

Appendix A - Likelihood calculations

In this appendix we show further details of the methods. We discuss:

- How a Gompertz (linear, Gaussian) state-space model for the multiple observation data can be factored into a state-space model for the estimated mean at each time and a piece for the independent sampling variance and site effects, if any. This simplifies likelihood calculations for the full ML method.
- How, based on this factorization, means and standard errors for population estimates can be plugged into the state-space model as an approximation, yielding the pseudo ML method.
- How, when site effects are treated as fixed effects in a linear model for the multiple surveys, their maximum likelihood estimates can be taken directly from the linear model and plugged into the Kalman Filter. This simplifies optimization of the likelihood for full ML by providing site effect values easily.
- Why the pseudo ML method can work in more general settings where the raw data are not normally distributed. This is used for the real-data example, where a negative binomial log-linear model is used to analyze the replicated data, yielding means and standard errors that are plugged in to the state-space model.

Factorization of the state-space model

The full likelihood for the model defined by (1) and (2) may be written as

$$\begin{aligned}
 L(a, c, \tau, \mathbf{b}, \sigma_{1:T}, \omega | \mathbf{y}_{1:T}) &= \int p(\mathbf{y}_{1:T} | x_{1:T}, \eta_{1:T}) p(x_{1:T}) p(\eta_{1:T}) d\eta_{1:T} dx_{1:T} \\
 &= \int \left\{ \prod_{t=1}^T \int p(\mathbf{y}_t | x_t, \eta_t) p(\eta_t) d\eta_t \right\} p(x_{1:T}) dx_{1:T}
 \end{aligned} \tag{A.1}$$

In this equation, $p(x_{1:T})$ is the probability density of the state dynamics, from equation (1), and $p(\mathbf{y}_t | x_t, \eta_t)$ is the probability density of the observations at time t from equation (2). When the data are fully independent, η_t is not needed. When the data are only partially independent, η_t is the shared sampling error for all data in year t , and must be integrated over as a random effect. The site effects, \mathbf{b} , are treated here as fixed effects with a sum-to-zero constraint (contrast): $\sum_i b_i = 0$. We have written $\sigma_{1:T}$ to allow the possibility of time-dependent sampling variance, but if sampling variance is constant this could be replaced by σ .

The first step is to write the normal distribution equation for $\mathbf{y}_t | x_t, \eta_t \sim N(\mathbf{1}(x_t + \eta_t) + \mathbf{b}, \sigma_t^2 I_K)$ and factor it into parts for \bar{y}_t , the estimated mean, and the rest. If site effects are not needed, the rest depends only on $S(\mathbf{y}_t)$, the sample standard deviations of \mathbf{y}_t . Otherwise, the rest depends on the sample standard deviations of $\mathbf{y}_t - \mathbf{b}$, $S(\mathbf{y}_t - \mathbf{b})$. The sample mean and standard deviation are sufficient statistics for independent normal variables, meaning that they contain all information in the data needed for the likelihood. (If site effects are needed, one way to simplify the equations would be to redefine \mathbf{y}_t as having \mathbf{b} subtracted, giving $\mathbf{y}_t | x_t, \eta_t \sim N(\mathbf{1}(x_t + \eta_t), \sigma_t^2 I_K)$. However, we do not do it that way in order to keep the equations as explicit as possible.) These sufficient

statistics may be used to decompose the densities $p(\mathbf{y}_t|x_t, \eta_t)$ as follows:

$$\begin{aligned}
p(\mathbf{y}_t|x_t, \eta_t) &= \frac{1}{(2\pi\sigma_t^2)^{K/2}} \exp\left(-\frac{1}{2\sigma_t^2} \sum_i (y_{it} - x_t - \eta_t - b_i - \bar{y}_t + \bar{y}_t)^2\right) \\
&= \frac{1}{(2\pi\sigma_t^2)^{K/2}} \exp\left(-\frac{1}{2\sigma_t^2} \sum_i (y_{it} - b_i - \bar{y}_t)^2\right) \exp\left(-\frac{K}{2\sigma_t^2} (\bar{y}_t - x_t - \eta_t)^2\right) \\
&= \left\{ \frac{1}{\sqrt{K}(2\pi\sigma_t^2)^{(K-1)/2}} \exp\left(-\frac{K-1}{2\sigma_t^2} S(\mathbf{y}_t - \mathbf{b})^2\right) \right\} \cdot \\
&\quad \left\{ \frac{1}{\sqrt{2\pi\sigma_t^2/K}} \exp\left(-\frac{K}{2\sigma_t^2} (\bar{y}_t - x_t - \eta_t)^2\right) \right\} \\
&= L_1(\sigma_t, \mathbf{b}|\mathbf{y}_t) \cdot p(\bar{y}_t|x_t, \eta_t)
\end{aligned} \tag{A.2}$$

The factor $L_1(\sigma_t, \mathbf{b}|\mathbf{y}_t)$ is defined as the contents of the first curly-braces. The second curly braces give $p(\bar{y}_t|x_t, \eta_t)$, the normal probability density of \bar{y}_t , with mean $x_t + \eta_t$ and variance σ_t^2/K .

Since $L_1(\sigma_t, \mathbf{b}|\mathbf{y}_t)$ does not depend on x_t or η_t , it may be moved outside of the integral in (A.1), yielding

$$\begin{aligned}
L(a, c, \tau, \mathbf{b}, \sigma_{1:T}, \omega|\mathbf{y}_t) &= \left\{ \prod_{t=1}^T L_1(\sigma_t, \mathbf{b}|\mathbf{y}_t) \right\} \int \left\{ \prod_{t=1}^T \int p(\bar{y}_t|x_t, \eta_t) p(\eta_t) d\eta_t \right\} p(x_{1:T}) dx_{1:T} \\
&= \left\{ \prod_{t=1}^T L_1(\sigma_t, \mathbf{b}|\mathbf{y}_t) \right\} \int p(\bar{y}_{1:T}|x_{1:T}) p(x_{1:T}) dx_{1:T} \\
&= \left\{ \prod_{t=1}^T L_1(\sigma_t, \mathbf{b}|\mathbf{y}_t) \right\} L_2(a, c, \tau, \sigma_{1:T}, \omega|\bar{y}_{1:T}).
\end{aligned} \tag{A.3}$$

Here the factor $L_2 = p(\bar{y}_{1:T})$ is simply a state-space model for the mean observation at each time. If η_t is not needed (i.e. the sampling is fully independent), then the state-space calculation L_2 uses σ_t^2/K as observation error variance for time t . If η_t is needed then the total observation error variance is $\omega^2 + \sigma_t^2/K$. This is the format that simplifies likelihood calculations for the full ML method, since the state-space calculation needs only univariate observations.

In cases beyond those studied here, e.g. when the population model is non-stationary the likelihood factor L_2 could be replaced by a REML version to potentially improve inferences (Staples et al. 2004).

Plugging in means and standard errors from a linear model for the replicated data

In equation (A.3), the product of the L_1 terms is proportional to the normal distribution likelihood one would have from a linear model for the observations y_{it} with fixed effects for year and site. This has a total exponent of $\sum_t \sum_i (y_{it} - b_i - \bar{y}_t)^2$. Thus, this is just the likelihood for a linear model, $y_{it} = \mu_t + b_i + \epsilon_{it}$, where μ_t are year effects, b_i are site effects subject to $\sum_i b_i = 0$, and $\epsilon_{it} \sim N(0, \sigma^2)$ are residuals.

The estimated year effects from this linear model will simply be \bar{y}_t . The standard error of \bar{y}_t will be an estimate of σ_t/\sqrt{K} , which is used in L_2 . If sampling variance is assumed constant

across years, σ , then the standard error for all \bar{y}_t 's will be an estimate of σ/K . Thus, when the full raw data are not available but means and standard errors are available, it makes sense to plug them into L_2 and drop the L_1 factors. This is an approximation because L_2 does contain σ_t , and hence the time-series aspect does have information about sampling variance, but experience indicates the information from replicates is often much stronger and hence should be retained. When η_t is used to allow additional variation, then ω^2 must be estimated from maximizing L_2 . In effect this amounts to estimating observation variance for the state-space model of the means (L_2), subject to the constraint that it is at least as large as the plug-in estimate of σ^2/K .

Note that the limitation in going further with using plug-in estimates is that we typically lack information on uncertainty in the standard errors themselves. Typically analysis of survey data reports a mean abundance index, its uncertainty in the form of a standard error, and no more. Hence, we have not assumed that any more information will be available. In order to more fully combine information from the multiple observations and the time-series aspect of the problem, one needs more information, which typically would mean access to the full replicated data and an appropriate likelihood model for it.

Easy estimation of site effects under full ML

The factor $L_2 = p(\bar{y}_{1:T})$ does not depend on \mathbf{b} since by definition $\bar{b} = 0$. Thus, based on the above description of the L_1 terms as the likelihood for a linear model with year and site effects, the maximum likelihood (or, equivalently, least squares, in this case) estimates of b_i can be obtained from this linear model and kept constant for maximizing the other parameters. This simplifies the optimization step of the full ML, since there are many site effect parameters.

If residual observation error is homogeneous ($\sigma_t = \sigma$ for all t) and L_2 does not depend on \mathbf{b} , the maximum likelihood estimate of \mathbf{b} is easy to write down. It is independent of the other parameters and, for component i , is given by $1/T \sum_t y_{it} - 1/(TK) \sum_{i,t} y_{it}$. If the residual observation error variance is not homogeneous, the maximum likelihood estimate of \mathbf{b} for given σ_t is given by $(\sum_t (1/\sigma_t^2))^{-1} \sum_t y_{it}/\sigma_t^2 - \sum_t (K/(\sigma_t^2))^{-1} \sum_i y_{it}/\sigma_t^2$. Substituting \mathbf{b} for this estimate in the likelihood (A.3) one may avoid numerical maximization over \mathbf{b} also in this case.

If replicates from some sites are missing in some years, L_2 will depend on \mathbf{b} because then $\bar{y}_t|x_t \sim N(x_t + \bar{b}_t, \sigma_t^2/(K - M_t) + \omega_t^2)$ where \bar{b}_t is the mean of the site effects over the sites that were observed in year t and M_t is the number of sites for which observations are missing in year t . The factors L_1 will change only in that K will be replaced by $K - M_t$.

The pseudo ML approach for general population estimates

The above work was based on linear models with normally distributed errors. However, the pseudo ML method does not require that the raw data are normally distributed, but rather that the population estimates are. In general, population estimates may come from a capture-mark-recapture model, a distance sampling model, or a generalized linear model for count data, for example. For many such procedures, population estimates will be at least approximately normal for large sample sizes even if the raw data are not. This is, for instance, the case with most procedures based on maximum likelihood estimation, such as the negative binomial regression of Breeding Bird Survey data in the main paper. Apart from normality, a second assumption of the pseudo ML method is that deviations between true and estimated population sizes are independent across years. This may not hold exactly for non-normal models or for unbalanced designs, especially if nuisance parameters that are shared across time are included. However, if the design is reasonably well balanced, dependence across time will presumably not be a major

issue. Considering these points we suggest that the pseudo ML method may provide a good approximation to full analyses in many cases in practice.

References

Staples, D., F. Taper, M. L. and B. Dennis. 2004. Estimating population trend and process variation for PVA in the presence of sampling error. *Ecology* 85:923-929.