ICES CM 2004/FF:22

Comparison of the effects of forced errors in survey data between an age and an age-and-length structured model of Northeast Arctic Cod

Daniel Howell, Institute of Marine Research, Bergen
Daniel Howell: Institute of Marine Research, P. O. Box 1870 Nordnes, N-5817 Bergen, Norway. tel: +47 55238500, fax: +47 55238681, e-mail: daniel.howell@imr.no

Keywords: Northeast Arctic Cod, appropriate complexity, population model, Gadget, XSA.

Abstract

It is known that many fisheries data sets contain several different sources of error. It is likely that models with different structures will, in general, be affected in different ways by these errors. It is therefore important to examine what those differences are, both in order to select models of "appropriate" complexity, and to interpret discrepancies in situations where multiple models are run on the same datasets.

This paper examines the Northeast Arctic Cod (*Gadus morhua*). A series of experiments have been conducted introducing additional, known, errors into some of the survey data sets. These errors have been selected in order to mimic problems known to occur in practice. The modified data sets have then been used as inputs into two different models. The two models are currently both run as part of the assessment process for Northeast Arctic Cod. One is a relatively simple, XSA age-structured model, and one a more complex biologically-detailed age-length based model constructed using Gadget. The effects on the model results of these known forced errors are then compared.

Introduction

It is increasingly the case that multiple different models, of different design and complexity, are available for application to a given problem. It may be that a single model must be selected, in which case it becomes a question of selecting the "appropriate" complexity for the situation at hand (e.g. Costanza and Sklar, 1985, Håkanson 1995). In other situations the approach taken is to run more than one model on the same fish stock, and compare the results. In both cases it is useful to know how the different possible models would react to different known, or anticipated, sources of error in the data. It can be expected that the response of different models will depend on the interaction between the fish stock dynamics, the type and magnitude of error involved, and the structure of the model employed. This paper represents an attempt to examine the response of models from two different classes, one age based (XSA) and one age-and-length based (Gadget). The level of complexity is different in the two models, with the age-and-length model employing a higher degree of complexity and biological realism (Anon 2003, Begley and Howell 2004). Gadget is also a much newer model, and it is therefore useful to compare the response of a Gadget model to known errors to that of the better known XSA model.

Data

The ideal approach to a problem of this kind would be to use simulated data sets. Error free, biologically realistic data, of known structure and properties, could be used as input into both models, producing an optimised reference solution. Specified errors could then be added to the data and the modelling repeated for each different, known, error. The differences between the models produced could then be studied, and related to the induced errors. However no such data set currently exists that can be used in this way. One could be generated using Gadget, but using this as input into a new Gadget model gives a situation where the model is well specified to the data, and this situation is unlikely to arise in practice. Furthermore having one model, but not the other, well specified to the data would undermine any comparisons that might be made.

The approach chosen here is therefore to use existing real-world data and induce additional errors. This approach has the disadvantage that the 'truth' in the population is never known. Thus it is only possible to analyse the difference between several different error prone situations, where only a part error structure is known. However it allows for experiments to be conducted in such a way that they correspond to real world situations and problems. For instance a situation where additional error is introduced into a single data series (e.g. discarding affecting the reported catch in length), or a single or small number of years (e.g. an anomalously high survey index in one year) is known to occur in real-world data. The existence of this error is often well known, even if the details are not always understood. It is possible to replicate this situation, and analyse and contrast the effects of different possible errors on several different models.

The case study chosen here is that for the Northeast Arctic cod. A large, detailed collection of data sets is available covering this stock, some of them covering a long time period. The current practice is for the Arctic Fisheries Working Group covering this stock to run several different models during the assessment process. This therefore makes an ideal setting to compare the differing responses of the different models, as well as ensuring that the results of such an investigation will be relevant in practice.

Models

The experiments have been conducted using two different models for the Northeast Arctic cod. The XSA model used during the AFWG (ICES 2004b), and a variant of the Gadget closed life-cycle cod model described in this volume (Fræysa et. al. 2002, Bogstad et. al. 2004). Both of these models were used at the Arctic Fisheries Working Group in 2004 (ICES 2004b). The two models employ very different methodologies. Gadget conducts a forward simulation of the population dynamics of a stock, and then uses all available data sets to optimize the parameters of the simulation model. XSA is a variant of the Virtual Population Analysis (VPA) model (Darby and Flatman, 1994), and is a backwards simulation. The XSA program "tunes" (optimizes) the solution to fit the supplied fleet data sets. The level of complexity in the two models is also different, with Gadget being more complex, and including a higher degree of biological realism than XSA. Both models use a range of data sets, but the current work concentrates on only one. In this preliminary analysis a single survey has been selected for these experiments. The survey chosen is Joint Norwegian-

Russian winter survey on the Northeast Arctic cod (Jakobsen et. al. 1997). The survey is conducted in January-March (mostly in February) of each year, and is targeted at immature fish. Although there have been changes in gear and area coverage over time, the survey has remained unchanged since 1994, the start of the time period examined in this paper. This survey is used in the AFWG XSA cod model as a tuning fleet, and as one of the data sets used during optimisation of the Gadget model. In both cases the survey is an important, but not overwhelming, source of data to the model.

The Gadget cod model employed here is that presented at the AFWG in 2004 (ICES 2004b), and the same as the closed life cycle model (Bogstad et. al. 2004) except that a value for recruitment of age one fish is estimated for each year rather than a spawning relationship being used. This was done for simplicity, and to avoid using a new, experimental, model for the comparisons undertaken here. The model was run from 1985 to 2004, with the period 1994-2003 examined here. The model considers immature and mature cod, with length-based growth, maturation and cannibalism-induced mortality. Residual natural mortality is modelled as an age based process. Fishing is conducted by two different aggregated commercial fleets (one trawl and one gill net), each with it's own length selectivity. The following data sets from the Barents Sea are used by the model; joint winter bottom trawl survey; joint winter acoustic survey; Lofoten acoustic survey; Russian bottom trawl survey; International 0-group cod survey; commercial catch in numbers, tons, and by length; Capelin abundance estimate. The model deals with the change in gear size in the Joint winter survey by splitting the affected surveys into two separate components, and fitting to each separately. XSA is the main assessment model used for Northeast Arctic Cod at the Arctic Fisheries Working Group (ICES 2004b), and is largely based on the same data sets. The model does not correct for the changes in mesh size in the Joint Winter survey in 1994, but the period with a larger mesh size is before that examined here.

Experiments

A general investigation of the responses of different classes of models to different possible data errors is needed. This study represents a start on this work, and concentrates on a single source of errors. The error considered here is that of a single year with an anomalously high survey index result. Such an occurrence is relatively common in real-world data series (e.g. ICES 2003), and may have a significant effect on the modelled population. It is not clear what the exact effects in a Gadget model would be, or how these would compare to the effects of errors on a VPA/XSA type model. In particular it would be useful to know if errors in some years would produce a more serious effect on the modelled population than others, and what the dynamics of the response are. An experiment was therefore conducted in which the survey index for a single year was artificially inflated by 50% and the model optimised. This was repeated for each year in the model separately, and the results analysed. The whole process was conducted on two both models, and the results compared.

Diagnostics

Within the Gadget model a weighted likelihood score is produced, and is used during optimisation. However this is not a suitable measure to use to compare the experiments conducted here. Introducing an error in the data will result in an increased likelihood score. Because some years have more data than others this increase will vary between years, as a result of the data structure, even before the effects on the modelled population are considered. Equally a variety of residuals in the XSA model can be studied, or residual plots produced, but not all of these can be directly compared with the results of the Gadget model. Correlations between the XSA results and various survey indices can be investigated, but the induced change in the main survey makes interpreting these results difficult. The two models produce different levels of detail and complexity n their outputs. As a result of it's greater structural complexity Gadget is capable of outputting a higher level of detail the XSA model. However both models produce directly comparable outputs on the overall stock trends in numbers, biomass, and fishing pressure. It has therefore been decided to concentrate on the effects on the final modelled population, and in particular the biomass of the spawning and total stocks. This ensures a realistic comparison of the key outputs of the two models, and investigates the effects of the induced errors on the most important model output from a practical fisheries management perspective. For an error in the survey in a given year the biomass throughout the simulation can be examined. Graphs can be produced for data errors in different years, allowing for comparisons between the different experiments conducted here. It should be noted that the aim here is not to identify the added errors in the data and adjust for them; rather it is to see how the two models respond to those errors.

Results

The reference runs for the gadget and XSA models are shown in Figures 1 and 2. The numbers of the spawning stock and the 3+ cod (Figure 2) and the biomass of spawning and 3+ fish (Figure 3) at January 1st of each year are shown for the duration of the simulation. Both models show the same population trends, and have similar sized stocks in both numbers and biomass. Total and spawning stock biomass and numbers are high for most of the 1990s under both models, declining to a low in 1999 and 2000, followed by a rise in the present decade. It can be seen that the two models are in good agreement for biomass (Figure 3), with the exception that the current model suggests that the high stock levels in the mid 1990s resulted in an even higher spawning stock biomass than that predicted by the XSA. The models are also in good agreement for stock numbers through time (Figure 2).

The results of the experiments are presented in Figures 3-6. Each line represents a complete model simulation conducted with a 50% in the winter survey values in the relevant year. The variation in results due to the induced extra errors is smaller than the year-to-year variation in stock size during the simulation. The results of the experiments are therefore presented as deviations from the standard reference runs for each model.

XSA results

The biomass of the total stock and spawning stock through time for each of the experiments are shown in Figures 3 and 4. Both the spawning stock and the total stock biomass show the same trends, though with slightly different percentage changes from the reference run. In all cases the maximum discrepancy, positive or negative, is highest in the last year of the model run (2003). The addition of errors in any years of the winter survey tuning fleet produces very little effect on the population size in the early part of the model (prior to 1998-1999). This is a result of the nature of VPA/XSA models, which have strong convergence in the early part of the time series, with the greatest potential for variation in the later years. It can be seen that increasing the magnitude of the winter survey in 2001, 2002 and 2003 produce increased stock biomasses in the later part of the run (1999-2003). Higher values in the survey for years before 1999 produce a slight decrease in the modelled biomass in the latter part of the model. This is due to the fact that the survey is acting as a survey index, where an increase in one year is the same as a decrease in all other years. An increase in an early year therefore produces an apparent downward trend in population size in later years.

Gadget results

The response of the gadget cod model to the data errors is markedly different to that in the XSA model. Adjusting the 1994 survey produces significantly different results from all the other years. This is the first year of the winter survey, and it is clear from Figures 5 and 6 than adjustments in this first year can have a noticeable impact on the overall population model.

For the total stock a clear pattern can be seen in the period 1994-1999. The modelled population is increased in the year of the data error, and this increase may persist at a lower level the following year. This is compensated for with a slight decrease in the stock prior to, and following the increase. However this trends breaks down in 2000, when stock levels reached a low value (Figure 5). It is likely that other information (from the other surveys and the catch) indicated at a stock size any higher than that predicted by the reference run would not be realistic, and this was therefore rejected during the optimisation procedure.

As with the XSA model and increase in the survey index in an early year results in a downward trend in stock sizes in the final years of the survey. An increase in the later years (2001, 2002 and 2003) produces a rise in stock numbers in the final years of the simulation, although they do not have any clear effect on the stock in the early part of the simulation.

The impact on the spawning stock is much less obvious and clear cut (Figure 7). The Joint Norwegian Russian Winter survey focuses on the immature portion of the stock. Where the error occurs in a year with a relatively large year class (1995, 1996, 1997) the increase in the recruitment for that year produces a higher number of mature fish overall. It is likely that the model is seeking a compromise in the proportion of larger fish between the year of the induced error, and the unaltered years.

Comparison

Although the two models examined here have very different structure and levels of complexity, they show similar responses in the terminal years of a simulation. An artificial increase in the survey value in the last few years produces higher predicted total stock biomasses in the last years of the model run. In both models an increase of 50% in the single survey studied here produced increases of up to 8% in the estimates for the last year of the model run. Conversely an increase in an early year produced a reduction in the predicted total biomass for the most recent years. This decrease was more marked in gadget (up to 9%) than in the XSA model (2%). The XSA model produces almost identical responses for total and spawning stock biomass, even though the affected survey is targeted at immature fish. The Gadget model, with its ability to model the maturation process, gives different responses for immature and mature fish. The Gadget model also shows a dynamic process throughout the simulation, with the greatest response being concentrated around the year of the induced error. The XSA model, in contrast, shows the greatest response in the last year of model time.

Mitigating data errors in a Gadget model

In this paper the effects of data errors on model results have been examined. In practice some of these errors will be identified and dealt with during the development and testing of the models. Because Gadget provides a great deal of flexibility in how data is used during optimisation, there are more possibilities for handling suspect data than simply excluding it.

An erroneous year in the survey index data can have an impact on the simulated population produced from a Gadget model. Because of the way Gadget creates a simulation through time the data error can effect the whole simulation, not just the years around the error. In general the effects on the model are fairly small, although they can be large enough to distort or obscure the actual trends occurring in the modelled stock. Using multiple surveys can reduce the errors introduced into the modelled population, as the model will attempt to find a solution best fitting the whole suite of data sets used. However there are situations in which a number of different survey indices can all give anomalous results in a single year (ICES 2002), in such a case having multiple surveys will not reduce the effects of the error.

If it is suspected that a problem has occurred with collection of survey data there are several possible remedies. The data may be used "as is", especially if there are other data sources unaffected by the error, and the affected data source contributes only a small amount to the overall likelihood score. The year of data may simply be excluded if it is suspected that the relative frequency of different classes has been affected. However if the problem is with the level, but not the length structure, of the data then placing that year's data in a separate likelihood component may remove the problem. The distribution of length classes within the year will be preserved, but the year will make no direct impact on any long-term trends within the model.

Summary and Conclusions

The preliminary work conducted here indicates that different categories of models currently used in fisheries assessments respond to data errors in fundamentally different ways. A VPA/XSA type model produces responses in the last part of the time series, and responds to increased survey index values by increasing the terminal population for increases in later survey years, and slightly decreasing it for high survey index values in early years. Altering a single tuning survey in this way has almost no effect on the predicted population in the earlier part of the model run. Gadget responds to a single year error by increasing the population size in that year, if the overall stock dynamics permit such an increase. The increase persists, at lower levels, for several years after the errors. Earlier and later years show a slight compensatory reduction in stock biomass.

In a number of situations more than one model, or class of models, is run on a stock. It is therefore useful to know the likely response of each class of model to different known or suspected errors in the data.

Further work

The most obvious extension of this work is to examine random errors in the data. By using multiple replicate data sets, each with an additional random error component, the differing response of the two models to such situations can be examined. The errors could be purely random and unbiased in nature, or they could add bias to reflect processes such as discarding or misreporting of catches. Only one of the data sets has been modelled here. The work should be extended to examine different surveys and the commercial catch data. It may be that the different models exhibit different degrees of sensitivity to errors in different data sets. Other classes of models exist (e.g. Huse and Ottersen 2003), and where these are used in an assessment context it would be valuable to extend the methodology described here to incorporate as many different models as possible.

The gadget model employed here considers cod of age 1+, but does not include a closed life cycle. Instead a recruitment value is estimated for each year. A comparison between the closed and non-closed life cycle gadget models could be run to examine the effects of the different dynamics of the two model formulations.

References

Anon 2003. Development of structurally detailed statistically testable models of marine populations. Pp 135-160. QLK-CT1999-01609

Begley, J., and Howell, D. 2004 An overview of Gadget, the Globally applicable Area-Disaggregated General Ecosystem Toolbox, Vigo 22-25 September. ICES C.M. 2004/FF:13

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

Costanza, R. and Sklar, F. H., 1985. Articulation, accuracy and effectiveness of mathematical models: a review of freshwater wetland applications. Ecological Modelling, 27: 45 – 68

Darby, C.D. & Flatman, S. 1994. Virtual Population Analysis: Version 3.1 (Windows/DOS) *User Guide*. Information Technology Series, MAFF Directorate of Fisheries Research, Lowestoft, (1): 85 pp.

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 55: 87-101.

Huse, G., and Ottersen, G. 2003. Forecasting recruitment and stock biomass of Northeast Arctic cod using neural networks. Sci. Mar., 67 (suppl. 1): 325-335.

Håkanson L., 1995. Optimal size of predictive models. Ecological Modelling, 78: 195-204.

ICES 2002. Report of the working group for Norwegian spring-spawning herring, October 2002. ICES CM 2002/ACFM:19

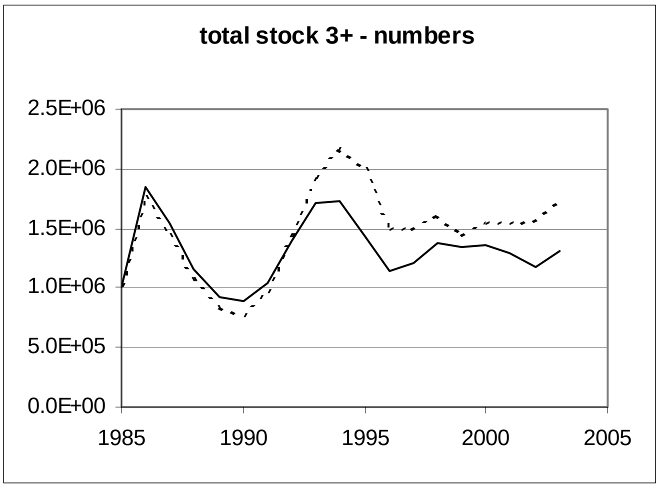
ICES 2004a. Report of the Study Group on Growth, Maturity and Condition in Stock Projections. Aberdeen, Scotland, 19-23 January 2004. ICES C.M. 2004/D:02, 66 pp.

ICES 2004b. Report of the Arctic Fisheries Working Group, Copenhagen 4-13 May 2004. ICES C. M. 2004/ACFM:28, 475 pp.

Jakobsen, T., Korsbrekke, K., Mehl, S., and Nakken, O. 1997. Norwegian combined acoustic and bottom trawl surveys for demersal fish in the Barents Sea during winter. ICES CM 1997/Y:17.

Figures

a)



b)

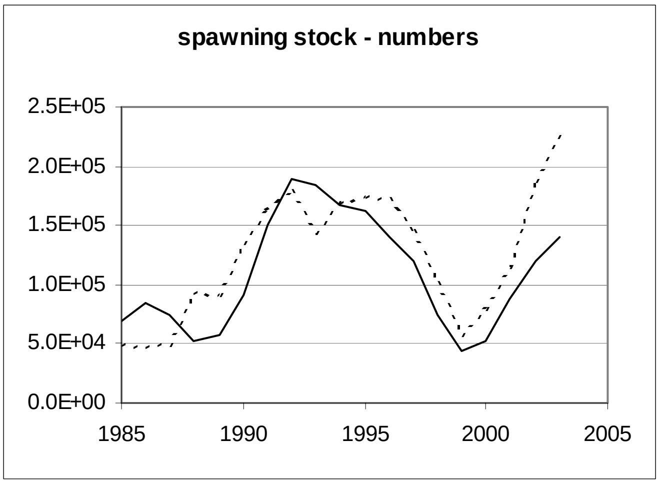
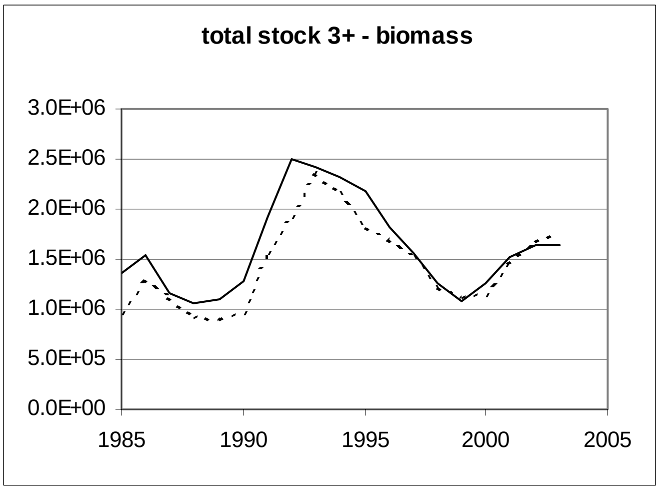


Figure 1. Number of fish of age 3+ (a) and in the spawning stock (b), in million fish for the references (unaltered) runs. Solid line indicates the Gadget model, dotted line is the 2004 AFWG XSA assessment.

a)



b)

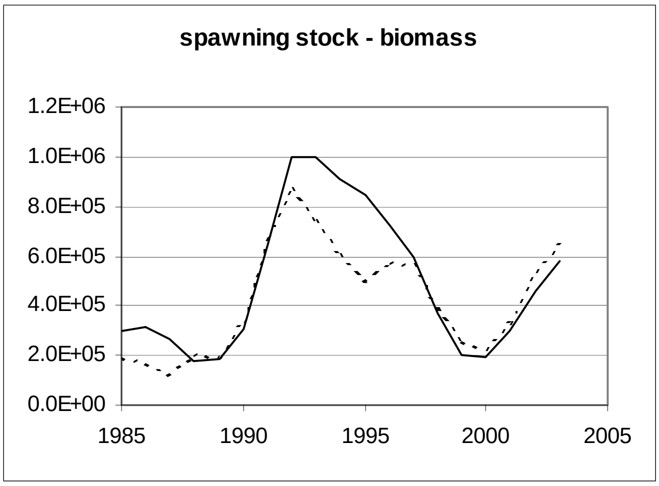


Figure 2. Total biomass fish of age 3+ (a) and in the spawning stock (b), in tonnes for the reference runs, with no added errors. Solid line indicates the Gadget model, dotted line is the XSA model.

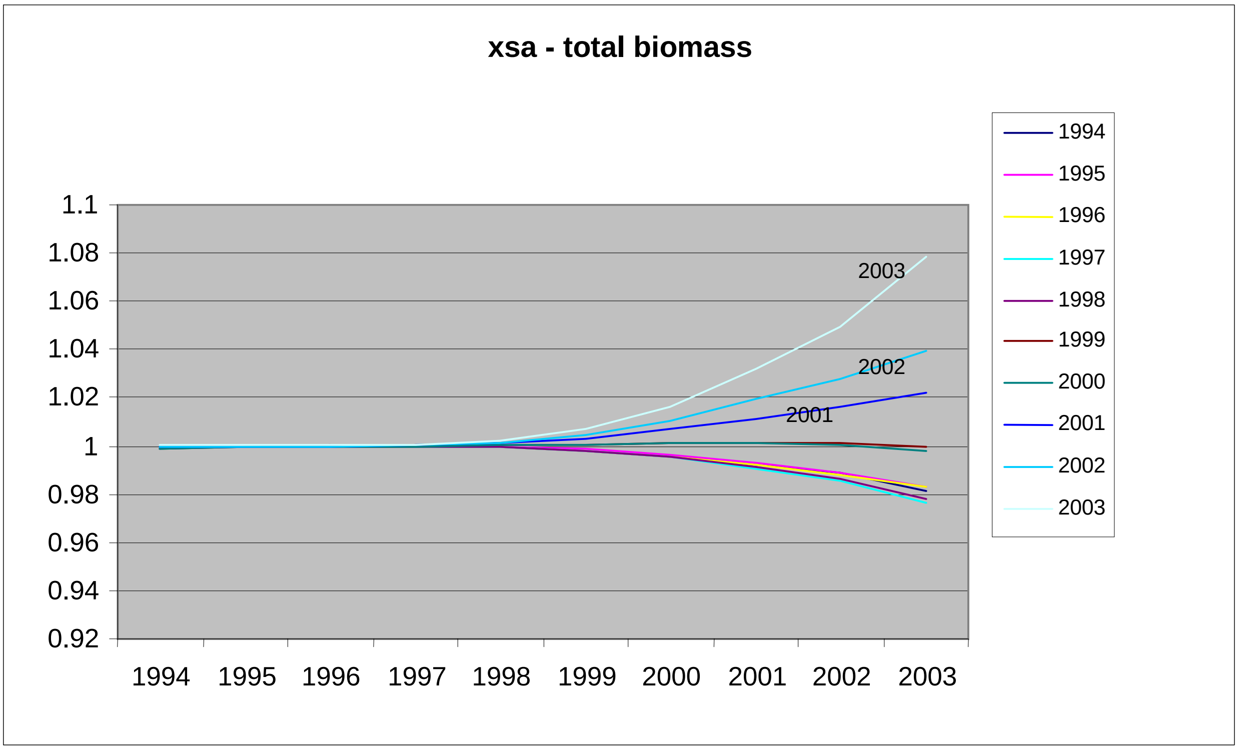


Figure 3. Variation from reference run biomass for the total stock through time for the XSA model. Each line represents a separate simulation, with a 50% increase in the Joint winter survey in the year stated.

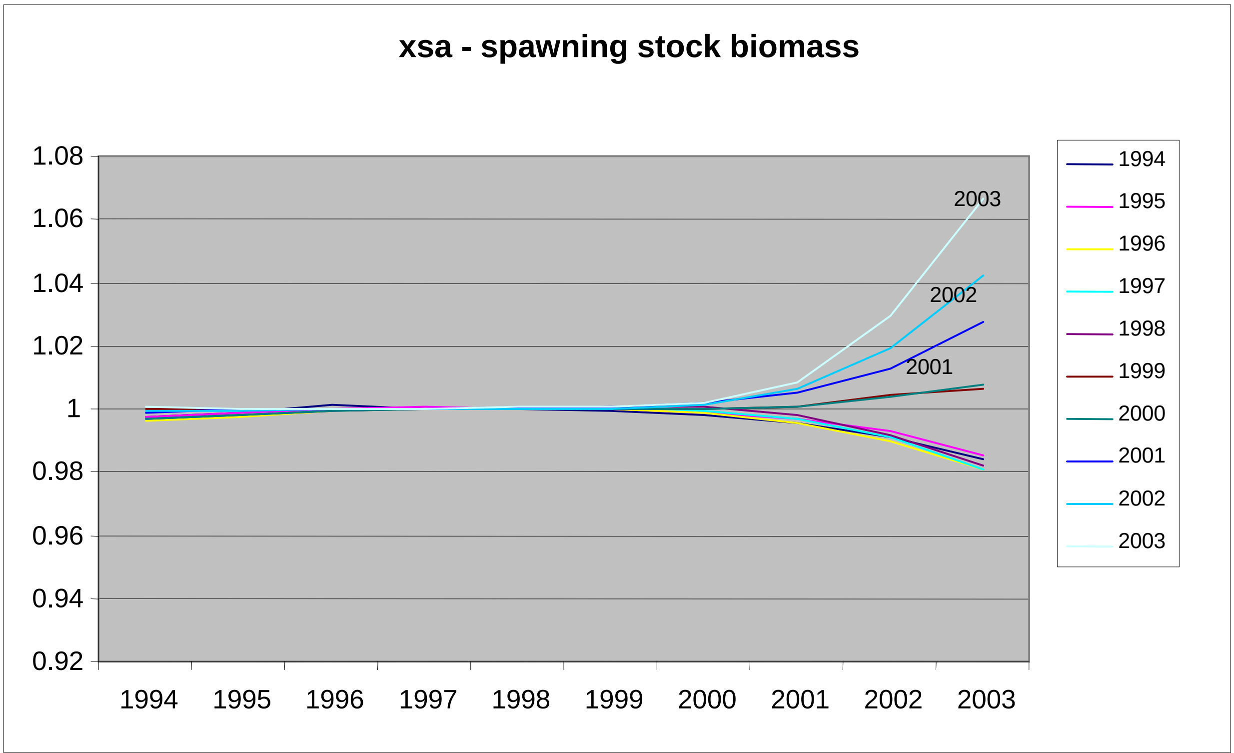


Figure 4. Variation from reference run biomass for the spawning stock through time for the XSA model. Each line represents a separate simulation, with a 50% increase in the Joint winter survey in the year stated.

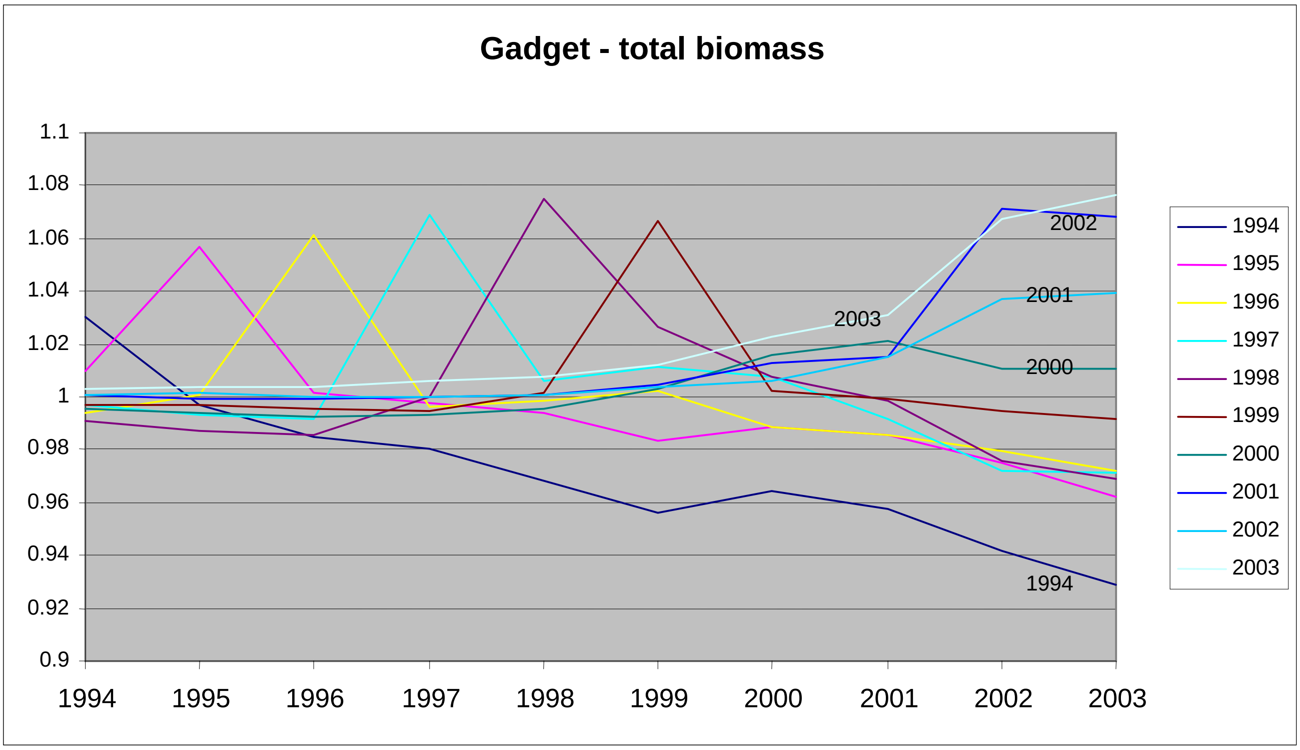


Figure 5. Variation from reference run biomass for the spawning stock through time for the Gadget model. Each line represents a separate simulation, with a 50% increase in the Joint winter survey in the year stated.

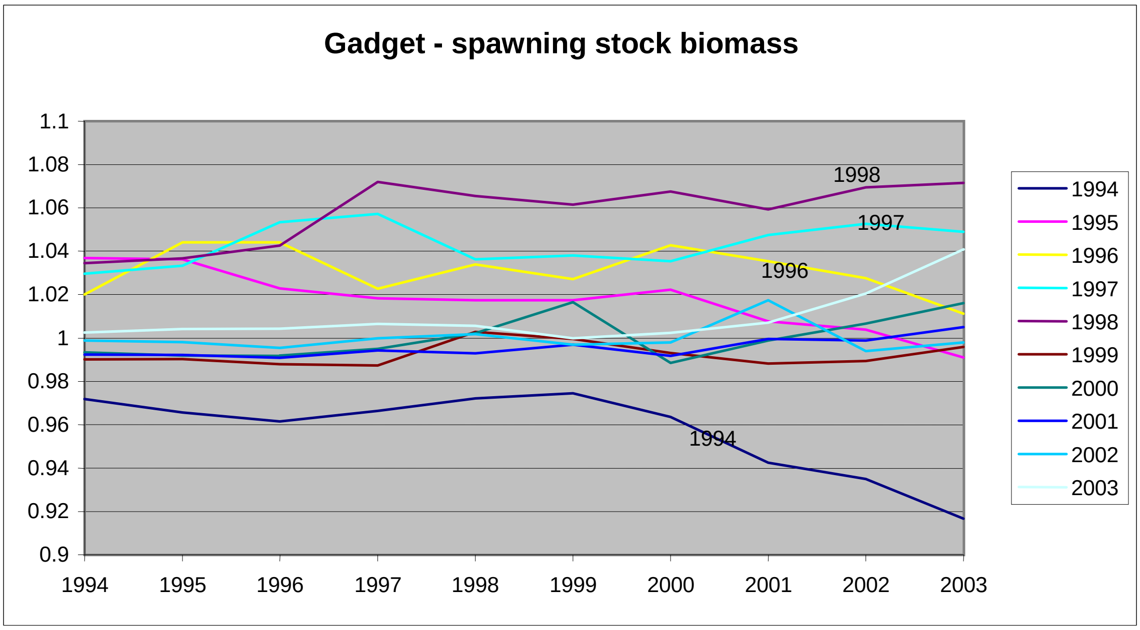


Figure 6. Variation from reference run biomass for the spawning stock through time for the Gadget model. Each line represents a separate simulation, with a 50% increase in the Joint winter survey in the year stated.