An overview of Gadget,
the Globally applicable Area-Disaggregated General Ecosystem Toolbox

James Begley[1]
and Daniel Howell[2]

[1] **Marine Research Institute, Reykjavik, Iceland**
[2] **Institute of Marine Research, Bergen, Norway**

Abstract

Gadget is the Globally applicable Area-Disaggregated General Ecosystem Toolbox. Gadget is a powerful and flexible framework that has been developed to model complicated statistical marine ecosystems within a fisheries management and biological context, and can take many features of the ecosystem into account. Gadget allows the user to include a number of features of the ecosystem into the model: One or more species, each of which may be split into multiple components; multiple areas with migration between areas; predation between and within species; growth; maturation; reproduction and recruitment; multiple commercial and survey fleets taking catches from the populations. Gadget works by running an internal forward projection model based on many parameters describing the ecosystem, and then comparing the output from this model to observed measurements to get a likelihood score. The model ecosystem parameters can then be adjusted, and the model re-run, until an optimum is found, which corresponds to the model with the lowest likelihood score. This iterative, computationally intensive process is handled within Gadget, using a robust minimisation algorithm. Gadget has successfully been used to investigate the population dynamics of stock complexes in Icelandic waters, the Barents Sea, the North Sea, the Irish and Celtic Seas and the Sofala Bank fishery of Mozambique. This paper describes the structure and main components of an ecosystem model developed using the Gadget framework.

1 Introduction

Gadget is a software tool that has been developed to model marine ecosystems, including both the impact of the interactions between species and the impact of fisheries harvesting the species. Gadget simulates these processes in a biologically realistic manner, and uses a framework to test the development of the modelled ecosystem in a statistically rigorous manner (Anon 2002, Anon 2003).

Gadget can run complicated statistical models which take many features of the ecosystem into account. Gadget works by running an internal model based on many parameters, and then comparing the data from the output of this model to observed data to get a goodness-of-fit likelihood score. The parameters can then be adjusted, and the model re-run, until an optimum is found, which corresponds to the model with the lowest likelihood score. The Gadget framework consists of three parts:

- a parametric model to **simulate** the ecosystem

- statistical functions to **compare** the model output to data

- search algorithms to **optimise** the model parameters

The files required for a Gadget model are all in plain text ASCII format, enabling them to be easily generated, read and edited. A data warehouse program exists that can output selected data in the correct format for Gadget (Anon 2003, Kupča 2004). A graphical front end is in development to produce the files required for the simulation model (Anon 2004).

Gadget is a freely available, open source program. The software and associated documentation can be downloaded and used free of charge, and is available from the Gadget website at `www.hafro.is/gadget`. Since the source code is freely available it is possible to examine the program to ensure that the program is correctly implementing the model selected by the user. Furthermore users can add features to the code as required, providing the flexibility to adapt the core program to a wide range of fisheries situations. These improvements can then be submitted to the maintainers of the program for inclusion in future releases of the code.

Gadget has been been developed from Bormicon (Stefánsson and Pálsson 1997) and Fleksibest (Frøysa et al. 2002), which in turn were influenced by the Multspec model (Bogstad et al. 1997). The recent development of

Gadget has successfully been used to investigate the population dynamics of various stock complexes in Icelandic waters, the Barents Sea, the North Sea, the Irish and Celtic Seas, and the Sofala Bank fishery of Mozambique.

2   Simulation Using Gadget

A Gadget model is a parametric forward-simulation model of an ecosystem, typically consisting of various fish populations and their interactions. Such a model simulation will track the changes to the fish populations due to both biological actions (eg growth) and also the interaction between populations (eg predation).

Gadget works by keeping track of the number N_{smalrt}, and mean weight W_{smalrt}, of fish in a "population cell". The population cell consists of all the fish of a given species (s), and maturation stage (m), with a specified age (a) and length (l), living in a specified region (r), on a specified timestep (t). Mathematically, this is simply a pair of numbers (over each of the six indices) that can be adjusted by various functions to simulate the effect of biological processes on the population cell.

Clearly, for any moderately complex model, this will involve the storage of a large amount of data, and a number of steps have been taken to reduce this. Gadget runs a forward simulation, Markovian, model so that the population at time $t + 1$ is only dependant on the population at time t, and not any earlier timestep, which means that Gadget only has to store the population from the previous timestep. Also, Gadget combines the species and maturation stage of the fish into a single index (termed "population group" or, perhaps erroneously, "stock") since it is assumed that differing maturation stages of a species are biologically different (eg there is a different growth rate for an immature fish compared to a mature fish) so that they can effectively be considered as different species by the simulation.

2.1   Stocks

A Gadget "stock" or "population group" is a group of fish which are all considered to share similar biological characteristics (e.g. growth, mortality, maturation ogive). A typical such stock could be an entire species, all mature fish in a species, or even all mature females from one region in a species.

Each stock is defined by specifying the length groups, age groups, and length weight relationship to be used, along with the functions that are to be used to simulate the biological processes that affect the stock. The oldest age group and the longest length group of the stock are both treated as plus groups. A full description of each of the available functions that are available to simulate biological processes in Gadget is outside the scope of this paper, and interested readers are referred to the Gadget User Guide (Begley 2004) for a full list.

2.1.1 Growth

Modelling the change in length, and corresponding mean weight, due to growth is an important process in a marine ecosystem model, and there are a number of growth functions that are included within the Gadget framework. All the growth functions work by calculating a mean increase (in either length or weight) and then implementing this increase by moving a number of fish from one length group up to longer length groups.

For fish in a given length cell, the mean increase in either weight, or length, is calculated according to a growth equation (for example by a function based on a Von Bertalanffy growth equation). The corresponding increase in length, or weight, is then calculated, based on the length weight relationship defined for the stock.

The mean increase in length and weight is then translated into a statistical distribution of actual growths around that mean increase. This distribution has the effect of splitting the calculated increase up into a number of discrete length groups, so that a proportion of the fish increase in length by zero length groups, some by one length group, some by two and so on, such that the mean of all the increases is equal to the calculated mean growth. The statistical distribution that is used by Gadget is the beta-binomial distribution, which gives a high degree of flexibility from a single parameter.

2.1.2 Predation

For any model of an ecosystem that has more than one stock, it is usually of interest to model the interaction between these stocks, typically by considering the consumption of a prey by a predator. When this is extended to consider an ecosystem with many stocks, each predator can consume more than one prey and each prey can be consumed by more then one predator.

For each predator-prey relationship, the preference of a predator from length group L for a prey from length group l is modelled using a suitability function, $S(L, l)$. Multiplying this suitability value by the prey biomass will then give the prey abundance available for each predator to consume. There is a maximum amount $H(L)$ that each predator can consume (this is typically obtained from laboratory experiments), which is multiplied by the fraction of the total food abundance that the prey accounts for, to give the biomass of prey that each predator would want to consume, assuming that enough food is available to meet the needs of all predators. This is then scaled by multiplying by the number of predators in length group L.

The effect of the predation is calculated for each predator-prey relationship in turn, and stored without actually applying this effect to the ecosystem until all the predation for that timestep has been calculated. Then the overall mortality induced in the prey species by all the predators is calculated, and prey biomass is removed from the model. If the total consumption of any prey by all the predators amounts to more than the biomass of prey available, then the model runs into "understocking". In this case, the consumption by the predators is adjusted so that no more than 95% of the available prey biomass is consumed, and a penalty is applied to the likelihood score obtained from the simulation.

2.1.3 Maturation

For any model that has many stocks modelling different maturation stages of the same species it is important to be able to simulate the maturation process by moving fish from one stock to another. This is done by calculating the proportion of each age-length cell of the immature stock that will mature on a given timestep, and then moving this proportion into the corresponding age-length cell for the mature stock.

The proportion of the immature stock that will mature is calculated from a maturity ogive that is calculated from parameters specified by the user. The maturation probabilities may be age and/or length dependant, and, by using multiple stocks, maturation may also be made sex-dependant. Note that when moving the fish from one stock to another, the total number of fish in the ecosystem is kept constant, and that there is no mortality associated with the maturation process.

2.1.4 Recruitment and Reproduction

In order to track the development of an ecosystem over time, it is necessary to add new fish into the model to replace the fish that are removed (eg by commercial fishing). Gadget has two mechanisms to simulate this - either by adding recruits into the simulation or by modelling a closed life-cycle system.

Adding recruits to the stock takes a simplistic approach. The total number of recruits on any given timestep is specified (either as an absolute number or a parameter), along with the mean length and standard deviation of the length. A Normal distribution is then used to calculate a length distribution for the recruits, and a length - weight relationship is used to calculate a mean weight for each of the length cells of the recruits. These new recruits are then simply added to the youngest age group of the stock. The model may then be optimised to select the values for each recruitment parameter that produces the population giving the best fit to the available data.

An alternative approach that can be taken is to model a closed life-cycle system, where the number of recruits to a stock is governed by the spawning population. The total number of recruits is calculated (either based on a function of the fecundity, or the spawning stock biomass, of the 'parent' stock). The mean length and standard deviation of the length of the recruits is specified by the user, enabling a Normal distribution for the lengths of the recruits to be calculated, and these new recruits can then be added to the simulation in a similar manner to that described above. This spawning process can also affect the parent stock, and the mortality and weight loss attributable to the spawning process is also modelled by Gadget.

2.1.5 Migration

For any model that has a stock that is found on more than one region within the simulation, it is important to be able to simulate the migration of the stock as the fish move from one region to another. This is done in Gadget by calculating the proportion of each age-length cell of the stock that will move from one region to another on a given timestep, and then removing this proportion from the original region and adding it to the corresponding age-length cell for the destination region.

The proportion of the stock that will migrate from one region to another is

calculated from migration matrices specified by the user. These migration matrices M_{ij} specify the proportion of the stock that will migrate from region i to region j on each timestep.

2.1.6 Other Processes

At the end of each year in the simulation, all the fish that are currently in the simulation will increase in age. This is done by moving all fish up one age group. Note that the oldest age group is modelled in Gadget as a plus group, and so this age group doesn't increase in age.

In addition to the mortality induced by some of the other processes (in particular predation), Gadget allows for a proportion of the stock to be removed due to other natural causes. This is simply modelled by specifying the proportion of each age group that will be removed from the simulation on each timestep.

2.2 Fleets

A Gadget "fleet" is treated as a simplified predator of the various prey stocks. Thus a fleet can be thought of a "stock" that has a single age group, a single length group, which doesn't grow, mature, migrate, recruit or reproduce. The only process that is of interest is the predation, where the fleet will 'consume' an amount of the prey stocks.

Gadget has two methods that can simulate effect on the ecosystem of the removal of fish due to fishing effort - either by specifying the biomass of the fish that is caught by the fleet (as recorded by landings data), or by specifying the fishing effort of the fleet. Both methods have similarities with the predator-prey interaction described in section 2.1.2 above, simplified slightly since the predator has only one length group.

The effect of specifying the biomass of the fish that is caught by the fleet is to fix the 'consumption' by the predator of the prey to the specified biomass. Since there is only one length group for the predator, this consumption can be shared amongst the length groups of the prey, using the suitability function as a multiplier. This approach to the simulation of the predation is termed "TotalFleet", and requires the landings data to be known.

Alternatively, by specifying a fishing effort parameter, the 'consumption' by the predator of the prey is fixed to a given proportion of the available

prey biomass, after taking the suitability function into account. This approach to predation is termed "LinearFleet", and doesnt require any landings data, so it can be used for a prediction model running the simulation into the future.

3 Comparisons Using Gadget

For a Gadget simulation, the data requirements are minimal - either none or just the overall catch for the fleets for each timestep, depending on the approach taken to model the fishing effort. However, for the model to be statistically testable, likelihood data is required. The data that is used depends on what aspects of the simulation are of particular interest, and could be length distributions, age length keys, survey abundance indices, mean length or weight at age, stomach content data or returns from tagging experiments. Unlike most fisheries models, catch at age data is not required, though it can be used if available.

Each data set that is to be used is assigned to a likelihood component. This specifies the statistical relationship to be used when comparing model results to the data. There are a number of likelihood components that are available to compare a Gadget simulation to likelihood data. A number of factors will influence the choice of the likelihood components, but the most important two factors to consider are availability and relevance of the data that is used to make the comparisons.

To make best use of the data that is available, it is sometimes necessary to aggregate sparse data, or even to exclude data from some timesteps completely. Gadget will only make a data comparison for timesteps that have data available, but a zero entry (from a sparse data source) is taken to mean zero, and not data missing due to poor sampling, which can lead to the comparison being unrepresentative of the data. The likelihood data does not need to be continuous, and the level of aggregation used doesn't need to be the same for all likelihood components used in the model.

Various likelihood types are used to define the various likelihood components that can be used to calculate the "goodness of fit" of the Gadget model to the available data. Each likelihood component will calculate a likelihood score for that individual component. A weighted sum of all the likelihood scores is then used to calculate an overall likelihood score. A full description of each of the available likelihood components that can be

used in Gadget is outside the scope of this paper - interested readers are referred to the Gadget User Guide for a full list (Begley 2004).

3.1   Distribution Data

A common source of data for fisheries studies is age- or length- distribution data, either from sampling the commercial catch or from government surveys. This data can be used in Gadget by specifying a "CatchDistribution" likelihood component, and the distribution data can either be aggregated into age groups (giving a distribution of length groups for each age), length groups (giving a distribution of age groups for each length) or into age-length groups. The likelihood score that is calculated gives some measure as to how well the distribution from the model fits to the distribution observed in the sampling process.

Various different functions to compare the distribution data are available in Gadget, ranging from a simple sum-of-squares comparison (as shown in equation 1 below) through to more complicated multivariate Normal distributions. The choice of equation used to compare the distribution data is left to the user.

$$\ell = \sum_{time} \sum_{regions} \sum_{ages} \sum_{lengths} \left(P_{tral} - \pi_{tral} \right)^2 \tag{1}$$

where P is the proportion of the distribution from the data sample, and π is the corresponding proportion from the model sample, for that time/ region/age/length combination

3.2   Abundance Data

Another common source of data that can be used in a Gadget model is information from abundance indices that are calculated from standardised government surveys. This data can be used by specifying a "SurveyIndices" likelihood component, and the abundance data can either be aggregated into age groups or length groups. This likelihood component is used to compare the relative abundance of a stock in the Gadget simulation to the indices calculated from a standardised survey for that stock.

The likelihood score is calculated by fitting a linear regression line to the difference between abundance indices calculated within the model and abundance indices calculated directly from standardised survey data, as given by equation 2. The model will calculate an index for the population,

and then fit a linear regression line between these calculated indices and those input from data files.

$$\ell = \sum_{time} \left(I_t - (\alpha + \beta \widehat{I_t}) \right)^2 \tag{2}$$

where I is the abundance index from the standardised survey, and $\widehat{I}$ is the corresponding index calculated in the Gadget model.

The exact format of this linear regression equation can vary, depending on survey index data available. It is possible to take the log of the indices and the modelled data before fitting the linear regression line. The slope and intercept of the linear regression line are controlled by the parameters alpha and beta, and it is possible to fix these to specified numbers, or let Gadget calculate these to get the best fit to the modelled data.

3.3   Statistical Data

An alternative use for data obtained by sampling the commercial catch or government surveys is to calculate the mean of some biological measurement and then compare this mean value to the corresponding value obtained from the simulated ecosystem data. This data is used in Gadget by specifying a "CatchStatistics" likelihood component, and is typically used to compare mean length at age, or mean weight at age.

The model will calculate the mean length (or weight) of the stock that is caught according to the model parameters, and aggregate this into specified age groups. The likelihood score is calculated as a function of the difference between the mean data calculated by the model and that caught according to the landings data and specified in the data file. Various functions are available for the user to make this comparison, of the form shown in equation 3 below.

$$\ell = \sum_{time} \sum_{regions} \sum_{ages} \left(\frac{(x_{tra} - \mu_{tra})^2}{\sigma_{tra}^2} N_{tra} \right) \tag{3}$$

where x is the sample mean length from the data, and μ is the corresponding mean length calculated from the model, σ is the standard deviation of the length, calculated from the model and N is the sample size for each time/region/age combination

3.4 Stomach Data

Data obtained from analysing the stomach contents of various predators can be used to give an indication of the diet composition of the stock. This data is used in Gadget by specifying a "StomachContent" likelihood component, and is typically used to compare the ratios of different prey stocks found in the diet of the predator. Care is needed when making this comparison, since the data will give information on the stomach content at the time of capture, where as the simulation can currently only give information about the modelled consumption of the prey.

The likelihood score is calculated by comparing the ratio of the consumption of different preys by a predator in the model to the ratio of the preys found in the stomach contents data specified in the input file. Various functions are available for the user to make this comparison, of the form shown in equation 4 below.

$$\ell = \sum_{time} \sum_{regions} \sum_{prey} \left(P_{trp} - \pi_{trp} \right)^2 \tag{4}$$

where P is the proportion of the stomach content from the data sample, and π is the corresponding proportion from the model sample, for that time/region/prey combination

3.5 Tagging Data

Data obtained from analysing the returns information from tagging experiments can be used to give an indication of the migration patterns of the stock. This data is used in Gadget by specifying a "Recaptures" likelihood component, and also specifying the tagging experiments that have taken place (Hannesson et al 2004).

The likelihood score is calculated by comparing the recaptures data obtained from the tagging experiment with an expected recapture value. The model will calculate an expected recaptures value for the tagging experiment, and then use a Poisson function (as shown in equation 5 below) to compare this to the recaptures obtained from the tagging experiment, as given in the data files.

$$\ell = \sum_{time} \sum_{regions} \sum_{length} \left(N_{trl} + \log \nu_{trl}! - N_{trl} \log \nu_{trl} \right) \tag{5}$$

where N is the recapture sample size from the tagging experiment, and ν is the corresponding sample size from the model, for that time/area/age/length combination

4   Optimisation Using Gadget

The overall likelihood score gives a single measure as to how well observations from the modelled ecosystem fit to the data that has been provided as likelihood components. Varying the value of the parameters that have been used to simulate the ecosystem will cause the fit to the data to change, and the overall likelihood score will also change to reflect this. By using a search algorithm to iteratively adjust the parameters, it is possible to find a set of parameter values that give the lowest likelihood score, and thus correspond to the best fit of the model to the data. The functional form of the model (number of population groups, choice of growth equation etc.) remain fixed, while the parameters are optimised. This computationally intensive search process is handled by Gadget, using several robust algorithms, which can be used either on their own or by combining them into a single hybrid algorithm.

4.1   Hooke & Jeeves

Hooke & Jeeves (Hooke and Jeeves 1961) is a simple and fast local minimisation algorithm. From an initial starting point the algorithm takes a step in various 'directions', by adjusting one parameter at a time, and conducts a new model run at each new point. If the new likelihood score is lower than the old one then the algorithm uses the new point as it's best guess, and re-starts from this new point. However, if the new likelihood score is higher, then the algorithm returns to the old point, and tries again by adjusting a different parameter. The search proceeds in series of these steps, each step slightly smaller than the previous one. When the algorithm finds a point which it cannot improve on with a small step in any direction, it accepts this point as being the 'minimum', and records this point to a data file before exiting.

4.2   Simulated Annealing

Simulated Annealing (Corona et al. 1987) is a global minimisation algorithm. From an initial starting point the algorithm takes a random step in various 'directions', by adjusting one parameter at a time, and conducts a

new model run at each new point. If the new likelihood score is lower than the old one then the algorithm stores the new point as it's best guess, and re-starts from this new point. However, if the likelihood score is higher then the algorithm may still accept this point, based on the probabilistic "Metropolis Criteria", and thus the algorithm can escape from a local minimum. The algorithm exits when a stable point is found which cannot be improved on with a small step in any direction, and the Metropolis Criteria rejects all the steps away from the current best point. The best point is then accepted as being the 'minimum', and is recorded to a data file, and the algorithm exits.

The Metropolis Criteria will accept a move to a point with a higher likelihood score based on a function of both the size of the move and a parameter termed "temperature", and is given in equation 6 below:

$$
\begin{aligned}
M &= e^{\dfrac{-\Delta F}{T}} \\
P &= \begin{cases} 1 & \text{if } M > r \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{6}
$$

where ΔF is the change in likelihood score, T is the "temperature" of the algorithm, and r is a random number between 0 and 1.

Note that when the "temperature" is very high ($T \rightarrow \infty$), the Metropolis Criteria will always accept any move, and the Simulated Annealing algorithm will simplify to a form of a random search algorithm. Conversely, when the temperature is very low ($T \rightarrow 0$), the Metropolis Criteria will always reject any move, and the Simulated Annealing algorithm will simplify to a form of a local search, similar to Hooke & Jeeves. By slowly reducing the temperature of the algorithm, the number of moves that are accepted by the Metropolis Criteria are reduced and the algorithm will find a minimum.

5   Summary

The Gadget software provides a "tool box" for constructing, running and optimising statistical marine ecosystem models. It is open source, and freely downloadable, with documentation and examples available online. The software consists of three parts; a biologically realistic parametric forward-simulation model; a suite of statistical functions to compare the

model output to data; an optimisation suite to adjust the parameters of the simulation model to give the best possible fit to the data.

Gadget provides the user with a choice of functional forms for the different biological processes within the model. The suite of statistical functions allows for a wide variety of different data sources to be incorporated into the modelling process in a statistically rigorous manner. There is a high degree of flexibility in the utilisation of different data sources, allowing for different time coverage and aggregation levels between the data sources. A hybrid optimisation algorithm is incorporated within Gadget, combining the wide-area searching capabilities of Simulated Annealing with the fast local convergence of Hooke & Jeeves. By linking the Gadget software with the custom designed data warehouse it is possible to produce new or modified Gadget models with minimum effort.

Gadget models have been, and are being, constructed for a variety of ecosystems, ranging from arctic cod to sub-tropical shrimp. The models have been used to investigate the biology of fish population and ecosystems as well as in an assessment context. The flexible framework provided by Gadget provides an important tool for generating statistically rigorous, biologically realistic, age and length structured models.

6   References

Anon. 2002. Development of Structurally Detailed Statistically Testable Models of Marine Populations (dst2). QLK5-CT1999-01609. Progress Report for 1 January to 31 December 2001. Marine Research Institute Report No. 87, Marine Research Institute, Reykjavik, Iceland.

Anon. 2003. Development of Structurally Detailed Statistically Testable Models of Marine Populations (dst2). QLK5-CT1999-01609. Progress Report for 1 January to 31 December 2002. Marine Research Institute Report No. 98, Marine Research Institute, Reykjavik, Iceland.

Anon. 2004. ACFM Report of the Working Group on Methods of Fish Stock Assessments, ICES CM 2004/D:03 Ref. G. 11-18 February 2004 Lisbon, Portugal.

Begley, J. 2004. Gadget User Guide, available online from the Gadget web page. `http://www.hafro.is/gadget`.

Bogstad, B., Hiis Hauge, K., and Ulltang, Ø. 1997. MULTSPEC - A Multi-Species Model for Fish and Marine Mammals in the Barents Sea. Journal of Northwest Atlantic Fisheries Science. vol 22: pp317-341.

Bogstad, B., Howell, D., and Åsnes, M.N. 2004. A closed life-cycle model for Northeast Arctic cod. ICES C.M.2004/K:26.

Corana, A., Marchesi, M., Martini, C., and Ridella, S. 1987, Minimising Multimodal Functions of Continuous Variables with the 'Simulated Annealing' Algorithm. ACM Transactions on Mathematical Software vol 13: pp262-280.

Frøysa, K. G., Bogstad, B., and Skagen, D. W. 2002. Fleksibest - an age-length structured fish stock assessment tool with application to Northeast Arctic cod (Gadus morhua L.). Fisheries Research vol 55: pp87-101.

Hannesson, S., Jakobsdóttir, A., Begley, J., Taylor, L.A., and Stefánsson, G. 2004. On the use of tagging data in statistical multi-species multi-area models of marine populations. ICES CM 2004/FF:30.

Hooke, R., and Jeeves, T.A. 1961. Direct Search Solution of Numerical and Statistical Problems. ACM Journal. vol 8: pp212-229.

Kupča, V. 2004. A standardised database for fisheries data. ICES CM 2004/FF:15.

Stefánsson, G., and Pálsson, Ó.K. 1997. BORMICON. A Boreal Migration and Consumption Model. Marine Research Institute Report 58, Reykjavik.

Taylor, L.A., Begley, J., Kupča, V., and Stefánsson, G. 2004. A Simple implementation of Gadget for cod in Icelandic waters. ICES CM 2004/FF:23.