FIF

# Some variants of age based models

Einar Hjörleifsson

# If we only have $C_{ay}$



- If there are no other available data for a stock than catch at age one could attempt to fit the model to catches alone.
- May need a extra "stabilizer", and the brave one may assume that fishing mortality does not change much between consecutive years:

$$
\min SSE = SSE_C + SSE_F
$$

$$
= SSE_C + \sum_y \left(\ln F_y - \ln F_{y-1}\right)^2
$$

**Use with extreme caution!**

FIF

# Examples of some “variants” on modelling selection pattern

# Modeling selection patterns



- Empirical fit:
  - In the setup here we have sorted to fitting of a fixed selection patterns, estimating each Sa separately (empirical fit)
    - For simplification we assumed that Sa in the two oldest age groups are equal to 1.00
- However many variants available in various software packages. Examples are:
    - Logistic function (AMCI, …)
    - Double (or quadruple) logistic function (West coast stuff)
    - Double half Gaussian
    - Thompson's exponential-logistic model (NFT StatCam)

# Modeling selection patterns



- Empirical fit:
  - Although using a functions may reduce the number of parameters to be estimated, we more often observe selection patterns that do not follow any parametric function.
  - In practice we thus most frequently estimate the selection of each age group separately. This requires two things:
    - Rescaling: The selection pattern is a relative value and need thus to be normalized so that some reference age or age range the value is ONE
    - Constraint on the oldest groups: Normally the observations do not contain enough information so some constraint on the selection on the oldest age groups is needed. The simplest assumption is to set the selection of the oldest age(s) equal to the average of some younger ages.
      - Note that this constraint does not necessarily imply that the selection pattern cannot go down with age.

# Rescaling selection pattern

- The selection at age (say) for the current year is finally obtained by scaling the $F_{a,y}$ so that the average selection over the reference age(s) is ONE:

$$s_{a.y} = \frac{F_{a,y}}{\frac{\sum_{a=a1}^{a=a2} F_{a,y}}{(a2-a1+1)}}$$

- Alternatively one could put a constraint on the selection parameters in the minimization:

$$\frac{\sum_{a=a1}^{a=a2} s_{a,y}}{(a2-a1+1)} = 1.00$$

  - a1: Lower reference age group
  - a2: Upper reference age group

# Constraint on the oldest age groups

- The simplest form of the constraint on the oldest age group is to assume that the selection pattern in the oldest age group(s) is the same as the prior age groups:

$$s_{a.y} = s_{a-1,y}$$

- Another way would be to set a penalty in the objective function on the amount of "curvature" in the selection pattern by age:

$$\rho_1 = \lambda_{\rho 1} \sum_{y} \sum_{a=a5}^{a=a6-2} \left( s_{a,y} - 2s_{a+1,y} + s_{a+2,y} \right)^2$$

  - a5 and a6: Lower and upper age group over which constraint is applied

# Excel implementation



**CONSTRAINTS**

| Name | |
|---|---|
| Average selection age 6 to 8 | 0.81 |

**PARAMETERS**

| Name | Ln(parameter) | Switches | Parameter |
|---|---|---|---|
| ln s1 | -4.90 | | 0.007 |
| ln s2 | -3.60 | 1 | 0.027 |
| ln s4 | -2.50 | 1 | 0.082 |
| ln s4 | -1.60 | 1 | 0.202 |
| ln s5 | -0.90 | 1 | 0.407 |
| ln s6 | -0.40 | 1 | 0.670 |
| ln s7 | -0.10 | 1 | 0.905 |
| ln s8 | -0.15 | 1 | 0.861 |
| ln s9 | -0.20 | 0 | 0.861 |
| ln s10 | -0.25 | 0 | 0.861 |
| ln s11 | -0.30 | 0 | 0.861 |

**RESCALING**

1. Decide on the reference age range
2. Calculate average selection pattern for that age range
3. Set the constraint in solver for the average to be equal to 1

**CONSTRAINT**

1. Set the sa for the oldest age group equal to that of the younger age groups. Could use switches to control this. If switch =1, estimate the value. If switch=0, use the value "above".

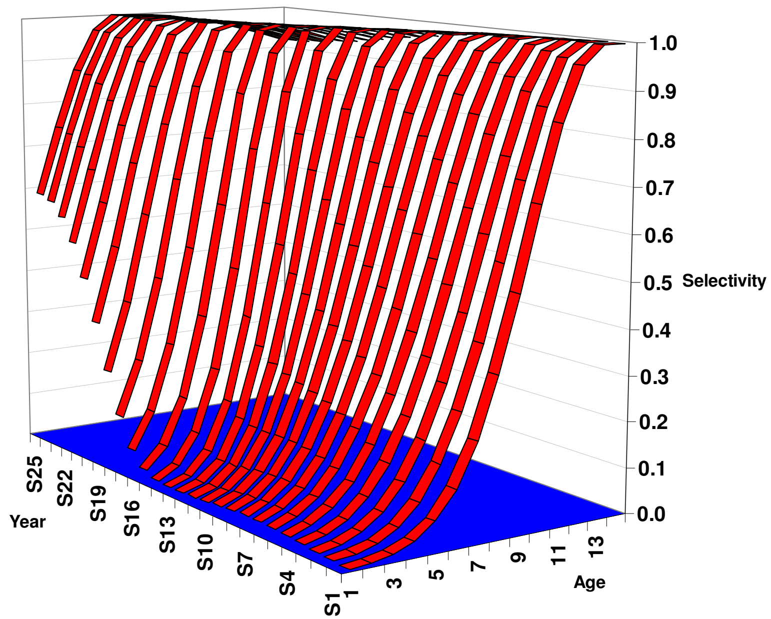
FIF

# What is this random walk stuff?

# Changes in selection pattern



- The fisheries of a stock we are dealing with rarely has a fixed selection pattern through the time period we are dealing with.




- How do we account for this in our catch at age models?

# Time-varying parameters

einar

- In manual describing models we may see something like:
  - Changes in selectivity occur each year through a random walk for every age:

$$S_{a,y+1} = S_{ay} e^{-\varepsilon_{a,y}}$$

    - where

$$\varepsilon_{a,y} \sim \mathrm{N}(0,\sigma^2)$$

    - and

$$\min SSE = \lambda_x \sum_{a,y} \varepsilon^2_{a,y}$$

# What does it mean?



$$L_x = \lambda_x \sum_{a,y} \varepsilon_{a,y}^2 = \lambda_x \sum_{a,y} \frac{\left(\ln S_{a,y} - \ln S_{a,y-1}\right)^2}{2\sigma^2} + \ln(\sigma)$$

- Conceptually this means that the selectivity in year y can vary from the selectivity in year y-1 and that the deviation between the years takes on a lognormal distribution with a standard deviation of sigma.
  - Note that the above is a log-likelihood formulation. For all practical purposes here, it is the same thing as the minimum sums of square concept.

# Where does this fit into statistics?



- Note the difference between:

$$\underset{\text{minimize}}{} SSE = \sum \left( \ln S_{a,y} - \ln S_{a,y-1} \right)^2$$

- and

$$\underset{\text{minimize}}{} SSE = \sum \left( \ln U_{a,y} - \ln \hat{U}_{a,y} \right)^2$$

  - In the latter we are minimizing the difference between observed and predicted data, in the former we are minimizing difference between two parameters, that are both predicted in the model!!
  - How many parameters are we actually estimating?

# Implementing change in selectivity



- Could do:

$$\underset{\text{minimize}}{} SSE_{A_{full}} = \sum \left( \ln A_y^{full} - \ln A_{y-1}^{full} \right)^2$$

$$\underset{\text{minimize}}{} SSE_{\sigma_L} = \sum \left( \ln \sigma_y^L - \ln \sigma_{y-1}^L \right)^2$$

$$\underset{\text{minimize}}{} SSE_{\sigma_R} = \sum \left( \ln \sigma_y^R - \ln \sigma_{y-1}^R \right)^2$$

- Have never seen this implemented exactly as above, but lets do it just to get the feel for it this random walk.

# Excel implementation for $\sigma_{L,y}$



**MINIMIZATION STUFF**

| Sum of squares | | λ |
|---|---|---|
| C@A | 71.29 | 1 |
| U@A - Survey 1 | 13 | 1 |
| ….. | | 1 |
| Penalty σL | 0.1839 | 10 |
| | | |
| Total SSE | 86.3919 | |

**PARAMETERS**

| Name | Ln(parameter) | Switches | Parameter |
|---|---|---|---|
| ln σL 1900 | 2.52 | | 12.47 |
| ln σL 1901 | 2.56 | 1 | 13.00 |
| ln σL 1902 | 2.52 | 1 | 12.39 |
| ln σL 1903 | 2.39 | 1 | 10.86 |
| ln σL 1904 | 2.20 | 1 | 9.01 |
| ln σL 1905 | 2.02 | 1 | 7.54 |
| ln σL 1906 | 1.95 | 1 | 7.00 |
| ln σL 1907 | 2.03 | 1 | 7.61 |
| ln σL 1908 | 2.21 | 1 | 9.13 |
| ln σL 1909 | 2.40 | 1 | 10.98 |
| ln σL 1910 | 2.52 | 1 | 12.46 |
| ……. | | | |

**PENALTIES**

| |
|---|
| |
| 0.04 |
| -0.05 |
| -0.13 |
| -0.19 |
| -0.18 |
| -0.07 |
| 0.08 |
| 0.18 |
| 0.18 |
| 0.13 |

1. **Set up a penalty area**
2. **Calculate the log deviation for consecutive years**

$$\left(\ln\sigma_{Ly} - \ln\sigma_{Ly-1}\right)$$

3. **Square the deviation, sum the stuff, multiply with λ**

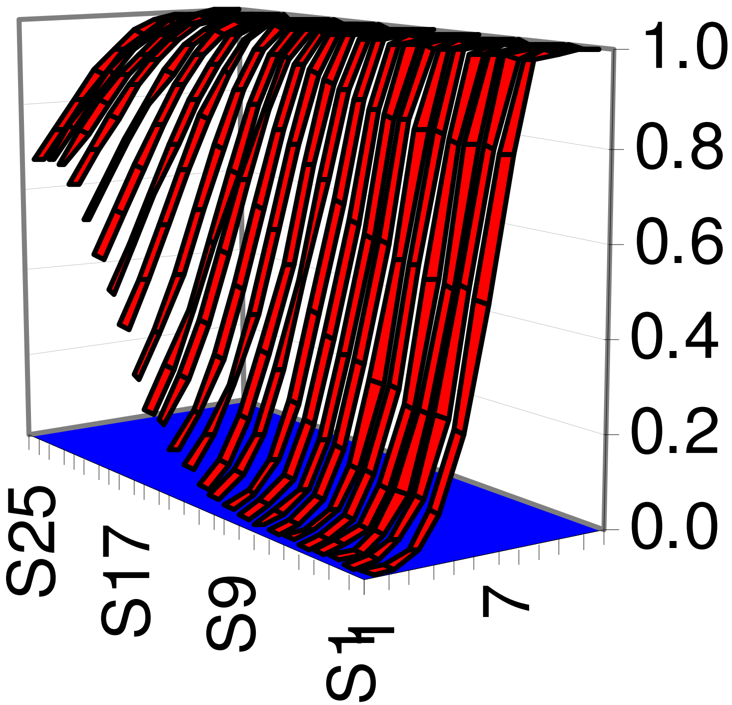
$$\lambda_{\sigma_L}\sum_y\left(\ln\sigma_{Ly} - \ln\sigma_{Ly-1}\right)^2$$

4. **The total objective function becomes:** **$_{\min}SSE = SSE_{c@a} + SSE_{u@a} + \ldots + SSE_{\sigma L}$**

# Random walk in σL



**Modelled selection pattern by year and age**



# Random walk: recursive updating (AMCI)

# Recursive updating



- In AMCI there is a different approach towards modelling random walk in the selection pattern. This requires 3 steps in each year (except the first year):
  - Step 1: Fishing mortality derived directly from the observed catches in the year in question.
  - Step 2: Fishing mortality derived from the weighted average of above and fishing mortality derived from the estimated selection pattern of the previous year and the estimated fishing mortality this year.
  - Step 3: The rescaled compromised fishing mortality from these two different approaches becomes the estimated selection pattern of the year in question.

# Recursive updating: Step 1

- If Nay and Cay are known one can estimate Fay from the catch equation:

$$C_{a,y}^{OBS} = \frac{\tilde{F}_{a,y}}{\tilde{F}_{a,y} + M_{ay}} \left(1 - e^{-(\tilde{F}_{a,y} + M_{ay})}\right) N_{ay}$$

  - Use F-tilda here to distinguish it from the conventional way of obtaining F in a catch at age model where the predicted catch determines the F value. Could refer to F-tilda as the “observed” fishing mortality.
    - Note that F-tilda is in a sense equivalent to the F obtained in a VPA analysis, where the observed catch is assumed to be without error. In the statistical catch at age world, the F value is directly related to the predicted catch via the catch equation.
- Since Fay cannot be calculated by elementary expression, could use Popes approximation:

$$N_{a+1,y+1} \approx N_{a,y} e^{-M_{a,y}} - C_{a,y}^{OBS} e^{-0.5 M_{a,y}}$$

$$\tilde{F}_{a,y} \approx -\ln\left(1 - \frac{C_{a,y}^{OBS}}{N_{a,y}} e^{0.5 M_{a,y}}\right)$$

# Recursive updating: Step 2



- Another estimate of Fay is available obtained by assuming that the selection pattern this year is the same as that of last year. Thus:

$$F_{a.y} = s_{a,y-1} F_y$$

  - Note that this is just the separable assumption, i.e. that selection pattern does not change between the two years.
  - Note that the Fy is directly related to the predicted catch.
- We can thus calculate $F_{temp}$ as a compromise between the two estimates of fishing mortalities by:

$$F_{a,y}^{TEMP} = \left(1-\alpha_{a,y}\right) s_{a,y-1} F_y + \alpha_{a,y} \tilde{F}_{a,y}$$

- If $\alpha$=0 this becomes a separable model, if $\alpha$=1 this is a forward based VPA model (no errors in the observed catches).

# Recursive updating: Step 3

- The selection at age (say) for the current year is finally obtained by scaling the $F_{temp}$ so that the average selection over the reference age is ONE:

$$s_{a.y} = \frac{F_{a,y}^{temp}}{\dfrac{\sum_{a=a1}^{a=a2} F_{a,y}^{temp}}{(a2-a1+1)}}$$

  - a1: Lower reference age group
  - a2: Upper reference age group

# Recursive updating – some note on α



- The algorithm has some similarity with a Time Series Analysis (TSA), where α would correspond to the Kalman gain vector.
- In TSA framework, the gain is (at least conceptually) the ratio between process variation and observation noise and is estimated in the model.
  - Don't ask me how, please – I don't know how TSA does this!
- In AMCI α is just input by the user. The choice of α should be guided by some perceived understanding of the ratio of the process vs. observation noise.
  - Process variation: Model error
  - Observation noise: Error in the catch at age matrix

FIF

# Modelling catch at age and survey indices at age as proportion

# Treatment of catch information I

- NFT-ASAP (as many other models) split the yearly catch into:
  - The total yield in weight:

$$Y_{a,y} = C_{a,y} W_{a,y} \tag{6}$$

  - Proportion by age in numbers

$$P_{a,y} = \frac{C_{a,y}}{\sum_{y} C_{a,y}} \tag{7}$$

Fleet subscript (g) ignored here

# Treatment of catch information II



- The objective function for catch thus becomes:
  - The yield in weight (log normal):

$$L_1 = \lambda_1 \left[ \ln\left( \sum_a Y_{a,y} \right) - \ln\left( \sum_a \hat{Y}_{a,y} \right) \right]^2 \tag{16}$$

  - Proportion by age in numbers (multinomial)

$$L_2 = \sum_y \lambda_{2,y} \sum_a P_{a,y} \ln\left( \hat{P}_{a,y} \right) - P_{a,y} \ln\left( P_{a,y} \right) \tag{17}$$

Fleet subscript (g) ignored here

# Treatment of survey indices



- In ASAP survey information can be split into selectivity and catchability:

$$\hat{I}_{a,y} = \mathrm{q}\sum_{\mathrm{a}} \mathrm{s}_{\mathrm{a}} N^{*}_{a,y} \qquad (11)$$

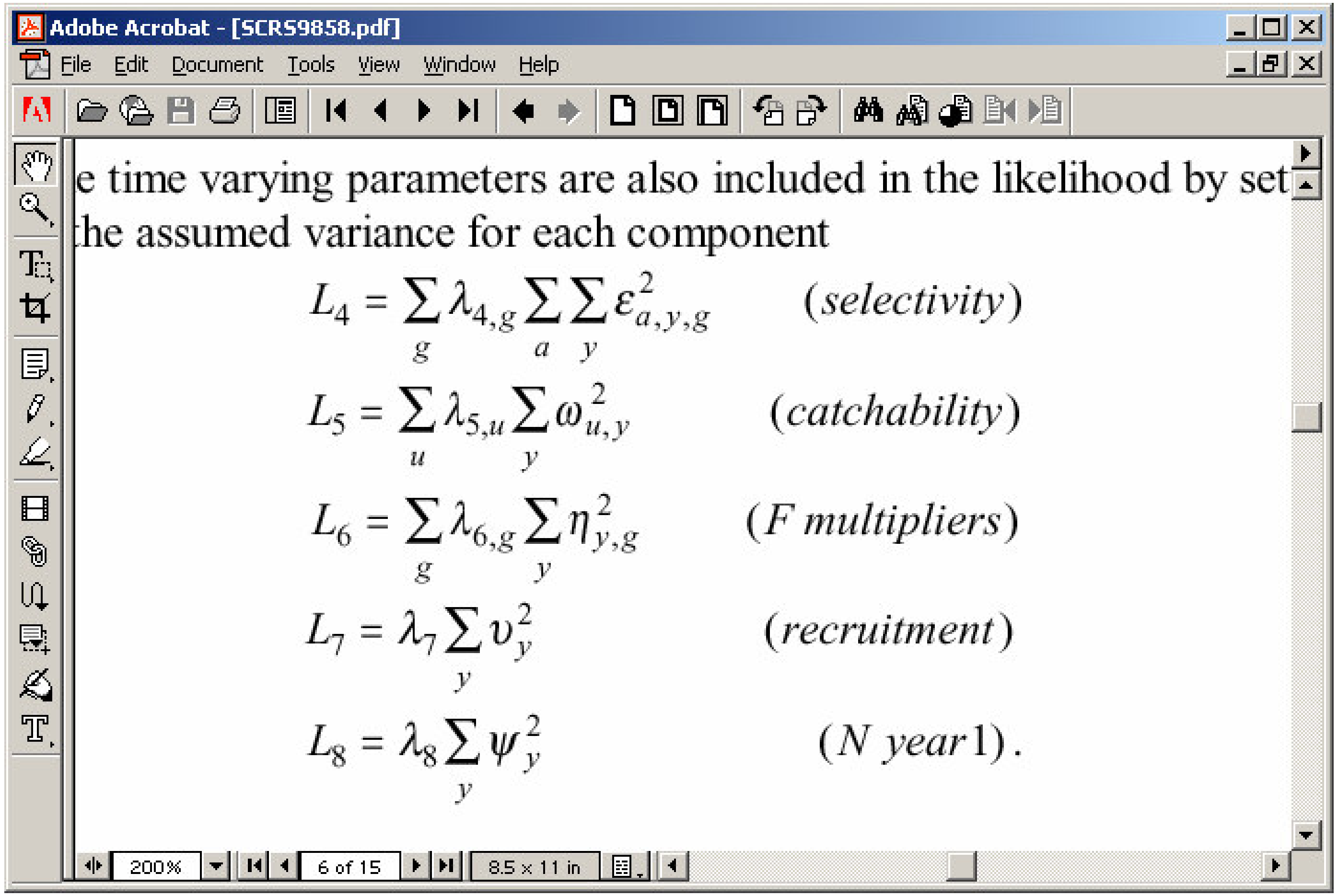
- Can be either number or biomass

- This is also what is done in Stock Synthesis and in Coleraine.

Fleet subscript (g), ignored here

# Lots of random walks

einar




e time varying parameters are also included in the likelihood by set
the assumed variance for each component

$$L_4 = \sum_g \lambda_{4,g} \sum_a \sum_y \varepsilon^2_{a,y,g} \qquad (selectivity)$$

$$L_5 = \sum_u \lambda_{5,u} \sum_y \omega^2_{u,y} \qquad (catchability)$$

$$L_6 = \sum_g \lambda_{6,g} \sum_y \eta^2_{y,g} \qquad (F\ multipliers)$$

$$L_7 = \lambda_7 \sum_y \upsilon^2_y \qquad (recruitment)$$

$$L_8 = \lambda_8 \sum_y \psi^2_y \qquad (N\ year\, 1).$$

# A final practical note



- Take flexible parts in many software packages as a sign of warning. Some software have more “switches” than you may have dreamt of.
- The “manuals” are sometimes very rudimentary. Even setting up a fixed separability model is a challenge in some of these software packages.
- Before using any off-the-shelf software, make sure you understand the underlying principles and assumptions. At least make sure that you know how to turn off options that you don’t want.

- If some things look unfamiliar try to emulate it on you own. Most of the time complex formulas or implementations can be demystified if one finds the time.

- First and foremost: Know your data!
  - If the data are good, the method does not matter
  - If the data are  poor, the method does not matter
  - (Un)fortunately most of the data are “in-betweenies”!

FIF

The VPA model

# The VPA map

einar

9. Solver

Store $q_y$ and $N_{term}$ here

1. Parameters

| | | |
|---|---|---|
| Year Age -----> <br> 2. Selection pattern <br> **Not modelled** | Year Age -----> <br> 3. Fishing mortality | Year Age -----> <br> Natural mortality |
| Year Age -----> <br> 4a. Population numbers | Year Age -----> <br> 4b. Nay at survey 1 time | Year Age -----> <br> 4c. Nay at survey 2 time |
| Year Age -----> <br> 5. Predicted catch <br> **Not modelled** | Year Age -----> <br> 6a. Predicted survey 1 indices | Year Age -----> <br> 6b. Predicted survey 2 indices |
| Year Age -----> <br> Observed catch at age | Year Age -----> <br> Observed survey 1 indices | Year Age -----> <br> Observed survey 2 indices |
| Year Age -----> <br> 7. Catch residuals <br> **Not modelled** | Year Age -----> <br> 8a. Survey 1 residuals | Year Age -----> <br> 8b. Survey 2 residuals |
| Year Age -----> <br> 7. Catch residuals squared <br> **Not modelled** | Year Age -----> <br> 8a. Survey 1 residuals squared | Year Age -----> <br> 8b. Survey 2 residuals squared |

Color code
Population model
Observation mo
????
Measurements
Objectives
Parameters

# The parameters of the VPA

- Terminal N values for each true cohort
  - If you don't have the suvey values for all cohorts "alive" in the terminal year one needs to do something else for those year classes (same of course applies to statistical catch at age model).
- $q_a$ values for each survey age index

- Terminal N here is $N_{a,1925}$ for all true ages
  - These are stored in the parameter area and transferred to the proper place in the Nay matrix

# VPA: Oldest true age (age 10)

- Different approach because don't generally have good (any) survey informations for the oldest ages.
- Generally make assumptions about F in the oldest age being similar as on younger age groups, taken values from younger ages:
  - $F_{a,y} = F_{a-1,y}$
- Populations numbers for oldest true age then derived from the inverted catch equation:

**a=10**

$$N_{ay} = \frac{C_{ay}}{F_{ay} / \left(F_{ay} + M_{ay}\right)\left(1 - e^{-\left(F_{ay} + M_{ay}\right)}\right)}$$

# The x-tools: A bit of a warning messages



- The purpose of the x-tools are:
  - Provide a tool to enhance the understanding of simulations
  - Provide a tool to that may help to understand the various implementations of the statistical catch at age assessment models (note the plural).
  - Provide tools that may address various hypothesis.
  - Provide tools that may show how observation errors may affect our perception of the stock.
  - Provide tools that may show how seriously wrong we may go if the observation data is biased or if we make wrong model assumptions.
- If you find a programming bug then you found IT!
  - Just make sure it is not in your head
- Don’t use the x-tools on real things for real purposes.
- And again, remember that the real world is much more complicated than set up in a simple simulator

# My knowledge base


- Mohn & Cook 1992. Introduction to sequential population analysis. NWAFO Sci. C. Studies 17.
- Hilborn & Walters 1992. Quantitative Fisheries Stock Assessment. Chapman and Hall.
- Burgman, Ferson & Akcakaya 1994. Risk Assessment in Conservation Biology, Chapman & Hall.
- Gallucci, Saila, Gustafson & Rothschild 1995. Stock assessment: Quantitative methods and applications for small-scale fisheries. CRC.
- Hilborn & Mangel 1997. The Ecological Detective. Princeton.
- Quinn & Deriso 1999. Quantitative Fish Dynamics. Oxford University Press. ISBN 0-19-507631-1
- Haddon 2001. Modelling and Quantitative Methods in Fisheries. CRC. ISBN 1-58488-177-1
- Lassen & Medley 2001. Virtual population analysis. A practical manual for stock assessment. FAO Fisheries Tech. Pap. 400.
- Punt & Hilborn 2001. BAYES-SA. Bayesian stock assessment methods in fisheries. User's manual. FAO Computerized Information Series No 12.
- Berkson, J.M., Kline, L.L., Orth, D.J. (eds.) 2002. Incorporating uncertainty into fisheries models. AFSS 27, Bethesda, Maryland.


- Practice
- Personal communication: Höski and Dankert Skagen