SCRS/00/104

Col. Vol. Sci. Pap. ICCAT, 52 (3) : 1094-1115 (2001)

EVALUATING THE RELATIVE MERITS OF ALTERNATIVE METHODS TO WEIGHT DIFFERENT TIME SERIES OF ABUNDANCE INDICES IN STOCK ASSESSMENT

Murdoch McAllister[1], Elizabeth A. Babcock[2] and Ellen K. Pikitch[2]

SUMMARY

*When more than one time series of abundance indices is used in an assessment model such as a surplus production model or VPA, it is necessary to determine the relative weights of the various series and observations in the series. Each weight is typically represented by the inverse of a parameter for the variance of the deviation between each model predicted and time series observation. The weights may be either fixed beforehand or estimated. This assigning of weights has long been a contentious issue in many stock assessments because there is no universally accepted protocol and the resulting policy recommendations can depend strongly on the weighting method applied. This paper reviews and reformulates the key methodological considerations in the assigning of weights, summarizes the various pros and cons of some alternative weighting methods and attempts to identify alternatives that most adequately address the various considerations. This paper evaluates the relative merits and potential effects on stock assessment results of applying nine alternative methods to weight different abundance series. For simplicity, the numerical evaluation will be undertaken by fitting a simple surplus production model to the catch rate data used in the 1998 assessment of western Atlantic bluefin tuna (*Thunnus thynnus*). Because other stock assessment methods used by ICCAT, such as ADAPT VPA, rely heavily on the tuning of modeled trends in abundance to the available abundance series data and their relative weightings, it should be possible to readily generalize the results to these other methods.*

RÉSUMÉ

Lorsque l'on utilise plus d'une série temporelle d'indices d'abondance dans un modèle d'évaluation tel que le modèle de production excédentaire ou la VPA, il est nécessaire de déterminer le poids relatif des diverses séries et des observations dans les séries. Chaque poids est typiquement illustré par l'inverse d'un paramètre pour la variance de la déviation entre chaque modèle prévu et l'observation des séries temporelles. Les poids peuvent être fixés d'avance ou estimés. Cette attribution de poids est depuis longtemps une question débattue dans de nombreuses évaluations de stock, du fait qu'il n'y a pas de protocole universellement admis, et que les normes recommandées qui en découlent peuvent dépendre à un degré important de la méthode de pondération qui est appliquée. Le présent document examine et formule de nouveau les considérations clés sur la méthodologie de l'attribution de poids, récapitule les divers avantages et inconvénients de quelques alternatives de pondération et trente de définir quelles sont les alternatives qui traitent le mieux des diverses considérations. Le document évalue les mérites relatifs, et les effets potentiels sur les résultats des évaluations de stock, de neuf alternatives méthodologiques pour la pondération de différentes séries d'abondance. Pour simplifier les choses, l'évaluation numérique sera entreprise en ajustant un simple modèle de production excédentaire au taux de capture utilisé dans l'évaluation de 1998 du thon rouge (Thunnus thynnus) *de l'Atlantique. Du fait que d'autres méthodes d'évaluation de stock utilisées par l'ICCAT, telles que la VPA ADAPT, dépendent beaucoup du calibrage des tendances modélisées de l'abondance aux séries disponibles de données sur l'abondance et leur poids relatif, il devrait être possible d'étendre aisément les résultats à ces autres méthodes.*

[1] T. H. Huxley School of Environment, Earth Sciences and Engineering, Imperial College, Rm 402 RSM Building Prince Consort Road, London, SW7 2BP UK
[2] Wildlife Conservation Society, Marine Conservation Program, 2300 Southern Blvd., Bronx, NY 10460 USA

RESUMEN

Cuando se usa más de una serie temporal de índices de abundancia en un modelo de evaluación, como el modelo de producción excedente o el VPA, es necesario determinar cuales son los pesos relativos de las diversas series y observaciones en las series. Cada peso relativo está típicamente representado por la inversa de un parámetro para la varianza de la desviación entre cada predicción y observación sobre serie temporal del modelo. Los pesos relativos pueden ser fijados de antemano o bien estimados. Esta asignación de pesos relativos viene siendo desde hace tiempo un contencioso en muchas evaluaciones de stock, ya que no hay un protocolo universalmente aceptado y la política resultante respecto a recomendaciones puede depender mucho del método de ponderación aplicado. Este documento examina y hace una nueva formulación de las consideraciones metodológicas clave en la asignación de pesos relativos, resume las ventajas y desventajas de algunos métodos de ponderación alternativos e intenta identificar las alternativas mas adecuadas para tratar las diferentes consideraciones. El documento evalúa las ventajas relativas y las posibles repercusiones que sobre los resultados de la evaluación de stock tendría la aplicación de nueve métodos alternativos para ponderar diferentes series de abundancia. Para simplificar, la evaluación numérica se hará ajustando un sencillo modelo de producción excedente a los datos de tasa de captura empleados en la evaluación de stock de atún rojo (Thunnus thynnus) *del Atlántico oeste hecha en 1998. Debido a que otros métodos de evaluación de stock empleados por ICCAT, tales como el ADAPT VPA, se basan sobre todo en el ajuste de las tendencias modeladas de la abundancia a las series de datos de abundancia disponibles y sus pesos relativos, debería ser posible hacer una aplicación general de los resultados a estos otros métodos.*

KEYWORDS

Accuracy, mathematical models, numerical analysis, stock assessment, recruitment, long-term changes, fishery management, Bayesian methods

INTRODUCTION

When more than one time series of abundance indices is used in an assessment model such as a surplus production model or VPA, it is necessary to determine the relative weights of the various series and observations in the series. Each weight is typically represented by the inverse of a parameter for the variance of the deviation between each model predicted and time series observation and may be either fixed beforehand or estimated.

The assigning of these weights has long been a contentious issue in many stock assessments (Schnute and Hilborn 1993; ICCAT 1999; Geromont and Butterworth 2000). This is especially because there is no universally accepted protocol for it and because the resulting policy recommendations can depend strongly on the weighting method applied. This paper reviews and reformulates the key methodological considerations in the assigning of weights, summarizes the various pros and cons of nine alternative weighting methods and attempts to identify alternatives that most adequately address the various considerations.

This paper also evaluates potential effects on stock assessment results of applying each of several alternative methods to weight different abundance series. Some of these have been proposed and applied in recent ICCAT assessments and meetings (e.g., ICCAT 1999; Geromont and Butterworth 2000). For simplicity, the numerical evaluation will be undertaken by fitting a simple surplus production model to the catch rate data used in the 1998 assessment of western Atlantic bluefin tuna. Because other stock assessment methods used by ICCAT rely heavily on the tuning of modeled trends in abundance to the available abundance series data and their relative weightings, the results should be readily generalizable to these other methods. The input and output weightings, estimates of the trends in abundance, the estimates of key model parameters and their confidence intervals, and the model deviance (goodness of fit (GOF)) provided by each weighting scheme are compared between the different weighting methods.

Some general methodological considerations for assigning weights to different data series

In this paper we define "weights" or "weightings" for different data points or time series as the inverse of the variance in deviations between the natural logarithms of observed and model-predicted data points. This would assume that the likelihood function of the data is lognormal (see below). If some other likelihood function were applied, then the weightings could be defined in terms of the inverse of the square of the coefficients of variation (CVs) of these deviations without taking the natural logarithms (mathematically very similar to the above). These quantities are intuitive because they effectively weight the deviations between observations and model predictions in the estimation of stock assessment model parameters. The larger the variance or CV, the lower the weight of a data point and data series.

If the different series were reasonably long, e.g., at least 15 years, and obtained using unbiased and consistently applied sampling methods, and the stock assessment model was accurately specified, then the variances for each time series could be estimated by fitting the model to all of the combined data. However, this is almost never the case. In many different stock assessments, due to many imperfections in sampling design and the application of fishery dependent abundance indices, there is large uncertainty over whether the time series can provide an unbiased estimate of the trend in abundance of the part of the population it is supposed to track.

Additionally, one of the largest uncertainties in stock assessment is not over the values of model parameters, but over the structure of the stock assessment model itself. For example, one of the most common uncertainties involves how to model the catching power of fishing fleets. With such structural uncertainties, the issue of weighting becomes even more important since an inappropriate choice of weightings can potentially amplify the biases in the model structures applied. Relying on estimation to assign weights could give too much weight to more precise fishery dependent indices that do not accurately account for temporal increases in catchability and too little to imprecise but unbiased fishery independent ones, in a stock assessment model that incorrectly assumes that catchability is constant over time.

Even if there is only one time series of abundance, the matter of weighting different observations is not trivial. For example, if sufficiently detailed information exists, coefficients of variation (CVs) derived from the analysis of research survey or CPUE data may be provided for individual observations. If some of these are very large and some are very small there is no universal protocol on how these should be used in parameter estimation, if at all, when a stock assessment model is fitted to the time series (Geromont and Butterworth 2000).

The problem of agreeing on a method to weight different data series and data points in a stock assessment can become most severe when different methods lead to different policy recommendations. This can occur prominently when different time series for supposedly the same segment of the assessed population suggest opposing trends in abundance. It may occur more subtly when there are several different time series covering different age groupings of the same population. But the end result is the same: different methods support different management policy options as in the 1998 assessment of western Atlantic bluefin tuna (ICCAT 1999).

Some nagging questions that arise regarding this problem include the following.

- *How can different weighting methods give rise to such different results in the first place?*

The answer to this question is that there are major mechanical differences between the methods that give rise to differences in estimation results when different time series are not entirely consistent in what their trends suggest about population dynamics. We try to illustrate how these can occur below in the example.

- *Why have scientists so frequently differed so strongly on which weighting method to apply?*

The answer to this question is complex and there are many different factors involved. However, scientists should agree that their choice of a method should **not** be made to depend on the particular result that it favours. Instead, it is likely that in each stock assessment, there are a variety of objective scientific considerations that can be addressed such that the choice of a method to apply is based mainly on the overall scientific credibility of the method.

- *What can be done to provide a more objective and scientific basis to help scientists come more easily to an agreement on which weighting method to apply?*

Until now little has been done to address this question. Due to this lack of guidance, it is predictable that strong differences among scientists will continue to occur regarding this issue. To attempt to provide some headway in improving the scientific basis for choosing a method for weighting different time series in stock assessment, this latter question is the focus of this paper. A list of considerations for choosing a weighting method is outlined in the following section.

Some general steps for the objective choice of a weighting method

While the choice of a weighting method should be made before running the stock assessment model, there are some additional considerations for the application of a weighting method to see that it is applied sensibly and behaves sensibly once applied. The first four steps are ones to be taken before evaluating the stock assessment data. The latter four steps deal with the treatment of different data series once the weighting method has been chosen and diagnostics to test for anomalous estimation behaviour that might lead to the reconsideration of a weighting method.

Step 1: Agree to adopt a method for weighting before viewing the stock assessment results

The identification of potential methods to apply to assign weightings to different data series should be based on objective scientific considerations before the stock assessment model is fitted to the data. The particular stock assessment results obtained from the different weighting methods, i.e., whether they are more or less conservative, should **not** be used to determine which weighting method to adopt for providing policy advice. However, anomalous estimation behaviour resulting from the application of a candidate weighting method (see below) would require that it be rejected and/ or modified.

Step 2: Identify the alternative weighting methods that ***could*** *be applied*

Based on previous stock assessments of the same stock and experience of the scientists present, identify several potential alternative methods to assign weightings for the alternative methods.

Step 3: Identify the relative scientific merits of each alternative weighting method

Perhaps the most important overall consideration is how likely each method is to avoid bias in the estimation of trends in abundance. For example, if there is a strong objective evidence suggesting that some time series are less biased than others, the method should permit this to be accounted for and prevent such time series from being down-weighted. Such down-weighting could occur for example from a mis-specification of stock assessment model structure as mentioned above. The method should allow the incorporation of, e.g., GLM, estimates of precision in data points in each of the time series that were obtained before running the stock assessment model. The method should be well-suited to the amount of data available – for example, if it requires the estimation of variances for each time series, the time series should be long enough for this. Also, each method should be expected to give an accurate assessment of the uncertainty in the stock assessment, in terms of uncertainty over the values of stock assessment model parameters. This could be possible if the method allowed the time series variances to be adjusted depending on the goodness of fit (GOF) of the model to the observations. The method should avoid anomalous estimation behaviour, for example, inadvertently placing most of the weight on one of several different time series, especially when expert judgment would suggest otherwise. Perhaps the most objective approach, if time were available, would be to simulation test each alternative method to evaluate the potential bias and precision in stock assessment model parameter estimates that might result from each alternative weighting method. See Appendix 1 for a more detailed outline of these suggested criteria. A check list of these is provided below.

Checklist of criteria for the choice of a weighting method

(i) *Does the method incorporate GLM or other estimates of the relative reliability of data points for a particular data series?*

(ii) *Does the method avoid according unrealistically high precision (low variance) and high weight to some data series?*

(iii) *If empirically determined input variances are applied in the weighting method, can the method allow for readjustments of these weightings based on the goodness of fit of the model to the data*?

(iv) *Does the weighting method allow the incorporation of expert judgement on the relative reliability of each series as an index of abundance?*

(v) *Does the weighting method prevent the most reliable data sets from being down weighted?*

(vi) *Is the weighting method conducive to there being relatively few data points in the time series?*

(vii) *Could the weighting method be expected to result in an accurate assessment of parameter uncertainty?*

(vii) *Is the method easily put into practice?*

Step 4: From the above considerations, choose the method with the highest overall scientific merits.

The reasons for choosing one particular weighting method need to be reported in the stock assessment document. It may not be possible to eliminate all of the other candidate methods. However, this is a clear instance where it is desirable to adopt a baseline method for weighting alternative data series. Without one, the choice of the weighting method is then passed on to fishery managers who are often less well-qualified to make this choice.

Step 5: Based on knowledge about each data series, rank the different series according to their reliability as unbiased indices of trends in abundance.

See Appendix 1, criteria (4) for some suggestions about how this might be done objectively. The weighting method chosen should not provide results that contradict this ranking. If it did, then this would suggest that the weighting method chosen and possibly the stock assessment model structure and assumptions adopted need to be reconsidered.

Step 6: Identify and segregate sets of data series that suggest contradictory trends in abundance.

It is common in stock assessment for different time series to suggest contradictory trends in abundance. Several have pointed out that stock assessment methodologies that average across data series that suggest contradictory trends in abundance should be avoided (Richards 1991; Schnute and Hilborn 1993; Punt and Hilborn 1997). Before applying a stock assessment method to different input data, it is thus important to identify which different sets of data, if any, suggest contradictory trends in abundance. One approach to identifying such data is to run the stock assessment separately on each individual time series and to see whether the confidence intervals on trends in abundance overlap. If they do not, then this would indicate contradictory data.

Step 7. Apply the weighting method chosen and stock assessment method only to data sets that appear to indicate non-contradictory trends in abundance, unless the statistical method is specifically designed to deal with contradictory data.

Some have suggested statistical methods that incorporate such data in a way that accounts for contradictory trends in abundance, for example, by producing bimodal probability distributions for trends in abundance (Schnute and Hilborn 1993; McAllister et al. 1999). If there is not time to develop and apply such methods, it has been suggested that the stock assessment be conducted separately on the opposing sets of data and that the results be treated as separate hypotheses. However, the issue of how stock assessment scientists should weight such alternative hypotheses before presenting them to managers would still need to be addressed (Butterworth et al. 1996; McAllister et al. 1999).

Step 8: Check the goodness of fit (GOF) of the stock assessment model to the alternative time series. After the stock assessment model has been fitted to the data, it is always sensible to evaluate the GOF of the model to the data. For example, is there strong autocorrelation in model residuals? If there is, this would suggest model mis-specification. If there was little that could be done in the immediate stock assessment to address the possibility of model misspecification, then it would be important to evaluate the GOF of the model to each of the different time series. The use of statistics for model deviance, such as chi-square tests, would indicate time series with poor GOF. If the time series deemed to be the most credible as indices of trends in abundance had the poorest GOF, then this would suggest that the weighting method applied needs to be reviewed. Also, if the model parameters estimates are at or near their boundary conditions, then the model may be mis-specified and the weighting method may need to be reviewed.

METHODS

Overview of some alternative methods to assign weights to different time series

A large variety of methods have been suggested and applied in stock assessment to assign weightings to alternative data series. Nine alternatives are listed below and the resulting $\log_e$ of the likelihood functions (Ln*L*) provided. Unless otherwise specified q was estimated as:

$$\hat{q}_j = \exp\left(\frac{1}{n}\sum_y \ln(I_{j,y}) - \ln(\hat{B}_y)\right)$$

where j refers to the subscript for the series and y to that for years, $I_{j,y}$ is the observed index of abundance for series j in year y, $\hat{q}_j$ is the model predicted constant of proportionality for time series j, and $\hat{B}_y$ is the model predicted abundance in year y

Method 1. No weighting or inputted equal weighting (s^2) (used in VPA)

$$\ln L = -\sum_j \sum_y \frac{\left[\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right]^2}{2s^2}$$

The value for σ^2 is often chosen based on previous experience, e.g., in the 1998 western Atlantic BFT assessment it was 0.4^2.

Method 2. Weighted by the MLE estimate of variance $\hat{\sigma}_j^2$ for each series. This is similar to iterative re-weighting (see below).

$$\text{Ln } L = -\sum_j \sum_y \left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2}{2\hat{\sigma}_j^2} + \ln(\hat{\sigma}_j) \right]$$

$$\hat{\sigma}_j^2 = \frac{1}{n} \sum_y \left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2$$

Method 3. The inverse variance method with annual observations proportional to the inputted annual CV^2 and the average variance for each series equal to the MLE estimate (NMFS 1998).

$$\ln L = -\sum_j \sum_y \frac{0.5}{c_j CV_{j,y}^2 \hat{\sigma}_j^2} \left[\log\left(\frac{I_{j,y}}{q_j B_y} \right) \right]^2 - 0.5 \log\left(c_j CV_{j,y}^2 \hat{\sigma}_j^2 \right)$$

where

$$\hat{q}_j = \exp\left(\frac{\sum_y \left(\ln(I_{j,y}) - \ln(\hat{B}_y)\right) / c_j CV_{j,y}^2 \hat{\sigma}_j^2}{\sum_y 1 / \left(c_j CV_{j,y}^2 \hat{\sigma}_j^2 \right)} \right)$$

and c_j is a constant for each series whose value is chosen such that the average variance for each data series equals its estimated average variance, $\hat{\sigma}_j^2$.

Method 4. Iterative re-weighting. The sigma for each series is treated as a free parameter

$$\text{Ln } L = -\sum_j \sum_y \left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2}{2\hat{\sigma}_j^2} + \ln(\hat{\sigma}_j) \right]$$

Method 5. Input variances re-adjusted by expert judgement, $\sigma'^2_{j,y}$, plus an estimated scale parameter $\hat{\sigma}_A^2$

$$\text{Ln } L = -\sum_j \sum_y \left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2}{2\left(\sigma'^2_{j,y} + \hat{\sigma}_A^2\right)} + \ln\left(\sqrt{\sigma'^2_{j,y} + \hat{\sigma}_A^2} \right) \right]$$

$$\hat{q}_j = \exp\left(\frac{\sum_y \left(\ln(I_{j,y}) - \ln(\hat{B}_y)\right) / \left(\sigma'^2_{j,y} + \hat{\sigma}_A^2\right)}{\sum_y 1 / \left(\sigma'^2_{j,y} + \hat{\sigma}_A^2\right)} \right)$$

The average input variance of each time series is set according to the scientists' expert judgement on the relative reliabilities of the different series as indices of trends in abundance. The GLM or other estimates of variances for individual data points in a time series are incorporated by making the assigned variance the sum of some unaccounted for variance, a_j, plus the GLM variance for each data point. Thus, an average variance, $\overline{\sigma}_j^2$, is assigned by expert judgement to each series, j. The input variances, $\sigma'^2_{j,y}$, for each data point are initially adjusted by the estimated sample variances, $\sigma^2_{j,y}$, e.g., from GLM analysis, assuming that the sample variances and the unaccounted for sources of variance are additive.

$$\sigma'^2_{j,y} = a_j + \sigma^2_{j,y}$$

We solve for the constant a_j for each data series such that the mean assigned variance results:

$$\bar{\sigma}^2_j = \frac{1}{n}\sum_{y=1}^{n_j}\left(a_j + \sigma^2_{j,y}\right)$$

Therefore the constant for each series, a_j is obtained by:

$$a_j = \bar{\sigma}^2_j - \frac{1}{n}\sum_{y=1}^{n_j}\left(\sigma^2_{j,y}\right)$$

Thus each input variance is given by:

$$\sigma'^2_{j,y} = \bar{\sigma}^2_j - \frac{1}{n}\sum_{y=1}^{n_j}\left(\sigma^2_{j,y}\right) + \sigma^2_{j,y}$$

Method 6. This method has often been applied in catch-age analysis (Quinn and Deriso 1999; Sullivan et al. 1999; Parma 2000). Input variances re-adjusted by expert judgement, $\sigma'^2_{j,y}$ (same as that in method 5 above), but multiplied by a scale parameter $\hat{c}$

$$\mathrm{Ln}\ L = -\sum_j\sum_y\left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j\hat{B}_y)\right)^2}{2\left(\hat{c}\sigma'^2_{j,y}\right)} + \ln\left(\sqrt{\hat{c}\sigma'^2_{j,y}}\right)\right]$$

$$\hat{q}_j = \exp\left(\frac{\sum_y\left(\ln(I_{j,y}) - \ln(\hat{B}_y)\right)/\hat{c}\sigma'^2_{j,y}}{\sum_y 1/\left(\hat{c}\sigma'^2_{j,y}\right)}\right)$$

Method 7. Input variance: simply dividing by the inputted variances or CVs (often done in the VPA):

$$\ln L = -\sum_j\sum_y\frac{\left[\ln(I_{j,y}) - \ln(\hat{q}_j\hat{B}_y)\right]^2}{CV^2_{j,y}}$$

Method 8. Additional variance method (Geremont and Butterworth 2000). $\hat{\sigma}^2_{A,j}$ is an estimable parameter for each series

$$\mathrm{Ln}\ L = -\sum_j\sum_y\left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j\hat{B}_y)\right)^2}{2\left(\hat{\sigma}^2_{j,y} + \hat{\sigma}^2_{A,j}\right)} + \ln\left(\sqrt{\hat{\sigma}^2_{j,y} + \hat{\sigma}^2_{A,j}}\right)\right]$$

$$\hat{q}_j = \exp\left(\frac{\sum_y \ln(I_{j,y}) - \ln(\hat{B}_y) \Big/ \left(\hat{\sigma}^2_{j,y} + \hat{\sigma}^2_{A,j}\right)}{\sum_y 1/\left(\hat{\sigma}^2_{j,y} + \hat{\sigma}^2_{A,j}\right)}\right)$$

Method 9. Inverse variance weighting with a variance input for each year, analogous to method 7.

$$\text{Ln } L = -\sum_j \sum_y \left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2}{2\sigma^2_{j,y}} + \ln\left(\sigma_{j,y}\right)\right]$$

$$\hat{q}_j = \exp\left(\frac{\sum_y \left(\ln(I_{j,y}) - \ln(\hat{B}_y)\right) \Big/ \left(\sigma^2_{j,y}\right)}{\sum_y 1/\left(\sigma^2_{j,y}\right)}\right)$$

See Appendix 1 and Table 1 for a comparison of the relative merits of each of these alternatives.

A Numerical Evaluation of the Alternatives: Application to western Bluefin tuna

Three methods to determine weights for different data series have been used in previous ICCAT bluefin tuna assessments and are as follows. (1) Equal input weighting - each data point in each series is given a fixed equal weight that is decided upon before fitting the assessment model to the data (method 1 above). (2) Inverse variance weighting (also called input weighting) - each data point in each series is weighted by the inverse of the variance estimated for it by GLM standardization (method 7). (3) Iterative re-weighting (also called maximum likelihood) - the relative weight of each time series is estimated by fitting the assessment model to the time series and other data to maximize the likelihood of the fit (method 4).

Geromont and Butterworth (2000) suggested a fourth method called "additive variance" (method 8) that is intermediate between inverse variance and iterative re-weighting. The weightings that are inputted to the assessment model are externally estimated as variances for each year and series (as in input weighting). Then, the assessment model is applied to estimate an additional variance for each series (similar to iterative re-weighting). While this method avoids many of the problems of equal input weighting (ignoring the model goodness of fit for the data) and iterative re-weighting (e.g., all of the weight going to one series), it still allows the estimation algorithm applied to change the relative weight of each series, depending on the relative consistency of each series with the assumed model structure (Appendix 1). This would be undesirable if there were strong a priori reasons for weighting one series more highly than another and there was uncertainty over the correctness of the assumed model structure.

A fifth alternative (method 6), has often been applied in catch-age analysis and is currently applied by the International Pacific Halibut Commission (Quinn and Deriso 1999; Sullivan et al. 1999; Parma 2000). This is constant relative weighting determined by expert judgement and a single estimated scale parameter, an extension of the additional variance method. This involves inputting the variances for data points in each alternative abundance series based partly on the output of GLM standardization and partly on expert judgment. The relative weightings of the different series could be initially determined by GLM but modified by expert judgement based on past experience with other similar data sets, additional knowledge about how the data were obtained and key properties not accounted for in GLM (e.g., their spatial and temporal coverage). For example, in the 1998 WA BFT assessment, the equal weighting approach assigned input CVs of 0.4 to all of the data series. This value, squared and averaged over all data series, should be at least as large as the average σ^2 from GLM analyses (since those from the latter often underestimate those based on GOF). Additional knowledge about each series could motivate adjustments to the average CV for each series upwards or downwards from the global average depending on knowledge about how the data for it were collected and its overall reliability as an index of abundance. The within series relative weightings for annual observations could be determined entirely by GLM (as shown above). Then, instead of estimating a separate additional variance for each series, a single scale parameter that multiplied each of the input variances would be estimated by fitting the assessment model to all of the abundance series and other data. This scale parameter would jointly expand or decrease (within pre-determined limits) the inputted variances in all of the series, depending on the overall goodness-of-fit of the model to all of the different data series. This would allow the inputted relative weights, that are based on the relative reliability of each data series, to remain the same, while allowing for empirically-based adjustments to the overall variance of the fit. The relative merits of this and other methods are further addressed in Appendix 1 and summarized in Table 1.

Details of Numerical Evaluation

We evaluated the effect of applying the nine alternative methods for weighting on stock assessment results. Using maximum likelihood methods, we fitted a discrete time Schaefer model to a selected set of data series for western Atlantic bluefin tuna (Table 2, Fig. 1). The data series selected were ones that were judged to represent the fishable biomass of this stock. They included all the series that included fish greater than 10 years of age (ICCAT, 1998, Table 6). All of these indices are fishery dependent, and their lengths vary from 8 to 18 years. In an actual assessment, it would be necessary to determine the relative reliability of the various series based on their area coverage and other considerations. For the sake of this exercise, we assumed that the series were equally believable; for methods 5 and 6, the seven indices were given equal average input weights. None of the series appeared to suggest contradictory trends in abundance a priori (Fig. 1), though there was considerable variability in the apparent trends.

The inputted versus outputted weights, estimates of trends in abundance and estimates of the Schaefer model parameters r and K were obtained for each weighting method. Additionally, the deviance in the residuals from each weighting method were computed and evaluated using chi-square tests (Gelman et al. 1995) for each series separately and in combination. Values greater than 0.99 or less than 0.01 indicate that the model or the model's likelihood function is mis-specified. Deviance is given by (Gelman et al. 1995):

$$\chi^2_{n-1} = \sum_i \sum_y \frac{\left(x_{i,y} - E(x_{i,y} \mid q)\right)^2}{\mathrm{var}(x_{i,y} \mid q)}$$

where n is the number of data points, $x_{i,y}$ is the observed value of index i in year x, $E(x_{i,y}|q)$ is the expected value of the index i in year y, given the estimated parameters q, (i.e. $E(x_{i,y}|q)$ is the predicted value of index i in year y), and $\mathrm{var}(x_{i,y}|q)$ is the variance of $x_{,y}$. The variances were:

$$\mathrm{var}(x_{i,y} \mid q) = \left(\exp(s_{i,y}{}^2) - 1\right) \cdot E(x_{i,y} \mid q)^2$$

where $\sigma_{i,y}{}^2$ is the model's estimate.

RESULTS

The seven data series together indicated a general declining trend throughout the time series of the fishery (Fig. 1). The nine weighting methods generally gave two kinds of fits, either K was estimated around 64000 MT and r around 0.34 (Method 1,5,6,7 and 9), or the model fit a very high K, and low r (Table 3). In the latter case, the best estimate of r was at the lower constraint of 0.01 (Method 2,3,4 and 8). However, these CPUE series, combined with the catch data were not well described by a Schaefer model for either the low *r* or intermediate *r* fits. The methods that input a variance and did not update it (Methods 1, 7 and 9) show either very high or very low total deviance values, implying that the model is mis-specified (Table 4). Methods that estimate the variance within the model showed fairly similar total deviance. However, the relative fits of the various series between methods imply that some of the data sets may be contradictory. For example, methods that fit a low *r* have lower deviances with the JLL Area wt series, while methods that fit a high *K* have lower deviances with most of the other series. If time allowed, it would be worthwhile to fit the data sets separately, or at least separate the series that favor a low *r* from the series that favor a low *K*.

The methods estimated different parameter values because the different methods assigned very different weights to the various series. Table 5 shows the average weights applied to each series as inputs, and Table 6 shows the average weights applied to each series under the fitted model. The relative weights of the series remain the same for methods 1, 7 and 9, which enter input variances and do not update them, and for method 6, which multiplies the input weights by a scale parameter. For method 5, the output weights were the same as the input weights, but this would not be the case if the input weights were different among series and the estimated added variance was greater than zero. The methods that input variances (7 and 9) placed a very high weight on the JLL Area Wt series, which had been assigned CVs of 0.2 in the absence of variance estimates- much lower than the CVs estimated for the other six series (Table 2). Thus, this series was given 27% of the weight, while Larval and US LL GOM, which had very high CVs (Table 2) were given only 8% and 7%, respectively, of the weight. Conversely, the methods that could change the relative fit of each series tended to downweight the JLL Area weight series, and increase the weight on JLL GOM and US RR. Methods 2, 3, 4 and 8 gave the JLL Area Wt series 1-2% of the weight. Method 5 did not greatly change the relative weights of the series.

Similar information is shown in Table 7, which shows the average input sigma for each series, and Table 8 which shows the average estimated sigma in each series. These results indicate that the methods vary in the accuracy of their estimates of sigma, and thus in their estimates of the parameters (i.e. they cannot all be right).

The guidelines presented above for choosing a method suggest that the chosen method be reconsidered if the weighting ranks of the data sets are different from the agreed upon ranking of the reliability of the data sets. By this criteria, methods 2, 3, 4 and 8 would not be deemed acceptable, since they greatly downweight JLL Area wt and overweight JLL GOM. The input weighting methods (7 and 9) greatly overweight the JLL Area wt series; however, this problem could be corrected by giving the JLL Area wt series an assumed variance more similar to the observed variances in the other series.

The guidelines suggest that contradictory data should not used. Because the Schaefer model had difficulty fitting both the Can GSL and JLL GOM series and the JLL Area Wt and Larval series, it would be worth running the model with each series separately to determine which series are contradictory.

Finally, the guidelines suggest examining the goodness of fit of the model under the chosen method. Although none of the methods fitted the data really well, only methods 7 and 9 provided an exceptionally bad fit in terms of model deviance (Table 4). However, methods 2, 3, 4 and 8 fit an *r* value at its lower limit, a value that appears to be inconsistent with the biology of the fish.

DISCUSSION

This paper provides a set of guidelines that could be used to help fishery scientists to more easily agree on a method to assign weightings to alternative data series in a stock assessment. It is suggested that the method should be chosen before the stock assessment results are obtained. Seven scientific criteria are suggested that can be applied to evaluate the relative merits of each candidate method

(Table 1). An evaluation of 9 different methods that could be applied to the western bluefin stock shows that the methods vary widely with regards to the criteria. Perhaps the most serious characteristic to avoid is the possibility that the method will inadvertently assign the highest weights to the series deemed by the scientists to be the most likely to be biased and the least weight to the series deemed least likely to be biased. This is possible in methods such as iterative reweighting where the variances for each time series is estimated solely on the basis of the goodness of fit of the stock assessment model to the time series. Time series that reflect more precise time trends in abundance will be given higher weight than less precise series that may be less biased. Even with input weighting methods, care must be taken that none of the series have been given implausibly high weights (as with JLL Area Wt above), since this causes these series to dominate the fit.

The weighting method chosen might be reconsidered only if the stock assessment results obtained are biologically or statistically unreasonable. This could occur if:

- The estimates of key population parameters such as the intrinsic rate of increase, are anomalous (i.e. biologically impossible or at one of the boundary conditions).
- The model deviance is anomalously large, especially for series believed to be the most credible.
- Time series judged to be the most reliable are downweighted and vice versa.
- Some time series are inadvertently given most of the weight.
- Confidence intervals are unrealistically narrow.

RECOMMENDATIONS

The recently proposed additional variance method (Geromont and Butterworth 2000) relies on the goodness of fit (GOF) of the stock assessment model to each data set to determine its "additional" variance and thereby its overall relative weighting. This is attractive because the method for allocating weightings becomes strictly mechanical without any intervention from stock assessment scientists. However, relying entirely on GOF to determine weightings can amplify biases in estimated trends, particularly when the data series with the smoothest trends and that are consistent with model predictions are not the most accurate. thus, this "hands-off "approach to determining relative weightings, while appearing objective and reducing debate over model inputs, may often produce anomalous stock assessment results. Instead, a more "interventionist " approach to determining relative weightings which relies on knowledge outside of that contained in the data themselves could help to avoid such undesirable outcomes. It should be noted that if expert judgement were to be applied, the method would only be transparent if the judgement applied was clearly explained and documented.

ACKNOWLEDGEMENTS

This work was made possible by a grant from the David and Lucile Packard Foundation and the Pew Charitable Trust, and by the Niarchos Fellowship. Thanks to Carryn Cunningham and Shelley Clarke for commenting on an earlier draft.

REFERENCES

BUTTERWORTH, D.S., Punt, A.E. and Smith, A.D.M. 1996. On plausible hypotheses and their weightings, with implications for selection between variants of the Revised Management Procedure. Rep. int. Whal. Commn 46: 481-491.

GELMAN, A., J. B. Carlin, H. S. Stern and D. B. Rubin. 1995. Bayesian Data Analysis. Chapman and Hall. New York.

GEROMONT, H.F. and Butterworth, D.S. 2000. Possible extensions to the ADAPT VPA model applied to western north Atlantic bluefin tuna, addressing in particular the need to account for "additional variance". ICCAT Doc. SRCS/00/33.

ICCAT. 1999. Collective Volume of Scientific Papers. XLIX (2).

MCALLISTER, M.K., Starr, P.J., Restrepo, V., and Kirkwood, G.P.. 1999. Formulating quantitative methods to evaluate fishery management systems: what fishery processes should be modeled and what trade-offs should be made? ICES Journal of Marine Science 56: 900-916.

NATIONAL MARINE FISHERIES SERVICE. 1998. Report of the Shark Evaluation Workshop. National Oceanic and Atmospheric Association. National Marine Fisheries Service. Southeast Fisheries Science Center. 3500 Delwood Beach Road. Panama City, FL 32408.

PARMA, A.M. 2000. Bayesian approaches to the analysis of uncertainty in the stock assessment of Pacific halibut. In J. M. Berkson, L.L. Kline and D.J. Orth (eds.) Incorporating uncertainty into fisheries models. (accepted).

PUNT, A.E. and R. Hilborn. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Rev. Fish Biol. and Fisheries. 7: 35-63.

QUINN II, T. J. and DERISO, R.B. 1999. Quantitative fish dynamics. Oxford University Press. Oxford.

RICHARDS, L.J. 1991. Use of contradictory data sources in stock assessments. Fish. Res. 11: 225-238.

SCHNUTE, J.T. and Hilborn, R. 1993. Analysis of contradictory data sources in fish stock assessments. Can. J. Fish. Aquat. Sci. 52: 2063-2077.

SULLIVAN, P.J., Parma, A.M., and Clark, W.G. 1999. The Pacific halibut stock assessment of 1997. International Pacific Halibut Commission Scientific Report 79, 84 pp.

APPENDIX 1: A LIST OF CRITERIA THAT CAN BE USED TO HELP SCIENTISTS OBJECTIVELY CHOOSE A METHOD TO WEIGHT ALTERNATIVE DATA POINTS AND DATA SERIES IN A STOCK ASSESSMENT

There are a variety of scientific criteria that could be applied to help scientists objectively choose a method to weight alternative data points and data series in a stock assessment. These are mentioned in the section on general considerations and specified and discussed in more detail in this appendix. A recent paper by Geromont and Butterworth (2000) makes reference to three of these. We reformulate these to make them more generally applicable and provide a few additional criteria. These criteria should be used jointly in evaluating alternative candidates for a weighting method, though some criteria are more important than others.

(1) *Does the method incorporate GLM or other estimates of the relative reliability of data points for a particular data series?*

For example, CPUE data points for years for which there are few commercial fishing vessels may be less precise and the "random-effects GLM estimates of variance" (Geromont and Butterworth 2000) given to these years may be larger. Some weighting methods such as input weights (methods 7 and 9) and the additional variance method (method 8) (Geromont and Butterworth 2000) can directly incorporate such variance estimates from GLM on data points in each time series. Methods 3, 5, and 6 also allow this. Thus data points with low sampling precision can be down-weighted.

(2) *Does the method avoid according unrealistically high precision (low variance) and high weight to some data series?*

This has been pointed out as a classic problem with the iterative re-weighting method (method 4) (Geromont and Butterworth 2000). This is also a problem for methods 2 and 3 in this paper, as the mean variance for each time series is estimated in the stock assessment. It is potentially a problem for method 8, because it allows the estimation of a variance term for each time series in addition to the ones inputted for each time series. If the inputted variances are low relative to the additional estimated variances, unrealistically high weight can still be accorded to some time series. This is undesirable because it can give rise to a strong negative bias in the precision in parameter estimates. If there are many different time series, this tends to increase the chance that any one or a few of them, just by chance will provide an anomalously good fit to the model. Most seriously, bias in stock assessment results from an "inappropriate choice of the resource assessment model" (Geromont and Butterworth 2000) can be amplified by this phenomenon. The best fitting time series might not necessarily be unbiased in the abundance trends that they suggest.

(3) *If empirically determined input variances are applied in the weighting method, can the method allow for readjustments of these weightings based on the goodness of fit of the model to the data*?

This would be desirable if some of the inputted variances did not incorporate the sources of variance in the data that were not accounted for by the stock assessment model or were not directly comparable. For example, variances for data points derived from research survey data are often not comparable with those derived from GLM analysis of CPUE data. Factors that are controlled for in research surveys are often not controlled for in catch rate indices and the assumption of random sampling from the entire population, though applied in analyses of both types of data, are not always entirely applicable to the latter. Also, GLM treatments of different CPUE time series are often not comparable because they apply different sets of explanatory variables or factors such as only month versus month and area (Geromont and Butterworth 2000). In such cases, the random effects GLM estimates of variances from these different GLMs will not be comparable. Where fewer factors are incorporated, the variance estimates for a given series will be more negatively biased (Geromont and Butterworth 2000). The result, if the weighting method relied solely on these inputted variances, would be that such series would be given undue weight in the stock assessment. This has long been identified as a classic problem for the input weight method that relies solely on sampling or GLM estimated variances (methods 7 and 9) (McAllister 1995; Geromont and Butterworth 2000). Thus some methods (e.g., 8) have been suggested that allow for readjustments of such input weightings based on the goodness of fit between the model and the data. Method 8 introduces an additional variance parameter to be estimated for each time series. If the input variances in one of the series, are negatively biased, the fitting of the stock assessment model to the set of series should *in theory* pick this up and provide a positive estimate of the additional variance term for this series. Questions raised during a discussion of this method in the ICCAT Methods Working Group meeting in May 2000, Madrid suggest criteria and these are reformulated below.

(4) *Does the weighting method allow the incorporation of expert judgement on the relative reliability of each series as an index of abundance?*

While it appears scientifically objective to rely entirely on estimation methods to determine the weightings of different time series (e.g., methods 2, 3, 4, 8), several conditions are required by this approach to provide unbiased stock assessment results. The structure of the stock assessment model applied must not be mis-specified, each of the data series is equally reliable as an unbiased trend in abundance, and the data sets are sufficiently long to precisely estimate variances. However, in most stock assessments these conditions almost never hold conjointly. If any one of them does not hold, then biased stock assessment results can be produced. Take for example, the considerable difference in the reliability as indices of abundance between fishery independent and fishery dependent data. For reasons given above, these differences are often not accurately reflected in the input weightings estimated, e.g., from GLM. The GLM variances for the CPUE data can often be less than sample variances for research survey data points, despite the greater reliability of the latter. If the trend is smoother in the CPUE series, as it often can be, then most stock assessment models will tend to fit this time series best and the stock assessment model estimate of the variance for the CPUE trend will be lowest and most weight will be accorded to it. Without some intervention by the stock assessment scientists, methods 2, 3, 4, and 8 would not be able to avoid this undesirable result. As bias in estimated trends is often a far more serious a problem than imprecision, it seems desirable and sensible to adopt a weighting methodology that allows for sensible interventions and constraints on the otherwise "hands-off" approach to weighting. This forms the fourth criterion: Does the weighting method allow the incorporation of expert judgement on the relative reliability of each series as an index of abundance? A single baseline set of relative input weightings could be adopted by the stock assessment scientists before running the stock assessment model. This would take into consideration (i) whether the index was fishery independent, and (ii) the overall scientific credibility of the time series as an index of relative abundance. More work is needed on developing an objective basis to come up with the latter. Guidelines should account for the spatial and annual temporal coverage of an index with respect to the spatial range and annual movements of the fish stock and the age classes indexed. Although it may be impossible to reach a consensus on relative input weightings by expert judgement, a baseline set is desirable. This is because scientists are typically better qualified to determine the relative scientific credibility of different data series than fishery managers. It would be absolutely essential for the reasons behind assigning the relative weightings to be documented.

(5) *Does the weighting method prevent the most reliable data sets from being down weighted?*

Even if all of the above criteria are met, the manner in which the goodness of fit to the stock assessment model updates the inputted weights will affect whether and to what extent the relative weightings determined by expert judgement are updated. In doing so, it is desirable that the weighting method should incorporate sensible constraints to prevent the time series originally deemed to be the most reliable from being down-weighted. Otherwise, what's the point of using expert judgement in the first place? Methods 5 and 6, which satisfy criteria 1 to 4, differ in this respect. In method 5, poorer goodness of fit will lead to all of the time series having more and more similar final weightings. In contrast, method 6 preserves the inputted relative weightings but still allows adjustments to the magnitudes of all of the weightings according to goodness of fit to the stock assessment model.

(6) *Is the weighting method appropriate for relatively few data points in the time series?*

If weightings are estimated based on goodness of fit to the stock assessment model, it is usually necessary for there to be at least 15 or more data points in the series to permit estimation of the variance. Often this is not the case in stock assessment. It would be desirable thus, for the weighting method, applied to incorporate sensible constraints and allow a sensible intervention if there are relatively few data points in a data series. Methods 5 and 6 allow this.

(7) *Could the weighting method be expected to result in an accurate assessment of parameter uncertainty?*

This is unlikely if the goodness of fit to the stock assessment model is ignored by the weighting method as in methods 1, 7, and 9 and especially if the weightings are determined only by GLM analysis; this will tend to underestimate the model fit variances and thus parameter uncertainty. If the weighting method relies on goodness of fit to the stock assessment model but is entirely "hands off", this is also unlikely for reasons mentioned under criteria 4 above. For example, to be valid, these methods rely on the structure of the stock assessment model applied to not be mis-specified, each of the data series to be equally reliable as an unbiased trend in abundance, and the data sets to be sufficiently long to precisely estimate variances.

(8) *Is the method easily put into practice?*

Are sufficient data, scientific information and expertise available to sensibly apply the method?

Table 1. Some of the relative merits of the alternative weighting methods. See the methods section for a more detailed description of the alternative weighting methods. GOF stands for goodness of fit of the stock assessment model to the data series. See Appendix 1 for an outline of the different selection criteria.

Method	1. Incorporation of GLM estimates of variance for data points	2. Avoids according unrealistically high precision to some data series	3. Allows for GOF readjustments of input weightings	4. Allows expert judgment of the relative reliability of each data series	5. Most reliable data series prevented from being down-weighted	6. Conducive to relatively few data in the time series (e.g., < 15 years)	7. Giving accurate assessment of parameter uncertainty
1. Equal input weightings	No	Yes	No	No	N/A	Possibly	Unlikely
2. Weighting by MLE	No	No	Yes	No	No	No	Unlikely
3. Input variance multiplied by a scale parameter estimated for each series	Yes	No	Yes	No	No	Not necessarily	Unlikely
4. Iterative reweighting	No	No	Yes	No	No	No	Unlikely
5. Input variance modified by expert judgement plus a scale parameter	Yes	Yes	Yes	Yes	Yes, if estimated scale parameter is not much larger than inputted variances	Yes	Possibly
6. Input variance modified by expert judgement multiplied by a scale parameter	Yes	Yes	Yes	Yes	Yes	Yes	Possibly
7. Wt = 1/Input CV^2	Yes	Yes	No	No	N/A	N/A	No
8. Input variance plus a scale parameter estimated for each series	Yes	Yes, if estimated scale parameters are not much larger than inputted variances	Yes	No	No	No	Requires unbiased time series and assessment model
9. Input fixed estimated variances with no updating of them	Yes	Yes	No	No	N/A	N/A	No

Table 2. CPUE indices used in the model. Only series that indexed mature fish and had been used in the baseline VPA were included (ICCAT 1998).

Name	Can GSL	CV	Can SWNS	CV	JLL GOM	CV	JLL Area wt	CV	Larval	CV	US LL GOM	CV	US RR >195	CV
Age Range	13+		7-13		10+		4-10+		8-10+		8-10+		8-10+	
Indexing	Numbers		Numbers		Numbers		Numbers		Biomass		Numbers		Numbers	
1962							3.673	0.2						
1963							1.621	0.2						
1964							0.904	0.2						
1965							0.811	0.2						
1966							0.959	0.2						
1967							0.169	0.2						
1968							0.053	0.2						
1969							0.099	0.2						
1970							0.004	0.2						
1971														
1972														
1973														
1974					0.968	0.266								
1975					0.534	0.205								
1976					0.666	0.207								
1977					0.913	0.216			2.435	0.433				
1978					0.876	0.225			5.824	0.272				
1979					1.287	0.283								
1980					1.158	0.265								
1981	6.072	0.289			0.553	0.239			1.277	0.432				
1982	5.578	0.285							1.514	0.311				
1983	7.673	0.273							1.235	0.308			1.923	0.242
1984	4.414	0.290							0.653	0.802			1.136	0.256
1985	1.643	0.307											0.714	0.277
1986	1.829	0.350							0.261	0.606			0.712	0.602
1987	0.963	0.526							0.445	0.507	0.880	0.292	0.711	0.340
1988	1.805	0.407	2.947	0.332					1.946	0.326	0.400	0.451	0.828	0.287
1989	2.112	0.412	4.907	0.304					0.798	0.439	0.700	0.311	0.853	0.296
1990	1.048	0.396	4.050	0.304					0.474	0.200	0.850	0.309	0.755	0.293
1991	1.975	0.426	3.195	0.304					0.365	0.594	1.000	0.301	0.961	0.276
1992	2.874	0.357	2.845	0.243					0.614	0.382	0.240	0.529	0.912	0.274
1993	2.153	0.322	1.062	0.236					0.667	0.615	0.230	0.562	0.681	0.369
1994	0.629	0.380	2.177	0.249					0.720	0.405	0.100	0.900	0.683	0.384
1995	2.950	0.286	1.851	0.235					0.465	0.585	0.170	0.683	1.031	0.305
1996	0.767	0.287	0.650	0.244					1.458	0.630	0.180	0.696	1.772	0.344
1997	0.823	0.287	0.750	0.246					0.619	0.448	0.310	0.536	1.327	0.333

Table 3. Maximum likelihood estimates of K and r from each weighting method.

Method	K	r
1. Equal input weightings	64758	0.34
2. Weighting by MLE	194676	0.01
3. Input variance multiplied by a scale parameter estimated for each series	183856	0.01
4. Iterative reweighting	194676	0.01
5. Input variance plus a scale parameter	69234	0.30
6. Input variance multiplied by a scale parameter	66024	0.33
7. Wt = 1/Input CV^2	64119	0.35
8. Input variance plus a scale parameter estimated for each series	190853	0.01
9. Input fixed estimated variances with no updating of them	64119	0.35

Table 4. Sum of residual deviance for each method and data series.

Method	Can GSL	Can SWNS	JLL GOM	JLL Area wt	Larval	US LL GOM	US RR >195	Total
1. Equal input weightings	6.67	1.32	1.01	27.46	6.32	3.43	3.71	49.92
2. Weighting by MLE	20.22	8.46	8.87	3.23	27.67	12.70	25.90	107.05
3. Input variance multiplied by a scale parameter estimated for each series	27.62	8.66	8.00	3.24	40.65	32.14	34.32	154.63
4. Iterative reweighting	20.37	8.48	8.94	3.26	27.82	12.72	25.93	107.52
5. Input variance plus a scale parameter	13.05	2.60	2.12	61.26	13.65	5.96	8.59	107.23
6. Input variance multiplied by a scale parameter	12.58	2.37	2.06	56.71	16.73	5.13	9.39	104.97
7. Wt = 1/Input CV^2	114.67	21.50	26.73	1143.99	83.58	42.59	92.58	1525.64
8. Input variance plus a scale parameter estimated for each series	19.13	8.44	9.11	3.23	25.84	10.36	27.99	104.11
9. Input fixed estimated variances with no updating of them	82.95	27.43	33.37	1167.28	75.03	14.34	85.90	1486.29
Number of points in series	17	10	8	9	18	11	15	88
Chi squared 95% significance level (df=n-1)	7.96	3.33	2.17	2.73	8.67	3.94	6.57	66.50

Table 5. Average of the weights applied to data points in each series, normalized so that the weights sum to one, as input to the model.

Method	Can GSL	Can SWNS	JLL GOM	JLL Area wt	Larval	US LL GOM	US RR >195
1. Equal input weightings	0.14	0.14	0.14	0.14	0.14	0.14	0.14
2. Weighting by MLE	NA	NA	NA	NA	NA	NA	NA
3. Input variance multiplied by a scale parameter estimated for each series	0.10	0.16	0.20	0.27	0.08	0.07	0.12
4. Iterative reweighting	NA	NA	NA	NA	NA	NA	NA
5. Input variance plus a scale parameter	0.14	0.14	0.14	0.14	0.14	0.14	0.14
6. Input variance multiplied by a scale parameter	0.14	0.14	0.14	0.14	0.14	0.14	0.14
7. Wt = 1/Input CV2	0.10	0.16	0.20	0.28	0.07	0.06	0.12
8. Input variance plus a scale parameter estimated for each series	0.10	0.16	0.20	0.27	0.08	0.07	0.12
9. Input fixed estimated variances with no updating of them	0.10	0.16	0.20	0.27	0.08	0.07	0.12

Table 6. Average of the weights applied to data points in each series, normalized so that the weights sum to one, from the fitted model.

Method	Can GSL	Can SWNS	JLL GOM	JLL Area wt	Larval	US LL GOM	US RR >195
1. Equal input weightings	0.14	0.14	0.14	0.14	0.14	0.14	0.14
2. Weighting by MLE	0.12	0.16	0.30	0.01	0.10	0.11	0.20
3. Input variance multiplied by a scale parameter estimated for each series	0.12	0.23	0.23	0.01	0.12	0.18	0.11
4. Iterative reweighting	0.15	0.20	0.37	0.01	0.13	0.13	0.25
5. Input variance plus a scale parameter	0.14	0.14	0.14	0.14	0.15	0.15	0.14
6. Input variance multiplied by a scale parameter	0.13	0.13	0.13	0.13	0.16	0.19	0.13
7. Wt = 1/Input CV2	0.10	0.16	0.20	0.28	0.07	0.06	0.12
8. Input variance plus a scale parameter estimated for each series	0.12	0.17	0.28	0.01	0.10	0.13	0.18
9. Input fixed estimated variances with no updating of them	0.10	0.16	0.20	0.27	0.08	0.07	0.12

Table 7. Average of the sigmas for the points in each data series, input to the model**.**

Method	Can GSL	Can SWNS	JLL GOM	JLL Area wt	Larval	US LL GOM	US RR >195
1. Equal input weightings	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. Weighting by MLE	NA	NA	NA	NA	NA	NA	NA
3. Input variance multiplied by a scale parameter estimated for each series	0.34	0.27	0.24	0.20	0.45	0.49	0.32
4. Iterative reweighting	NA	NA	NA	NA	NA	NA	NA
5. Input variance plus a scale parameter	0.50	0.50	0.50	0.50	0.50	0.50	0.50
6. Input variance multiplied by a scale parameter	0.50	0.50	0.50	0.50	0.50	0.50	0.50
7. Wt = 1/Input CV2	0.39	0.38	0.38	0.47	0.35	0.25	0.40
8. Input variance plus a scale parameter estimated for each series	0.34	0.27	0.24	0.20	0.45	0.49	0.32
9. Input fixed estimated variances with no updating of them	0.34	0.27	0.24	0.20	0.45	0.49	0.32

Table 8. Average of the sigmas estimated for the points in each data series, from the fitted model.

Method	Can GSL	Can SWNS	JLL GOM	JLL Area wt	Larval	US LL GOM	US RR >195
1. Equal input weightings	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. Weighting by MLE	0.58	0.50	0.37	1.86	0.63	0.62	0.45
3. Input variance multiplied by a scale parameter estimated for each series	0.56	0.39	0.39	1.84	0.67	0.57	0.61
4. Iterative reweighting	0.58	0.50	0.37	1.86	0.63	0.62	0.45
5. Input variance plus a scale parameter	0.77	0.77	0.77	0.77	0.77	0.77	0.77
6. Input variance multiplied by a scale parameter	0.78	0.78	0.78	0.78	0.78	0.78	0.78
7. Wt = 1/Input CV2	0.35	0.27	0.24	0.20	0.48	0.54	0.34
8. Input variance plus a scale parameter estimated for each series	0.58	0.48	0.38	1.85	0.65	0.60	0.48
9. Input fixed estimated variances with no updating of them	0.34	0.27	0.24	0.20	0.45	0.49	0.32

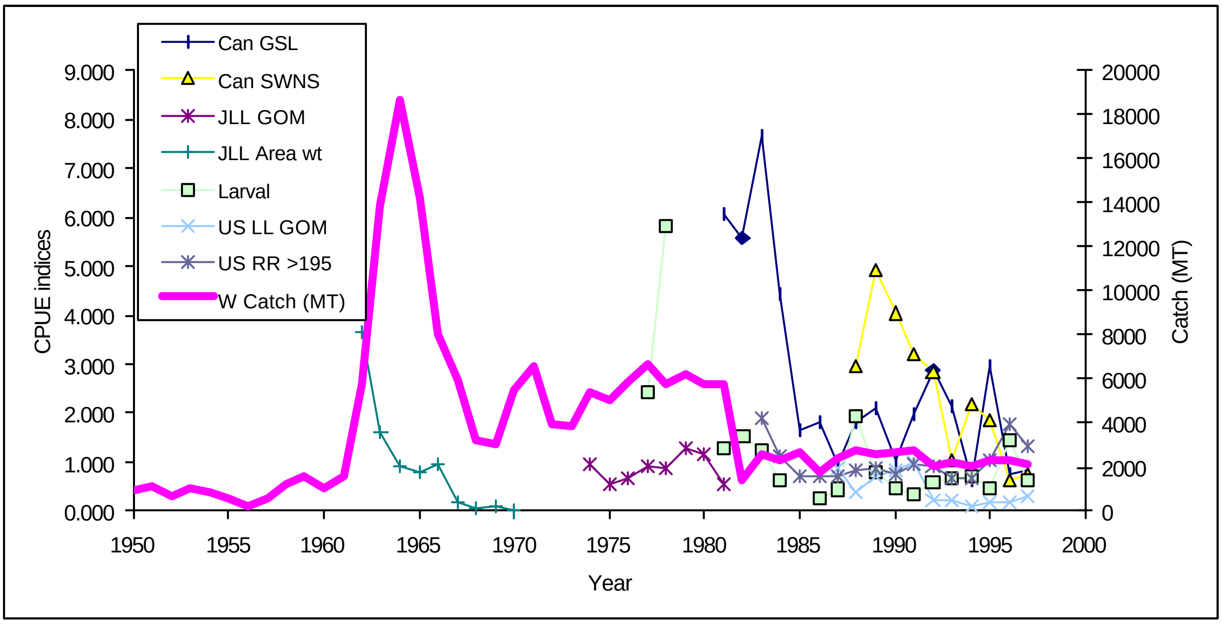

Figure 1. Indices of abundance and catch rate indices used in model.

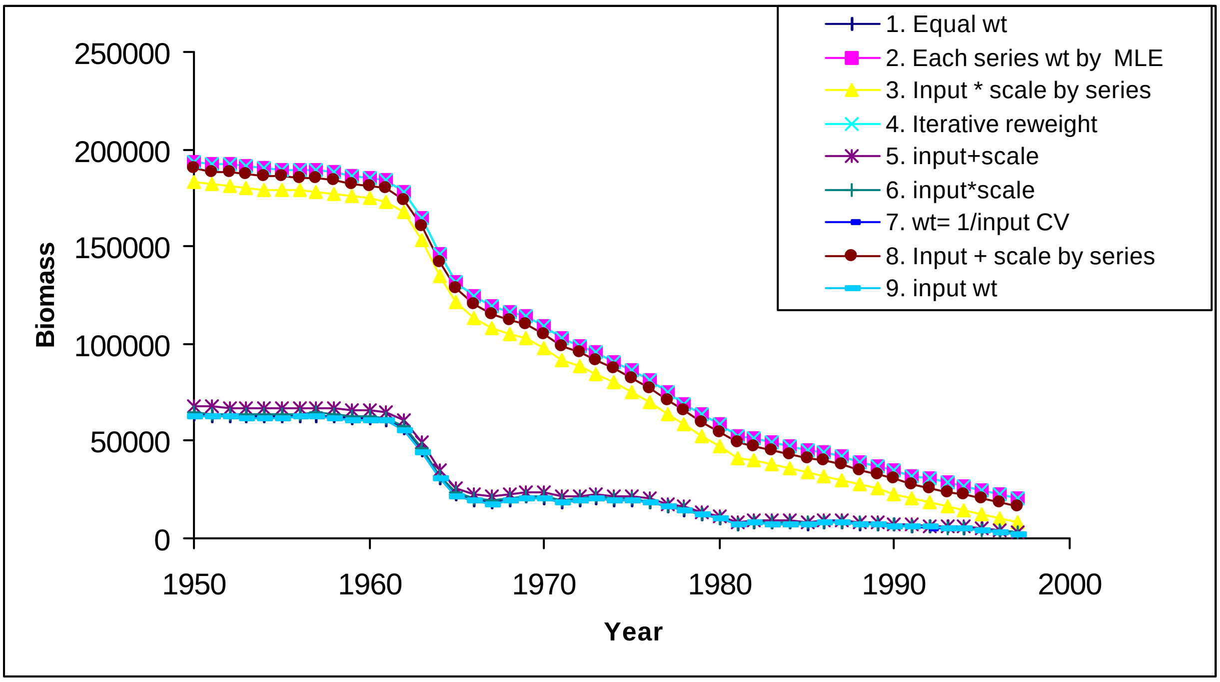

Figure 2. Biomass trends estimated by each weighting method.