure interface gives the choice to search by predator or predator/prey for all iterations, as is also the case in the search interface within EwE [1].

A variety of output results for further analysis can be chosen from the fitting procedure user interface: these are; no results, Ecosim results (similar to all the output results obtained from a usual Ecosim run), aggregated results (a smaller sub group of Ecosim results which include biomass, yield, mortality, vulnerabilities and anomaly shape) and both Ecosim and aggregated. Once all iteration searches are completed a csv file of the runs and fitting results is produced. Ecosim results are stored in separately named iteration folders. Aggregated results are stored in one folder and the file names identify the providence of the results. Anomaly shape results for all iteration searches are also stored in one csv file when Ecosim or aggregated result output is selected.

The fitting procedure can be used in three ways: command line, stand-alone and as an integrated plug-in within the EwE software suite. Iteration searches in the stand-alone version are faster than that of the integrated version. The stand-alone version has a user interface (Fig. 2) and has all the functionality of the integrated plug-in. The integrated plug-in will be available in the next release of the EwE software suite due for release in February 2016.

3. Example of the automated model fitting procedure

3.1. Iteration numbers and run times

The West Coast of Scotland model [5] is used as an example to demonstrate the fitting procedure. As stated previously there are 48 time series of catch or biomass, and therefore the fitting procedure calculates that a maximum of 47 parameters can be estimated. These parameters can be either vulnerabilities or PP

spline points (up to the number of years in the time series). Thus 2–23 spline points (as there are 23 years in the time series, and one cannot have less than 2 spline points) or 47 vulnerabilities and a combination of these are possible. The fitting procedure calculates the number of observations (n in the AIC formula) in the time series as 1063 and that there is in total 1658 iteration searches. There is 829 iteration runs for both Baseline and Fishing. The time taken for the fitting procedure to complete all iteration runs was approximately 3.5 days. The number of different search iterations and the time taken for the fitting procedure to complete each are shown in Table 2.

3.2. Results

The fitting procedure user interface highlights the best fit model (the iteration with the lowest AICc value) which in this case includes fishing and 39 vulnerabilities and 7 PP anomaly spline points as shown in Fig. 2. Fig. 3 presents the AICc result output for all iteration searches. This shows a comparison between the baseline and fishing iterations. Baseline (without fishing) iterations have higher AICc values than fishing iterations which indicates that fishing is the main driver in this ecosystem. Although the fitting procedure can produce various output results, in the case of the West Coast of Scotland model, aggregated result output was selected. The fitting procedure can assist in the analysis of results by providing output that can be easily converted into different visualisations.

4. Impact

This fitting procedure will aid in the analysis of a variety of EwE models implemented by the high number of worldwide EwE registered users and encourage new users to the EwE software. The fitting procedure automates the Ecosim model fitting procedure described by calculating all alternative hypotheses and suggests the 'best-fit' model. This procedure is useful for exploratory analysis. It allows for broad exploration of the parameter space and can be useful if the drivers of the system are not clearly elucidated. This procedure will help identify key drivers in the system. The quality of the results obtained by this procedure is dependent on the quality and quantity of data used in the fitting and best practice in the parameterisation of the initial Ecopath with Ecosim modelling [3,8].

The choice in 'best-fit' model is still subjective as it only reflects the statistical fit, therefore the user should take care in selecting the best model on the basis of its fit to data, supporting ecological knowledge and reasoned justification [3]. There are other parameters within an Ecosim model that could also be estimated to find the 'best-fit' model such as the diet matrix [9] which the fitting procedure does not deal with and the user should also consider. In addition, this procedure does not address the issue of confirmatory analysis. Here the user assumes that the data is perfect and observation error is not resolved. For truly confirmatory analyses, suitable for management advice to policy makers, a limited number a priori candidate hypotheses should be tested (i.e. only a limited number of parameters estimated) and compared to see which hypothesis explains the data

Table 2
The number of iterations and time taken to complete for each of the 8 alternative hypotheses of the West Coast of Scotland model. The stand-alone version of the fitting procedure was run on a PC running Windows 7 with an i7 core and 8 GB RAM.

Hypothesis	No. of iterations	Time (min)	No. of parameters estimated
Baseline	1	1	0
Baseline and trophic effects (vulnerabilities)	47	82	1–47
Baseline and PP anomaly	22	9	2–23
Baseline and trophic effects (vulnerabilities) and PP anomaly	759	2333	a
Fishing	1	2	0
Fishing and trophic effects (vulnerabilities)	47	139	1–47
Fishing and PP anomaly	22	31	2–23
Fishing, trophic effects (vulnerabilities) and PP anomaly	759	2672	a

[a] A combination of vulnerabilities and spline points up to the maximum number of estimated parameters (47).

best. Thus, running the automated stepwise fitting procedure should not be used without forethought.

As stated by [3] the key strength of the procedure is that it allows users to explore a range of possible ecological mechanisms that may contribute to ecosystem changes on decadal scales. Automating what was before a manual ad-hoc procedure reduces human error and accelerates the fitting process. Reducing the occurrences for human error will give more confidence in the results produced. Speeding up this repetitive process will give more time for explorative analysis of results and choosing the best model. This will help produce better models that will help guide policy makers.

5. Future work

This procedure uses a specific formulation of the AIC which is based on the work done by [3] and it balances model fit against overparameterisation. However, it is probably more suited to univariate analysis as it penalises the incorporation of too many time series or too long a time series, while in manual fitting more and longer time series data help constrain the model. Thus in future it would be useful to incorporate the likelihood formulation of the AIC as used by [10] or other multivariate model formulations. In addition the specific estimation of K in this program is based on the number of time series included, not the number of data points, as we assume that a time series is autocorrelated. This might be an assumption too far, and in future the value of K could be decided by the user, and not calculated by the procedure.

In addition, in EwE it is possible to weight the time series used for fitting based on its quality [11] and these weightings could be included in future programs, or these could be varied to undertake an ensemble approach to fitting EwE models.

6. Conclusion

The Stepwise Fitting Procedure automates testing of alternative hypotheses used for fitting EwE models to observation data [3]. The fitting procedure allows users to easily explore a range of possible ecological mechanisms that may contribute to ecosystem changes. It will assist in the knowledge of empirical investigations and help confront data issues [3]. The main benefit the fitting procedure provides is reducing the
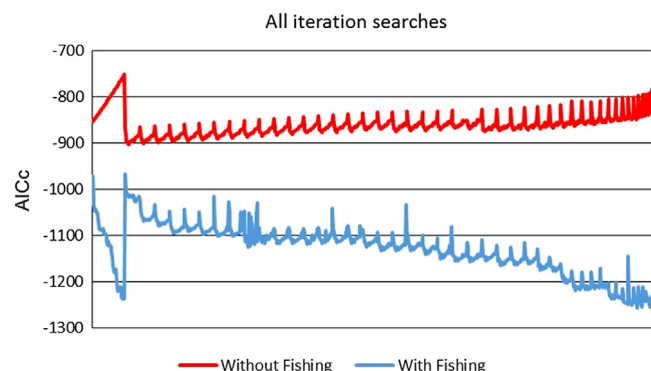

Fig. 3. The AICc values for all iteration searches. The red line indicates baseline values and the blue line indicates fishing iteration values. The *x*-axis is the order in which the fitting procedure carried out the iterations, which was the same for both baseline and fishing. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

occurrences of human error and aids in finding the 'best-fit' model. This allows the user to concentrate on the ecological reasoning behind the model parameters and the justification of the 'best fit' model rather than on the model fitting procedure itself.

Acknowledgements

References

[1] Christensen V, Walters CJ. Ecopath with Ecosim: Methods, capabilities and limitations. Ecol Modell 2004;172(2–4):109–39.

[2] Coll M, Akoglu E, Arreguín-Sánchez F, Fulton EA, Gascuel D, Heymans JJ, Libralato S, Mackinson S, Palomera I, Piroddi C, Shannon LJ, Steenbeek J, Villasante S, Christensen V. Modelling dynamic ecosystems: venturing beyond boundaries with the Ecopath approach. Rev Fish Biol Fish 2015;413–24.

[3] Mackinson S, Daskalov G, Heymans JJ, Neira S, Arancibia H, Zetina-Rejón M, Jiang H, Cheng HQ, Coll M, Arreguin-Sanchez F, Keeble K, Shannon L. Which forcing factors fit? Using ecosystem models to investigate the relative influence of fishing and changes in primary productivity on the dynamics of marine ecosystems. Ecol Modell 2009; 220(21):2972–87.

[4] Ahrens RNM, Walters CJ, Christensen V. Foraging arena theory. Fish Fish 2012;13:41–59.

[5] Alexander K, Heymans S, Magill S, Tomczak M, Holmes S. Investigating the recent decline in gadoid stocks in the west of Scotland shelf ecosystem using a food-web model. ICES J Mar Sci 2015;72:436–49.

[6] Christensen V, Walters CJ, Pauly D, Forrest R. Ecopath with ecosim version 6 user guide, vol. 281. Vancouver, Canada: Fish. Centre, Univ. Br. Columbia; 2008. p. 1–235.

[7] Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974;19(6):716–23.

[8] Burnham KP, Anderson DR. Multimodel inference: Understanding AIC and BIC in model selection. Sociol Methods Res 2004;33(2): 261–304.

[9] Ainsworth CH, Walters CJ. Ten common mistakes made in Ecopath with Ecosim modelling. Ecol Modell 2015;308(8):14–7.

[10] Gaichas SK, Aydin KY, Francis RC. What drives dynamics in the Gulf of Alaska? Integrating hypotheses of species, fishing, and climate relationships using ecosystem modeling. Can J Fish Aquat Sci 2011;68(9).

[11] Heymans JJ, Coll M, Link J, Mackinson S, Steenbeek J, Walters C, Christensen V. Best practice in Ecopath with Ecosim food-web models for ecosystem-based management. Ecol Modell. http://dx.doi.org/10.1016/j.ecolmodel.2015.12.007.