FISHERIES RESEARCH SERVICES

SURBA 3.0: Technical Manual (first draft)

Coby L. Needle

FRS Marine Laboratory Aberdeen

August 22, 2005

Contents

1 Data requirements

1.1 Data formats

- Lowestoft VPA format (Darby and Flatman 1994).
- Plus-groups are removed from abundance indices. This is because multiple surveys may be used, and plus-groups can cause problems in this situation. For example, consider a stock for which two surveys are available: I_1 with ages 1–6, and I_2 with ages 1–6+. There is no correct way to deal with this and retain all data. If age 6 is treated as a plus-group for both surveys, then it will be too small for I_1. If age 6 is treated as a single age in both surveys, then it will be too large for I_2. The only general solution is to remove the plus-group. This will potentially remove fish from the index, which can lead to bias if the plus-group is large. It is therefore necessary to ensure for survey-based analyses that the plus-group indices are as small as possible.

2 The separable model

2.1 Basis

The basis of SURBA is a simple survey-based separable model of mortality. This was first applied to European research-vessel survey data by Cook (1997, 2004), but the underlying model has a long history in catch-based fisheries stock assessment (e.g. Pope and Shepherd 1982, Deriso et al. 1985, Gudmundsson 1986, Johnson and Quinn II 1987, Patterson and Melvin 1996): see Quinn II and Deriso (1999) for a summary. The separable model used in SURBA assumes that total mortality $Z_{a,y}$ in age a and year y can be expressed as

$$Z_{a,y} = s_a f_y, \tag{2.1}$$

where s_a and f_y are respectively the *age* and *year* effects of mortality. Note that this differs from the usual assumption in that total mortality Z is the quantity of interest, rather than fishing mortality F. Then, given $Z_{a,y}$, abundance $N_{a,y}$ can be derived as

$$\begin{aligned} N_{a,y} &= N_{a_0,y_0} \exp\left(-\sum_{m=a_0}^{a-1}\sum_{n=y_0}^{y-1} Z_{m,n}\right) \\ &= r_{y_0} \exp\left(-\sum_{m=a_0}^{a-1}\sum_{n=y_0}^{y-1} Z_{m,n}\right) \end{aligned} \tag{2.2}$$

where a_0 and $y_0 = y - a - a_0$ are respectively the age and year in which the fish measured as $N_{a,y}$ first recruit to the observed population. Thus the abundance at each age and year of a cohort is given by the recruiting abundance (denoted by $r_{y_0-A+a_0} = N_{a_0,y_0}$) of the relevant cohort modified by the cumulative effect of mortality during its lifetime.

2.2 Estimation

The parameters to be estimated when fitting the model are $\mathbf{\Theta} = [\mathbf{s}, \mathbf{f}, \mathbf{r}]$, where:

$$\begin{aligned} \text{Age effects} &: \quad \mathbf{s} = [s_a] = [s_{a_0}, s_{a_0+1}, \ldots, s_A] \\ \text{Year effects} &: \quad \mathbf{f} = [f_y] = [f_{y_0}, f_{y_0+1}, \ldots, f_Y] \\ \text{Cohort effects} &: \quad \mathbf{r} = [r_{yc}] = [r_{y_0-A+a_0}, r_{y_0-A+a_0+1}, \ldots, r_{Y-a_0}] \end{aligned}$$

Here A is the oldest age and Y is the last year in the dataset. The data available may include:

$$\begin{array}{rcl}
\text{Age-structured abundance indices} & : & \mathbf{I} = [I_{a,y,i}] \\
\text{Biomass abundance indices} & : & \mathbf{J} = [J_{y,j}] \\
\text{Stock weights-at-age} & : & \mathbf{W} = [W_{a,y}] \\
\text{Proportion mature-at-age} & : & \mathbf{Mat} = [\mathrm{Mat}_{a,y}] \\
\text{Proportion mortality before spawning} & : & \mathbf{PZ} = [\mathrm{PZ}_{a,y}] \\
\text{Age-structured index catchabilities} & : & \mathbf{q} = [q_{a,y,i}] \\
\text{Age-structure index weightings} & : & \boldsymbol{\omega} = [\omega_{a,y,i}] \\
\text{Age-structured index timings} & : & \boldsymbol{\rho} = [\rho_i] \\
\text{Biomass index weightings} & : & \boldsymbol{\nu} = [\nu_{y,j}] \\
\text{Penalty term weighting} & : & \lambda
\end{array}$$

Here $i \in [1, 2, \ldots, NI]$ indexes age-structured series, while $j \in [1, 2, \ldots, NJ]$ indexes biomass series. SURBA assumes that at least one age-structured survey index is available: biomass indices are optional. If present, it is assumed that biomass indices are measured at spawning time, following the convention used in such models as ICA (Patterson and Melvin 1996).

The fitting procedure is as follows. Abundance indices (age-structured and biomass) are mean-standardised using

$$I'_{a,y,i} = I_{a,y,i} \left(\tfrac{1}{A-a_0+1+Y-y_0+1} \sum_{m=a_0}^{A} \sum_{n=y_0}^{Y} I_{m,n,i} \right)^{-1}, \tag{2.3}$$

$$J'_{y,j} = J_{y,j} \left(\tfrac{1}{Y-y_0+1} \sum_{n=y_0}^{Y} J_{n,j} \right)^{-1}, \tag{2.4}$$

so that the mean of each index over all ages and years is 1.0. Given estimated parameters $\hat{\mathbf{s}}$, $\hat{\mathbf{f}}$ and $\hat{\mathbf{r}}$, fitted age-structured abundance indices are calculated using

$$\hat{Z}_{a,y} = \hat{s}_a \hat{f}_y \tag{2.5}$$

and

$$\hat{N}_{a,y} = \hat{r}_{y_0-a-a_0} \exp \left(- \sum_{m=a_0}^{a-1} \sum_{n=y_0}^{y-1} \hat{Z}_{m,n} \right) \tag{2.6}$$

and

$$\hat{I}'_{a,y,i} = q_{a,y,i} \hat{N}_{a,y}. \tag{2.7}$$

Note that both $\hat{I}'_{a,y,i}$ and $\hat{N}_{a,y}$ refer to the stock as measured at January 1 of the year in question. In order to compare these fitted values with observations, SURBA backshifts the observed indices from the time of observation to the start of the year (January 1st), thus:

$$I'^{*}_{a,y,i} = I'_{a,y,i} \exp \left(\rho_i \hat{Z}_{a,y} \right). \tag{2.8}$$

Then the sum-of-squares to be minimised for NI age-structured indices is

$$\mathrm{SSQ}_I = \sum_{i=1}^{\mathrm{NI}} \sum_{a=a_0}^{A} \sum_{y=y_0}^{Y} \omega_{a,y,i} \left(\ln I'^{*}_{a,y,i} - \ln \hat{I}'_{a,y,i} \right)^2 . \tag{2.9}$$

The procedure for biomass indices is somewhat different. Since these indices refer to all mature fish, there is no sensible way to shift a biomass index back to the start of the year on the basis of age-structured mortality up until the time of the survey. Rather than shifting biomass indices back, SURBA shifts SSB estimates forward to the time of the survey (assumed to be spawning time) before comparing observations with fits. Forward-shifted SSB is estimated as

$$\hat{B}_y = \sum_{a=a_0}^{A} \hat{N}_{a,y} W_{a,y} \mathrm{Mat}_{a,y} \exp\left(-\mathrm{PZ}_{a,y} \hat{Z}_{a,y} \right) , \tag{2.10}$$

and the fitted biomass indices are then

$$\hat{J}'_{y,j} = q_{y,j} \hat{B}_y^{k_j} \tag{2.11}$$

where k_j is the power relationship for the biomass index (currently it is assumed that $\mathbf{k} \equiv \mathbf{1}$). The sum-of-squares to be mimimised for NJ biomass indices can then be written as

$$\mathrm{SSQ}_J = \sum_{b=1}^{NJ} \sum_{y=y_0}^{Y} \nu_{y,j} \left(\ln J'_{y,j} - \ln \hat{J}'_{y,j} \right)^2 . \tag{2.12}$$

The final element of the sum-of-squares is a term which penalises inter-annual variation in the estimated year-effect $\hat{\mathbf{f}}$ (Cook 1997), and which is defined as

$$\mathrm{SSQ}_\lambda = \lambda \sum_{y=y_0}^{Y-2} \left(\hat{f}_y - \hat{f}_{y+1} \right)^2 . \tag{2.13}$$

Then the overall sum-of-squares to be minimised is

$$\mathrm{SSQ} = \mathrm{SSQ}_I + \mathrm{SSQ}_J + \mathrm{SSQ}_\lambda \tag{2.14}$$

2.3 DERIVED VALUES

As presented above, the SURBA model is indeterminate and has no unique solution. For this reason, elements of $\mathbf{s}$ and $\mathbf{f}$ must be fixed beforehand. If A is the oldest age, Y is the last year, and a_r is a reference age chosen by the user, then the following parameter values are derived from other parameters (and are therefore not directly estimated):

$$s_{a_r} = 1.0 \tag{2.15}$$

$$s_A = \hat{s}_{A-1} \tag{2.16}$$

$$f_Y = \frac{1}{3} \sum_{y=Y-3}^{Y-1} \hat{f}_y \tag{2.17}$$

2.4 UNCERTAINTY

SURBA estimates a variance-covariance matrix for the estimated parameters $\hat{\boldsymbol{\Theta}}$. However, the variance and associated confidence limits of derived estimated population characteristics such as mortality $\hat{Z}_{a,y}$, abundance $\hat{N}_{a,y}$ and SSB $\hat{B}_y$ are more directly relevant to users. To generate derived variances, SURBA uses the delta method (e.g. Seber 1982, Oehlert 1992).

From above, we have seen that the SURBA model has a number of parameters, namely $\mathbf{s}$, $\mathbf{f}$ and $\mathbf{r}$, and that the parameter vector for the model is denoted by

$$\begin{aligned}\boldsymbol{\Theta} &= [\mathbf{s}, \mathbf{f}, \mathbf{r}] \\ &= [s_{a_0}, \ldots, s_A, f_{y_0}, \ldots, f_Y, r_{y_0-A+a_0}, r_{y_0-A+a_0+1}, \ldots, r_{Y-a_0}]\end{aligned}$$

Denote further the estimated variance-covariance matrix of the estimated parameters by $\hat{\text{Var}}\left[\hat{\boldsymbol{\Theta}}\right]$, and let $\mathbf{G}$ denote the transformation to the required model output $\hat{\boldsymbol{\Phi}}$ (which could be mortality or SSB, for example). Further, let $\delta\mathbf{G}$ be the matrix of partial derivatives of $\mathbf{G}$ with respect to the estimated parameters. The delta method then uses a first-order Taylor expansion to approximate the variance-covariance matrix of $\hat{\boldsymbol{\Phi}}$ via

$$\hat{\text{Var}}[\hat{\boldsymbol{\Phi}}] = \hat{\text{Var}}\left[\mathbf{G}(\hat{\boldsymbol{\Theta}})\right] = \delta\mathbf{G}(\hat{\boldsymbol{\Theta}})\hat{\text{Var}}\left[\hat{\boldsymbol{\Theta}}\right]\delta\mathbf{G}(\hat{\boldsymbol{\Theta}})^T.$$

This approximation will be close in general if the assumption of lognormally-distributed errors is reasonably accurate: otherwise, it may be misleading. Standard errors for the components of $\hat{\boldsymbol{\Phi}}$ are obtained by taking square roots of the diagonal elements of this matrix. The estimated parameter correlation matrix $\hat{\text{Corr}}[\hat{\boldsymbol{\Phi}}]$ may also be informative, and is derived from the variance-covariance matrix using

$$\hat{\text{Corr}}[\hat{\Phi}_i, \hat{\Phi}_j] = \frac{\hat{\text{Cov}}\left[\hat{\Phi}_i, \hat{\Phi}_j\right]}{\sqrt{\hat{\text{Var}}\left[\hat{\Phi}_i\right]\hat{\text{Var}}\left[\hat{\Phi}_j\right]}}.$$

Inferred variance estimates

Here I give examples of the calculations required to infer (via the delta method) the variances of population summary statistics and other values produced by SURBA. The full list of estimates produced by SURBA is given in Table 2.1, which also includes the variance estimates used for the derived values listed in Section 2.3 . Note that hats ($\hat{}$) have been dropped throughout for clarity. As a first example, consider mortality $Z_{a,y}$ in age a and year y, where

$$\boldsymbol{\Phi}|_{a,y} = \mathbf{G}(\boldsymbol{\Theta})|_{a,y} = Z_{a,y} = s_a f_y.$$

Then

$$\delta\mathbf{G}(\boldsymbol{\Theta})|_{a,y} = \begin{bmatrix} \frac{\delta Z_{a,y}}{\delta s_a} & \frac{\delta Z_{a,y}}{\delta f_y} \end{bmatrix} = \begin{bmatrix} f_y & s_a \end{bmatrix}$$

and

$$\text{Var}\left[Z_{a,y}\right] = \begin{bmatrix} f_y & s_a \end{bmatrix} \begin{bmatrix} \text{Var}\left[s_a\right] & \text{Cov}\left(s_a, f_y\right) \\ \text{Cov}\left(s_a, f_y\right) & \text{Var}\left[f_y\right] \end{bmatrix} \begin{bmatrix} f_y \\ s_a \end{bmatrix}$$

so that

$$\text{Var}\left[Z_{a,y}\right] = f_y^2 \text{Var}\left[s_a\right] + 2 s_a f_y \text{Cov}\left(s_a, f_y\right) + s_a^2 \text{Var}\left[f_y\right].$$

As a second example, consider abundance $N_{2,y}$ at age 2 in year y, which can be written as

$$\mathbf{\Phi}|_{2,y} = \mathbf{G}(\mathbf{\Theta})|_{2,y} = N_{2,y} = r_{y-1-a_0} e^{-s_1 f_{y-1}}$$

Then

$$\begin{aligned}
\delta\mathbf{G}(\mathbf{\Theta})|_{2,y} &= \begin{bmatrix} \frac{\delta N_{2,y}}{\delta r_{y-1-a_0}} & \frac{\delta N_{2,y}}{\delta s_1} & \frac{\delta N_{2,y}}{\delta f_{y-1}} \end{bmatrix} \\
&= \begin{bmatrix} e^{-s_1 f_{y-1}} & -r_{y-1-a_0} e^{-s_1 f_{y-1}} & -r_{y-1-a_0} e^{-s_1 f_{y-1}} \end{bmatrix}
\end{aligned}$$

and

$$\text{Var}\left[N_{2,y}\right] = \begin{bmatrix} e^{-s_1 f_{y-1}} & -r_{y-1-a_0} e^{-s_1 f_{y-1}} & -r_{y-1-a_0} e^{-s_1 f_{y-1}} \end{bmatrix} \times$$
$$\begin{bmatrix} \text{Var}\left[r\right] & \text{Cov}\left(r, s_1\right) & \text{Cov}\left(r, f_{y-1}\right) \\ \text{Cov}\left(r, s_1\right) & \text{Var}\left[s_1\right] & \text{Cov}\left(s_1, f_{y-1}\right) \\ \text{Cov}\left(r, f_{y-1}\right) & \text{Cov}\left(s_1, f_{y-1}\right) & \text{Var}\left[f_{y-1}\right] \end{bmatrix} \times \begin{bmatrix} e^{-s_1 f_{y-1}} \\ -r_{y-1-a_0} e^{-s_1 f_{y-1}} \\ -r_{y-1-a_0} e^{-s_1 f_{y-1}} \end{bmatrix}$$

so that

$$\begin{aligned}
\text{Var}\left[N_{2,y}\right] = e^{-2s_1 f_{y-1}} \big(&\text{Var}[r_{y-1-a_0}] + \text{Cov}(r_{y-1-a_0}, s_1) + \text{Cov}(r_{y-1-a_0}, f_{y-1}) + \\
&r^2_{y-1-a_0} \left(\text{Cov}(r_{y-1-a_0}, s_1) + \text{Cov}(r_{y-1-a_0}, f_{y-1}) + \text{Var}[s_1] + 2\text{Cov}(s_1, f_{y-1}) + \text{Var}[f_{y-1}]\right)\big)
\end{aligned}$$

In most cases, confidence limits about a summary estimate $\hat{\Phi}_i$ can be approximated with

$$\hat{\text{CL}}_{\Phi_i} = \hat{\Phi}_i \pm 2\sqrt{\hat{\text{Var}}\left[\hat{\Phi}_i\right]}. \tag{2.18}$$

The exceptions in SURBA are those estimates which are generated as monotonic transformations of parameter estimates. For example, recruitment in year y is given by

$$\hat{R}_y = \ln \hat{r}_y.$$

In these cases SURBA uses $\hat{\text{Var}}\left[\hat{r}_y\right]$ to generate confidence intervals on the log scale, which are then back-transformed to the required arithmetic scale. Thus the confidence interval for recruitment $\hat{R}_y$ would be given by

$$\hat{\text{CL}}_{R_y} = \exp\left(\ln \hat{R}_y \pm 2\sqrt{\hat{\text{Var}}\left[\ln \hat{R}_y\right]}\right).$$

NOTES

- A *debug* error occurs when reading in a second (or subsequent) dataset. This only appears to happen when run under Windows XP, and it seems safe to choose the Ignore option in the error-report dialogue.
- Word must be open when a SURBA run is started for the Copy Plot function to work correctly.
- PDF output is possible from plots, by first setting PDF as the default print driver in Print Manager.

Value	Parameter	Variance
Age effect on oldest age s_A	s_{A-1}	$\mathrm{Var}\left[s_{A-1}\right]$
Year effect in last year f_Y	$\frac{1}{3}\sum_{i=Y-3}^{Y-1} f_i$	$\frac{1}{9}\left(\mathrm{Var}\left[f_{Y-3}\right]+\mathrm{Var}\left[f_{Y-2}\right]+\mathrm{Var}\left[f_{Y-1}\right]\right.$ $\left.+2\mathrm{Cov}\left(f_{Y-3},f_{Y-2}\right)+2\mathrm{Cov}\left(f_{Y-3},f_{Y-1}\right)++2\mathrm{Cov}\left(f_{Y-2},f_{Y-1}\right)\right)$
Mortality $Z_{a,y}$	$s_a f_y$	$f_y^2\mathrm{Var}\left[s_a\right]+2s_a f_y\mathrm{Cov}\left(s_a,f_y\right)+s_a^2\mathrm{Var}\left[f_y\right]$
Mean mortality $\bar{Z}_y$	$\frac{1}{a_2-a_1+1}\sum_{a=a_1}^{a_2} Z_{a,y}$	$\frac{1}{a_2-a_1+1}\sum_{i=a_1}^{a_2}\sum_{j=a_1}^{a_2} f_y^2\mathrm{Cov}\left(s_i,s_j\right)+s_j f_y\mathrm{Cov}\left(s_i,f_y\right)+s_i f_y\mathrm{Cov}\left(s_j,f_y\right)+s_i s_j\mathrm{Var}\left[f_y\right]$
Abundance $N_{a,y}$	$r_{y-a-a_0}\exp\left(-\sum_{m=a_0}^{a-1}\sum_{n=y_0}^{y-1} Z_{m,n}\right)$	$e^{-2s_{a-1}f_{y-1}}\left(\mathrm{Var}[N_{a-1,y-1}]+\mathrm{Cov}(N_{a-1,y-1},s_{a-1})+\mathrm{Cov}(N_{a-1,y-1},f_{y-1})+\right.$ $N_{a-1,y-1}^2\left(\mathrm{Cov}(N_{a-1,y-1},s_{a-1})+\mathrm{Cov}(N_{a-1,y-1},f_{y-1})+\mathrm{Var}[s_{a-1}]+\right.$ $\left.\left.2\mathrm{Cov}(s_{a-1},f_{y-1})+\mathrm{Var}[f_{y-1}]\right)\right)$
SSB B_y	$\sum_{a=a_0}^{A} N_{a,y}W_{a,y}\mathrm{Mat}_{a,y}$	$\sum_{i=a_0}^{A}\sum_{j=a_0}^{A} W_{i,y}\mathrm{Mat}_{i,y}W_{j,y}\mathrm{Mat}_{j,y}\mathrm{Cov}\left(N_{i,y},N_{j,y}\right)$

Table 2.1: Estimated parameters and variances for summary statistic generated by SURBA. Note that hats (ˆ) have been dropped throughout for clarity.

Bibliography

Cook, R. M. (1997). Stock trends in six North Sea stocks as revealed by an analysis of research vessel surveys, *ICES Journal of Marine Science* **54**: 924–933.

Cook, R. M. (2004). Estimation of the age-specific rate of natural mortality for Shetland sandeels, *ICES Journal of Marine Science* **61**: 159–164.

Darby, C. D. and Flatman, S. (1994). Lowestoft VPA Suite Version 3.1 User Guide. MAFF: Lowestoft.

Deriso, R. B., Quinn II, T. J. and Neal, P. R. (1985). Catch-age analysis with auxiliary information, *Canadian Journal of Fisheries and Aquatic Sciences* **42**: 815–824.

Gudmundsson, G. (1986). Statistical considerations in the analysis of catch-at-age observations, *Journal du Conseil Internationl pour l'Exploration de la Mer* **43**: 83–90.

Johnson, S. J. and Quinn II, T. J. (1987). Length frequency analysis of sablefish in the Gulf of Alaska, *Technical Report UAJ-SFS-8714*, University of Alaska, School of Fisheries and Science, Juneau, Alaska. Contract report to Auke Bay National Laboratory.

Oehlert, G. W. (1992). A note on the delta method, *American Statistician* **46**: 27–29.

Patterson, K. R. and Melvin, G. D. (1996). Integrated Catch At Age Analysis Version 1.2, *Scottish Fisheries Research Report*. FRS: Aberdeen.

Pope, J. G. and Shepherd, J. G. (1982). A simple method for the consistent interpretation of catch-at-age data, *Journal du Conseil International pour l'Exploration de la Mer* **40**: 176–184.

Quinn II, T. J. and Deriso, R. B. (1999). *Quantitative Fish Dynamics*, Oxford University Press, Oxford.

Seber, G. A. F. (1982). *The Estimation of Animal Abundance*, 2nd edn, Griffin, London.