SURPLUS PRODUCTION MODELS

1. Surplus production model

State equation

The deterministic state equation of the production model (1969) is:

$$B_{t+1} = B_t + g(B_t) - C_t,$$

where B_t is the biomass at the start of the year t, C_t is the catch during year t and g(B_t) is the surplus production function that determines the overall change in biomass due to growth, recruitment and natural mortality.

In particular, in the Pella-Tomlinson model, the surplus production function is as follows

$$g(B_t) = \frac{r}{p} B_t \left(1 - \left(\frac{B_t}{k}\right)^p\right),$$

where r is the intrinsic growth rate parameter, k is the carrying capacity and p is the asymmetry parameter. When $p = 1$ the model is the Schaeffer's model.

If fishing begins in year 1, it is common to assume that the virgin biomass is equal to the carrying capacity ($B_1 = k$).

From these equations it follows that:

$$B_{msy} = k\left(\frac{1}{1+p}\right)^{1/p}$$

$$F_{msy} = r\frac{1}{1+p}$$

$$Y_{msy} = rk\left(\frac{1}{1+p}\right)^{1+1/p}$$

The parameter ϕ that is within the interval (0,1) is defined in ASPIC as

$$\phi = \left(\frac{1}{1+p}\right)^{1/p}$$

Observation equation

The deterministic observation equation is:

$$I_t = qB_t\ ,$$

where I_t is the relative biomass index and q is the catchability coefficient. There might be more than one index. If the biomass index is absolute, the catchability coefficient is fixed at 1.

Note that catches are assumed to be observed without error.

2. Maximum likelihood version

The model can be fitted by maximum likelihood by assuming a log-normal error distribution in the observation equation as follows:

$$\log(\hat{I}_t) \sim \mathrm{Normal}(\log(I_t), \sigma_I^2) .$$

The observed relative biomass index is denoted by $\hat{I}_t$ and σ_I^2 represents the variance (in log-scale) of the observation equation for the abundance index. This means that the coefficient of variation (in natural scale) of the abundance index is $\mathrm{cv}(\hat{I}_t) = \sqrt{\exp(\sigma_I^2) - 1}$.

The function to maximise is the likelihood function of the observations which is given by:

$$L = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_I^2}} exp\left\{-\frac{1}{2\sigma_I^2}\left(\log(\hat{I}_t) - \log(q) - \log(B_t)\right)^2\right\}.$$

Or equivalently, the logarithm of the likelihood:

$$l = \log(L) = -\frac{n}{2}\log(2\pi\sigma_I^2) - \frac{1}{2\sigma_I^2}\sum_{t=1}^{n}\left(\log(\hat{I}_t) - \log(q) - \log(B_t)\right)^2 .$$

By looking at the partial derivatives with respect to q and σ_I^2 we get closed-form estimates of the catchability coefficient q and the variance of the observation equation σ_I^2.

$$\begin{aligned}
\frac{\partial l}{\partial q} = -\frac{1}{2\sigma_I^2}\sum_{t=1}^{n} 2\left(\log(\hat{I}_t) - \log(q) - \log(B_t)\right)\left(-\frac{1}{q}\right) = 0 \\
\sum_{t=1}^{n}\left(\log(\hat{I}_t) - \log(q) - \log(B_t)\right) = 0 \\
-n\ \log(q) + \sum_{t=1}^{n}\left(\log(\hat{I}_t) - \log(B_t)\right) = 0 \\
q = \exp\left\{\frac{1}{n}\sum_{t=1}^{n}\left(\log(\hat{I}_t) - \log(B_t)\right)\right\}
\end{aligned}$$

And

$$\begin{aligned}
\frac{\partial l}{\partial \sigma_I^2} = -\frac{n}{2}\frac{2\pi}{2\pi\sigma_I^2} + \frac{1}{2(\sigma_I^2)^2}\sum_{t=1}^{n}\left(\log(\hat{I}_t) - \log(q) - \log(B_t)\right)^2 = 0 \\
-n\sigma_I^2 + \sum_{t=1}^{n}\left(\log(\hat{I}_t) - \log(q) - \log(B_t)\right)^2 = 0 \\
\sigma_I^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\log(\hat{I}_t) - \log(q) - \log(B_t)\right)^2
\end{aligned}$$

The rest of parameters to be estimated are B_1, r, k and p. It's important to note that the parameters are bounded because they have to fulfil that the biomasses are positive, i.e. they have to be larger than the catches that are assumed to be observed without any error.

In case of having more than one index, the above equations for the catchability coefficient q and the variance of the observation equation σ_I^2 for each of the indices still apply.

Software available to fit this type of models are CEDA and ASPIC. The model can be implemented in ADMB.

It is recommended to do a grid search over B_1, r, k and p to explore the domain of plausible parameter values.

Additional information from the parameters:

$$B_1 \leq K$$

Relationship between r, Fmsy, Z, etc from general papers like Zhou, might also be useful.

The carrying capacity can be set initially around 80% of the total catch observed (see ASPIC manual).

3. Bayesian version

A Bayesian state-space model version including observation and process errors can be constructed.

The state equation includes log-normal process error. A different parameterization $P_t = B_t/k$ instead of B_t.

$$\log(\hat{I}_t) \sim \text{Normal}(\log(q) + \log(k) + \log(P_t), \sigma_I^2)$$

$$\log(\hat{P}_1) \sim \text{Normal}(\log(P_1), \sigma_P^2)$$

$$\log(\hat{P}_{t+1}) \sim \text{Normal}\left(\log\left(P_t + \frac{r}{p} P_t (1 - P_t{}^p) - \frac{C_t}{k}\right), \sigma_P^2\right)$$

The parameters to be estimated are $P_1, r, k, q, p, \sigma_I^2$ and σ_P^2.

The prior distributions are as follows

$$\log(r) \sim Normal\ (\mu_{\log(r)}, \psi_{\log(r)})$$

$$\log(k) \sim Normal\ (\mu_{\log(k)}, \psi_{\log(k)})$$

$$\log(p) \sim Normal\ (\mu_{\log(p)}, \psi_{\log(p)})$$

$$\log(q) \sim Normal\ (\mu_{\log(q)}, \psi_{\log(q)})$$

$$P_1 \sim Uniform\left(a_{P_1}, b_{P_1}\right)$$

$$\psi_I \sim Gamma\left(a_{\psi_I}, b_{\psi_I}\right)$$

$$\psi_P \sim Gamma\left(a_{\psi_P}, b_{\psi_P}\right)$$