AGENDA

1. Opening, adoption of the Agenda and meeting arrangements.
2. Manual for CPUE standardization
3. Review of methods to address species targeting and gear/species overlap during CPUE standardization
4. Influence of life history characteristics, environmental variability and gear selectivity on Status Determination with respect to the Convention objectives
5. Other matters
6. Recommendations
7. Adoption of the report and closure

LIST OF PARTICIPANTS

SCRS CHAIRMAN
Scott, Gerald P.
SCRS Chairman, NOAA Fisheries, Southeast Fisheries Science Center Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, Florida 33149
Tel: +1 305 361 4261, Fax: +1 305 361 4219, E-Mail: gerry.scott@noaa.gov

CONTRACTING PARTIES

BRAZIL
Minte-Vera, Carolina V.
Universidade Estadual de Maringá, Av. Colombo, nº 5790 - Bloco H90, Mirangá, PR
Tel: +55 44 3261 4622, Fax: , E-Mail: cminte@nupelia.uem.br

CANADA
Neilson, John D.
Head, Large Pelagics Projects, Population Ecology Section, St. Andrews Biological Station, Fisheries and Oceans, Canada, 531 Brandy Cove Road, St. Andrews, New Brunswick
Tel: +1 506 529 4922, Fax: +1 506 529 5862, E-Mail: neilsonj@mar.dfo-mpo.gc.ca

EUROPEAN COMMUNITY
Abascal Crespo, Francisco Javier
Instituto Español de Oceanografía, Centro Oceanográfico de Canarias, Apdo. Correos 1373, 38005 Santa Cruz de Tenerife, ESPAÑA
Tel: +34 922 549 400, Fax: +34 922 549 554, E-Mail: francisco.abascal@ca.ieo.es

De Bruyn, Paul
AZTI - Tecnalia, Herrera Kaia Portualdea z/g, 20110 Pasaia, Gipuzkoa, ESPAÑA
Tel: +34 943 004800, Fax: +34 943 004801, E-Mail: pdebruyn@pas.azti.es

Di Natale, Antonio
Director-AQUASTUDIO, Via Trapani, nº 6, 98121, Messina, Sicilia, ITALY
Tel: +39 090 346 408, Fax: +39 090 364 560, E-Mail: adinatale@acquariodigenova.it

Gaertner, Daniel
I.R.D. UR nº 109 Centre de Recherche Halieutique Méditerranéenne et Tropicale, , Avenue Jean Monnet - B.P. 171, 34203, Sète Cedex, FRANCE
Tel: +33 4 99 57 32 31, Fax: +33 4 99 57 32 95, E-Mail: gaertner@ird.fr

Ortiz de Urbina, Jose Maria
Instituto Español de Oceanografía, C.O de Málaga, , Apartado 285 - Puerto Pesquero s/n, 29640, Fuengirola, Málaga, ESPAÑA
Tel: +34 952 47 1907, Fax: +34 952 463 808, E-Mail: urbina@ma.ieo.es
Ortiz de Zárate Vidal, Victoria
Ministerio de Educación y Ciencia, Instituto Español de Oceanografía, Promontorio de San Martín s/n, 39012, Santander, Cantabria, ESPAÑA
Tel: +34 942 29 10 60, Fax: +34 942 27 50 72, E-Mail: victoria.zarate@st.ieo.es

Pereira, Joao Gil
Universidade dos Açores, Departamento de Oceanografia e Pescas, 9900 Horta, PORTUGAL
Tel: +351 292 200 431, Fax: +351 292 200 411, E-Mail: pereira@ua.pt

Rodríguez-Marín, Enrique
Instituto Español de Oceanografía, C.O. de Santander, Promontorio de San Martín s/n, 39004 Santander, Cantabria, ESPAÑA
Tel: +34 942 29 10 60, Fax: +34 942 27 50 72, E-Mail: rodriguez.marin@st.ieo.es

JAPAN
Takeuchi, Yukio
National Research Institute of Far Seas Fisheries, Fisheries Research Agency of Japan, Mathematical Biology Section- Pelagic Resource Division, 7-1, 5 Chome Orido, Shizuoka Shi, Shimizu-ku
Tel: +81 543 36 6039, Fax: +81 543 35 9642, E-Mail: yukiot@fra.affrc.go.jp

MOROCCO
Idrissi, M’Hammed
Chef, Centre Régional de l’INRH á Tanger, , B.P. 5268, 90000, Drabeb, Tanger
Tel: +212 39 325 134, Fax: +212 39 325 139, E-Mail: mha_idrissi2002@yahoo.com;m.idrissi.inrh@gmail.com

TURKEY
Karakulak, Saadet
Faculty of Fisheries, Istanbul University, Ordu Cad. Nº 200, 34470, Laleli, Istanbul
Tel: +90 212 455 5700/16418, Fax: +90 212 514 0379, E-Mail: karakul@istanbul.edu.tr

UNITED STATES
Cass-Calay, Shannon
NOAA Fisheries, Southeast Fisheries Center, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, Florida 33149
Tel: +1 305 361 4231, Fax: +1 305 361 4562, E-Mail: shannon.calay@noaa.gov

Goodyear, Phil
1214 North Lakeshore Drive, Niceville, Florida 32518
Tel: +1 850 897 2666, Fax: +1 850897 2666, E-Mail: philgoodyear@cox.net

Schirripa, Michael
NOAA Fisheries, Southeast Fisheries Science Center, 75 Virginia Beach Drive, Miami, Florida 33149
Tel: +1 305 361 4288, Fax: +1 305 361 4562, E-Mail: michael.schirripa@noaa.gov

OBSERVERS

Non-governmental Organizations

Federacion of European Aquaculture Producers-FEAP
Caruana, Saviour
Tarxien Road, Ghaxaq, MALTA
Tel: +356 2180 9460, Fax: +356 2180 9462, E-Mail:saviour@mobileemailvodafone.com.mt

**********

ICCAT SECRETARIAT
C/Corazón de María, 8 – 6º planta; 28002 Madrid
Tel: +3491 4165600; Fax: +3491 4152612; E-Mail: info@iccat.int

Restrepo, Victor
Pallarés, Pilar
Appendix 3

LIST OF DOCUMENTS

SCRS/2009/027 Conveying the overall situation of the ICCAT stocks through the use of face plots. Schirripa, M.J. and V. R. Restrepo


SCRS/2009/030 Common trends model in catch per unit of effort for the tropical tunas. Gaertner, D.

SCRS/2009/031 An exploration of targeting variables in the Canadian swordfish longline CPUE. Stacey D. Paul and John D. Neilson

Appendix 4

POTENTIAL SOURCES
FOR A MANUAL ON CPUE STANDARDIZATION

**MODEL DIAGNOSTICS** (Section provided by M. Ortiz)

The common practice of model fitting involves the selection of a model class appropriate to the data, fitting and results presentation. Comparisons between alternative models are normally related to “goodness” of fit to the data, and the selected model is commonly ‘the best fit’ according to some specific criteria, such as log likelihood, AIC or similar statistics. However, seldom there is a report of the model diagnostics and evaluation of “appropriateness” of the model implicit assumptions to the data structure. Selection of the model class is usually the first step and decisions should take into account the nature of the data and objective(s) of the analysis mainly. But, even after a “proper” selection of the class model, the data themselves may show inconsistencies in the fitting process. These inconsistencies can arise from: a) the data as whole show systematic departures of the model assumptions, and or b) that a subset of observations shows discrepancies from the rest of the data.

Therefore model checking should be an integral part of the model fitting. Based on diagnostics the modeler will introduce a verification step in the process assuring that the selected model is the appropriate one given the data analyzed. McCullagh and Nelder (1989) classified the model checking techniques into two groups: a) informal, those that relay on subjective human decision to determine patterns or better, departure from expected patterns. And b) formal techniques that imply a wider model (where the “selected current model” is a subset) with higher number of parameters. In this case, the current model passes the check if it can demonstrate that the extra parameter(s) in the wider model did not improve the fitting of the data.

**Model checking elements**

Model checking was originally developed for classical linear models, then McCullagh and Nelder (1989) extended it to Generalized Linear Models (GLMs). Analysis of residuals are the primarily element for model checking, however other components in model checking include the fitted values, the linear predictors, the residual variance, the dispersion parameter and the elements of the projection (‘hat’) matrix.

The classical residual definition; \( r = y - \mu \) is often called “response residual” in GLMs (R, S-Plus, SAS) and is simply the difference between the observed value and the corresponding predicted by the model. Depending upon the GLM error model distribution, residuals are often transformed such its variance is constant and have a “normal like” behavior for model checking purposes, these residuals are called “standardized residuals” and include: i) the “Pearson residuals”, where response residuals are scaled by the estimated standard deviation of the
response variable \( r_p = \frac{y - \mu}{\sqrt{V(\mu)}} \). ii) The ‘studentized’ standardized residuals, where response residuals are scaled by the corresponding element of the “hat” matrix \( r_z = \frac{y_i - \mu_i}{\sqrt{1 - h_i}} \). iii) And, the Anscombe residual, where the residual is replaced by a function \( A(y) \) that makes \( A(Y) \) “as normal as possible”, where \( A \) is a transformation function specific for each type of error distribution. For example for the Poisson distribution the Anscombe residual is estimated as \( r_A = \frac{2^{1/2} y^{2/3} - \mu^{2/3}}{\mu^{1/6}} \).

Other types of residuals are available from GLMs, including: iv) Deviance residuals, where each residual reflects the contribution to the total Deviance of the GLM fit, \( r_D = \text{sign}(y - \mu) \sqrt{d_i} \), such \( \sum r_D^2 = D \). Of course, for each error distribution the Deviance formulation varies (pp 34 McCullagh and Nelder 1989). In the case of the Poisson distribution, the deviance residuals are \( r_D = \text{sign}(y - \mu) \left\{ 2(y \log \left( \frac{y}{\mu} \right) - y + \mu) \right\}^{1/2} \). Both the Pearson and Deviance residuals can be “studentized” scaling them by the variance and/or dispersion parameter, and the corresponding “hat” element, such \( r_P' = \frac{y\hat{\mu}}{\sqrt{\Phi V(\hat{\mu})(1-h)}} \) for the standardized Pearson residual, and \( r_D' = \frac{r_D}{\sqrt{\Phi(1-h)}} \) for the standardized deviance residual. v) Working residuals, these residuals correspond to the “linear” last step of the iterative fit solution of the GLM, where the \( y \) values are in the scale of the link function and the \( \mu \) represent the linear predictors of the GLM model component.

According to McCullagh and Nelder (1989) in general deviance residuals are preferred for model checking procedures because of its distributional properties are more similar to the linear regression models.

Systematic departures from model assumptions

The following diagnostics try identifying inconsistencies between the ‘model selected’ and the data structure as a whole. We want to confirm the assumptions regarding a) the error or variance function, b) the link function selection, and c) possible omitted linear factors.

Residual Plots

Residual plots generally include plots of standardized deviance residuals against the linear predictors (\( \eta_i \)) of the model, or if not available, the fitted values transformed to the constant information scale of the error distribution assumed. In the later case, the transform values are estimated as

- \( \hat{\mu} \) for Normal error distribution (Gaussian)
- \( 2 \sqrt{\hat{\mu}} \) for Poisson error distribution
- \( 2 \sin^{-1} \sqrt{\hat{\mu}} \) for Binomial error distribution
- \( 2 \log \hat{\mu} \) for gamma error distribution
- \( -2\hat{\mu}^{-1/2} \) for inverse Gaussian error distribution.

The expected pattern of this plot is a distribution of residuals with constant range and with mean of zero. If there is large number of observations, it is recommended to use smoother functions to facilitate the interpretation of patterns. Departures of the expected pattern, for example curvature of the mean, may indicate incorrect choice of the link function, wrong scale of one or more covariates (linear components in model), or omission of higher order terms in the model. A non-constant range of residual indicates an incorrect variance function assumption in the model. This plot is not informative for binary data (0, 1).

A plot for checking the variance function assumption is the absolute residuals against fitted values in a constant-information scale. The expected pattern is no trend with constant distribution on the horizontal scale. A positive trend indicates that the variance function increases too slowly compare to the mean values of the data, in contrast a negative trend indicates that the variance function increases much faster than the mean values of the data.
A plot for checking the link function assumption is the adjusted dependent variable (y) against the linear predictor of the GLM model \( (\eta_{i}) \). The expected pattern in a straight line, curvatures in the plot indicate a low or high power in the exponential link assumption. This plot is not informative for binary data.

A plot for checking the scale of factors/covariates is the standardized residuals against an explanatory variable(s) of the linear predictor formula. The expected pattern is also a constant range distribution with a mean distribution about zero. Departures indicate possible missing interactions or quadratic terms in some factor(s), incorrect scale for the explanatory variable or improper link function. An alternative plot is also the partial residual plot for each factor/covariate. Where the partial residual is estimated as \( u = y - \hat{\eta} + \hat{\gamma}x \) where \( y \) is the adjusted dependent variable, \( \hat{\eta} \) the linear predictor, and \( \hat{\gamma} \) is the parameter estimate for the explanatory variable \( x \). The expected pattern is a linear trend, departures may indicate the wrong scale of the factor, however it may also be due to wrong scale of another correlated factor, in this case multiple residual plots may help to identify common issues.

**Isolated departures from the model**

The prior section addressed mainly overall data inconsistencies with the selected model assumptions, once a final model can be confirm the next step is to evaluate particular points in the data especially those that appear in difference with the remain bulk of the observations. In this regard, residual analysis focus on the influence and/or “leverage” of a given observation, or their inconsistency with the general trend and the effect(s) on the parameters estimated by the model.

**Leverage**

For linear models the measure of leverage is given by the diagonal elements of the ‘hat’ matrix that project the data onto the fitted values. The leverages \( h_i \) represent the influence of a given point in the fit. In classical linear models they are a function of \( X \) only, but in GLM, they also depend on the weights assigned by the Iterative Re-weighting Least Square (IRWLS) algorithms for solving the GLM models. Large value of \( h_i \) indicates that the fit may be sensitive to the response observation \( i \). A plot of leverage values indicating those values of \( h_i > 2p/n \), or approximately \( h_i > 2 \), where \( h_i = nh_i/p \) is an informative tool.

**Influence**

Influence is normally measure as weighted combination of the changes of estimates with and without a given extreme point. Cooke (1977) introduce an statistic, ‘Cooks distances’ that approximates the residual scaled difference between the model fit with observation response for case \( i \) and the model fit without the observation \( i \). The approximation by Williams (1987) is \( (y - \bar{\mu})\sqrt{(1 - h_i)r_i^2} + h_i r_i^2 \), where \( r_i^2 \) is the studentized deviance residual, and \( r_i^2 \) is the studentized Pearson residual.

A plot of Cook’s distances will show the most influential observation(s) in the model fit. Q-Q plots of residuals are standard plots for checking linear models or GLM with normal error assumptions; however for other error distributions it should not expect a normal distributed residual pattern, particularly of the response residuals. For outlier detection or influential observations, half-normal plots are also informative. This type of plot will identify in one tail of the distribution those observations with large influence or leverage. Half normal plots of the leverages, Cook’s distances and studentized residuals provide an informal check for observations diagnostics.

**Mixed Models GLM Diagnostics**

In mixed models things get complicated because a) removing data points affects fixed effects and covariance parameter estimates, b) in repeated/ longitudinal studies interested in multivariate influence and not impact of isolated points.

**Objective:** influence analysis in GLMM. Objective: determine which cases are influential and the manner in which they are important to the analysis.

**General Linear Mixed models**

\[ Y = X\beta + Z\gamma + \epsilon \]

\( Y = (n \times 1) \) vector observed data; \( X (n \times p) \) fixed-effects design matrix or rank \( k \), \( Z \) is \( (n \times g) \) random-effects design matrix, \( \gamma \) is \( (g \times 1) \) vector of random effects normal with mean 0 and variance \( G \), \( \beta \) is \( (p \times 1) \) vector fixed
parameters, and ε is (n x 1) vector of model errors with mean 0 and variance R. Parameters are β and all unknowns in the variance matrices G and R (covariance parameters usually refer as θ).

**Parameter estimation in SAS**

Restricted log likelihood of Y given the fixed effects matrix X. Fixed effect profiled form the residual log likelihood and if possible the residual variance \( \sigma^2 \). -2 this function is minimized via a ridge-stabilized Newton-Raphson algorithm (iterative). On convergence, estimates of the fixed effects and predictions of the random effects γ are obtained by solving the mixed model equations. These quantities are:

\[
\hat{\beta} = (X'V(\hat{\theta})^{-1}X)^{-1}X'V(\hat{\theta})^{-1}Y \\
\hat{\gamma} = G(\hat{\theta})Z'V(\hat{\theta})^{-1}(y - X\hat{\beta}) \quad \text{where} \quad V(\theta) = \text{var}[Y] = ZG(\theta)Z' + R(\theta)
\]

Both β and γ depend on the estimates of the covariance parameters. In mixed models data can be considered as conditional or unconditional. If you are interested in the single realization of the data, then you are interested in the conditional distribution \( Y|\gamma \). But if you are interested in the ‘population’ effect, then you are interested in the unconditional Y, called the marginal formulation.

Estimates of fixed effects β depend on the estimates of the covariance parameters θ. If interested in the influence of observation(s) in the analysis, you must determine whether it influence only on the covariance θ or both the fixed and covariance.

In longitudinal or repeated measures analyses you are not interested in a single observation, but mostly in the influence of a ‘set’ of measurements from a single subject. Computation of case deletion diagnostics are only possible if removing single observations and assuming that the covariance parameters are NOT affected by the removal of the observation in question.

In ordinary least squares (OLS) residual sum to zero, not in GLMM necessarily, also leverage can be negative, not in OLS.

**Residuals**

A difference between an observed quantity and its estimated/predicted value. In mixed models, there are marginal residuals (\( r_m \)) and conditional residuals (\( r_c \)).

\[
r_m = y_i - x_i'\hat{\beta} \quad ; \quad r_c = y_i - x_i'\hat{\beta} - z_i'\hat{\gamma}
\]

Studentized residuals are those residuals divided by an estimate of its standard deviation. It that estimate of σ is independent of the i-th observation, it is called external studentization, if the observation i-th contributes to the standard error computation, then it is called internally studentization. These are the available residuals in Proc Mixed SAS (9.1 up)

<table>
<thead>
<tr>
<th>Residual</th>
<th>Marginal</th>
<th>Conditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>( r_m = y_i - x_i'\hat{\beta} )</td>
<td>( r_c = r_{mI} - z_i'\hat{\gamma} )</td>
</tr>
<tr>
<td>Studentized</td>
<td>( r_{mS} = \frac{r_{mI}}{\sqrt{\text{var}[r_{mI}]}}, )</td>
<td>( r_{cS} = \frac{r_{cI}}{\sqrt{\text{var}[r_{cI}]}}, )</td>
</tr>
<tr>
<td>Pearson</td>
<td>( r_{mp} = \frac{r_{mI}}{\sqrt{\text{var}[r_{mI}]}}, )</td>
<td>( r_{cp} = \frac{r_{cI}}{\sqrt{\text{var}[r_{cI}]}}, )</td>
</tr>
<tr>
<td>Scaled</td>
<td>( \hat{c}^{-1}r_m )</td>
<td>( r_{cI} )</td>
</tr>
</tbody>
</table>

**Influence diagnostics in GLMM**

Not always possible, for example if removal of observation and the new G matrix is not positive definite. But informative, indicate that observation is important qualitatively. Basic procedure:

1. Fit the model to the data and obtain estimates of all parameters
2. Remove one or more data points from the analysis and compute updated estimates of model parameters
3. Based on full- and reduced-data estimates, contrast quantities of interest to determine how the absence of the observations changes the analysis.
Influences statistics computed by Proc Mixed:

Overall influence measured as the likelihood distance (or likelihood displacement). Compute the full data parameter estimates $\hat{\beta}$ and estimates based on the reduced data, $\hat{\beta}(U)$. Then the likelihood and restricted likelihood distances are obtained as

$$LD(U) = 2\left\{l(\hat{\beta}) - l(\hat{\beta}(U))\right\}$$
$$RLD(U) = 2\left\{l_p(\hat{\beta}) - l_p(\hat{\beta}(U))\right\}$$

Important to note that $l(\hat{\beta}(U))$ is NOT the likelihood obtained by fitting the model to the reduced data set, but by evaluating the likelihood function based on the full data set (all observations) at the reduced-data estimates. If this global measure suggests that the points in $U$ are influential, next it is to determine the nature of such influence. Points affect: estimates of fixed effects, estimates of the precision of the fixed effects, estimates of the covariance parameters, estimates of the precision of the covariance parameters, or fitted and predicted values.

Change in parameter estimates

Proc Mixed able to compute summary statistics that capture the change in the entire parameter vector (fixed effects and covariance parameters), based on Cook’s D and the multivariate DFFITS statistic. For both statistics, we are concern about large values (indicating that the change in the parameter estimate is large relative to the variability of the estimate).

$$D(\beta) = (\hat{\beta} - \hat{\beta}(U))^\prime var[\hat{\beta}] (\hat{\beta} - \hat{\beta}(U))/\text{rank}(X)$$
$$MDFITS(\beta) = (\hat{\beta} - \hat{\beta}(U))^\prime var[\hat{\beta}(U)] (\hat{\beta} - \hat{\beta}(U))/\text{rank}(X)$$

Change in precision of estimates

Although point estimates may no change, precision (variance of estimates) and change largely. Proc Mixed compute functions of the trace and determinants of the variance matrices based on the full-data and the reduced-data estimates. The benchmarks of no influence are zero for the covariance traces and one for the covariance ratio. These can be computed for $\beta$ and $\theta$.

Effects on fitted and predicted values

Proc Mixed computes the statistic PRESS that measures influence on fitted and predicted values, as the difference between the observed value and the predicted (marginal) mean, where the predicted values is obtained without the observations in question.

$$\hat{e}_i(U) = y_i - x_i'(\hat{\beta}(U))$$
$$PRESS(U) = \sum_{i\in U} \hat{e}_i(U)$$
EXAMPLE OF A REPORT CARD FOR EVALUATION OF ALTERNATIVE INDICES OF ABUNDANCE.
FROM SEDAR-CPUE WORKSHOP MIAMI NOV-2008. (Provided by S. Cass-Calay)

DESCRIPTION OF THE DATA SOURCE

1. Fishery Independent Indices

A. Describe the survey design (e.g. fixed sampling sites, random stratified sampling), location, seasons/months and years of sampling.

B. Describe sampling methodology (e.g. gear, vessel, soak time etc.)

C. Describe any changes in sampling methodology (e.g. gear, vessel, sample design etc.)

D. Describe the variables reported in the data set (e.g. location, time, temperature, catch, effort etc.).

E. What species or species assemblages are targeted by this survey (e.g. red snapper, reef fish, pelagic).

F. Describe the size/age range that the index applies to. Include supporting figures (e.g. size comp) if available.

2. Fishery Dependent Indices

A. Describe the data source and type of fishery (e.g. commercial handline, commercial longline, recreational hook and line etc.).

B. Describe any changes to reporting requirements, variables reported, etc.

C. Describe the variables reported in the data set (e.g. location, time, temperature, catch, effort etc.).

D. Describe the size/age range that the index applies to. Include supporting figures (e.g. size comp) if available.

METHODS

1. Data Reduction and Exclusions

A. Describe any data exclusions (e.g. gears, fishing modes, sampling areas etc.). Report the number of records removed and justify removal.

B. Describe data reduction techniques (if any) used to address targeting (e.g. Stephens and MacCall, 2004; gear configuration, species assemblage etc).

C. Discuss procedures used to identify outliers. How many were identified? Were they excluded?
2. Management Regulations (for FD Indices)

- A. Provide (or cite) history of management regulations (e.g. bag limits, size limits, trip limits, closures etc.).
- B. Describe the effects (if any) of management regulations on CPUE.
- C. Discuss methods used (if any) to minimize the effects of management measures on the CPUE series.

3. Describe Analysis Dataset (after exclusions and other treatments)

- A. Provide tables and/or figures of number of observations by factors (including year, area, etc.) and interaction terms.
- B. Include tables and/or figures of number of positive observations by factors and interaction terms.
- C. Include tables and/or figures of the proportion positive observations by factors and interaction terms.
- D. Include tables and/or figures of average (unstandardized) CPUE by factors and interaction terms.
- E. Include annual maps of locations of survey sites (or fishing trips) and associated catch rates OR supply the raw data needed to construct these maps (Observation, Year, Latitude, Longitude (or statistical grid, area), Catch, Effort).
- F. Describe the effort variable and the units. If more than one effort variable is present in the dataset, justify selection.
- G. What are the units of catch (e.g. numbers or biomass, whole weight, gutted weight, kilograms, pounds).

4. Model Standardization

- A. Describe model structure (e.g. delta-lognormal).
- B. Describe construction of GLM components (e.g. forward selection from null etc.).
- C. Describe inclusion criteria for factors and interactions terms.
- D. Were YEAR*FACTOR interactions included in the model? If so, how (e.g. fixed effect, random effect)? Were random effects tested for significance using a likelihood ratio test?
- E. Provide a table summarizing the construction of the GLM components.
- F. Summarize model statistics of the mixed model formulation(s) (e.g. log likelihood, AIC, BIC etc.).
**MODEL DIAGNOSTICS**

*Comment: Other model structures are possible and acceptable. Please provide appropriate diagnostics to the CPUE indices working group.*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Include plots of the chi-square residuals by factor.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Include plots of predicted and observed proportion of positive trips by year and factor (e.g. year*area)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Report overdispersion parameter and other fit statistics (e.g. chi-square / degrees of freedom)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Include histogram of log(CPUE) or a histogram of the residuals of the model on CPUE. Overlay the expected distribution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Include plots describing error distribution (e.g. Studentized residuals vs. linear predictor).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Include QQ-plot – (e.g. Student deviance residuals vs. theoretical quantiles), Overlay expected distribution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Include diagnostic plot for variance function (e.g. square root of std residuals vs. fitted values). Overlay expected distribution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Include diagnostic plot for link function (e.g. linear response variable vs. linear predictor). Overlay expected distribution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Include plots of the residuals by factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Report overdispersion parameter and other fit statistics (e.g. chi-square / degrees of freedom).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Include plots describing error distribution (e.g. Studentized residuals vs. linear predictor).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Include QQ-plot – (e.g. Student deviance residuals vs. theoretical quantiles), Overlay expected distribution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Include diagnostic plot for variance function (e.g. square root of std residuals vs. fitted values). Overlay expected distribution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Include diagnostic plot for link function (e.g. linear response variable vs. linear predictor). Overlay expected distribution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Include ROC curve to quantify goodness of fit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Include plots describing error distribution (e.g. Studentized residuals vs. linear predictor).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Include QQ-plot (e.g. Student dev. residuals vs. theoretical quantiles), Overlay expected distribution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The feasibility of this diagnostic is still under review.*

**Working Group Comments:**
MODEL DIAGNOSTIC (CONT.)

D. Include diagnostic plot for variance function (e.g. square root of std residuals vs. fitted values). Overlay expected distribution.
E. Include diagnostic plot for link function (e.g. linear response variable vs. linear predictor). Overlay expected distribution.

IF MULTIPLE MODEL STRUCTURES WERE CONSIDERED:
(Note: this is always recommended but required when model diagnostics are poor.)

1. Plot of resulting indices and estimates of variance
2. Table of model statistics (e.g. AIC criteria)

<table>
<thead>
<tr>
<th>Date Received</th>
<th>Workshop Recommendation</th>
<th>Revision Deadline</th>
<th>Author and Rapporteur Signatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Submission</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revision</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Justification of Working Group Recommendation**
REVIEW OF INDEX WEIGHTING SCHEMES FOR STOCK ASSESSMENTS. (Section provided by C. Porch)

Stock assessment algorithms typically incorporate data on stock structure and mortality from a variety of sources including landings, length or age composition, and indices of relative abundance derived from fishery catches or fishery-independent surveys. Usually some data sets are less informative (more uncertain) than others and an important part of the stock assessment process is to determine how to take these differences into account. For example, one may wish to place greater emphasis on a large sample from the entire study population than on a small sample from a fraction of the study population. The challenge, of course, lies in objectively determining the degree to which each data source should be emphasized (weighted). At present, there is no universally accepted method for meeting this challenge, with the predictable result that the choice of weighting schemes often depends as much on who is participating in the assessment as it does on the types of data involved. As this has long been a contentious issue in many forums (Schnute and Hilborn 1993; ICCAT 1999), it is important for SEDAR participants to understand the strengths and weaknesses of the various weighting schemes before adopting one for a particular assessment. To that end, the chapter reviews the weighting schemes that have been developed specifically for indices of abundance and suggests criteria for selecting between those schemes during SEDAR data and assessment workshops.

Measuring uncertainty

Uncertainty can be thought of in terms of observation errors (degree to which the observed value of an index differs from its true value) and process errors (degree to which the true value of an index differs from that of the study population). Observation errors arise due to incomplete or nonrandom sampling and are typically a function of sample size. Process errors arise, for example, when the fraction of the study population that occupies the area sampled for the index varies through time. If these errors are effectively random over the study period, then one may consider them to be unbiased representations of the study population (although they may still be highly imprecise). In that case it is advantageous to have multiple indices representing the same quantity or, when possible, to combine several uncertain indices into one (see Conn’s chapter). Systematic observation or process errors, of course, can lead to bias.

Known biases of a substantive nature should obviously be corrected during the index standardization process or else the index should be discarded. The relevant question here is what to do with indices that are suspected of some unknown bias. In many assessments the available data are too sparse to afford the luxury of simply discarding suspicious indices. For such cases, we maintain that unknown biases among multiple independent data sets may be regarded (hopefully) as canceling out on average and therefore may be treated as another degree of imprecision. It is therefore appropriate simply to deemphasize such data during the fitting process in some proportion to the degree to which bias is suspected (more on this later). Given these caveats, the next order of business is to select a common metric for measuring the uncertainty of each candidate index of abundance.

Most modern stock assessment models are cast in a probabilistic framework, in which case the most natural measure of the uncertainty in an index is its variance. For example, consider two independent normally distributed indices $X_1$ and $X_2$ with the same expectation but different variances ($\sigma_1^2$ and $\sigma_2^2$). The probability of observing two given values $x_1$ and $x_2$ is

\[ p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[ -\frac{(x_1 - \mu)^2}{2\sigma_1^2} - \frac{(x_2 - \mu)^2}{2\sigma_2^2} \right] \]

The maximum likelihood estimator for $\mu$ is the value that maximizes equation 1, or equivalently, minimizes its negative logarithm:

\[ -\ln p(x_1, x_2) = \ln(2\pi\sigma_1\sigma_2) + \frac{(x_1 - \mu)^2}{2\sigma_1^2} + \frac{(x_2 - \mu)^2}{2\sigma_2^2} \]

The solution when $\sigma_1^2$ and $\sigma_2^2$ are known is easily obtained by simple calculus:
As can be seen, the influence of each index series on the final estimate of \( \hat{\mu} \) is directly proportional to the variance associated with that index.

Some authors have adopted somewhat simpler notation where weight coefficients \( (w) \) are specified directly in the likelihood function, e.g.,

\[
-\ln p(x_1, x_2) \propto w_1(x_1 - \mu)^2 + w_2(x_2 - \mu)^2
\]

In this example it is easy to see that the \( w \)'s are merely the reciprocals of the variances and that minimizing (4) or (2) will yield identical results as long as the variances (and \( w \)'s) are constant. In keeping with most statistical texts, we find the variance notation to be more intuitive, particularly when different error structures are mixed such that the relative values of the \( w \) coefficients are no longer comparable. Moreover, the weighting coefficient formulation is impractical when the values of the weights are to be estimated internally in the model.

**Review of index weighting schemes**

The weights applied to the indices of abundance used can have a profound effect on the outcome of an assessment. To illustrate this, consider the following simple model:

\[
x_{1,t} = q_1(N + Gt) + \varepsilon_{1,t}
\]

\[
x_{2,t} = q_2(N + Gt) + \varepsilon_{2,t}
\]

where \( x_{1,t} \) and \( x_{2,t} \) are values of independent indices representing an underlying population that is growing linearly with time \( t \) from a known starting abundance \( N \). If the errors \( \varepsilon_{1,t} \) and \( \varepsilon_{2,t} \) are approximately normal-distributed with equal variances, then maximum likelihood estimates for \( q_1, q_2, \) and \( G \) may be obtained by minimizing

\[
-\ln p(\bar{x}_1, \bar{x}_2) = \sum_t \ln(2\pi\sigma^2) + \frac{(x_{1,t} - q_1(N + Gt))^2}{2\sigma^2} + \frac{(x_{2,t} - q_2(N + Gt))^2}{2\sigma^2}
\]

Otherwise, the appropriate likelihood function for unequal variances is

\[
-\ln p(\bar{x}_1, \bar{x}_2) = \sum_t \ln(2\pi\sigma_1, \sigma_2 \sigma^2) + \frac{(x_{1,t} - q_1(N + Gt))^2}{2\sigma_1^2} + \frac{(x_{2,t} - q_2(N + Gt))^2}{2\sigma_2^2}
\]

In both (6) and (7) the over-arrow denotes the vector of time series observations for each index. It is easy to show that the estimates for \( q_1, q_2, \) and \( G \) obtained from the equal variance formulation (6) will be the same whether \( \sigma^2 \) is estimated or fixed to an arbitrary constant (which is the least squares solution). It is also generally true that the equal-variance (6) and bi-variance (7) estimators will perform similarly when applied to situations where the true variances of each index are the same (and the underlying model is correct). However, if the true variances are substantially different, the assumption of equal variances (application of 6) can lead to biased estimates.

**Figure 1a** and **Table 1** compare the results when the equal-variance (6) and bi-variance (7) estimators were applied to simulated indices generated according to (5) when the true values of \( \sigma_1^2 \) and \( \sigma_2^2 \) were set to 64 and 13, respectively. The bi-variance estimator performed best, producing slightly better parameter estimates and a much lower value for the likelihood function. This would seem to confirm that there is some advantage to estimating separate variances for each index rather than assuming they are the same, however the experiment was predicated on having the correct underlying model. Consider now the same set of experiments when the indices of abundance were equally biased, but in opposite directions such that the annual rate of increase for index 1 is 0.2 and the annual rate for index 2 is 1.8 (compared to the true rate of 1.0). In that case the bi-variance
estimator again produced a significantly lower likelihood function value than the equal-variance estimator (82.6 versus 96.6), and on that basis an investigator might be tempted to prefer it. However, closer inspection reveals that the bi-variance estimator produced this low likelihood value by severely down-weighting the less precise index 1 and focusing on the more precise index 2 (Figure 1b). As a result, the MLEs for $q_2$ and $\mu_2$ were very close to the true values and the predicted values for index 2 matched the observations very well, whereas the MLE for $\mu_1$ was double the true value and the predicted values for index 1 did not match the observations at all. Hence, the estimate of the primary parameter of interest ($G$) was 1.8; matching the true slope of index 2, but much higher than the underlying population. In contrast, the equal variance estimator gave equal credence to the two biased indicators and essentially averaged them, in this case resulting in a much less biased estimate of $G$.

Table 1. Parameter estimates from the equal-variance and bi-variance estimators applied to data generated from model 5.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>TRUE</th>
<th>Unbiased indices</th>
<th>Conflicting indices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Equal-var.</td>
<td>Bi-variance</td>
</tr>
<tr>
<td>$G$</td>
<td>1</td>
<td>0.98</td>
<td>1.01</td>
</tr>
<tr>
<td>$q_1$</td>
<td>2</td>
<td>2.04</td>
<td>1.99</td>
</tr>
<tr>
<td>$q_2$</td>
<td>1</td>
<td>1.02</td>
<td>0.99</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>64</td>
<td>46</td>
<td>64</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>13</td>
<td>46</td>
<td>13</td>
</tr>
</tbody>
</table>

The above example above was contrived to make the points that (1) weighting matters and (2) the estimation of variances is inextricably linked to the assumption that the model underlying the estimator is correct. While one variance scheme may appear better on the basis of the values of the likelihood function or other metric, it may in fact exacerbate potential biases by favoring solutions which disregard indices that are difficult to reconcile with the model but nevertheless accurate. For this, and other reasons that will become apparent, there is no widely accepted protocol for choosing a weighting scheme. Therefore it behooves the analyst to be familiar with both the indices being used and with the strengths and weaknesses of the various weighting options. The remaining discussion presents six basic approaches to index weighting. Most of the weighting schemes developed to date are essentially special cases of these six categories. The implementation of each scheme depends, of course, on the type of error structure assumed. The normal and lognormal error structures seem to be the most common choices in the literature; the latter in particular because it places less weight on outliers and because the variance of many indices tends to increase with the mean. Other error structures (e.g., gamma, Poisson, negative binomial) have also been considered, but the basic principles are the same. For the sake of brevity, equations are shown only for the normal and lognormal cases.

Equal Weighting. The simplest weighting scheme is the so-called “equal weighting” option, where all observations are essentially assigned the same variance (i.e., all indices $j$ in all years $y$) so that the likelihood function has the form:

$$ L = \frac{\sum_{j,y} (x_{j,y} - \mu_{j,y})^2}{2\sigma^2} + \ln \sigma \quad \text{normal} $$

$$ \frac{\sum_{j,y} (\ln x_{j,y} - \mu_{j,y})^2}{2\sigma_{\ln x}^2} + \ln \sigma_{\ln x} \quad \text{lognormal} $$

Note that $\mu$ is the expectation in the case of the normal distribution and the median in the case of the lognormal distribution. Also, the variance term in the normal case is on an arithmetic scale whereas in the lognormal case it is on a logarithmic scale (i.e., the variance of $\ln x$). Hence, the implications of equal weighting are somewhat different between the two error structures; constant variance in the normal case and constant coefficient of variation (CV) in the lognormal case (recall $\sigma_{\ln x}^2 = \ln(1+CV^2)$). It is also important to note that in the case of a normal distribution, the annual values for each index should be rescaled so that their means are similar (e.g., by dividing the original series by their respective means), otherwise series with larger absolute values will receive more weight in the fitting process than series with low values. In the lognormal case the scaling is implicit and the results will be the same whether the series is rescaled or not. In equal-weighting models where the uncertainty is limited to the indices of abundance (e.g., catch is assumed known without error), the estimates of $G$ will not depend on the value of the variance parameter (or whether it is estimated or not), but will depend on
the choice between lognormal and normal errors.

**Simple maximum likelihood weighting.** A simple extension of the equal weighting procedure is to assume the variance parameter for a given index is constant from year to year, but may differ among indices:

\[
L = \left\{ \begin{array}{ll}
\sum_{j,y} \left( \frac{(x_{j,y} - \mu_{j,y})^2}{2\sigma_{j,y}^2} + \ln \sigma_j \right) & \text{normal} \\
\sum_{j,y} \left( \frac{(\ln x_{j,y} - \mu_{j,y})^2}{2\sigma_{\ln x,j,y}^2} + \ln \sigma_{\ln x,j} \right) & \text{lognormal}
\end{array} \right.
\]

In this case the variance parameters may be fixed to some externally derived values (a special case of input variance weighting discussed below) or estimated internally in the model. Maximum likelihood estimates of the variances have been derived in three common ways. An early method was referred to as iterative re-weighting because it consisted of a two step process where the likelihood was minimized using fixed values for each variance parameter, after which the variances were recomputed using the standard MLEs.

\[
\begin{align*}
\sigma_j^2 &= \frac{1}{n} \sum_y (x_{j,y} - \mu_{j,y})^2 & \text{normal} \\
\sigma_{\ln x,j,y}^2 &= \frac{1}{n} \sum_y (\ln x_{j,y} - \mu_{j,y})^2 & \text{log normal}
\end{align*}
\]

The new estimates of variance would be substituted back into the likelihood and the minimization procedure reinitiated until the estimates of variance converged (essentially an application of the EM algorithm). Subsequently, it was realized that the concentrated likelihood approach, whereby the formulas in (10) are directly substituted into (9), was more efficient and yielded the same results. Alternatively, equivalent MLEs may be obtained by use of the search algorithm (minimizing 9) in the same manner as for the other estimated parameters (Porch 2002). Note that annual variances cannot be estimated using these techniques because the number of index parameters would exceed the number of index data points.

One potential pitfall of simple maximum likelihood weighting is that some index series may be severely down-weighted, as illustrated in the example. This is especially problematic if the assessment model is mis-specified and happens to match up (incorrectly) with some indices more than others. Simulations by Legault and Porch (2001) that were conditioned on data for Atlantic bluefin tuna (with various minor model mis-specifications) indicated no clear pattern in the bias or uncertainty of estimates between the equal weighting and maximum likelihood schemes.

**Input-variance weighting.** This common alternative to equal weighting simply involves specifying the variance parameters based on considerations external to the stock assessment model. For example, estimates of the variance of an index may be derived for each year during the standardization process (e.g., by application of a general linear model). To the extent that they are reasonable estimates of year to year changes in index uncertainty, incorporating such externally derived variances can lead to more accurate assessments (Maunder and Starr 2003).

\[
L = \left\{ \begin{array}{ll}
\sum_{j,y} \left( \frac{(x_{j,y} - \mu_{j,y})^2}{2\sigma_{j,y}^2} + \ln \sigma_{j,y} \right) & \text{normal} \\
\sum_{j,y} \left( \frac{(\ln x_{j,y} - \mu_{j,y})^2}{2\sigma_{\ln x,j,y}^2} + \ln \sigma_{\ln x,j,y} \right) & \text{log normal}
\end{array} \right.
\]

Note that, in principle, it is possible to estimate the annual variance parameters for each index if the standardization procedure is conducted simultaneously with the stock assessment (Maunder 2001). In practice, however, the process is prohibitively time consuming.
Some scientists have advocated developing input weights based entirely on expert opinion, arguing that issues such as spatial coverage and potential for bias outweigh consideration of observation error. For example, Suzuki (2001) presented a framework where indices for bluefin tuna would be graded (on a scale of 1 to 5) on each of three factors: spatial coverage, temporal coverage, and the likely importance of changes in the operational characteristics of the index over time that are not accounted for during the index standardization process. Each index would then be weighted in accordance with the sum of those three scores (i.e., the variances would be set equal to the reciprocal of those scores).

Additional-variance weighting. This alternative is based on the premise that the variance of an index may be decomposed into a component reflecting observation errors, such as might be obtained from the GLM standardization, and an additional component characterizing the uncertainty in the index as a reflection of the true abundance trends, as might occur when the proportion of the stock sampled by the index varies from year to year (Wade 1996, Punt et al. 1997, Geromont and Butterworth 2001). To the extent that these two error sources are independent, one might expect their variances to be additive:

\[
L = \frac{1}{2} \sum_{j,y} \left( \frac{(x_{j,y} - \mu_{j,y})^2}{\sigma_{j,y}^2 + \chi_j^2} + \ln(\sigma_{j,y}^2 + \chi_j^2) \right) \quad \text{normal}
\]

\[
L = \frac{1}{2} \sum_{j,y} \left( \frac{(\ln x_{j,y} - \mu_{j,y})^2}{\sigma_{\ln x,j,y}^2 + \chi_j^2} + \ln(\sigma_{\ln x,j,y}^2 + \chi_j^2) \right) \quad \text{log normal}
\]

where \( \chi_j^2 \) is the additional variance term which may vary among indices and can be estimated by maximum likelihood methods analogous to those discussed in connection with simple maximum likelihood weighting (equations 9 and 10). This method mitigates the pitfalls of equal weighting (e.g., giving equal credence to observations based on little data) and simple maximum likelihood weighting (e.g., all of the weight going to a few series), but still depends on the relative consistency of each series with the underlying model structure. The approach would be undesirable if there were strong a priori reasons for weighting one series more highly than another and there was uncertainty about model structure.

Multiplicative-variance weighting. This approach is similar to the additional variance approach above except that the scaling parameter is multiplied by the year-specific input values rather than added to it:

\[
L = \frac{1}{2} \sum_{j,y} \left( \frac{(x_{j,y} - \mu_{j,y})^2}{c_j \sigma_{j,y}^2} + \ln(c_j \sigma_{j,y}^2) \right) \quad \text{normal}
\]

\[
L = \frac{1}{2} \sum_{j,y} \left( \frac{(\ln x_{j,y} - \mu_{j,y})^2}{c_j \sigma_{\ln x,j,y}^2} + \ln(c_j \sigma_{\ln x,j,y}^2) \right) \quad \text{log normal}
\]

In many applications the year-specific variances are initially determined by GLM and then the multiplicative parameters \( c_j \) are adjusted to account for expert judgments on key factors not accounted for in the GLM (see Quinn and Deriso 1999; Sullivan et al. 1999; 2002). In the most recent assessment of Atlantic and Gulf king mackerel, for example, the \( c_j \) values were adjusted such that the average coefficient of variation (CV) attributed to each index was the same, but the relative year to year changes in CV from the GLM were preserved (Cass-Calay et al., 2008).

The multiplicative-variance weighting approach is perhaps less intuitive than the additional variance approach in terms of distinguishing process and observation errors. It is useful, however, for cases where the available estimates of observation variance (\( \sigma_{j,y}^2 \)) are trusted as measures of relative precision from one year to the next, but not as measures of the uncertainty across indices. It is of course also possible to estimate the \( c_j \) values analogous to the simple maximum likelihood method, subject to the same limitations (most notably that some indices may be severely down-weighted, particularly if the assessment model is mis-specified).

Variance as a function of the mean. This last approach is based on the premise that the variance of catch per unit effort (or survey counts) tends to be correlated with the mean
\[ L = \begin{cases} 
\frac{1}{2} \sum_{j=1}^{J} \frac{(x_{j,y} - \mu_{j,y})^2}{f_j(x_{j,y})} + \ln f_j(x_{j,y}) & \text{normal} \\
\frac{1}{2} \sum_{j=1}^{J} \frac{(\ln x_{j,y} - \mu_{j,y})^2}{f_j(\ln x_{j,y})} + \ln f_j(\ln x_{j,y}) & \text{log normal}
\end{cases} \]

where \( f \) is some arbitrary function such as Taylor’s power law: \( f = ax_j^b \). In the trivial case where \( x \) follows a lognormal process with constant CV, the power function reduces to \( b = 0 \) and \( a = \ln(1+CV^2) \). More generally, the relationship (parameters \( a \) and \( b \)) could be established outside the stock assessment model by regressions of variance against the mean. However, it has been shown that such correlations tend to be spurious (see Porch 1998). Another possibility is to estimate the parameters internally, in which case one may expect similar issues to the multiplicative variance case. Regardless of whether the parameters of the mean-variance relationship are estimated internal or external to the assessment model, further bias may occur by overemphasizing observations \( x_{j,y} \) that happen to be very low and it may be prudent to make the procedure more robust by adding a small constant.

Criteria for choosing between index schemes

There are no firm criteria for determining when to use one weighting scheme over another (but see McAllister et al. 2001 for a useful review). While several simulation studies have been conducted, they have all been rather limited in scope and tend to focus on comparing a few special cases of the weighting schemes discussed above. Nevertheless, there are several guidelines that can be posited.

A priori criteria

It is important for the choice of weighting schemes to be rooted in arguments that are based on the nature of the data at hand rather than how they affect the outcome of the assessment. Otherwise, the working group may be accused of succumbing to divisive arguments predicated on producing a desired stock status rather than the best available science. To this end, the working groups should discuss the following points:

1. Are year to year variations in uncertainty likely to be substantial and measurable for a given index? If so:
   - Are they reliable only in a relative sense (as is sometimes the case with model-based estimators such as GLM standardizations).
   - Are they also reliable in terms of absolute magnitude (e.g., a designed-based estimate from a survey that covers the entire range of the stock)

2. Does the level of uncertainty likely vary substantially among indices? If so:
   - Is the available expertise sufficient and able to reach consensus ranking?
   - Is the candidate index-weighting estimation method well-understood (widely-practiced, simulation-tested, and easily applied in the current assessment framework)?

A posteriori criteria

While it is important that the initial choice of a weighting method be made before running the stock assessment model, it is also important to ensure that the weighting method behaves sensibly once applied. In particular, the working group should ask the following questions once the method has been applied:

1. Does the method accord unrealistically high or low variance to some indices?
   - What level should be deemed unrealistic?
   - Do the estimates of variance strongly disagree with expert judgment on the relative reliability of each series as an index of abundance

2. Are estimates statistically defensible?
   - Examine goodness of fit statistics (e.g., Chi-square deviance statistic)
   - Examine model selection criteria (e.g., AICc)
A flow chart is offered in Figure 2 to help facilitate discussions relating to the above criteria. Of course in any given application there may be gray areas. Accordingly, the recommendations offered in the flow chart are intended to be viewed as guidance rather than hard and fast rules.

**Acknowledgments**

I thank Paul Conn, Joe O’Hop and Michael Prager for their helpful suggestions.

**Literature Cited**


Figure 1. Predicted trends in population growth using the equal-variance (red) and bi-variance estimators (blue) compared to the true values (diamonds) and ‘observed’ values for index 1 (circles) and index 2 (triangles). The top and bottom panels refer to the cases when the ‘observed’ values are generated without bias (Fig. 1a) or generated with equal but opposite biases (Fig. 1b).
Index weighting key

Estimates of variance reliable for all years

Estimates of variance not available or do not fully represent primary sources of uncertainty

Overall uncertainty levels deemed similar for all indices used in the assessment model
- Uncertainty levels similar for all years
- Annual variance estimates reliably measure relative uncertainty between years
- Annual variance estimates reliably measure observation errors only
- Annual variance estimates unavailable or uninformative

Overall uncertainty levels differ between indices used in the assessment model
- Expert consensus reached concerning relative weights to be assigned for each index
- Annual variance estimates reliably measure relative uncertainty between years
- Annual variance estimates reliably measure observation errors only
- Annual variance estimates unavailable or uninformative
- Relationship between variance and mean is well defined

Expert consensus not reached (or not preferred) concerning relative weights to be assigned for each index
- Annual variance estimates reliably measure relative uncertainty between years
- Annual variance estimates reliably measure observation errors only
- Annual variance estimates unavailable or uninformative
- Relationship between variance and mean is unclear

Figure 2. Key depicting the recommended chain of decisions with various levels of information on the uncertainty in the candidate indices of abundance. Numbers in parentheses represent the equations associated with each method in the text above. Note that the key is intended for general guidance towards developing initial weighting schemes. Assessment scientists should also consider a posteriori diagnostics (as discussed earlier) as well as the interactions between selected index weighting schemes and the weighting schemes applied to other types of data used in the assessment model (e.g., total catch or age composition).

---

1 The single variance parameter representing all indices may be estimated or fixed based on expert consensus.
2 Scaled so that the mean variance is the same for all index series.
3 Scaled so that the mean variance of each index is consistent with expert opinion.
4 Mean variance of each index estimated by maximum likelihood.
5 In the trivial case of a lognormal process with constant CV, this option is equivalent to a lognormal model with simple maximum likelihood weighting.
Appendix 5

ANALYSIS OF THE VARIABILITY IN SSBMSY IN ICCAT STOCKS FOR WHICH RECENT AGE-BASED ASSESSMENTS WERE AVAILABLE AND ESTIMATION THE EXPECTED RANGE OF VARIABILITY RESULTING FROM A LUMPED BIOMASS FORM OF ASSESSMENT.

Introduction

The working group noted that an important result of SCRS/2009/29 was that the variability in SSB relative to the expected level of SSB$_{MSY}$ was insensitive to the proxy F level used for MSY calculations and the selectivity pattern modeled. This result is similar to the finding in SCRS/1998/120. In view of this, the WG decided to examine the expected variability in SSB$_{MSY}$ in ICCAT stocks for which recent age-based assessments were available. The WG also decided to examine the variability in fishable biomass based on the age-structured analyses used for N SWO, BET and YFT to approximate the expected range of variability resulting from a lumped biomass form of assessment.

Methods

Stocks examined: N-ALB, BET, YFT, N-SWO, W-BFT, E-BFT

For each stock, one thousand bootstrap projections were made using PRO-2BOX (Porch, 2002). The projection input files were those used during the most recent assessment (Table A.1) with the following exceptions:

1) To allow comparability between species, predicted recruitment was modeled using resampling of the observed recruitments (see example below – note that inputs that appear in bold italics are not used by PRO2-BOX if resampling is selected). If the stock assessment had excluded the some observations to avoid projecting poorly estimated recruitment estimates, these exclusions were preserved.

<table>
<thead>
<tr>
<th>Zone/Stock</th>
<th>1</th>
<th>6</th>
<th>400000</th>
<th>3500</th>
<th>-0.6</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>epsilon for autocorrelated recruitment in last year</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rho- autocorrelation coefficient</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>std error of random devs in recruitment</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sb parameter (=ssb hinge for two-line)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>recruitment parameter (= rmax for two-line)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>spawner-recruit curve (1=Bev &amp; Holt, 2=Ricker, 3-5=2-line, 6=resample Obs R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2) All stocks were projected using two scenarios, 1) constant F at F$_{MAX}$ and 2) constant F at 75% of F$_{MAX}$.

Projections were run from the terminal year of the assessment to at least 2100 in order to achieve an equilibrium condition. In all cases, the QUOTA.TXT file was as follows:

```plaintext
**Enter the number of catch/effort scenarios to run
2
#-------------------------------------------------------------------------------------
**SCENARIO 1 Constant F = Fmax to equilibrium
#-------------------------------------------------------------------------------------
**Enter the total allowable catch (mt) for each projection year
 1 1 200*1000000000 ### A sufficiently large catch to be non-limiting
-1
**Enter the fixed fully-selected fishing mortality rates by year
 1 1 200*-1 ### -1 results in projection at Fmax
-1
#-------------------------------------------------------------------------------------
**SCENARIO 2 Constant F = 75%Fmax to equilibrium
#-------------------------------------------------------------------------------------
**Enter the total allowable catch (mt) for each projection year
 1 1 200*1000000000 ### A sufficiently large catch to be non-limiting
-1
**Enter the fixed fully-selected fishing mortality rates by year
 1 1 200*-0.75 ### -0.75 results in projection at Fmax
-1
```
3) In two cases, E-BFT and W-BFT two VPA runs were conducted using different time periods to estimate the stock recruitment relationship (SRR). Both runs were used to inform management. For these stocks, two analyses were conducted that resampled recruitment over the time interval consistent with the VPA assessment runs.

   a) W-BFT: Resample recruitment 1971-2004 (B&H SRR, High Rec)
   b) W-BFT: Resample recruitment 1976-2004 (2-Line, Low Rec)
   c) E-BFT: Resample recruitment 1970-2004 (High Rec)
   d) E-BFT: Resample recruitment 1990-2004 (Low Rec)

Results and Discussion

For each stock, the variability in equilibrium SSB is illustrated in Figs. 1-8. Each figure contains the following information:

- **Panel A)** Probability density function and cumulative frequency of equilibrium SSB/SSB_{FMAX} when projected at constant F=\text{FMAX}
- **Panel B)** Probability density function and cumulative frequency of equilibrium SSB/SSB_{FMAX} when projected at constant F=75\%\text{FMAX}
- **Panel C)** Comparison of panels A and B
- **Panel D)** Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=\text{FMAX}
- **Panel E)** Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=75\%\text{FMAX}

In the cases of stocks for which age-structured assessments are not the primary basis for management advice (e.g. N SWO, BET), or lumped biomass model results were combined with age-structured results to produce management advice (YFT), it was recommended the variability in fishable biomass based on the age-structured analyses be used to approximate the expected range of variability resulting from a lumped biomass form of assessment. These results are shown in Figs. 9-11. The figures are identical in structure to those described above.

The working group also estimated the probability of falling below expected equilibrium biomass ratios due to natural (modeled) variability while fishing at the F_{MSY} (or proxy) level and 75\% F_{MSY}. These results are summarized in Section 4 of the report of the ICCAT Methods Working Group.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Year Assessed</th>
<th>Description of Run</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>YFT</td>
<td>2008</td>
<td>Run 5*</td>
<td>SCRS/2008/016</td>
</tr>
</tbody>
</table>

* YFT: Management Advice was developed using combined results of VPA and production models.
** N-SWO, BET: A production model was used as the base model to provide management advice.
*** N-ALB: Management advice was developed using combined results of VPA and MFCL.
Figure 1. Summary of variability in SSB for yellowfin tuna (YFT). A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium SSB/SSB_{MAX} when projected at constant F=F_{MAX}. B) PDF and CF of equilibrium SSB/SSB_{MAX} when projected at constant F=75%. C) Comparison of panels A and B. D) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=F_{MAX}. E) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=75% F_{MAX}. 

<table>
<thead>
<tr>
<th>Year</th>
<th>F_{MAX}</th>
<th>60%</th>
<th>80%</th>
<th>100%</th>
<th>120%</th>
<th>140%</th>
<th>160%</th>
<th>180%</th>
<th>200%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>75% F_{MAX}</th>
<th>60%</th>
<th>110%</th>
<th>160%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.42</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SSB/E(SSB_{MSY})</th>
<th>F_{MAX}</th>
<th>mean</th>
<th>median</th>
<th>cv</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>0.097</td>
<td>0.998</td>
<td>0.097</td>
</tr>
<tr>
<td>1.42</td>
<td>1.42</td>
<td>0.091</td>
<td>1.42</td>
<td>0.091</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SSB/E(SSB_{MSY})</th>
<th>0.75 F_{MAX}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>0.091</td>
</tr>
</tbody>
</table>
Figure 2. Summary of variability in SSB for bigeye tuna (BET). A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium SSB/SSBMAX when projected at constant F=FMAX. B) PDF and CF of equilibrium SSB/SSBMAX when projected at constant F=75% . C) Comparison of panels A and B. D) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=FMAX. E) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=75% FMAX.
Figure 3. Summary of variability in SSB for northern albacore tuna (N-ALB). A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium SSB/SSB_{MAX} when projected at constant F=F_{MAX}. B) PDF and CF of equilibrium SSB/SSB_{MAX} when projected at constant F=75\% F_{MAX}. C) Comparison of panels A and B. D) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=F_{MAX}. E) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=75\% F_{MAX}.
Figure 4. Summary of variability in SSB for north Atlantic swordfish (N-SWO). A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium SSB/SSB_{FMAX} when projected at constant F=F_{MAX} B) PDF and CF of equilibrium SSB/SSB_{FMAX} when projected at constant F=75%. C) Comparison of panels A and B. D) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=F_{MAX}. E) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=75%M_{FMAX}. 

mean = 1.0
median = 0.999
cv=0.052

mean = 1.53
median = 1.53
cv=0.046

0.75F_{max}
F_{max}
Figure 5. Summary of variability in SSB for western bluefin tuna “2-Line/ Low Recruitment” model (W-BFT-R71+).  A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium SSB/SSB$_{F_{\text{MAX}}}$ when projected at constant $F=F_{\text{MAX}}$. B) PDF and CF of equilibrium SSB/SSB$_{F_{\text{MAX}}}$ when projected at constant $F=75\%F_{\text{MAX}}$. C) Comparison of panels A and B. D) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant $F=F_{\text{MAX}}$. E) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant $F=75\%F_{\text{MAX}}$. 
Figure 6. Summary of variability in SSB for western bluefin tuna “Beverton&Holt/ High Recruitment” model (W-BFT-R76+). A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium SSB/SSB_{MSY} when projected at constant F=F_{MAX}. B) PDF and CF of equilibrium SSB/SSB_{FMAX} when projected at constant F=75%F_{MAX}. C) Comparison of panels A and B. D) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=F_{MAX}. E) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=75%F_{MAX}.
Figure 7. Summary of variability in SSB for eastern bluefin tuna “High Recruitment” model (E-BFT-R70+). A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium SSB/SSB$_{F_{\text{MAX}}}$ when projected at constant $F=F_{\text{MAX}}$. B) PDF and CF of equilibrium SSB/SSB$_{F_{\text{MAX}}}$ when projected at constant $F=75\%F_{\text{MAX}}$. C) Comparison of panels A and B. D) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant $F=F_{\text{MAX}}$. E) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant $F=75\%F_{\text{MAX}}$. 
Figure 8. Summary of variability in SSB for eastern bluefin tuna “Low Recruitment” model (E-BFT-R90+). A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium SSB/SSB_{MAX} when projected at constant F=F_{MAX}. B) PDF and CF of equilibrium SSB/SSB_{F_{MAX}} when projected at constant F=75\%. C) Comparison of panels A and B. D) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=F_{MAX}. E) Trajectories of equilibrium SSB during the last 20 years of the projection interval when projected at constant F=75\%F_{MAX}.
Figure 9. Summary of variability in fishable biomass (FB) for yellowfin tuna (YFT). A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium FB/FB_{F\text{MAX}} when projected at constant F=F_{\text{MAX}}.
B) PDF and CF of equilibrium FB/FB_{F\text{MAX}} when projected at constant F=75\% F_{\text{MAX}}. C) Comparison of panels A and B. D) Trajectories of equilibrium FB during the last 20 years of the projection interval when projected at constant F=F_{\text{MAX}}. E) Trajectories of equilibrium FB during the last 20 years of the projection interval when projected at constant F=75\% F_{\text{MAX}}.
Figure 10. Summary of variability in fishable biomass (FB) for bigeye tuna (BET). A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium FB/\(FB_{\text{FMAX}}\) when projected at constant \(F=\text{FMAX}\). B) PDF and CF of equilibrium FB/\(FB_{\text{FMAX}}\) when projected at constant \(F=75\%\). C) Comparison of panels A and B. D) Trajectories of equilibrium FB during the last 20 years of the projection interval when projected at constant \(F=\text{FMAX}\). E) Trajectories of equilibrium FB during the last 20 years of the projection interval when projected at constant \(F=75\%\text{FMAX}\).
Figure 11. Summary of variability in fishable biomass (FB) for north Atlantic swordfish (N-SWO). A) Probability density function (PDF) and cumulative frequency (CF) of equilibrium FB/FB_{MAX} when projected at constant F=F_{MAX}. B) PDF and CF of equilibrium FB/FB_{FMAX} when projected at constant F=75%. C) Comparison of panels A and B. D) Trajectories of equilibrium FB during the last 20 years of the projection interval when projected at constant F=F_{MAX}. E) Trajectories of equilibrium FB during the last 20 years of the projection interval when projected at constant F=75%F_{MAX}.

Appendix 5

MATTERS TO BE CONSIDERED IN FUTURE WORKS

The Group considered that oceanographic factors might strongly affect the spatial patterns and behaviour of tuna schools, and feed into stock status evaluation uncertainty by influencing pre-spawning and spawning season patterns differently over time. According to the available knowledge, it seems that relevant changes in the distribution pattern of the high concentration of bluefin tuna had occurred in the Mediterranean Sea (mostly between from 1996 to 2005), inducing a remarkable change in the fishing pattern and distribution of the fleets. The influence of the environment, including oceanographic factors, is actually not included in the models being
used and this fact should be properly taken into account. Efforts to get at least basic information about the environmentally related distribution variables may help to better quantify uncertainty in stock status evaluations for some species.

The Group also considered that the use of the available information on the fleet spatial distribution and concentrations (VMS aggregated data by gear and time strata) is expected to add significant understanding of fishing strategies coincident with environmental changes and fish movements. Combining VMS with remotely sensed environmental information would also contribute to an improved understanding of fish and fishery dynamics. Methods to fully utilize VMS and auxiliary data for support of stock status evaluations should be a focus for future application in providing advice to the Commission on stock status and availability.