SCRS/1992/127

Col.Vol.Sci.Pap. ICCAT, 40 (2) : 498-502 (1993)

# RECENT DEVELOPMENTS IN EXTENDING THE ASPIC PRODUCTION MODEL

*Prager, M. H.*
*NMFS, Southeast Fisheries Science Center, 75 Virginia Beach Drive, Miami, Florida, USA*

## SUMMARY

Recent work to extend the ASPIC non-equilibrium production model has resulted in several theoretical changes. Many of these have been translated into FORTRAN code and are beginning to be tested and used in assessments. Some of these new options include accumulation of residuals in effort rather than yield; the ability to tune to year-end (as well as start-of-year and mid-year) indices and estimates of abundance; the use of iterative re-weighting to estimate statistical weights when fitting more than one series; and the incorporation of the iterative re-weighting into the bootstrap, so that uncertainly about the weighting can be incorporated into the bootstrap results. Accumulation of the residuals in effort provides two additional benefits: it provides a simple, natural way to accommodate missing values of effort (or an additional time series of yield without recorded effort), and it allows computation of bootstrapped projections based on catch limits.

## RESUME

Les travaux récents pour étendre le modèle de production ASPIC ne postulant de conditions d'équilibre ont abouti à plusieurs changements théoriques. Plusieurs d'entre eux ont été traduits en code FORTRAN et commencent à être testés et utilisés dans les évaluations. Certaines de ces nouvelles options comprennent l'accumulation de résidus en effort plutôt qu'en rendement; la capacité d'ajuster en fin d'année (ainsi qu'en début d'année et milieu d'année) les indices et les estimations d'abondance; l'utilisation d'une repondération itérative pour estimer les poids statistiques lorsque l'on ajuste plus d'une série; et l'incorporation de la repondération itérative dans le bootstrap, de façon à ce que l'incertitude de la pondération puisse être incorporée dans les résultats itératifs. L'accumulation de résidus en effort apporte deux bénéfices supplémentaires: il donne un moyen simple, naturel pour ajuster les valeurs de l'effort qui manquent (ou une série temporelle additionnelle de rendement sans effort enregistré), et permet de calculer les projections itératives basées sur les limites de prise.

## RESUMEN

Las recientes tareas para ampliar el modelo de producción ASPIC en condiciones de no equilibrio, han producido numerosos cambios teóricos. Varios de ellos se han traducido al código FORTRAN y se les está empezando a ensayar y utilizar en evaluaciones. Algunas de estas nuevas opciones incluyen la acumulación de residuales de esfuerzo, más que en rendimiento; la capacidad para calibrar los índices y estimaciones de la abundancia hasta el final del año (así como a comienzos y mediados del año); el empleo iterativo de la reponderación para estimar ponderaciones estadísticas cuando se ajusta más de una serie; y la incorporación de la reponderación iterativa de reajuste a partir de submuestras ("bootstrap"), de forma que las incertidumbres de la ponderación puedan incorporarse en los resultados del reajuste a partir de submuestras ("bootstrap"). La acumulación de los residuales en el esfuerzo aporta dos beneficios adicionales: facilita una forma simple y natural de acomodar los valores de esfuerzo que faltan (o una serie temporal adicional de rendimiento sin esfuerzo registrado), y permite computar proyecciones iterativas de reajuste a partir de submuestras ("bootstrap") basadas en límites de captura.

## INTRODUCTION—REVIEW OF MODEL

The ASPIC production model (Prager 1992) is a non-equilibrium production model based on a quadratic differential equation describing the dynamics of stock biomass. Production models such as ASPIC are useful in assessments where age-structured data are not available or are highly uncertain. In addition, such models can give an additional view of the data for stocks customarily assessed with age-structure models. Recent work has resulted in several extensions of the ASPIC model. These extensions were developed with the objective of providing practical solutions to problems faced in some assessments, and some have been used in assessment meetings already.

In this introductory section of the paper, the basic equations of the model are reviewed, following the material in Prager (1992) but in a somewhat simpler notation. The balance of the paper is devoted to the recent extensions. Finally, the discussion section of the paper describes the progress that has been made to date in translation of these theoretical improvements into computer programs that can be used for assessment purposes.

### Basic Differential Equations

Most surplus-production models characterize a resource population as an undifferentiated biomass. The number of individuals present or harvested plays no part, nor is age or size structure considered. Rather, the population is described by its size in biomass (weight) units only. A quantity termed *surplus production* is then used to characterize population dynamics at different levels of population size. Surplus production is the algebraic sum of three major forces: recruitment, growth, and natural mortality, all expressed as rates in proportion (of biomass) per unit time (usually a year). The adjective "surplus" refers to the surplus of gains to the stock from recruitment and growth over losses from natural mortality; i.e., the net production. Here, surplus production is often termed simply "production."

In the logistic or Graham–Schaefer (Graham 1935; Schaefer 1954, 1957) form, which is used by ASPIC, a quadratic differential equation describes the rate of change of stock biomass $B_t$ due to production. In the absence of fishing, the population's rate of increase or decrease is a function of the current population size and two parameters $r$ and $K$:

$$\frac{dB_t}{dt} = rB_t - \frac{r}{K}B_t^2. \tag{1}$$

Lotka (1924) pointed out that the right side of equation (1) is simply the start of a Taylor expansion of an arbitrary function $\phi(B)$ passing through the origin.

In equation (1), $K$ can be considered to represent the maximum population size, or carrying capacity, and $r$ to represent the stock's intrinsic rate of increase (in proportion per unit time). Here, as in most production models, both are assumed constant. The same model can be written in many different ways. Indeed, a slightly different form is used later in this paper to simplify the notation.

Adding fishing mortality $F_t$ to the model, it becomes—

$$\frac{dB_t}{dt} = (r - F_t)B_t - \frac{r}{K}B_t^2. \tag{2}$$

### Solutions for Biomass and Yield

We now define $\alpha_t = r - F_t$ and $\beta = r/K$ to arrive at a simpler equivalent of (2):

$$\frac{dB_t}{dt} = \alpha_t B_t - \beta B_t^2. \tag{3}$$

Integration of (3) provides models of biomass and yield through time. We solve equation (3) for biomass $B_t$ under the assumption that $F_t$ is constant and that therefore $\alpha_t$ is constant. This is a very weak assumption, for if $F_t$ varies, time can be divided into short periods of constant or nearly constant $F$ and a solution found for each period. From here on, we assume periods of one year.

For the year beginning at $t = \tau$ and ending at $t = \tau + 1$, the integration of (3) gives

$$\hat{B}_{\tau+1} = \frac{\alpha_\tau B_\tau e^{\alpha_\tau}}{\alpha_\tau + \beta B_\tau (e^{\alpha_\tau} - 1)} \qquad \text{if } \alpha \neq 0 \text{, or} \tag{4a}$$

$$\hat{B}_{\tau+1} = \frac{B_\tau}{1 + \beta B_\tau} \qquad \text{if } \alpha = 0. \tag{4b}$$

Equation (4a) is the familiar logistic equation. It is undefined when $F = r$, and (4b) is used in its place. The condition $F = r$ is occasionally encountered while fitting the model, and if (4b) is not used when necessary, the computations can fail.

Modeling the yield during the same period involves another integration—

$$\hat{Y}_\tau = \int_{t=\tau}^{\tau+1} F_\tau \hat{B}_t \, dt, \tag{5}$$

where $\hat{B}_t$ is defined in equations (4a) and (4b). In this paper, $t$ is always used to denote an instant in time, and $\tau$ is used to refer to a specific period of time (a calendar year). Performing the integration in equation (5),

$$\hat{Y}_\tau = \frac{F_\tau}{\beta} \ln\left[1 - \frac{\beta \hat{B}_\tau (1 - e^{\alpha_\tau})}{\alpha_\tau}\right] \qquad \text{if } \alpha \neq 0 \text{, or} \tag{6a}$$

(6b) $\hat{Y}_\tau = \frac{F_\tau}{\beta} \ln\left[1 + \beta \hat{B}_\tau\right]$ if $\alpha = 0$.

Equation (6a) was apparently first given by Pella (1967). I have not found equation (6b) in previous literature.

Two additional quantities can be estimated for each year of data. From the definition of instantaneous rates, the estimated average biomass during year $\tau$ is

(7) $\bar{B}_\tau = \hat{Y}_\tau / \hat{F}_\tau$.

The estimated surplus production during year $\tau$ is estimated by mass balance, and is

(8) $\hat{P}_\tau = \hat{B}_{\tau+1} - \hat{B}_\tau + Y_\tau$.

When yield is equal to surplus production, the population is in equilibrium.

### Parameter Estimation

A procedure for estimating the parameters of the model from catch and effort data was described by Prager (1992), and will not be repeated. However, the following points are relevant—

- The fitting procedure is similar to one presented by Pella and Tomlinson (1969) in their description of the GENPROD model. The GENPROD model (which includes an estimated exponent) does not allow an analytical solution to its catch equation; but for ASPIC, that solution is given by equations (6).
- The fitting procedure does *not* make an equilibrium assumption. That is, it does not assume that the yield taken in each year is the equilibrium yield.
- The fitting procedure was recently described by Hilborn and Walters (1991), who termed it the "time-series method" of fitting surplus-production models.
- The method as described in Prager (1992) used as the basis of the loss function the sums of squares of residuals in the logarithm of yield, $\sum_{\tau=1}^{T}\left[\log(Y_\tau) - \log(\hat{Y}_\tau)\right]^2$, where $T$ is the total number of years in the data set.

### Other Model Features

Several extensions to the model were described by Prager (1992). These included—

- The use of bootstrapping to obtain approximate nonparametric confidence intervals and hypothesis tests on MSY, optimum effort, and the current status of the stock
- Revised loss functions that include weak constraints on the estimated value of the initial stock biomass
- The decomposition of fishing mortality into gear-specific components. This results in a model that can analyze several simultaneous or sequential series of catch–effort data on the same stock
- Estimation of changes in catchability over time, although Prager (1991) warned that these might be difficult to distinguish from changes in population abundance

## NEW EXTENSIONS TO THE MODEL

The additional extensions described here are a revised loss function; the use of iterative reweighting to obtain asymptotically optimum series weights; incorporation of the iterative reweighting into a bootstrap; tuning the model to year-end indices; and formulating projections based on catch limits. This section describes each extension in turn.

### Revised Loss Function

Equations (6) provide an estimate of yield conditional on the observed fishing effort and estimates of the model parameters $B_1$, $K$, $r$, and $q$. Rearranging the equations provides an alternative solution, an estimate of fishing mortality rate (or equivalently, fishing effort rate) conditional on the parameters and the observed yield. This solution is

(9a) $\hat{F}_\tau = \dfrac{\beta Y_\tau}{\ln\left[\dfrac{\beta \hat{B}_\tau (e^{\alpha_\tau} - 1)}{\alpha_\tau} + 1\right]}$ if $\alpha \neq 0$, or

(9b) $\hat{F}_\tau = \dfrac{\beta Y_\tau}{\ln\left[1 + \beta B_\tau\right]}$ if $\alpha = 0$.

Because $\alpha_\tau$ includes $F_\tau$, $F$ appears on both sides of equation (9a), and to obtain an estimate of $F_\tau$ an iterative solution is necessary. This is obtained by putting a starting guess $\tilde{F}_\tau$ into the right-hand side of the equation, solving, and substituting the result repeatedly until convergence is achieved. A logical starting guess is $\tilde{F}_\tau = Y_\tau / \hat{B}_\tau$.

Estimating the fishing mortality rate from the observed yield makes it possible to accumulate a loss function of residuals in effort or the logarithm of effort. From a statistical point of view, this is probably better than accumulating residuals in yield, because yield is usually measured more precisely than fishing effort (especially effort in constant units). A practical advantage of accumulating residuals in effort is that it greatly simplifies the analysis of data with a few missing values of effort.

Despite these advantages, this approach is more complex than using a loss function in yield. One reason is that the iterative solution of equation (9a) demands more computer time than the closed-form solution of equation (6a). The second reason involves a fundamental distinction between estimating yield and estimating effort. One the one hand, one can *always* estimate the yield that should result from a given starting biomass and a specified rate of fishing effort. On the other hand, even an infinite rate of effort cannot attain sufficiently high values of yield from some values of starting biomass. Under these circumstances, the catch equation [(9a) or (9b)] has no solution for effort conditional on yield, and a computational tactic must be developed for treating years in which the observed yield cannot be matched exactly. One that has been used elsewhere (Methot 1989, 1990) is placing a constraint on the maximum allowable value of $F$ (and consequently of $f$). When the estimate of $F$ reaches this constraint, a residual in yield is added to the loss function along with the usual residual in effort. This tactic is employed in the ASPIC computer program developed by the author.

### Iterative Reweighting

It has proved useful in stock assessments using this model to divide the annual fishing mortality among several different fisheries according to the observed data. This can reduce the need to standardize among gears before conducting the assessment. A question that often arises is what statistical weight should be placed on observations from each series. A simple answer is to weight each observation equally. However, if the series are weighted inversely to their variances, the least-squares solution is asymptotically a maximum-likelihood solution and thus asymptotically unbiased and of minimum variance. In general, the variance of each data series is estimated by the mean-squared error of the series. Asymptotically optimal weighting can be achieved by repeatedly using inverse-variance weights obtained from the MSEs and recomputing the model. This is a type of EM algorithm, and the estimated weights will eventually converge.

Although EM algorithms always converge, the convergence can be quite slow. In translating ASPIC into a computer program, this problem has been addressed through the use of an ad hoc convergence criterion. The criterion is based on the concept of angular deviation. The angle $\theta$ between two vectors $\vec{u}$ and $\vec{v}$, which in this case are two successive vectors of statistical weights, is given by the linear algebra formula

$$(10) \qquad \cos(\theta) = \frac{\vec{u}\cdot\vec{v}}{|\vec{u}|\,|\vec{v}|}.$$

Iterative reweighting is usually continued until $\theta$ is extremely small, approaching the limit of precision of the computer used. However, in practical cases of which I am aware, small changes in weighting usually result in very small changes in estimates of the parameters of interest. For that reason, the author's ASPIC computer program terminates reweighting when the angle is less than $0.01 \times 2\pi$ radians (3.6°). This criterion is usually reached in five or fewer iterations.

### Iteratively-Reweighted Bootstrap

How can a bootstrap capture the variability inherent in iterative reweighting? There is no single solution to this question; here, I describe one approach that seems within the bounds of statistical theory. The sequence of the analysis is as follows:

1. The analysis begins with an initial fit, which is iteratively reweighted as described in the preceding section. From this fit, normalized residuals (residuals adjusted by the appropriate series weights) and estimated values are saved.
2. In each realization of the bootstrap, each observed value of fishing effort is replaced by its estimated value, adjusted by a normalized residual. These residuals are randomly chosen with replacement from the set of saved residuals.
3. In each realization, iterative reweighting is performed prior to fitting. Because the saved residuals were normalized by the weights that were estimated iteratively during the initial fit, each series is expected to have the same variance. However, sampling variability will cause the estimated variances to differ. Thus, this step captures some of the variability inherent in estimating the variances from the observed data.
4. As in any bootstrapped analysis, results of each realization are saved. At the conclusion of the analysis, the multivariate distribution of results is used to summarize the parameter estimates, construct approximate nonparametric confidence intervals, and perform approximate nonparametric tests of hypotheses.

### End-of-Year Biomass Index

The ASPIC framework as detailed by Prager (1992) can "tune" its estimates to fishery-independent indices of abundance at the beginning of the year or representing average values for the year. This is accomplished by assuming that in year $\tau$ the index $I_\tau$ is related to the true biomass $B_\tau$ by an estimated proportionality coefficient $q_I$. The beginning-of-year comparison is made by forming a residual between $q_I I_\tau$ and the model's estimate of starting biomass $\hat{B}_\tau$. This average comparison is made by forming the residual between $q_I I_\tau$ and the model's estimate of average biomass $\hat{\bar{B}}_\tau$, as given in equation (7). A simple extension has been made to the ASPIC computer program to use an index that reflects abundance at year-end. In this case, a residual is formed between $q_I I_\tau$ and the estimated abundance at the start of the next year $\hat{B}_{\tau+1}$. Although this seems identical to the start-of-year case, making special provision for it can avoid missing data at the end of the series in some practical cases.

### Projections Based on TACs

The theory of production modeling allows projections to be made easily, and the solution embodied in equations (9) allows them to be based on catch quotas. By using the bootstrap, approximate nonparametric confidence intervals can be placed around projected biomass levels. This combination of elements can be used in the following way:

1. Define the catch strategy (e.g., the level of TAC) to be examined via the simulation.
2. Perform a bootstrapped analysis of the stock in question, saving results from each realization of the bootstrap. The major items to save are estimates of the parameters $r$ and $K$ and the estimate of biomass at the end of the final year.
3. Using the bootstrap results and equations (3), (4), and (9), project a simulated population forward for the years of the simulation. Assume that the TAC is taken in each year, and estimate the effort (and fishing mortality rate) needed to take it. This is done once for each realization of the bootstrap.
4. From this simulation, one obtains a projection of the population biomass for each realization of the bootstrap. The median of each year's projection is suggested as a representative projection; approximate nonparametric confidence intervals can be constructed from percentiles of the distribution.

## DISCUSSION

Except for the iteratively-reweighted bootstrap, each of the ideas discussed here has been translated into FORTRAN code, tested on simple simulated data, and used in assessments or assessment attempts. While the ideas do not represent revolutionary steps in modeling theory, they do represent sound, practical improvements to the ASPIC modeling framework. There are many possible paths that future efforts could take. A list of issues includes these:

- More extensive testing of various model options on simulated data, to clarify model performance in the presence of different levels of noise, trend, etc., in both the dependent and independent variables. A particular point of interest is under what conditions the bootstrapped median or the point estimate tends to be a better estimator of MSY and other quantities.
- Implementation of the iteratively reweighted bootstrap.
- Computation of $f_{0.1}$ for each data series, as another benchmark of interest.
- Possible reparameterization of the ASPIC model from using $r$ and $K$ to more orthogonal quantities. While realized parameter estimates should not be affected by this change, the change might speed computation considerably (Ratkowsky, 1983).
- Automated outlier elimination from the bootstrap.

## REFERENCES CITED


Graham, M. 1935. Modem theory of exploiting a fishery, and application to North Sea trawling. J. Cons. Int. Explor. Mer 10:264–274.

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment. Chapman and Hall, New York. 570 p.

Lotka, A. J. 1924. Elements of physical biology. Reprinted 1956 as "Elements of mathematical biology" by Dover Press, N.Y. 465 p.

Methot, R. M. 1989. Synthetic estimates of historical abundance and mortality for northern anchovy. Am. Fish. Soc. Symp. 6:66–82.

Methot, R. M. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. *In* Low, L.-L. (ed.), Proc. Symp. Appl. Stock Assess. Techn. Gadids, p. 259–277. ICNAF Bull. 50.

Pella, J. J. 1967. A study of methods to estimate the schaefer model parameters with special reference to the yellowfin tuna fishery in the eastern tropical Pacific Ocean. Doctoral dissertation, University of Washington.

Pella, J. J. and P. K. Tomlinson. 1969. A generalized stock procuction model. Bull. Inter-Am. Trop. Tuna Comm. 13:419–496.

Prager, M. H. 1991. User's manual for ASPIC: A stock-production model incorporating covariates. Miami Lab. Doc. MIA–91/92–20 (available from the author).

Prager, M. H. 1992. ASPIC—A surplus-production model incorporating covariates. ICCAT Coll. Vol. Sci. Pap. 37:218–229 (SCRS/91/24, rev.).

Ratkowsky, D. A. 1983. Nonlinear regression modeling. Marcel Dekker, New York. 276 p.

Schaefer, M. B. 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Bull. Inter-Am. Trop. Tuna Comm. 1(2):27–56.

Schaefer, M. B. 1957. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical Pacific Ocean. Inter-Am. Trop. Tuna Comm. Bull. 2:247–268.